



Use of the cDNA microarray technology in the safety assessment of GM food plants

Pedersen, Jan W.; Knudsen, Ib; Eriksen, Folmer Damsted; Kärenlampi, S.; Mikalsen, A.; Hammerling, U.; Kok, E.; Bak, S.; Nielsen, Henrik Bjørn

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Pedersen, J. W., Knudsen, I., Eriksen, F. D., Kärenlampi, S., Mikalsen, A., Hammerling, U., Kok, E., Bak, S., & Nielsen, H. B. (2003). *Use of the cDNA microarray technology in the safety assessment of GM food plants*. Nordic Council of Ministers. TemaNord No. 558

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Use of the cDNA microarray technology
in the safety assessment of GM food
plants

Use of the cDNA microarray technology in the safety assessment of GM food plants

Use of the cDNA microarray technology in the safety assessment of GM food plants

TemaNord 2003:558

© Nordic Council of Ministers, Copenhagen 2003

ISBN 92-893-0955-5

ISSN 0908-6692

Print: Ekspresen Tryk & Kopicenter

Copies: 360

Printed on paper approved by the Nordic Environmental Labelling.

This publication may be purchased from any of the agents listed on the last page.

Nordic Council of Ministers

Store Strandstræde 18

DK-1255 Copenhagen K

Phone (+45) 3396 0200

Fax (+45) 3396 0202

Nordic Council

Store Strandstræde 18

DK-1255 Copenhagen K

Phone (+45) 3396 0400

Fax (+45) 3311 1870

Website: www.norden.org

The Nordic Food Policy Co-operation

The Nordic Committee of Senior Officials for Food Issues is concerned with basic Food Policy issues relating to food and nutrition, food toxicology and food microbiology, risk evaluation, food control and food legislation. The co-operation aims at protection of the health of the consumer, common utilisation of professional and administrative resources and at Nordic and international developments in this field.

The Nordic Council of Ministers

was established in 1971. It submits proposals on cooperation between the governments of the five Nordic countries to the Nordic Council, implements the Council's recommendations and reports on results, while directing the work carried out in the targeted areas. The Prime Ministers of the five Nordic countries assume overall responsibility for the cooperation measures, which are co-ordinated by the ministers for cooperation and the Nordic Cooperation committee. The composition of the Council of Ministers varies, depending on the nature of the issue to be treated.

The Nordic Council

was formed in 1952 to promote cooperation between the parliaments and governments of Denmark, Iceland, Norway and Sweden. Finland joined in 1955. At the sessions held by the Council, representatives from the Faroe Islands and Greenland form part of the Danish delegation, while Åland is represented on the Finnish delegation. The Council consists of 87 elected members – all of whom are members of parliament. The Nordic Council takes initiatives, acts in a consultative capacity and monitors cooperation measures. The Council operates via its institutions: the Plenary Assembly, the Presidium and standing committees.

Contents

Kolofon	4
Contents	5
Preface	6
Executive summary	7
Dansk resumé	12
Yhteenveto	17
Chapter 1 Introduction and scope of the report.....	23
Chapter 2 Characterisation of food plants on the RNA level.....	26
2.1 Comparative approach.....	26
2.2.Traditional methods for the detection of differences in gene expression.....	28
2.3 Stress-related differences in gene expression.....	30
Chapter 3 Principles of the cDNA microchip technology.....	31
3.1 The microarray technology	31
3.2 Technological developments	35
3.3 Potential and bottlenecks of the technology	37
3.4. Sampling	38
3.5 Data analysis	41
3.6. Data storage	46
Chapter 4 Past and present experiences with the use of cDNA microarray technology in plant varieties grown under different (environmental) conditions.....	50
4.1.Model systems	50
4.2. Identification of stress-related genes in plants by use of microarrays	52
Chapter 5 Selection of individual cDNAs for characterisation of food plants	56
5.1. Selection of markers	56
Chapter 6 The role of gene expression microarray technology in the wider profiling of food plants	59
6.1 Proteomics	59
6.2. Metabolomics.....	60
6.3 Glycomics	62
Chapter 7 General discussion.....	64
Chapter 8 Conclusions.....	69
Glossary	72
References.....	75

Preface

The Nordic Committee of Senior Officials for Food Issues is a body under the Nordic Council of Ministers co-ordinating Nordic work in the field of foods. The Nordic Working Group on Food Toxicology and Risk Evaluation (NNT) has been given the responsibility by the Committee to promote co-operation and co-ordination among Nordic countries in matters relating to food toxicology and risk assessment.

Under this working group a project group was established in order to discuss and evaluate the usefulness of the DNA microarray technique in the safety assessment of genetically modified plants and other novel plants as well as the usefulness and possibility for making a database containing profiles of food plants. This includes a description of the state of the art, the technical challenges and a discussion of a more targeted microarray and the nature of the probes that may be relevant for such an array.

The report has been accepted by NNT in October 2003.

The project group agreed on a number of conclusions, which is summarised in chapter 8.

The project group consisted of the following members:

Jan Pedersen (Chairman)	Danish Veterinary and Food Administration	Denmark
Ib Knudsen	Danish Veterinary and Food Administration	Denmark
Folmer D. Eriksen	Danish Veterinary and Food Administration	Denmark
Sirpa Kärenlampi	Institute of Applied Biotechnology, University of Kuopio	Finland
Arne Mikalsen	Norwegian Institute of Public Health	Norway
Ulf Hammerling	National Food Administration	Sweden
Esther Kok	Institute for Food Safety, RIKILT	Netherlands
Søren Bak	The Royal Veterinary and Agricultural University	Denmark
Henrik Bjørn Nielsen	Technical University of Denmark, CBS	Denmark

The report was prepared by E.J. Kok, G.A. Kleter and J.P. van Dijk, RIKILT Institute of Food Safety, Bornsesteeg 45, PO Box 230, 6700 AE Wageningen, The Netherlands. The authors would like to thank ir. Angeline M.A. Van Hoef for her contribution to the manuscript.

Executive summary

Until 1997 there were no common regulations or guidelines within the European Union to assess the safety of food from novel plant varieties. This situation changed in 1997 when the European Regulation on Novel Foods and Novel Food Ingredients came into force (EU regulation (EC) No 258/97). This Regulation, commonly referred to as the Novel Foods Regulation, covers also foods and food ingredients containing, consisting of or produced from genetically modified organisms (GMOs). In July 2003 the European Council of Ministers have adopted the Regulation on Genetically Modified Food and Feed. This new regulation will replace the part of the Novel Foods Regulation dealing with GMOs and GMO-derived products (http://europa.eu.int/comm/food/fs/biotech/biotech08_en.pdf). According to the new regulation the scientific risk assessment will be carried out by the European Food Safety Authority along the lines of the guidance document for the risk assessment of genetically modified (GM) plants and derived food and feed (http://europa.eu.int/comm/food/fs/sc/ssc/out327_en.pdf).

An important part of the initial phase of the safety evaluation of all these novel GMO-derived plant products is an elaborate comparative compositional analysis, including analyses of key nutrients and anti-nutrients, of the newly developed plant variety and of the 'traditional' counterpart(s), if available, that was already on the market. Although this has been considered as an adequate procedure for evaluating the safety of GM plants, some concerns have been raised whether this targeted approach will cover all unintended effects that may occur and may have an influence on the consumer's health. As a result of these developments the interest in the development of informative non-targeted methodologies to screen for compositional differences and assess their toxicological relevance has rapidly increased. It was generally acknowledged that the time had come to evaluate new analytical approaches that may give more insight into possible changes in the plant's physiology compared to the current approach of analysing a limited set of key constituting compounds.

These so-called unbiased methodologies have the potential to screen for aberrations in a much wider range of metabolic routes compared to the targeted analyses. Non-

targeted approaches may relate to either the transcriptome, i.e. the total of initial transcription products (mRNA) per cell, or the proteome, the total of proteins per cell, or the metabolome, the total of metabolites per cell. Recently even the glycome, the total of sugar molecules per cell, has become subject of scientific scrutiny to detect specific changes. For all of these different cell constituents more or less advanced profiling methodologies have been developed. One of the most promising developments in this respect is the DNA microarray technology. Prior to the advent of the microarray technology methods were available to screen for differences in gene expression, but only a very limited set of informative probes could be analysed simultaneously in the mRNA population of interest. The microarray technology miniaturises well-established molecular biological and biochemical principles, thereby enabling to establish gene expression profiles of individual mRNA samples on the basis of thousands of different probes, representing numerous metabolic routes. In reality mRNA levels will not necessarily correlate closely to the abundance of the corresponding protein or its biological activity. However, for the near future the gene expression microarray may be the best developed system for the unbiased screening for metabolic changes in plants.

Initially two types of arrays were available: oligonucleotide arrays, where the oligo's were synthesised on a solid support, and cDNA or oligonucleotide arrays, where presynthesised probes were spotted on glass slides. Recently a number of alternative array systems have been developed that either increase the sensitivity of the system, reduce the time for hybridisation or require less expensive equipment.

The quality of the mRNA populations to be hybridised in microarray experiments is crucial for the reliability of the resulting data. The mRNA has to be of high-standard quality in order to be able to compare two subsequent samples. Therefore standardisation of the sampling and mRNA isolation protocols as well as the quality control procedure is required. As slightly different conditions in the food or food product will generally have large effects on the quality of the isolated mRNA, it is unlikely that the microarray technology can become standard for the safety evaluation of processed products. The technology will have its main application in the living organism: the development and safety evaluation of novel plant varieties, rather than in the evaluation of the derived products.

Because of the fact that any microarray experiment will result in a large number of data points, it is crucial that the experiment is well-designed taking into account a) available experimental, control and reference samples, b) statistical considerations and c) practical limitations such as the number of available mRNA samples and the series of arrays that can be handled simultaneously.

Data analysis and data mining form the last but very important phase in the experimental chain. In general, aspects of data storage and analysis should already be considered prior to the experiment taking place in order to prevent data loss and to be able to include the right references and not lose time afterwards by undirected data computing. Data analysis is usually performed by the subsequent steps of quantification of the signal, normalisation of the obtained data and data visualisation and interpretation. In recent years a large variety of software packages have become available that aim on any of these different aspects or combine subsequent stages. One drawback of computerising large data sets is that the output may be difficult to interpret in relation to potential health aspects. Small differences in experimental set-up will often result in differences in gene expression profiles that can be detected by the data analysis software but are unrelated to the scientific questions underlying the experiment.

Furthermore, it should always be kept in mind that gene expression analysis by the microarray technology will only provide indications on the differential gene expression pattern of individual genes that may be related to unintended effects as a result of the selected breeding strategy. Only in very specific cases will statistical analysis of the microarray results give full evidence of differential gene expression of a gene or a cluster of genes. Normally additional analyses by other techniques, such as real-time PCR or Northern blotting, will be required to confirm the observations. Furthermore, to test the biological and toxicological significance of the results, as DNA or RNA differences in itself are not a health issue, it will be necessary to confirm the results with experiments on other integrative levels, such as the protein or metabolite level. This requirement is partially due to the lack of knowledge on the relationship between gene expression and changes in the plant's physiology.

The first initiatives have been launched to harmonise the criteria for the storage of microarray data in globally approachable databases. It is clear that a good database design is crucial for the ability to use the wealth of data that will result from the growing number of array experiments worldwide. Basically, if the same (commercially available) array is used by many research groups this can be seen as a common experimental platform. If the results obtained with this array are stored in a centralised database, future in silico analyses of the available data may reduce the number of experiments needed to obtain the desired information.

Such worldwide initiatives are now starting for a limited number of model plants. Important model systems in plant (functional) genomics are *Arabidopsis thaliana*, rice and tomato. The number of microarray gene expression experiments that have been performed in plants in general is already considerable and still rapidly increasing. The pathways of interest in these studies differ, but the larger part is aimed towards the elucidation of stress-related pathways. Microarray studies have revealed, for example, unanticipated interrelationships between stress and other elicitors (plant hormone, ripening) that induce altered gene expression of the same pathways.

For the safety assessment of novel plant varieties, including GM plants, relevant metabolic routes may be nutrient-related pathways as well as metabolic pathways that can be used to screen for differences in the plants' basal physiology. Other metabolic routes of interest may be stress-related metabolic pathways as it is known that upregulation of these pathways may lead to an increased production of anti-nutrients, including natural toxins. It may have advantages to construct a specific food safety assessment array comprising these three different categories of probes.

Alternative 'omics' approaches comprise proteomics, analysis of the protein fraction, metabolomics, analysis of the pool of metabolites, and perhaps glycomics, analysis of the sugar molecules, in selected tissues or cell systems. Proteomics is the direct counterpart to transcriptomics. As any indication for differential gene expression normally will need to be confirmed by other approaches the best way to determine the biological effect may be to analyse the proteome for similar shifts in the protein profile. The current approach to analyse the proteome is usually restricted to a small subset of the proteome and it will be difficult and time-consuming to gain further insight into

possible changes in the entire proteome. For the proteome microarrays are also 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.

The metabolome represents the highest level of physiological interactions and consists of all metabolites in the tissue or cell system under investigation.

Metabolomics, i.e. the non-targeted study of the metabolome, can be a valuable addition to gene expression or proteome profiling as it provides insight into the extent of altered metabolism as a result of changes in gene expression or translation. The drawbacks of current metabolomics are similar to those of proteomics, i.e. the fact that only a subset of metabolites can be identified in a single analysis.

Glycomics is the most recent shoot on the 'omics' tree and presumably the most complicated one as glycosylation patterns change continuously. In future times glycomics may, however, also become valuable in hypothesis-driven risk assessment strategies for novel (GM) plant varieties.

In general, it can be concluded that the advent of GM plant varieties has rapidly enhanced the interest in the compositional characteristics of crop plant species. Profiling methodologies have the potential to screen more effectively for differences in novel plant varieties as they can, theoretically, cover a much wider range of different metabolic routes. The microarray technology has the major advantage that genome-wide gene expression arrays will become available for the first crop species in the near future. As a result of this it can be envisaged that gene expression profiling will enforce and refine current targeted safety assessment strategies for novel plant varieties. This is, however, likely to be a temporary phenomenon, until whole proteome and metabolome profiling methodologies will become available. Finally, direct selection of wholesomeness-related differences in novel plant varieties, either GM-derived or not, will only be achieved once we have obtained a full understanding of the physiology of the plant and its relationship to the health of the consumer.

Dansk resumé

Indtil 1997 var der ingen fælles reguleringer eller retningslinier i EU til vurdering af sikkerheden af fødevarer fra nye plantesorter. Dette ændredes i 1997 da den fælles EU rådsforordning om Nye Levnedsmidler og Nye Levnedsmiddelingredienser (kaldet Novel food rådsforordningen) trådte i kraft.

Denne regulering dækker bl.a. fødevarer og fødevaringredienser som indeholder, består af eller er produceret ved brug af genmodificerede organismer (GMO). I juli 2003 blev en ny regulering om genmodificerede fødevarer og foderstoffer vedtaget i EU's Ministerråd. Den nye regulering vil erstatte den del af Novel food rådsforordningen som vedrører de genmodificerede produkter (http://europa.eu.int/comm/food/fs/biotech/biotech08_en.pdf).

I følge den nye regulering skal den faglige risikovurdering foretages af det europæiske fødevarer agentur (EFSA) i overensstemmelse med de retningslinier som er vedtaget for risikovurdering af genetisk modificerede planter (GM-planter) og de afledede fødevarer og foderprodukter (http://europa.eu.int/comm/food/fs/sc/ssc/out327_en.pdf).

En vigtig del af denne risikovurdering af fødevarer fra nye genmodificerede planter er en indgående sammenligning af indholdsstoffer, herunder vigtige ernæringsstoffer og antinutritionelle stoffer, mellem den genmodificerede plante og den tilsvarende traditionelle plante. Selvom sammenligningen blev anset for værende tilstrækkelig for risikovurdering af GM-planter, har der været rejst tvivl om metoden nu også dækkede alle utilsigtede effekter som kunne forekomme og have indflydelse på forbrugerne sundhed. Derfor var der interesse for udvikling af nye ikke-målrettede metoder til sammenligning og screening af indholdsstoffer og vurdere den sundhedsmæssige relevans heraf. Det har generelt været erkendt, at tiden var inde til at vurdere nye metoder som kunne give bedre indsigt i ændringer i planters fysiologi i forhold til de nuværende metoder som ser på et begrænset antal indholdsstoffer. Disse nye (ikke fokuserede) metoder har potentialet til at undersøge for ændringer i et meget bredere område af metaboliske veje sammenlignet med de traditionelle sammenlignende analyser med begrænset antal målinger.

De nye ikke-fokuserede fremgangsmåder kan relateres til enten transskriptom, proteom eller metabolom som er det totale indhold af hhv. mRNA (transskript fra DNA), proteiner eller metabolitter i en celle. Senest er også glycome (indhold af suktermolekyler i en celle) kommet i fokus som mulig metode til måling af ændringer. For alle disse forskellige celle komponenter er mere eller mindre avancerede "profil metoder" blevet udviklet. Et af de mest lovende fremskridt i relation til dette er DNA mikroarray teknologien. Før fremkomsten af denne teknik var det kun muligt i praksis at foretage analyser af et begrænset antal mRNA uden større ressourceanvendelse. Mikroarray teknikken nedskalere velkendte molekylærbiologiske og biokemiske principper og muliggør derved simultan måling af mRNA fra flere tusinde gener repræsenterende mange metabolisme veje.

I virkeligheden vil mRNA niveauet ikke nødvendigvis være tæt forbundet med mængden af det tilhørende protein eller dets biologiske aktivitet. Imidlertid forventes gen ekspressions mikroarray på kortere sigt at være det bedste system til en bred og ikke-fokuseret måling for utilsigtede ændringer i en plante.

I starten var to arrays til rådighed: oligonucleotid arrays, hvor oligoer blev syntetiserede på et fast materiale og cDNA eller oligonucleotid arrays hvor præsyntetiserede prober sættes på en glasplade. Nu findes en række alternative arrays systemer som enten øger følsomheden, reducerer tiden til hybridisering eller kræver mindre dyrt udstyr.

For at to prøver kan sammenlignes i et mikroarray forsøg skal kvaliteten af mRNA være meget høj for at sikre pålideligheden af resultaterne. Derfor er standardisering for udtagning af prøver og mRNA isolering, så vel som kvalitetskontrol procedurer, en nødvendighed. Da selv en mindre ændring i forholdene af et fødevareprodukt kan have en stor effekt på kvaliteten af det oprensede mRNA, er det ikke sandsynlig at mikroarray teknikken kan anvendes som standard for risikovurdering af forarbejdede fødevarer. Teknikken forventes anvendt på levende organismer dvs. til udvikling og risikovurdering af nye plantesorter frem for til risikovurdering af de afledede (forarbejdede) produkter.

Eftersom mikroarray eksperimenter vil frembringe en stor mængde data, er det vigtigt at eksperimentet er godt designet og tager hensyn til a) tilgængelige eksperimentelle kontroller og reference prøver, b) statistiske overvejelser og c) praktiske begrænsninger såsom antal af tilgængelige mRNA prøver og arrays som kan håndteres samtidig.

Data analyse og data håndtering udgør den sidste, men meget vigtige fase i den eksperimentelle kæde. Generelt skal data opbevaring og analyse være overvejet før eksperimentets start for, at forbygge tab af informationer, at være i stand til at inkludere de rigtige referencer og at undgå at spilde tid på computer analyse med dårlige data. Data analyse foregår normalt ved trinvis at kvantificere signalet, normalisere data og visualisere og fortolke data. I de senere år er fremkommet en lang række forskellige software pakker til brug for disse opgaver. En ulempe ved computeranalyse af store mængder af data er, at output (uddata) kan være svært at fortolke i relation til potentielle sundhedsmæssige aspekter.

Små forskelle i det eksperimentelle setup kan ofte resultere i forskelle i genekspressions profiler som fanges i dataanalysen, men uden at de er relateret til de videnskabelige spørgsmål som er grundlaget for eksperimentet. Herudover skal det ikke glemmes, at analyse af genekspression ved mikroarray teknikken kun vil give indikationer om forskelle i genekspressions mønstre for individuelle gener og som måske er relateret til utilsigtede effekter som følge af den selektive forædlingsstrategi.

Kun i særlige tilfælde vil en statistisk analyse af et mikroarray resultat give et endeligt bevis for forskel i genekspression af et gen eller for en gruppe af gener. Normalt vil det efterfølgende være nødvendigt at analysere med andre metoder så som real-time PCR eller Northern blot, for at bekræfte observationerne. Da forskelle i DNA eller RNA i sig selv ikke er et sundhedsmæssigt problem, vil det endvidere være nødvendigt at undersøge den biologiske og toksikologiske relevans af disse forskelle, f.eks. ved måling på protein eller metabolit niveau. Denne nødvendighed er delvis begrundet i den manglende viden om relationerne mellem genekspression og ændringer i plantens fysiologi.

De første initiativer er taget for at harmonisere de kriterier for lagring af mikroarray data i globale tilgængelige databaser. Det er klart at et godt databasedesign er

afgørende for anvendeligheden af de store mængder af data som vil fremkomme af det voksende antal eksperimenter som foretages på verdensplan. Hvis samme (kommercielle) type af array bliver brugt af mange forskergrupper kan dette betragtes som en fælles eksperimental platform. Hvis de opnåede resultater fra en sådan array lagres i en centraliseret database, vil fremtidige computeranalyser af data måske kunne reducere antallet af eksperimenter, som er nødvendige for at opnå den ønskede viden. Sådanne initiativer er taget for et begrænset antal af modelplanter. Vigtige modelplanter for (funktionel) genomanalyse er *Arabidopsis thaliana*, ris og tomat.

Der er allerede foretaget et stort antal genekspressions eksperimenter baseret på mikroarrays og dette antal øges kraftigt. De undersøgte metabolismeveje varierer, men en stor del af eksperimenterne er rettet mod de stress relaterede synteseveje. Mikroarray studier har fx. afsløret uventede sammenhænge mellem stress og andre elicitorer (plantehormoner, modning) som inducerer ændring i genekspression i samme synteseveje. I forbindelse med risikovurdering af nye plante sorter, herunder genmodificerede planter, kan relevante synteseveje være de synteseveje som er ernæringsmæssigt relaterede eller synteseveje, som afspejler plantens basale fysiologi. Andre relevante synteseveje kan være de stress relaterede, da det vides at en opregulering af disse kan medføre til en forøgelse af de antinutritionelle stoffer herunder de naturlige toksiner. Det kan være relevant at lave en specifik DNA mikroarray til risikovurdering af fødevarer hvor arrayen indeholder prober for disse tre elementer.

Alternative "omics" omfatter proteomics, analyse af proteinerne, metabolomics, analyse af andre celle metabolitter, og måske glycomics som er analysen af suktermolekyler i cellerne. Proteomics kan betragtes som direkte afledet af transskriptom da proteinerne dannes fra mRNA. Da enhver indikation for ændret genekspression skal bekræftes med andre målinger, kan en bedre metode til måling af biologiske effekter være direkte at analysere for ændringer på proteinniveau ved brug af proteomics metoden. De nuværende metoder til proteom analyse er imidlertid begrænset til en mindre del af det totale proteom, og det vil være både vanskeligt og ressourcekrævende at opnå en bedre dækning af ændringer på proteinniveauet. Proteom mikroarray er under udvikling og baseret på interaktion mellem individuelle

proteiner og tilhørende substrater eller andre proteiner. Denne udvikling kan med tiden føre til en metode som stort set dækker alle proteiner og reducerer behovet for analyse af genekspression.

Metabolomet repræsenterer det højeste niveau af fysiologisk interaktion og består af alle alle metabolitter i den celle eller det væv som analyseres. Metabolom analyse dvs det ikke målrettede studie af metabolomet, kan være en vigtig tilføjelse til genekspression eller proteom analysen da den giver information om omfanget af de ændringer som følge af ændringerne i genekspression eller translation. Begrænsningerne med hensyn til metabolomics er de samme som for proteomics, at det kun er en mindre del af den totale indhold af metabolitter der kan analyseres for i en enkelt analyse.

Glycomics er sidste nye skud på "omics" stammen og er formodentlig den mest komplicerede da glycosylerings mønsteret er under konstant ændring. I fremtiden kan det ikke udelukkes at glycomics kan bidrage til den hypotese motiverede risikovurdering af nye (GM) plantesorter.

Generelt kan det konkluderes at introduktionen af gensplejsede planter har øget interessen for sammensætningen af indholdsstoffer i vore afgrødeplanter. Profil metoder har potentialet til at analysere effektivt for forskelle i nye plantesorter, idet de (teoretisk) dækker meget bredere de forskellige synteseveje. DNA mikroarray teknologien har den store fordel, at arrays dækkende hele genomer snart vil være tilgængelig for de første afgrødeplanter. Som følge heraf kan det forudses, at genekspressions profiler kan forstærke og forbedre de nuværende risikovurderings strategier for nye plantesorter. Dette er sandsynligvis kun et midlertidigt fænomen indtil hele proteom eller metabolom profiler bliver tilgængelige. Endelig vil en direkte selektion for sundhedsrelaterede relevante forskelle i nye plantesorter, herunder genmodificerede planter, først være mulig med bedre indsigt i planternes fysiologi og deres relation til sundhedsmæssige aspekter for forbrugerne.

Yhteenveto

Ennen vuotta 1997 ei Euroopan Unionin alueella ollut yhteisiä pelisääntöjä 'uuskasvilajikkeista' peräisin olevien elintarvikkeiden turvallisuuden arviointiin. Tällöin astui voimaan Euroopan parlamentin ja neuvoston asetus uuselintarvikkeista ja elintarvikkeiden uusista ainesosista. Tämä asetus, jota usein kutsutaan uuselintarvikeasetukseksi, kattaa myös geneettisesti muunnettuja organismeja (GMO) sisältävät tai niistä koostuvat tai valmistetut elintarvikkeet ja niiden ainesosat. Uuden Euroopan parlamentin ja neuvoston asetuksen geneettisesti muunnetuista elintarvikkeista ja rehuista arvioidaan astuvan voimaan alkuvuodesta 2004. Se tulee korvaamaan uuselintarvikeasetuksen niiltä osin kuin tämä koskee GMO:ja ja niistä valmistettuja elintarvikkeita

(http://europa.eu.int/comm/food/fs/biotech/biotech08_en.pdf). Uuden asetuksen mukaan Euroopan elintarviketurvallisuusviranomainen suorittaa tieteellisen riskinarvioinnin; tämä tapahtunee niiden suuntaviivojen mukaisesti, jotka on kirjattu ohjeistoon geneettisesti muunnettujen (GM) kasvien ja niistä valmistettujen elintarvikkeiden ja rehujen riskinarvioinnista (http://europa.eu.int/comm/food/fs/sc/ssc/out327_en.pdf).

Kaikkien uusien GMO-peräisten kasvituotteiden turvallisuuden arvioinnin erään tärkeän alkuvaiheen muodostaa yksityiskohtainen vertaileva koostumusanalyysi, jossa verrataan tärkeimpiä ravintoaineita ja ravintoarvoa heikentäviä aineita (antinutrientteja) uudessa kasvilajikkeessa ja, mikäli mahdollista, sen jo markkinoilla olevissa 'perinteisissä' verrokeissa. Vaikka tätä lähestymistapaa on pidetty sopivana GM-kasvien turvallisuuden arviointiin, on toisaalta epäilty, tuoko tämä täsmämenettely esille kaikki mahdolliset tahattomat muutokset, joilla saattaa olla vaikutusta kuluttajien terveyteen. Tähän liittyen on virinnyt nopeasti kiinnostus kehittää runsaasti tietoa antavia yleismenetelmiä koostumuserojen seulontaan ja niiden toksikologisen merkityksen arviointiin. Nyt onkin katsottu olevan korkea aika arvioida näitä uusia analyttisiä lähestymistapoja, jotka saattavat antaa paremman kuvan mahdollisista kasvin fysiologisista muutoksista kuin nykyinen lähestymistapa, jossa analysoidaan rajallinen joukko tärkeimpiä yhdisteitä.

Näillä ns. tasapuolisilla menetelmillä on mahdollista kattaa poikkeavuuksia paljon suuremmassa joukossa aineenvaihduntareittejä kuin täsmäanalyseillä. Yleismenetelmät viittaavat tavallisesti transkriptomiin (kaikki solun transkriptio- eli lähetti-RNA-tuotteet), proteomiin (kaikki solun proteiinit) tai metabolomiin (kaikki solun aineenvaihduntatuotteet eli metaboliitit). Viime aikoina on jopa glykomi (kaikki solun sokerit) otettu harkittavaksi erityisten muutosten havaitsemisessa. Kaikille edellämainituille solun komponenteille on nykyisin olemassa enemmän tai vähemmän kehittyneet yleismenetelmät, ns. profilointimenetelmät. Yksi lupaavimmista kehityksistä on tapahtunut DNA-mikrosirujen kohdalla. Jo ennen mikrosiruteknologian keksimistäkin oli olemassa menetelmiä geenien ilmentymisessä esiintyvien erojen seulontaan, mutta niillä pystyttiin analysoimaan kiinnostavasta lähetti-RNA-joukosta yhdellä kertaa vain pieni joukko hyödyllisiä tunnistimia. Mikrosiruteknologiassa on käytössä vakiintuneet molekyylibiologiset ja biokemialliset periaatteet miniatyyrimuodossa, jolloin pystytään tutkimaan yksittäisten lähetti-RNA-näytteiden ilmentymisprofiilia tuhansilla tunnistimilla ja näin analysoimaan lukuisia aineenvaihduntareittejä. Todellisuudessa lähetti-RNA-tasot eivät välttämättä korreloi kovinkaan hyvin vastaavan proteiinin määrän tai biologisen toiminnan kanssa. Tästä huolimatta geenien ilmentymistä osoittava mikrosiru saattaa olla lähitulevaisuudessa kehittynein systeemi kasvien aineenvaihdunnassa tapahtuneiden muutosten tasapuoliseen seulontaan.

Aluksi oli saatavilla kahdentyyppisiä siruja: oligonukleotidisirut, joissa oligonukleotidit tuotettiin kiinteälle alustalle, ja komplementaarinen-DNA- tai oligonukleotidisirut, joissa etukäteen tuotetuista tunnistimista tehtiin mikropisarot lasilevyille. Viime aikoina on kehitetty useita vaihtoehtoisia sirusysteemejä, jotka joko lisäävät systeemin herkkyyttä, lyhentävät hybridisaatioaikaa tai tulevat toimeen halvemmilla laitteistoilla.

Mikrosirukokeissa hybridisoitavien lähetti-RNA-populaatioiden laatu on keskeinen tekijä tulosten luotettavuuden kannalta. Jotta kahta peräkkäistä näytettä voitaisiin verrata toisiinsa, on lähetti-RNA:n laadun oltava huippuluokkaa. Siksi on välttämätöntä vakioida näytteenotto- ja lähetti-RNA:n eristysmenetelmät sekä laadunvarmennusmenettely. Koska hieman erilaisilla olosuhteilla on yleensä suuri vaikutus elintarvikkeista eristetyn lähetti-RNA:n laatuun, ei ole todennäköistä, että mikrosiruteknologiasta tulee rutiinimenetelmää prosessoitujen tuotteiden

turvallisuuden arviointiin. Teknologian pääsovellusalue tulee olemaan elävä eliö: uusien kasvilajikkeiden, eikä niinkään niistä saatujen tuotteiden, kehittäminen ja turvallisuuden arviointi.

Koska jokaisesta mikrosirukokeesta saadaan suuri määrä tietoa, on erittäin tärkeää, että koe on hyvin suunniteltu, ja huomioi a) käytettävissä olevat koe-, verrokki- ja vakiointinäytteet, b) tilastolliset vaatimukset ja c) käytännön rajoitukset, kuten lähetti-RNA-näytteiden lukumäärän ja mikrosirusarjan, joka pystytään yhdellä kertaa käsittelemään.

Tietojen analysointi ja louhinta muodostavat viimeisen, mutta tärkeän vaiheen kokeessa. Yleisesti ottaen tietojen säilytys ja analysointi pitäisi miettiä jo ennen kuin koetta aloitetaan, jotta välttyttäisiin tietojen häviämiseltä ja voitaisiin ottaa mukaan oikeat vakiointinäytteet eikä hukattaisi jälkikäteen aikaa suunnittelemattomaan tietojenkäsittelyyn. Tietoja analysoitaessa selvitetään yleensä signaalin suuruus, normalisoidaan saadut tulokset, sekä havainnollistetaan ja tulkitaan tulokset. Viime vuosina on tullut markkinoille suuri joukko ohjelmistoja, joiden tarkoituksena on helpottaa jotakin tai kaikkia edellämainittuja vaiheita. Yksi vaikeus suurten tiedostojen käsittelyssä on, että lopputulosta saattaa olla vaikea tulkita erityisesti mahdollisten terveyteen liittyvien seikkojen osalta. Pienet erot koeasetelmassa aiheuttavat usein sellaisia eroja geenien ilmentymisprofiileissa, jotka voidaan havaita ohjelmistoilla, mutta jotka eivät välttämättä liity kokeen tieteelliseen kysymyksenasetteluun.

Lisäksi tulisi aina pitää mielessä, että geenien ilmentymisen analysointi mirosirutekniikalla on ainoastaan viitteellinen antaessaan tietoa yksittäisten geenien erilaisesta ilmentymistavasta, joka kenties viittaa tahattomiin, valitun kasvinjalostusstrategian tuloksena syntyneisiin vaikutuksiin. Mikrosirutulosten tilastollinen analyysi antaa vain aivan poikkeustapauksessa varman todisteen geenin tai geeniryhmän erilaisesta ilmentymisestä. Normaalityapauksessa havainnot pitää varmistaa muilla tekniikoilla, kuten reaaliaika-PCR:llä tai Northern-blottauksella. Koska erot DNA:ssa tai RNA:ssa eivät suoranaisesti liity turvallisuuteen, on tulos biologisen ja toksikologisen merkityksen testaamiseksi lisäksi vahvistettava muuntotyypisin kokein, eli esim. proteiini- tai metaboliittitasolla. Tämä tarve johtuu osittain siitä, että ei tiedetä riittävästi geenien ilmentymisen suhteesta kasvin fysiologisiin muutoksiin.

Ensimmäiset yritykset harmonisoida kriteerit mikrosirutietojen säilömiseksi kaikkialta saatavissa oleviin tietokantoihin on tehty. On selvää, että hyvä tietokannan suunnittelu on välttämätöntä, jotta pystytään tehokkaasti käyttämään hyödyksi sitä valtavaa tietomäärää, joka syntyy maailmanlaajuisesti yhä kasvavassa mikrosiruanalyysien määrässä. Jos useat tutkimusryhmät käyttävät samaa (kaupallista) mikrosirua, tämä voisi periaatteessa toimia yhteisenä kokeen perustana. Jos tällä sirulla saadut tiedot varastoidaan keskitettyyn tietokantaan, voivat käytettävissä olevien tietojen *in silico* (puhtaasti tietokoneella tehdyt) analyysit tulevaisuudessa vähentää halutun informaation saamiseen tarvittavien kokeiden määrää.

Näitä kansainvälisiä aloitteita on käynnistymässä tietyillä mallikasveilla. Kasvien (funktionaalisen) genomiikan tärkeitä mallisysteemejä ovat lituruoho (*Arabidopsis thaliana*), riisi ja tomaatti. Geenien ilmentymiskokeita mikrosiruilla on tehty kasveilla jo huomattavan paljon, ja määrä kasvaa nopeasti. Mielenkiinto näissä kokeissa on kohdistunut monenlaisiin aineenvaihduntareitteihin, mutta merkittävä osa tähtää stressiin kytkeytyvien reittien selvittämiseen. Mikrosirututkimuksilla on esim. saatu selville odottamattomia yhtymäkohtia stressin ja muiden ärsykkeiden (kasvihormoni, kypsyminen) välillä, jotka indusoivat muutoksia samojen reittien geenien ilmentymisessä.

Uusien kasvilajikkeiden, mukaanlukien GM-kasvit, turvallisuuden arvioinnissa voivat olla tärkeitä aineenvaihduntareitit, jotka liittyvät ravintoaineisiin tai kasvin perusfysiologiaan. Muita kiinnostavia aineenvaihduntareittejä saattavat olla stressiin liittyvät reitit, koska tiedetään, että näiden reittien voimistuminen voi johtaa ravintoarvoa heikentävien aineiden, esim. luontaisten toksiinien, tuotannon lisääntymiseen. Saattaisikin olla järkevää kehittää erityinen elintarvikkeiden turvallisuuden arviointiin tähtäävä siru, jossa olisi tunnistimia näille kaikille kolmelle kategorialle.

Muita mahdollisesti käyttökelpoisia 'omiikoita' ovat proteomiikka eli proteiinien analysointi, metabolomiikka eli metaboliittien analysointi, ja ehkä glykomiikka eli sokerien analysointi sopivista solukoista. Proteomiikka liittyy läheisesti transkriptomiikkaan eli RNA-mikrosiruihin. Koska geenien ilmentymisessä havaittavat

erot joudutaan yleensä varmistamaan muilla menetelmillä, on paras tapa todeta biologinen vaikutus tutkimalla, näkykö vastaavia muutoksia proteomiprofiilissa. Tällä hetkellä proteomin analysointi on rajoittunut yleensä pieneen osaan proteomia, ja on vaikeaa ja työlästä saada lisätietoa mahdollisista muutoksista koko proteomissa. Proteomeille on myös kehitteillä mikrosiruja, jotka perustuvat yksittäisten proteiinien vuorovaikutuksiin substraattinsa tai toisten proteiinien kanssa. Tämä kehitys voi johtaa aikanaan mahdollisuuteen tutkia 'koko proteomia', jolloin tarve tutkia ensin geenien ilmentymisprofiilia vähenee.

Metabolomi koostuu tutkittavan solukon kaikista metaboliiteista ja edustaa korkeinta fysiologisten vuorovaikutusten tasoa. Metabolomiikka, eli metabolomin (kaikkien aineenvaihduntatuotteiden) tasapuolinen tutkiminen, voi olla arvokas lisä geenien ilmentymisen ja proteomin profilointiin, koska sen avulla saadaan tietoa muuttuneesta aineenvaihdunnasta, joka on seurausta geenien ilmentymisen tai proteiinien tuotannon muutoksista. Nykyisessä metabolomiikassa on sama heikkous kuin proteomiikassa, eli vain pieni osa metaboliiteista voidaan tunnistaa yhdessä analyysissä.

Glykomiikka on 'omiikoiden' uusin, ja luultavasti kaikkein monimutkaisin tulokas, koska glykosylaatiopatteri muuttuu jatkuvasti. Tulevaisuudessa tästä menetelmästä voi kuitenkin myös tulla arvokas lisä uusien (GM) kasvilajikkeiden hypoteeseihin perustuviin riskinarviointistrategioihin.

Yleisenä johtopäätöksenä voidaan todeta, että GM-kasvilajikkeiden kehittämisen myötä kiinnostus viljelykasvien koostumukseen on kasvanut nopeasti. Profilointimenetelmillä on mahdollista seuloa tehokkaammin eroja uusissa kasvilajikkeissa, koska näillä menetelmillä voidaan teoriassa kattaa entistä huomattavasti laajempi aineenvaihduntareittien joukko. Mikrosirutekniikan päätetu on, että koko genomin kattava geenien ilmentymissiru on lähitulevaisuudessa saatavilla ensimmäisille viljelykasveille. Niinpä onkin nähtävissä, että geenien ilmentymisen profiloinnilla voidaan tarkentaa nykyistä täsmäanalyysiin perustuvaa uusien kasvilajikkeiden turvallisuuden arviointistrategiaa. Tämä tulee kuitenkin luultavasti olemaan vain ohimenevä kehitysvaihe, kunnes saadaan käyttöön profilointimenetelmät koko proteomille ja metabolomille. Lopuksi on todettava, että

uusien kasvilajikkeiden, olivatpa ne geneettisesti muunnettuja tai eivät, terveellisyyteen liittyvien erojen suora toteaminen on mahdollista vain, kun kasvin fysiologia ja sen suhde kuluttajan terveyteen ymmärretään kokonaisuudessaan.

Chapter 1 Introduction and scope of the report

Until 1997 there were no general European guidelines for the safety assessment of newly developed plant varieties. In 1997 the European Regulation on Novel Foods and Novel Food Ingredients came into force (EU, 1997a). This Novel Foods Regulation, as it is generally called, comprises six different novel food categories, three of them relating to genetically modified organisms (GMOs) and one related to novel plant products that are not GMO-derived. This novel plant product category refers to all plant products that were not as such on the European market prior to the Novel Foods Regulation coming into force. Therefore, in principle, all new plant varieties or products derived thereof that were not yet on the EU market or traditional plant products that have been altered substantially, may be considered novel food products under this Regulation. The consequence of this may be that novel plant products will have to comply with the criteria as set out in the Novel Foods Regulation, meaning that the producer will have to compile a dossier to establish the safety of the plant products to be marketed. The information required for such a dossier was further detailed in the Commission Recommendation with relation to the Novel Foods Regulation (EU, 1997b).

This situation has recently been changed. On 22 July 2003 the Regulation on Genetically Modified Food and Feed has been adopted, replacing the part of the Novel Foods Regulation dealing with GMOs and GMO-derived food products. According to the new regulation the evaluation of new GMOs and GMO-derived products for food and feed purposes will be performed by the European Food Safety Authority along the lines of the EU guidance document for the risk assessment of GM plants and derived food and feed.

In all cases a thorough compositional analysis of novel plant products, whether GMO-derived or not, will be part of the food safety evaluation. As a result of these regulatory developments the interest for the development of informative methodologies to assess compositional differences has rapidly increased in the last few years. In former days, prior to the Novel Foods Regulation, limited methods of analysis were available for individual plant constituent compounds, e.g. for specific macro- or micronutrients and to a lesser extent for anti-nutrient compounds, such as natural toxins, and natural

variation in these compounds was documented in a rather fragmentary way. The regulatory developments have boosted research in this area of determining levels of food constituent compounds with relation to the safety and nutritional value of individual plant products, culminating in a.o. the OECD consensus documents with information on key substances for a number of economically important food crops (OECD, 2003).

Random insertion of gene(s) can inactivate or change expression of gene(s) already expressed, or could, theoretically, activate normally silent gene(s). International advisory bodies have therefore discussed since the early 1990s the development and evaluation of additional unbiased non-targeted methodologies of analysis that may supply additional information on such unpredictable, unintended changes in the plants' physiology that may remain undetected using targeted methods of analysis (OECD, 1993; OECD, 1996; OECD, 1998; FAO/WHO, 2000).

The non-targeted approach may relate to the transcriptome, the proteome, the metabolome, or even the glycome. For all of these cell constituents more or less advanced experimental profiling methodologies have been developed (Kuiper et al., 2002, Kuiper et al., 2003). One of the most promising developments to this end is the miniaturisation of well-established molecular biological and biochemical principles in the microarray technology. This technology makes it feasible to study differences in gene expression, in the transcriptome, on a scale and with a resolving power not seen before in the area of molecular analysis. The microarray technology enables the screening of significant numbers of different tissues and plant parts for potential unintended side effects in a.o. gene expression as a result of the breeding strategy. In this way known tissue-specific metabolic networks can be investigated for alterations, but also activation of normally silenced pathways in specific tissues or plant varieties would not remain undetected.

This report describes the information on gene expression profiles in plant products as obtained with traditional approaches and how the advent of the microarray technology has accelerated our insight into the gene expression in plant products that may have direct consequences to the way plant products for our food supply are evaluated. The report focuses on the different aspects of the application of the microarray technology

for the study of gene expression and the bottlenecks related to the large-scale application of the microarray technology in this field. Another interesting area of research is the use of the microarray technology to identify plant varieties and products derived thereof for the purpose of traceability. This field of research is based on the profiling of the genomic DNA of individual organisms rather than the mRNA as is the target in gene expression profiling. As the questions underlying this field of research are only partly related to the characterisation of plant products with relation to their safety assessment, the subject of this report, the DNA-profiling for identification is considered to be outside the scope of this report . The aim of this report is to discuss and evaluate the usefulness of the microarray technique in the safety assessment of genetically modified plants and other novel plants as well as the usefulness and possibility for making a database containing profiles of food plants. This includes a description of the state of the art, the technical challenges and a discussion of a more targeted microarray and the nature of the probes that may be relevant for such an array. Although the focus of the report is on GM plants, it is clear that the knowledge gained by the profiling of this specific group of novel plant varieties and their traditional counterparts may be helpful in the food safety analysis of any novel plant variety in the future.

Chapter 2 Characterisation of food plants on the RNA level

2.1 Comparative approach

The safety of biotechnology-derived crops is commonly assessed prior to their introduction into the field and the market. A consensus approach has been developed in international fora organised by a.o. the United Nations' Food and Agriculture Organisation and World Health Organisation (FAO/WHO, 2000; <http://www.fao.org/english/newsroom/news/2003/20363-en.html>), and the Organisation for Economic Co-operation and Development (OECD, 1998).

The generally acknowledged approach for assessing food safety of GM plants is based on a comparative approach which focuses on the determination of similarities and differences between the GM plant and its conventional counterpart. This comparison usually comprises phenotype, field behaviour, and composition of the crops under investigation. Preferably data are used from multiple locations and seasons, to account for environmental, climatological, and geographical influences. The assessment of the introduced genes and their products is usually a starting-point for the food safety evaluation.

FAO/WHO (2000), OECD (1996) and EU (1997) have recommended that comparative compositional analysis should be carried out on compounds (nutrients, anti-nutrients, and toxins) that are characteristic for the crop under study. For this purpose, the OECD Task Force on the Safety of Novel Foods and Feed has compiled consensus documents describing which crop components are recommended for comparative analysis as well as background levels of these components (table 1). At present, consensus documents on soybean, canola, maize, sugar beet, potato and wheat have been published, while additional documents on cotton, rice, sunflower, barley, forage legumes, and tomato are in preparation (OECD, 2003). Another very comprehensive information system exist on critically assessed compositional data on bioactive constituents (toxicants and health-protective compounds) in food plants and edible mushrooms. The database contains 300 major food plants and mushrooms on the European market and is based on national databases and data obtained from the EU NETTOX (NETTOX 1998) and BASIS projects (Gry *et al.*, 2002; <http://www.vfd2.dk/basis/>)

In addition, the ILSI / International Food Biotechnology Committee (IFBiC) is currently developing a comprehensive database on crop composition that will be accessible through the Internet. The data incorporated in the database will be checked for their quality with respect to field trial design, sample handling, and analytical methodologies used. A first compilation of maize data has been made, which contains data on 95 analytes analyzed in samples from company field trials that were carried out in 1995-2000 in US, EU, and Argentina (<http://www.cropcomposition.org>).

Crop	Product	Antinutrient / toxin
Maize	kernel	phytic acid raffinose furfural ferulic acid <i>p</i> -coumaric acid
Potato	tuber	glycoalkaloids allergens (patatin, soybean trypsin inhibitor) protease inhibitors lectins
Soybean	seed	phytic acid trypsin inhibitor lectins isoflavones
Oilseed rape	Oil	erucic acid
Sugar beet		--

Table1. Antinutrients and toxins recommended by OECD for analysis in food.

The larger part of the data on comparisons between GM varieties and their nearest comparator is dossier information and not freely accessible for scientific scrutiny. However, a number of comparative analyses have been published in the open literature. Examples of comparisons are reported in a 1998 Nordic report with data on 14 cases of GM plants evaluated within Europe (Nordic Council, 1998). An overview of literature on the composition of genetically modified crop plants in comparison to

traditional varieties until 2000 is provided by Kuiper et al. (2001). From this overview it can be seen that most studies are on canola and maize with a limited number of studies on potato, rice, soybean, squash, sugar beet and tomato. Most introduced gene sequences code for either herbicide or insect resistance. Kuiper et al. (2000) includes an inventory of publicly available reports on unintended effects in different GM plants (canola, potato, rice, soybean and wheat), that shows that these effects are diverse and can not in all cases be directly related to the function of the inserted genes.

In addition, several scientific reports provide examples of the comparative compositional analysis of commercial biotechnology-derived crops, such as soybeans (Nair et al., 2002) cotton (Nida et al., 1996), maize (Ridley et al., 2002; Sidhu et al., 2000), and potato (Rogan et al., 2000). Regierer et al. (2002) describe transgenic potato plants where the activity of plastidic adenylate kinase was modulated. In the resulting GM plants a substantial increase in the level of adenylates was observed as well as a significant increase in the starch level to 60% above that found in wild-type plants and in the concentration of several amino acids. In this case the changes could be linked to the modification although the exact mechanism remains to be elucidated.

In general, it is, however, difficult to compare different studies that are performed outside the more strict regulatory dossier requirements, as the set-up and analysis of individual comparative experiments may differ considerably. Information on differences in gene expression in GM plant varieties compared to either the direct parent line or any other relevant comparator is thus far limited as the generation of gene expression profiles is not part of the food safety evaluation procedure (yet).

2.2.Traditional methods for the detection of differences in gene expression.

'Traditional' methodologies to detect differences in gene expression include a.o. Northern blotting (Alwine et al., 1977), where mRNA populations are arranged by size, transferred to a filter and hybridised with the probe of interest to investigate the gene expression levels of this particular target. Arbitrarily primed PCR (Welsh and

McClelland, 1990; Welsh et al., 1992) and differential display (Liang and Pardee, 1992) are methodologies to amplify specific subsets of the mRNA populations under investigation for the purpose of comparison. Differential display has also been used to detect altered gene expression in genetically modified plant varieties compared to their traditional counterpart (Kok et al., 1998; Kok et al., 2001). Differential display will result in the amplification of specific subsets of mRNAs depending on the primer design. It was shown that the method is rather labour intensive, has to be repeated multiple times with different primer combinations in order to be sufficiently informative and is not very well suited for routine application to detect altered gene expression as part of a risk assessment strategy. Other approaches are SAGE (serial analysis of gene expression, Velculescu et al., 1995), where cloned concatamerised mRNA-fragments are sequenced to generate a profile of the mRNA population of interest, and the recently developed MPSS (massively parallel signature sequencing) technology. Both SAGE and MPSS make use of the combination of cloning cDNA fragments and subsequent sequencing. SAGE uses conventional cloning and sequencing strategies, whereas MPSS attaches digested, tagged polyA-cDNA fragments to beads that are coated with many copies of one out of millions of different 32-base oligonucleotides, prior to sequencing of the attached molecules. The advantage of this approach over the microarray approach is that it may be even more unbiased: gene expression using arrays will only provide information on the sequences printed on the array, whereas MPSS can give information on the entire transcriptome. MPSS is, however, less feasible as a high throughput system and can be regarded as complementary to array systems, as long as whole transcriptome arrays are still limitedly available, rather than as an alternative (Constans, 2002; Brenner et al., 2000).

In general, it can be stated that only in the last decade tools have become available to analyse individual plant varieties or derived tissues for differences in gene expression. Initial methodologies only allowed the analysis of specific subsets, whereas now the microarray technology has made it feasible to analyse the entire transcriptome. With this complete of set tools it is possible to select the right tool for any individual scientific question. For the purpose of the initial detection of unintended side effects of a genetic modification the microarray technology has the obvious advantage of enabling the screening of multiple metabolic routes and, possibly, the entire

transcriptome. For confirmation of detected differences, however, it will in general be advisable to more specifically analyse a single or a limited number of transcript(s).

2.3 Stress-related differences in gene expression

Until recent years there was, as mentioned before, a remarkable lack of interest in the composition of plants in general, including differences in gene expression. Plant breeding strategies were primarily based on other phenotypic aspects. Information on gene expression profiles in crop plant varieties is therefore only very limitedly available in the scientific literature. An exceptional area in this respect may be the stress-related physiological reactions within homeostasis. The issue of stress-related reactions in plants is of major importance for agriculture, given the yield-reducing impact of various stress conditions (drought, cold, saline, light, pathogens), and has therefore been in the focus of agricultural research for a number of years.

Several reviews highlight the state-of-the-art knowledge on gene expression influenced by stress in plants, e.g., cold temperatures and freezing (Thomashow et al., 2001; Ouellet, 2002), salt and drought (Zhu, 2002), as well as UV-B light (Jordan, 2002). For example, COR (cold-regulated)- and DRE (dehydration-responsive)- genes have been identified as stress-related and their activation has been associated with increased stress tolerance, such as cold acclimatisation (Thomashow et al., 2001). Besides gene expression, post-transcriptional and post-translational modifications of transcripts and proteins, respectively, have also been linked to stress tolerance (Jordan, 2002; Ouellet, 2002; Zhu, 2002).

Chapter 3 Principles of the cDNA microchip technology

3.1 The microarray technology

The cDNA microchip or microarray technology is based on miniaturisation of methodologies that were already available in molecular biology and biochemistry. The technology is based on the same principles for the specific binding of molecules that are recognised in a larger pool. For cDNA microarrays the principle is hybridisation, i.e. the specific binding of two DNA fragments that have complementary nucleotide sequences. This principle of hybridisation has formerly been used in methods for the detection of altered gene expression such as Northern or dot blotting, differential display and arbitrarily primed PCR, but blotting methods could only analyse single transcripts and PCR-based methods can analyse at the most limited subsets of the mRNA population in individual experiments.

With the advent of the microarray technology it has become feasible to study the larger part of the mRNA population in a single experiment. Where in classical gene expression studies the mRNA population was immobilised on a filter prior to hybridisation with a single labelled probe, the microarray technology made it possible to immobilise thousands of individual probes and screen series of fluorescently labelled mRNA populations of interest against those arrayed probes. The technology is semiquantitative, i.e. the fluorescent signals can be quantified and relative values between different slides can be calculated.

The spotted probes can be of different origins. They can be either newly synthesised oligonucleotides or pre-amplified fragments from sequence-identified or anonymous cDNA or EST (expressed sequence tag) libraries of any species of interest. EST fragments are often obtained using primers with a 5'-amino modification, whereas the oligonucleotides are directly synthesized with a similar modification in order to allow covalent binding to, for instance, free aldehyde groups on silylated glass slides. Oligonucleotide probes are often extended with a (oligo-dT) spacer fragment to avoid steric hindrance of the small molecules on the glass surface. Free reactive surface

groups (epoxy groups may be used instead of aldehyde groups) on the surface are subsequently chemically inactivated to avoid aspecific binding of DNA fragments. Franssen- van Hal et al. (2002) showed, however, that an effect of this amino-linker due to enhanced cross-linking to the silylated slides could only be observed with short PCR products (< 200 bp). Moreover, it was shown that the linker should preferably be present on the coding strand. It was also established that single-stranded DNA instead of double-stranded DNA will be more sensitive mainly due to the fact that more coding strands with the capacity to bind the labelled cDNA will be present in the spot. Finally, it was shown that redundancy in specific gene fragments in different spots should be avoided in order not to dilute the signal of the specific transcript product.

Following the fact that on microarrays the probes are immobilised, it is clear that the immobilised probe set determines the information that can be obtained using a specific array, whether this is a oligonucleotide or a cDNA array. It is therefore crucial to use array systems that will be able to provide the information that is sought after when designing an array experiment. This is less obvious than it may seem at first. Especially in the area of cDNA or EST arrays there are no arrays (yet) available that cover entire transcriptomes of individual (crop) plant species. The arrays that are available (table 2) contain ESTs or cDNAs of specific tissues of the plant in specific developmental stages. The information related to the spotted probes should be analysed carefully before designing experiments that aim to elucidate metabolic pathways in specific plant tissues. If the tissues were not included in the setting up of the libraries underlying the spotted probe set, it may well be that the information that can be obtained from the array is insufficient and it would be advisable to follow other routes (as well) to meet the aims of the project.

As the microarray technology is based on hybridisation of complementary strands there is the possibility that other mRNAs with a high level of homology with the probe, e.g. different members of the same gene family, will crosshybridise. It may be very difficult to distinguish similar mRNAs, especially if they function in the same metabolic pathways. This is also the case for fusion proteins as a result of the genetic modification: the derived mRNA may hybridise in the same way compared in the wild type if the homologous part is of sufficient length. Related effects to the plant's

physiology may, however, be detected. It is estimated that sequences with over 70% sequence homology over more than 200 nucleotides are likely to crosshybridise to some extent under standard conditions (Kane, 2000; Xu et al.; Richmond and Somerville, 2000).

Plant	Array description	Institute/company	Academic/industrial
Arabidopsis Thaliana	>24.000 genes, 25mer oligos, 11 oligos per gene	Affymetrix	Industrial
Arabidopsis Thaliana	21.500 genes, 60mer oligos	Agilent	Industrial
Arabidopsis Thaliana	26.090 genes, 70mer oligos	Qiagen	Industrial
Barley	53.030 genes, 25mer oligos, 11 oligos per gene	Affymetrix	Industrial
Maize	4.800-5.700 unique cDNA clones	Iowa State University & Arizona University	Academic
Potato	10.000 unique cDNA clones	TIGR	Academic
Rice, in combination with a rice blast fungus	13.500 fungus genes (complete genome), 7.000 rice cDNAs, 60mer oligos	Agilent	industrial
Tomato	12.000 ESTs	Cornell University, USA	Academic

Table 2. Commercially available plant microarrays

It is also essential to have some insight into the kinetics of hybridisation events in order to be able to optimally interpret the resulting data from an hybridisation experiment.

Richmond and Somerville (2000) distinguish three types of microarray experiments: 1) marker discovery experiments, to discover a limited number of highly specific marker genes for a specific cell type, stage of developmental or environmental condition; 2) biology discovery experiments where expression information on all genes provides an integrative view of gene expression under the selected conditions; and 3) gene-function discovery experiments which may be *in silico* experiments to use existing data to find additional information on the function of a specific gene on the basis of the expression profile under a range of different conditions. For the purpose of the safety assessment of GMO-derived plant products microarray experiments type 1 and 3 will likely be most relevant. Type 1 can either be based on hybridisations to a whole transcriptome array or to a focused array to detect unintended effects of a genetic modification on gene expression profiles. Type 3 experiments may be informative to assess the biological relevance of detected differences by comparing the results to similar experiments in comparable varieties. All type of experiments in the end determine ratios of expression for the individual genes spotted under the selected conditions in order to gain insight into physiological processes that were hitherto part of plant 'black box'.

The DNA microarray approach to analyse metabolic pathways in plant physiology has the potential to elucidate complex pathways that could never be revealed with the former single or limited gene approaches. The potential to this effect was first shown in the medical and microbiological area (Alizadeh et al., 2000), but has already shown its value also in the plant and food sciences (Aharoni et al., 2002; van Hal et al., 2000; Schaffer et al., 2000; Schenk et al., 2000, Maleck, 2000; Aharoni and Vorst, 2002). This will not only allow breeders to develop new plant varieties with improved characteristics by means of genetic modification, but may also improve the possibilities of marker assisted selection (MAS), early in the plant breeding schemes. This may be especially beneficial in the case of 'slow' growing plants and trees, thereby speeding up the breeding programmes for this type of crop species considerably (Schaffer et al., 2000). Other recent developments in DNA microarray technology, but outside the direct scope of this report, is the arraying of a gene expression cell system in order to rapidly identify gene products and their function (Ziauddin and Sabatini, 2001) and TSAA, translational state array analysis, a mRNA micro-array technique that discerns translated from non-translated mRNA based on

centrifugal separation of translated mRNA bound by many ribosomes and from poorly translated mRNA bound by few ribosomes in density gradients (Serikawa et al., 2003).

Finally, it should always be kept in mind that gene expression analysis by oligonucleotide or cDNA/EST arrays will in general only give indications on specific genes being related to unintended effects in specific pathways/ responses/ developmental stages etc. Only in very specific cases will statistical analysis of the microarray results give full evidence for differential gene expression of an individual or cluster of genes. Usually additional techniques will be required to confirm the observations by e.g. real time PCR or Northern blotting experiments and test the biological relevance of the measured differences by investigating the effects on protein or metabolite level. Final confirmation of the function of a not yet annotated gene may require other approaches, such as (bio)chemical analyses or the creation of (knock-out) mutants for the specific gene under investigation.

3.2 Technological developments

The microarray technology is still very much in development. Initially only two types of arrays were available. First of all the arrays where oligonucleotides are synthesised on a solid support, especially the Affymetrix system. By this in-situ synthesis very high densities of the oligonucleotides can be achieved, up to 250.000 spots/cm². For individual cDNAs 15 to 25 matching and (single-)mismatching oligo's are spotted to confirm any detected sequence (Lockhart and Winzeler, 2000). The second often used type of array is the array where either oligonucleotide or cDNA sequences are presynthesized pre-amplified and then spotted in array format. The oligonucleotides spotted in this way may be longer compared to the Affymetrix array, but usually do not exceed 70 bases.

There are, however, other developments in the area of the microarray technology, e.g.:

- Three-dimensional array systems, e.g. PAMgene arraysystem where DNA probes are attached inside the tunnels of a microporous material. Sample cDNA is pumped up and down through this material several times to facilitate the

hybridisation process. cDNA will hybridise to the probe and can be detected accordingly (<http://www.pamgene.com/index-ie.html>).

- Signal enhanced fluorescence arrays: the system is based on metal nanocluster layers between the solid surface and the cDNA probes. The bound fluorescent cDNA molecules are at such a distance of the nanocluster that the fluorescence is enhanced up to 200 times (Stich et al., 2001).
- Nanosphere technology: DNA fragments are hybridised to immobilised probes on the solid surface. In a second hybridisation step specific oligonucleotide probes that are linked to minuscule gold particles hybridise to the bound DNA fragments. The bound gold particles are then visualised through a reaction with silveriodide. A recent variant on this method uses the conductivity of the silver and gold molecules to generate an electric signal when the electrodes surrounding the spots are connected. The generated signal is measured and this approach increases the sensitivity of the approach significantly (Park et al., 2002).
- Suspension array technology: internally labelled fluorescent microspheres that are coated with oligomeres can bind labelled DNA target fragments. A mix of microspheres is allowed to hybridise with sample DNA and subsequently analysed by flow cytometry or optic fiber cables (Nolan and Sklar, 2002).
- Bead array counter (BARC): target DNA binds to DNA probes coated on magnetic beads. A second DNA probe attached to a magnetoresistive sensor also binds the target DNA. Bringing the magnetic beads in close proximity a signal is detected (Edelstein et al., 2000)
- Nanochip technology: probes are attached to spots with underlying electrodes. DNA is electronically directed to the spot to allow for hybridisation. A second fluorescent probe is passively hybridised to the bound target DNA in the spot (<http://www.nanogen.com>).

There are also important developments in relation to the materials that can be hybridised to the array. A bottleneck in this respect is often the amount of mRNA that is needed for a hybridisation experiment. A solution may be here to use total RNA (totRNA) samples as it has been shown that relatively small amounts of totRNA are necessary for hybridisation compared to the amount of mRNA that is needed for a hybridisation and would be obtained from a much larger quantity of totRNA. This may, however, in many cases not be sufficient to solve the problem of RNA quantities. The

combination of developments with relation to the possibilities of excision of selected single cells out of a tissue and reliable and reproducible amplification of the mRNA population that can be isolated from the cell is likely to make more specific hybridisation experiments feasible in the near future. It seems that the sensitivity of the technology will not be the limiting factor as it has been calculated that EST microarrays are able to detect one fragment in a pool of 100.000 – 500.000, enabling to detect mRNAs that are present in only a few copies per cell (Richmond and Somerville, 2000; Zhu and Wang, 2000).

3.3 Potential and bottlenecks of the technology

By using the DNA microarray known metabolic networks can be investigated for alterations in gene expression, but, depending on the metabolic routes represented on the array, also activation of normally silenced pathways in specific tissues or plant varieties could be detected. For the food safety assessment preferably edible as well as non-edible parts of the novel plant varieties should be tested as newly formed substances may be transported between tissues and plant parts.

Thorough elucidation of the relations between genes, proteins and metabolites will greatly enhance the power of microarray analysis in this respect. One point of consideration is that after mRNA expression, several mechanisms can influence the level of protein that is formed from the mRNA-borne genes. In other words, mRNA levels may not be representative of the protein levels finally formed. This is due to several factors that are known to affect the translation process that builds the amino acid sequences of nascent proteins from the open reading frames of the mRNA. Analyses conducted in yeast have shown a very poor correlation (coefficient < 0,5) between mRNA and protein levels (Pradet-Balade, B. et al 2001 and Washburn, M.P. et al. 2003). The time necessary for the translation of mRNAs into proteins may vary and this may result in a delayed shift in the proteome or metabolome compared to the transcriptome. Translation depends on the binding of various ribonucleic and proteinaceous molecules to the mRNA chain (Gallie, 2002). This process of translation initiation can be influenced, for example, by (temperature) stress in plants. Examples are furthermore known where upstream sequences influence the translation levels by two till three orders of magnitude (Meijer and Thomas, 2002).

Another well-known factor that contributes to translation efficiency is "codon bias", given that different codons that encode the same amino acid may not be translated with the same efficiency. Codon preference may vary between organisms, and will differ, for instance, between bacteria (AT-rich sequences) and plants (CG-rich sequences) (Nakamura et al., 2000).

In order to be able to apply the full potential of the microarray technology, it is necessary to very carefully design an array experiment. The wealth of data that will result from any microarray experiment makes it crucial to think very carefully about the experimental setup taking into account statistical considerations as well as practical limitations of the available arrays and samples for RNA isolation. Also aspects of data storage and analysis should already be considered prior to the experiment in order to prevent data loss and to be able to include e.g. the right references and not lose time afterwards by undirected data computing.

3.4. Sampling

The quality of the mRNA populations under investigation is crucial for the information that can be obtained. It is therefore necessary to determine the quality of the isolated RNA fragments both spectrophotometrically as well as by gel electrophoresis. The ratio values that can be obtained by spectrophotometric analyses will give information on the purity of the sample whereas gel electrophoresis will show the intactness of the RNA fragments. The mRNA populations under investigation can be labelled either by direct incorporation of fluorescently labelled nucleotides during a reverse transcription reaction or indirectly where either a reporter molecule is bound to the nucleotide analog that is incorporated in the initial reverse transcription phase or by attachment of reactive molecules to the backbone of the unlabelled RNA molecules. Alternatives to the indirect approach combine the two-step procedure with subsequent signal amplification strategies (Adler et al., 2000). Labelling two experimental mRNA samples with different fluorescent dyes with distinctive excitation and emission characteristics allows direct comparison of the two samples on a single slide. A second experiment with reversed labelling conditions (dye swap) should preferably confirm the results. For comparison of larger series of samples it will in general be

necessary to use multiple slides and to apply a procedure that allows for the correction of differences between slides and/or hybridisation conditions. Because the quality of the (tot)RNA is very important standardisation of protocols used is required (Zhu and Wang, 2000).

As long as such 'precision' hybridisation is not yet routinely applied, it should be kept in mind that the mRNA will usually be derived from complex tissues. It is therefore crucial to perform comparison microarray studies with very similar testing materials that are fully homogenised in order to facilitate interpretation of the results or as Richmond and Somerville (2000) put it: the resolving power of each microarray experiment will be determined by the biological variation in the plant samples and the technical variation associated with the microarray technology.

An important limitation in many tissues/products will be the isolation of sufficient mRNA molecules of sufficient quality for hybridisation to informative arrays. As stated, characterisation of mRNA populations of small-sized tissues will in general require an mRNA amplification step prior to the array-based characterisation, unless sufficient similar samples can be pooled to obtain the amount of mRNA required for a hybridisation series.

The same goes for foods and food products. The mRNA has to be of high-standard quality in order to be able to compare two subsequent samples. As slightly different conditions in the food or food product will in general have large effects on the quality and quantity of the mRNA that can be isolated, it is highly unlikely that the application of the microarray technology in processed products will become standard for their characterisation and/or safety evaluation. At most it may be an additional tool for the evaluation of less-processed products where high quality mRNA can be isolated and characterised.

In order to relate individual microarray experiments to each other it will be necessary to use references. Three alternative strategies can be applied:

- 1) Individual reference spots/genes on the array that are known to be similarly expressed under the conditions under investigation. To this end such genes

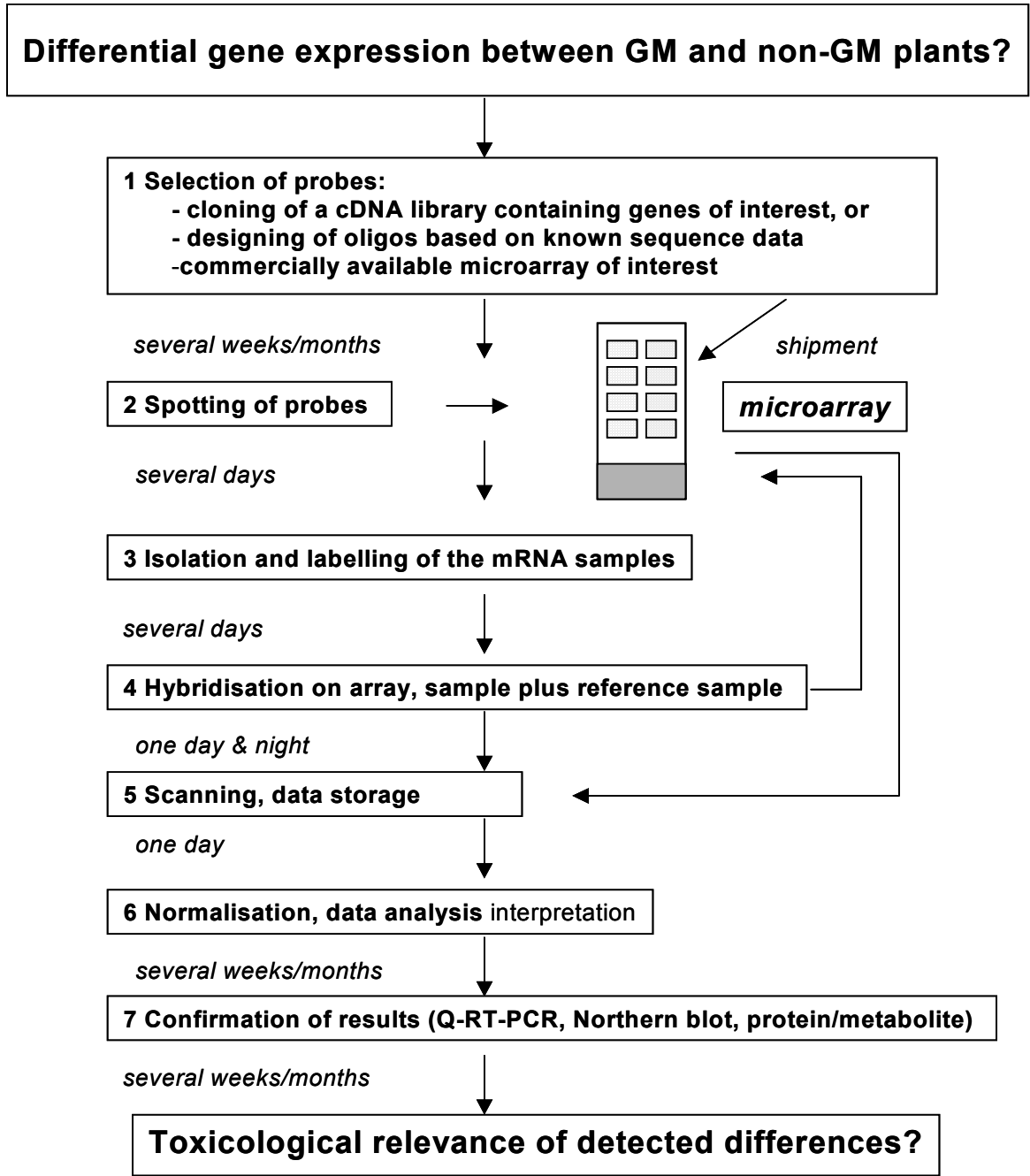


Figure 1. Schematic overview of GMO food safety microarray experiments

- must have been identified and the internal control spots should be manifold on the array or local differences within the array can not be controlled for;
- 2) A reference mRNA sample that is used on all microarrays of an experimental series where a different fluorescent dye is used to label the mRNAs under

investigation prior to hybridisation on the same series of arrays. This reference sample may be an additional sample or a pool of samples or a mix of all samples to be analysed.

These approaches are more or less experiment-specific and may result in a considerable number of negative reference spots, which may mean loss of experimental spots that can adequately be analysed. An alternative has been published by Sterrenburg et al. (2002). They propose

- 3) to use cRNAs of all spotted PCR products as an alternative common reference. Prior to labelling the PCR products are transcribed into single-stranded cRNA molecules and simultaneously amplified. Labelling can thus take place parallel to the labelling of the experimental sample. This approach results in high numbers of positive reference spots, reducing the loss of experimental data. In this way correction factors can be established for every individual spot of the array by comparing the signal obtained with the reference mRNA sample. Finally,
- 4) normalisation can be based on the fact that the overall expression between samples is not likely to change much. A correction factor is in this case determined by the overall fluorescent value.

3.5 Data analysis

Data analysis of microarray experiments is usually performed by the subsequent steps of quantification of the fluorescent signal, normalisation of the obtained data and data visualisation and interpretation. Quantification of the signal is performed with image analysis software programmes that can detect individual spots, have the possibility to exclude specific inferior spots and can calculate and subtract the background value for every individual spot. Background values can be determined in different ways that can include more or less background signal information from the area around either the spot, the block of spots or the entire array area. In general, it seems preferable to take a well-established background value on the basis of the fluorescence around the spot of interest, but this requires well-characterised spots with sufficient space in between to determine a background signal. Spots with signals below the background are generally discarded. Usually the image analysis software also has the possibility to do some limited data analysis and visualisation of the results. These programmes are still

in development and the possibilities for the user increase to decide on the optimal way to analyse the array results.

Normalisation of the data, in order to compare arrays with each other can be performed in different ways as explained at the end of section 3.4. Comparisons between different ways of linear normalisation in specific datasets may not show large differences in the resulting data analysis (Richmond and Somerville, 2000). More recent optimised alternative approaches use non-linear intensity-dependent methods for normalisation (Workman et al., 2002; Yang et al., 2002). Of interest is also the website of the Microarray Gene Expression Data Society (MGED) on normalisation (www.dnachip.org/mged/normalisation.html).

For microarray data visualisation and interpretation, different analysis tools are used to find the most relevant differences or resemblances in gene expression patterns, depending on the questions underlying the microarray experiment. When working with large datasets, as is the case with microarrays, simple interpretation of the data is not possible. In order to come to meaningful interpretations global insight in the structure of the data must be obtained first, followed by validation of the preliminary results. For the first part two main types of analysis tools are available, unsupervised and supervised methods.

Unsupervised methods look for trends or patterns in the dataset. They are of an exploratory nature and are most useful when differences are to be found, but when the nature of the differences is not known beforehand. This will usually be the case when, for instance, newly developed GM crops are investigated and compared to traditionally bred lines. Examples of unsupervised methods are cluster analysis, principle component analysis (PCA) and self-organising maps (SOMs).

Many forms of cluster analysis are available. Most software packages for microarray analysis offer clustering algorithms that use hierarchical clustering methods. This means that the algorithm looks at the samples as individuals. On the basis of the measured variables the similarity between individuals is calculated. Individuals that show high similarity are then grouped. Subsequently, groups of individuals are further

grouped together on the basis of similarity between groups, and so on until no more grouping can occur. The result is shown as a dendrogram (figure 2).

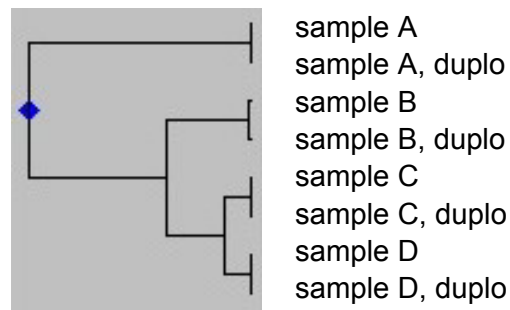


Figure 2. Example of a hierarchical clustering dendrogram. In this example the duplo experiments are clustered together. Sample A is most distinct from samples B, C and D. Within that group, C and D are most similar.

To explore which genes are correlated with this clustering of samples, the same algorithm is applied on the genes that were measured on the array. Genes with similar expression patterns will then be clustered together. Interpretation of the resulting data to come to plausible explanations in a biological sense will always be the final step. A number of options are available both for the way the distances, or similarities, between data points are calculated, as well as for the algorithm that groups the distances or similarities in clusters. For datasets that have distinct, clear features, most methods will produce similar dendrograms. However, if the differences are marginal compared to the noise in the dataset, different methods can produce quite different results. Insight into the matter of investigation, or additional experiments, will then be decisive for a meaningful interpretation.

PCA is a way to reduce the number of dimensions of a dataset while retaining as much of the variation as possible. If in a dataset 30 genes are measured, the dataset has 30 dimensions. The number of components is 29. In most datasets the first ten components will explain more than 90% of the variation, even if several thousand genes are measured. PCA calculates the components so that they explain the variation between the samples of the dataset, the first component explains the larger part of the variation, then the second, and so on. Each component consists of a series of numbers that act as weights for all the different genes. In the first component the highest weights are assigned to the genes that differ most between the samples of the

dataset. In the second component different weights are assigned, explaining as much of the remaining variation as possible, and so on. The major advantage of PCA is that differences in the dataset can be viewed at different levels. The first component of PCA can be compared with the results of a dendrogram in cluster analysis, but in addition in PCA the second and third etc. components can also be viewed, which is not possible with hierarchical clustering (figure 3). However, the interpretation of a PCA is also more intricate. First of all, the component that best separates the groups of interest has to be identified; this is not necessarily the first component. Then the genes have to be sorted according to the weight that was assigned to them in that component; these are likely to be the most interesting genes involved in the experiment. Understanding all the components and the variance they explain, may not always be feasible, and sometimes not even relevant.

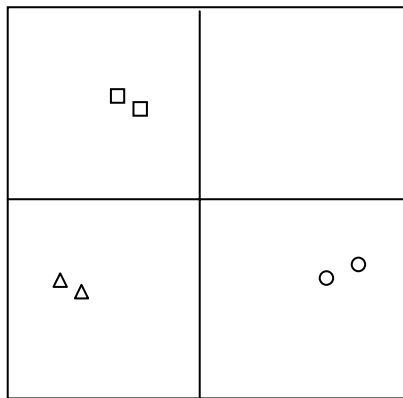


Figure 3. Example of a PCA analysis on the expression profiles of six samples (duplos of \triangle , \square and \circ). Plotted is the first component (X-axis) against the second component (Y-axis). Genes that are important in the first component separate the \circ samples from the \triangle and \square samples. Genes that are important in the second component separate the \square samples from the \circ and \triangle samples.

A SOM is an example of an artificial neural network (ANN). An ANN is an algorithm that is loosely based on the spatial organization of the human brain. In a SOM, the experiments can be divided in a user-defined number of nodes. Experiments that are most similar will be placed in the same node. In most software the variation in each node can be viewed, in order to see if this grouping is meaningful. The interpretation

of which genes are most important for the grouping is less transparent than for PCA and clustering.

Supervised methods are more useful for diagnostic types of analysis than for exploratory purposes and are less likely to play an important part in the evaluation of the food safety of GM crops initially. Supervised methods of data analysis involve the training of an algorithm with a number of different, well-established samples. The training is an iterative process, increasing the discriminative power of the algorithm with each round of training. The trained algorithm can then be applied to new datasets in order to determine the category of these test samples. This means that a number of conditions should be met in order for supervised methods to work. First of all, a substantial number of samples must be available with sufficient details regarding measured data and categorisation. There must be three sufficiently large sets: a training and a validation set for the training of the algorithm, and an independent test set to test the algorithm once it is trained. In all these sets all the groups must be present. Secondly, the variables that are measured must be able to distinguish the groups of interest. For instance, to determine the tissue identity of a particular unknown plant sample, the array must contain sufficient tissue-specific probes for the purpose. A known caveat for supervised methods is to know when to stop training the algorithm: too few rounds will decrease the discriminative power, while too many rounds result in training on the noise in the training set; this results in less discriminative power when unknown samples are analysed. A good supervised method should report on the predictive power of the algorithm that is the result of the training (percentage of false positives and negatives). Examples of supervised methods are Discriminant Function Analysis (DFA), Decision Tree Learning (DTL), Artificial Neural Networks (ANN) and Evolutionary Computing and the number of supervised methods is still increasing.

In general, any differences between samples that are found with these methods, especially unsupervised methods, should be considered potential differences. Therefore, the last part of data analysis must always be validation of some sort. If different approaches on analysis detect the same differences, chances are increased that these differences are significant between the groups of interest. However, it will always be necessary to go back to the raw data and perform statistics on the number

of potential differently expressed genes. Methods for this are ANOVA (analysis of variance) and univariate analysis. However, much debate is ongoing regarding the right statistical approach for microarray experiments (Hess et al., 2001; Nadon and Schoemaker, 2002; Becker, 2001). One of the major problems is the few samples analysed compared to the large number of variables (genes). Specific statistical routines will have to be developed to overcome this and other problems.

From this overview it is clear that the number of ways to analyse a set of thousands to millions of microarray data of a single experiment is sheer endless and this further underlines the necessity to also include the aspect of data analysis in the preparation of any microarray experiment.

3.6. Data storage

Data storage is essential throughout the data analysis route. It is clear that a good database design is crucial for the ability to use the wealth of data that will already result from a single experiment for future analysis. Furthermore, if the same (commercially available) array is used by many research groups this can be seen as a common experimental platform. If the results obtained with this array are stored in a centralised database, an in silico analysis of the available data may reduce the number of experiments that need to be actually carried out (Kehoe et al., 1999). Initiatives to come to harmonised approaches for microarray databases have already been launched (MIAME, Brazma, 2001; Brazma et al., 2001).

A timely setting up of informative databases is very important also in the light of the fact that (statistical) data analysis tools for microarray experiments are still very much in development. At this moment we can therefore only extract limited information from individual experiments and from the combination of different microarray experiments. With the progressing development of statistical microarray data analysis tools we may be able to use the experiments of today to solve the questions of tomorrow. It is also clear that to this end the current way of communicating research results will have to be revised. At this moment the results of (functional) genomics sequencing programmes and microarray experiments are already published for the larger part on the internet, with only summaries of the results in conventional scientific publications.

It is feasible that publicly accessible databases will become more and more an integral part of (electronic) scientific publications. The Microarray Gene Expression Data Society (MGED, <http://www.MGED.org>) is an international group of scientists that aims to facilitate the sharing of microarray data generated by functional genomics and proteomics experiments. The current focus is on establishing standards for microarray data annotation and exchange, facilitating the creation of microarray databases and related software implementing these standards, and promoting the sharing of high quality, well annotated data within the life sciences community.

A long-term goal of the group is to extend the mission to other functional genomics and proteomics high throughput technologies. At the moment the MGED is organised in four workgroups:

- 1) MIAME (minimal information about a microarray experiments) aims to formulate the minimum information about a microarray experiment required to interpret and verify the results;
- 2) MAGE aims to establish a data exchange format (MAGE-ML) and object model (MAGE-OM) for microarray experiments;
- 3) Ontologies (OWG, object working group) develops ontologies for microarray experiment description and biological material (biomaterial) annotation in particular;
- 4) Transformations develops recommendations regarding microarray data transformations and normalisation methods.

Recently, a number of scientific journals have made it compulsory for articles presenting microarray data to be MIAME compliant, indicating the importance of the MGED initiative. MIAME has two major sections: 1) array design description and 2) gene expression experiment description. The array design description deals with the description of the array as a whole and of each individual spot on the array. The gene expression experiment description was designed to provide relevant information on any set of microarray data that are related to each other, as is the case for arrays in a single experiment or article. Items for which information should be provided are:

- experimental design; e.g traditional vs. GMO comparison, number of hybridisations, number of replicates, the use of a common reference or not etc.
- samples used, extract preparation and labelling; detailed description of the experimental samples used, including method of isolation and labelling.

-hybridisation procedures and parameters; for instance, the composition of the reaction buffer, the blocking agent, the wash procedure, hybridisation conditions etc.
-measurement of data and specifications of data processing; providing information on the hardware and software used to generate the raw data, the raw data itself, the subsequent analysis, including quantitation and normalisation, and the software used.

The MGED internet site contains more information about the workgroups as well as links to publicly or commercially available databases for review or submission of microarray experiments as well as software for data analysis. MIAME was not intended as a rigid set of rules, but more as a set of guidelines, ensuring a minimum quality of microarray information when adhered to. As a result, many different forms of information about microarray experiments can be MIAME compliant. One example is MIAMExpress, a software tool publicly available via the European Bioinformatics Institute (<http://www.ebi.ac.uk>), and designed for public storage of microarray data. MIAMExpress is MIAME compliant and the scheme for the submission of data gives an overview of the MIAME criteria (figure 4). Future development will have to prove whether the public sharing of data is feasible and if the self-regulatory capacities of the scientific community are good enough for transparent and uniform presentation of the vast amounts of data that are being generated. The human genome project and public databases such as the one of the NCBI are good examples of successful efforts of public data storage. There are many more public databases that contain information on specific genes or proteins, focussing on different species, metabolic pathway, disease types etc. An overview of these databases was published in a special issue of Nucleic Acids Research (volume 31, issue 1, 2003), also available online (<http://nar.oupjournals.org/content/vol31/issue1/>).

Finally, it becomes increasingly important that microarray databases are directly linked to comparable databases of the other 'omics' technologies, as it is likely to become a standard approach to investigate metabolic pathways of interest simultaneously by different technologies, in order to be able to make optimal progress in the elucidation of the pathways. It is therefore important to start thinking already on the compatibility of initiatives such as of the MGED with other initiatives that aim to develop similar systems for NMR and/or MS data, for instance an initiative in the British FSA G02

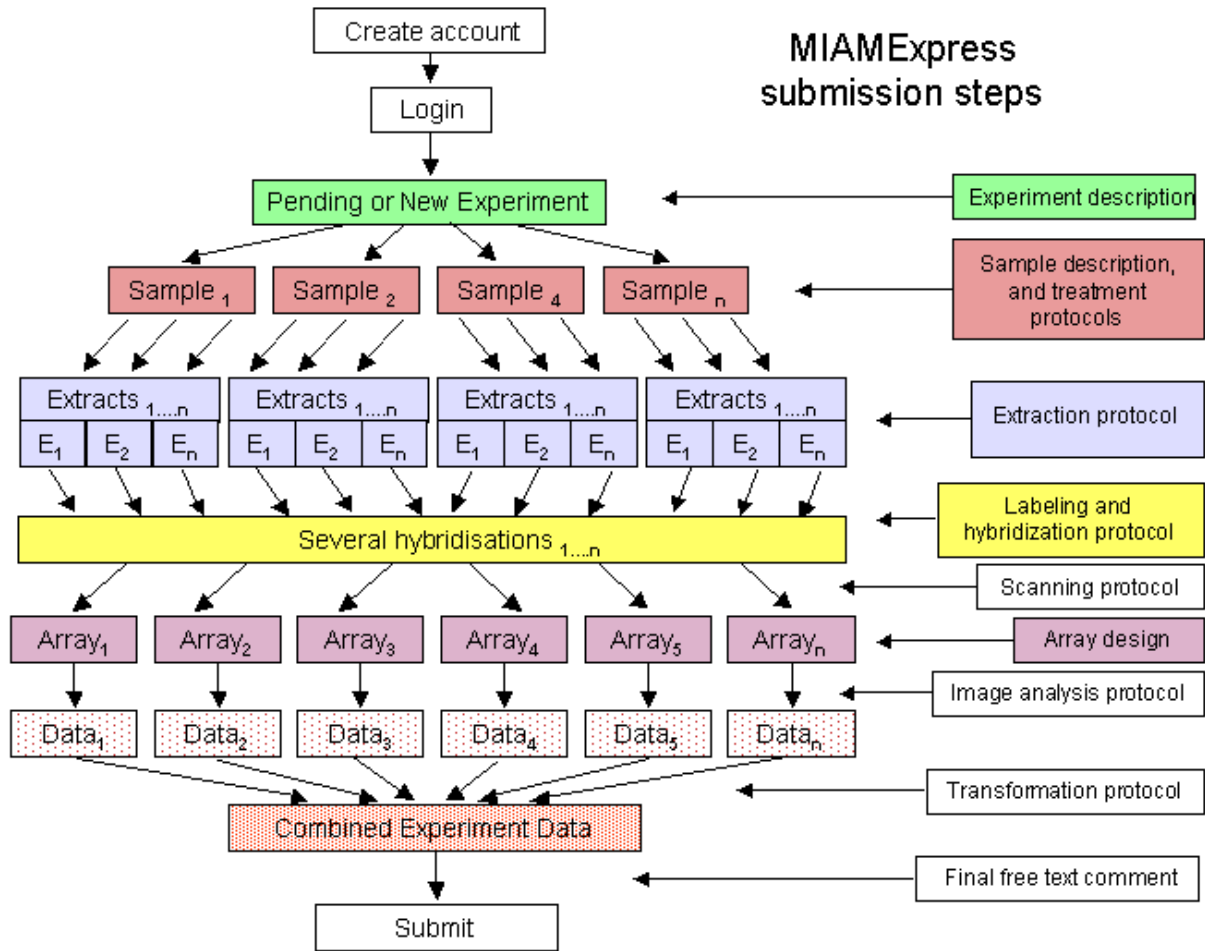


Figure 4 Example of the information required for MIAME compliant submission and publication of microarray data. The example shows the flowchart of the MIAMExpress software (publicly available) for the public storage of microarray data, developed by the European Bioinformatics Institute (EBI), and available on their website (<http://www.ebi.ac.uk>).

programme (<http://www.foodstandards.gov.uk/science/research/NovelFoodsResearch/g02programme/g02projectlist/g02006>). Although linkage will be a huge task, the advantages for the (food safety) analysis of novel plant varieties are evident as it is unlikely that the assessment will in the end be based on a single methodology.

Chapter 4 Past and present experiences with the use of cDNA microarray technology in plant varieties grown under different (environmental) conditions

4.1. Model systems

The plant that has been most often used as a model system for all 'omics' applications so far is *Arabidopsis thaliana* (Meinke et al., 1998). This is a weed that can be found all over the world and can easily be grown and transformed and is therefore ideal for genetic experiments. For 'omics' applications the additional advantage is that the genome is relatively small (130 M bp). It was the first plant genome to be fully sequenced (The Arabidopsis Genome Initiative, 2000). Functional analysis of the annotated genes shows that approximately half of the gene sequences could not be annotated, a quarter of all genes are believed to be involved in metabolism, transcriptional and cell growth and division –related activities, with the rest being related to other areas of cell activity. The collection of *Arabidopsis* cDNAs and ESTs is also one of the largest plant collections that are currently available and large research projects aim to elucidate the function of the identified genes, e.g. by large-scale insertional mutagenesis (Parinov and Sundaresan, 2000; Pereira, 2000, Rudd et al., 2003). For monocotyledonous plants rice has become the model system. The genome size of rice is 400 M bp (Zhu and Wang, 2000). The TIGR rice genome annotation resource provides a very valuable and extensive source of information on the rice genome as well as on ESTs from the important rice subspecies *japonica* and *indica* (Yuan et al., 2003; <http://www.gramene.org/>). The *japonica*-ESTs are mainly derived from GenBank, whereas the *indica*-ESTs are mainly derived from the Beijing Genomics Institute (BGI).

An informative overview of the use of microarrays in plants until 2001 is provided by Aharoni and Vorst (2001). From their overview it is clear that for almost all array experiments described thus far EST arrays were used, for the larger part *Arabidopsis* arrays, and in addition a limited number of arrays with rice, beans, maize, strawberry and *Petunia* ESTs. Furthermore it is shown that most array experiments relate to

experimental exposure to different environmental conditions, or to the analysis of differential gene expression as result of different types of mutagenesis strategies.

In rice microarray experiments were performed with relation to salt stress. A microarray was constructed with 1728 cDNA probes that were derived from cDNA libraries of salt-stressed roots. Hybridisations with series of samples that were taken at different time-points after a salt shock showed that differences in gene expression could already be detected after 15 minutes, peaking within one hour after the stress was applied in 10% of the selected transcripts (Kawasaki et al., 2001).

Tomato has been the model system of choice to study fruit ripening. Very important in this respect has been the availability of a number of well-characterised ripening mutants. To this end a large library of non-redundant EST clones has been generated at TIGR (The Institute for Genomic Research, Tomato NSF (National Science Foundation USA) EST project <http://www.tigr.org/tdb/lqi>) of approximately 28.000 clones. Experiments were performed with probes isolated from ripening tomatoes at a time-course spanning fruit development from 7 days post-anthesis to 15 days past breaker (first yellowing of the tomato) stage (Moore et al., 2002). Expression ratios in different stages of ripening were determined for four sequences known to be involved in tomato ripening. Ratios were shown to be the highest between the green and breaker stage of ripening.

The Institute for Genomic Research (TIGR) has developed a large potato EST library with sequences derived from nine different potato plant part libraries as well as libraries obtained after exposure to the late-blight pathogen (*Phytophthora infestans*) (Ronning et al., 2003). From over 60.000 EST sequences 19.892 unique sequences could be identified. For 43.7% of these sequences a putative function could be identified and 48 were found to be expressed in all nine libraries, indicating that these are constitutively expressed or 'housekeeping' genes. 21% of the sequences were uniquely expressed and detected only in a single library. Comparison of the potato EST library with TIGR Tomato Gene Index which is composed of over 150.000 tomato EST sequences showed that depending on the stringency applied, 70-80% of the unique sequences showed sequence similarity.

4.2. Identification of stress-related genes in plants by use of microarrays

The application of microarray technology to identify stress-related genes in plants is comparatively young and is starting to take shape. In the pioneer study of Schenk et al. (2000), microarray analysis has been performed on *Arabidopsis* under various stressful conditions induced by fungal infection (*Alternaria brassicola*) and plant hormones (salicylic acid, methyl jasmonate, or ethylene). These authors found that 705 out of the 2375 spotted genes respond to one or more treatments by up- or down-regulation, while 169 mRNAs were regulated by multiple treatments. The results of this study therefore show that a number of differentially expressed genes are involved with multiple stressful conditions. A number of reports on microarray analysis of gene expression plants exposed to stress have subsequently appeared, such as by Maleck et al. (2000), who identified groups of genes with common regulation patterns under 14 different conditions that would either induce or repress plant disease resistance mechanisms. Two reviews highlight the progress and prospects of microarray technology for investigation of differential expression induced by stress, in particular during plant defence (Reymond, 2001) and drought (Bray, 2002). In addition, examples of recent studies are summarised in the Table below.

As mentioned above, various classes of genes are known to be involved with either systemic or local stress-induced response by plants, covering a wide array of specific cellular functions (e.g. detoxification, metabolism, transport) and transcriptional regulation. By microarray analysis, it has been confirmed that the expression of these known genes, but also many novel- or previously unrelated- genes are affected by stressful conditions (Bray, 2002, Reymond, 2001).

Also the studies summarised in table 3 have shown the added value that microarrays have in elucidating the involvement of specific genes/pathways in stress-response. Rossel et al., (2002) and Chen et al. (2002) both describe, for example, that specific promoter elements are associated with a number of stress-induced genes. In two of the summarised studies, stress-responsive genes have been identified that are similarly affected under multiple (2) stressful conditions, which further corroborates the

specific relationship with stress (Swidzinski et al., 2002; Ozturk et al., 2002). In one case, stress-responsive genes show cross-relationships with ripening (Aharoni et al., 2002). In another case, gene expression affected by diurnal variation could be discerned from that affected by stress (Negishi et al., 2002).

Plant hormone-responsive genes are also stress-related under the various conditions of stress investigated in the studies listed in table 3 and previously reviewed (Reymond, 2001). In barley, for example, jasmonic acid- responsive genes are active after 6 hours of drought, while abscisic acid-responsive genes become active after 10 hours (Ozturk et al., 2002).

Transcriptional factors are also included in many microarray-supported studies of stress-response in plants. Chen et al. (2002), for example, have constructed microarrays of probes specific for transcriptional factors. As the results show, these transcriptional factors can be allocated to different groups, based on their involvement in abiotic- and biotic- stress responses. Transcriptional factors are attractive compared to specific genes, because they are more likely to provide an avenue to stress-tolerant plants in plant breeding.

Another important issue is the processing of microarray data related to stress-responsive gene expression. Mitra et al. (2002) describe an on-line database that is specifically focused at stress-research with microarrays in Arabidopsis. It is expected that this specialised database will facilitate exchange of results of microarray experiments by allowing for specific additional data to be incorporated with the submitted data (Mitra et al., 2002). With regard to normalisation, it should be taken into account that “housekeeping” genes are not suitable for this purpose, given the fact that the expression of these genes may also be affected by stress (Swidzinski et al., 2002).

Stress	Plant	Types of genes probed	Differentially expressed genes	Reference
High light	Arabidopsis	pigment biosynthesis antioxidant biosynthesis photosynthesis	Heat shock proteins (+) pigment biosynthesis (+) antioxidant (+) flowering (-/+) chlorophyll synthesis (-)	Rossel et al., 2002
Heat Senescence	Arabidopsis ¹	senescence hypersensitive response antioxidant metabolism cell maintenance mitochondrial proteins	senescence-related and -activated (+) ^{2,3} hypersensitive response (+) ² antioxidant enzymes (+) ² mitochondrial adenine transport (-) ²	Swidzinski et al., 2002
Abiotic ⁴ Biotic ⁵	Arabidopsis	stress-related (putative) transcription factors	salicylic acid-inducible (+) jasmonic- acid/ethylene inducible (+) inducible by additional factors (+)	Chen et al., 2002
Oxidative	Strawberry	ripening-regulated	detoxifying enzymes (+) protective enzymes (+) pathogenesis related proteins (+)	Aharoni et al., 2002
Iron deficiency	Barley	rice ESTs	(deoxy)mugineic acid synthesis/conversion (+) methionine synthesis and Yang cycle (+) cytoskeleton-related (+) transcription regulation (+) translation-related (+) vacuole-related (+) vesicle transport-related (+) enolases (+)	Negishi et al., 2002
Drought Salt	Barley	drought-stressed	jasmonic acid biosynthesis (+) ⁶ proline biosynthesis (+) ⁶ abiotic stress-related (+) ⁶ transport proteins (-) ⁶ alpha-amylase (-) ⁶ vacuolar processing enzyme (-) ⁶ unknown function (-) ⁶	Ozturk et al., 2002

Drought Shade	Maize	- maize ESTs ⁷ - non-targeted	- shade: abscisic-acid response (+) starch biosynthesis (-) - drought ear: cellular growth (-) pedicel: phenylpropanoid/lignin biosynthesis (-) seed except pedicel: non-classified (+)	Zinselmeier et al., 2002
Drought	Loblolly pine	resistance-related signal transduction stress-related	aquaporins/dehydrins (+) ⁸ heat shock proteins (+) ⁸ lignin biosynthesis (0) ⁸ waterstress-inducible chaperone (+) ⁸	Heath et al., 2002

(+) upregulated; (-) downregulated; EST = expressed sequence tag

¹ cells in suspension culture

² genes differentially expressed in both conditions (heat and senescence)

³ especially cysteine protease expression was noted

⁴ abiotic stress: cold, salt osmoticum, jasmonic acid

⁵ biotic stress: infection by bacteria, fungi, oomycetes, and viruses

⁶ genes differentially expressed in both conditions (drought and salt)

⁷ carbohydrate/hormone metabolism; cell cycle; stress response/signalling

⁸ regulation in mild stress, role for stress adaption given no- or negative regulation of same gene during severe stress

Table 3. Recent studies into stress-induced differential gene-expression in plants by microarray analysis

Chapter 5 Selection of individual cDNAs for characterisation of food plants

5.1. Selection of markers

With our increasing knowledge of the plant's physiology and the metabolic routes that lead towards the key nutrients and anti-nutrients in the plant, it becomes more and more feasible to select individual cDNAs, or parts thereof, to compose a new, highly informative microarray containing solely cDNAs that are directly relevant for the food safety characterisation of a new crop plant variety. For economically important food crops this may be possible in the coming few years. On the other hand, developments are going everfaster and full transcriptome microarrays will be available for the first few important food crops before long.

For the characterisation of the safety of new crop plant varieties basically three sets of genes can be thought of, i.e. 1) general markers or the so-called 'housekeeping' genes, 2) markers related to the nutritional and wholesomeness aspects of the plant and 3) markers related to antinutritional factors.

- 1) General cell metabolism markers. These are genes that are expressed in all different tissues and in the subsequent developmental stages of the crop plant and disturbance of the overall physiology of the plant will be reflected in the expression levels of these general markers. It will, however, be very delicate to establish such a set of general markers. It is known that constant expressors may be disturbed by different stress factors and detailed knowledge on their expression levels under different environmental stress conditions will be required in order to be able to interpret results from plant samples that may have experienced a form of stress during the growth period. Also, constant expression may vary in the different tissues and therefore a very clear definition of the tissues to be sampled will be required.
- 2) Health/nutrition-related markers are expression products of genes that are known to be involved in metabolic routes leading to macro- and micronutrients. At this moment we do not have a clear listing of plant constituents that can be considered nutrients, but for a number of plants the OECD consensus documents can be

considered a starting-point in this respect. Assuming that the nutrient-related metabolic routes of the crop species that are now being scrutinised using the microarray technology will be elucidated, a health/nutrition-related marker array of these species will become feasible. At the same time it is becoming clear that the networks, both in GM and traditional plant varieties, are very complex and small changes in ratios may lead to important shifts in e.g. secondary metabolites. Therefore it should be investigated whether it will be sufficient to monitor individual genes in important metabolic pathways or that it is necessary to determine relative levels between the different expression products involved. Before the precise interactions with the related proteins and, most important, the resulting metabolites are sufficiently known, it will require additional independent analysis of these substances.

- 3) For anti-nutrients/natural toxins and stress-related markers the same considerations can be made as for the health/nutrition-related markers, with the addition that the elucidation of the anti-nutrient metabolic routes lags far behind the elucidation of positive factors in economic crop plants. Despite some valuable overviews of secondary metabolite (Hadacek, 2002), in many less well-documented crop species there is very limited information on the presence and character of anti-nutrients, including natural toxins. In economically more important plant varieties the information on the nature of the anti-nutrients is usually available, but natural variation in these substances has only recently, basically since the advent of genetically modified organisms and the related regulatory requirements, obtained more attention. As a result of this in many cases there are no validated methods of analysis available to determine levels of the anti-nutrients in the plant. This will severely hamper the establishment of clear correlations between gene expression levels in plant samples and the related levels of natural toxins.

In general it can be stated for all three categories that an informative array can only be constituted of well-characterised cDNA sequences, both in the sense of the activity and function in physiological networks as with relation to the natural variation in gene expression in different tissue and under different (stress) conditions. Such a focused array may be more sensitive as the number of spots is limited, thereby optimising the conditions for hybridisation of individual sequences. Likewise global, whole genome

arrays may show a larger number of significant changes in gene expression that may not be biologically relevant, due to more varying conditions for individual hybridisation events. On the other hand theoretically the same information can be obtained from the whole genome array compared to the focused array without the risk of overlooking important changes in metabolic networks that were deemed less relevant in the initial selection of pathways to be investigated. Direct comparison of the two options in a number of different settings will be necessary in order to be able to make an informed choice.

Chapter 6 The role of gene expression microarray technology in the wider profiling of food plants

6.1 Proteomics

Proteomics is the direct counterpart to transcriptomics. Any indication for differential gene expression will need to be confirmed by other technologies. The most direct way of verifying the biological effect of altered levels of transcription products is the analysis of the subsequent translation step which leads to the formation of the related proteins. The main approach currently applied to analyse the proteome, i.e. the total of proteins per unit (cell, tissue or organism), is two-dimensional gel electrophoresis (2DGE) to separate the individual proteins in two-dimensional space by size and charge. Subsequently it is possible to excise individual protein spots and analyse them by mass spectrometry (Anderson and Mann, 2000; Kuiper et al., 2003). There are, however, limits to what 2DGE can analyse as, in general, only highly expressed proteins will be detected (Gygi et al., 2000). Also, Heazlewood and Millar (2003) noted that if peptide mass spectra are to be matched with genomic data from other species, MS/MS rather than MALDI-TOF should be employed to avoid uncertainties. The application of this technique to study differences in gene expression in GM plants versus their traditional counterpart is now also being tested in the European GMO CARE project (<http://www.entransfood.nl>). In this project samples of GM tomato and potato plant lines and their traditionally bred parent lines are analysed with a range of profiling techniques, including proteomics on the basis of 2DGE. To this end libraries are also developed of 2DGE images of e.g. the tomato in different stages of ripening. These libraries will serve as background data for the analysis and interpretation of differences in protein composition between the GM plant and the traditional parent line. Other approaches to analyse the proteome of GMOs are now investigated, e.g. 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>). Protein microarrays are now being developed on the basis of well-characterised protein-protein interactions, enzyme substrates and inhibitors, antigen-antibody and protein-ligand interactions. Theoretically protein

arrays can expand more easily the moment new information on the proteome becomes available and may reduce the time to set up new protein analysis systems, while increasing the reproducibility and potential for quantification. Routine use of informative protein microarray systems to investigate changes in the physiology of new plant varieties will, however, not be possible until a number of current problems with respect to array production and assay performance are solved (MacBeath et al., 2000; Templin et al., 2002).

6.2. Metabolomics

Analogous to the cascade of gene - gene product (enzyme) - metabolite, the metabolome is the level of cellular organization after the genome and proteome. The metabolome engenders the metabolites that occur within a biological entity. Metabolomics, i.e. the non-targeted study of the metabolome, can be a valuable addition to cDNA profiling in providing insight into the extent by which metabolites are affected by mRNA and protein expression and thereby contribute to phenotypic/genotypic variation. In fact, a recent review describes that due to the broad substrate specificity of enzymes, substrate availability is an important factor in metabolite formation, and hence the identity and composition of the metabolites (Schwab, 2003).

Various recent reviews describe the background of- and techniques used for- metabolomics (e.g., Phelps et al., 2002; Roessner et al., 2002; Sumner et al., 2003). From these reviews, it can be learned that metabolomics is a rapidly evolving field of scientific research, not only for plant physiology, but also medicine and microbiology. For plant metabolomics, several analytical techniques in particular are applied to obtain metabolite profiles of plant extracts. In most cases, these techniques combine chromatographic separation with broadly applicable spectrometric detection. Examples are gas chromatography coupled to mass spectrometry (GC-MS), liquid chromatography coupled to mass spectrometry (LC-MS), and liquid-chromatography coupled to nuclear magnetic resonance (LC-NMR). These techniques have their own particular advantages. Gas chromatography, for example, may be readily applied to small molecules that are volatile or that can be volatilized by derivatization. Liquid chromatography provides an alternative for larger molecules that cannot be properly

volatilized, as well as for heat-labile substances that disintegrate under the high-temperature conditions of gas chromatography. For mass spectrometric detection following liquid chromatography, interfaces are needed to bring the target substances from the liquid- into the gaseous- phase.

By GC-MS, Fiehn et al. (2000), for example, observed that the metabolomes of two ecotypes of *Arabidopsis thaliana* showed less resemblance to each other than each ecotype with a mutant that displayed metabolic alterations. In addition, Ward et al. (2003) applied proton(¹H)-NMR to extracts of *Arabidopsis thaliana* from various geographical origins. It was observed that differences in the levels of two metabolites (fumaric acid and glutamic acid) contributed most to differences between clusters of samples from different origins.

Furthermore, a recent report describes differences between patterns of NMR-profiles of plant extracts (Ott et al., 2003). NMR patterns were analysed that were induced by specific herbicides that had been applied to plants and thereby allowed to defer the modes-of-action of these herbicides.

Various reports describe the application of metabolomics to the identification of differences between genetically modified crops and their non-modified counterparts. Noteborn et al. (2000), for example, applied LC-NMR to genetically modified tomatoes modified with either the insecticidal Cry1Ab protein and with long-ripening characteristics (exogalactanase suppression). By subtraction of spectra of genetically modified- *versus* non-modified tomato extracts, they were able to identify glutamic acid and citric acid as the two metabolites that were altered in the long-ripening tomato compared to the parent line. Another example of metabolite profiling of crops with a more complex, "second generation" modification, is provided by Le Gall et al. (2003). Tomatoes were analysed that had been modified with transcription factors in order to enhance flavonoid synthesis in the fruit flesh by ¹H NMR. In addition to the appearance or elevated levels of flavonoids, also alteration in levels of other metabolites as observed, including amino acids, nucleotides/-sides, sucrose, malic acid, and citric acid (Le Gall et al., 2003). These reports and others on plant mutants and transformants demonstrate the applicability of metabolomics to characterisation of plants with altered genetic constitutions.

Different phases of the metabolomic assay techniques can in general be discerned, i.e. sampling, extraction, analysis (separation, detection, characterisation), and data processing, all of which need to be standardised/normalised in order to obtain reproducible outputs in terms of quality and quantities. Defernez and Colquhoun (2003), for example, determined the influence of peak shape, -height and -separation in ¹H-NMR fingerprints of metabolites from potato, tomato, and tea on the outcomes of the subsequent clustering of results through PCA. It was found that the PCA outcomes could vary between repeated experiments due to imperfections in signal detection. Ways to prevent this variation would include alternative sample preparation and peak alignment. While this study focused on NMR, the findings can be generalised towards other analytical techniques, including HPLC separation and infrared spectroscopy.

6.3 Glycomics

In addition to generally known 'omics' technologies, such as (functional) genomics, proteomics and metabolomics, another 'omics' technology may prove its value in the coming years. This is glycomics, the set of technologies that aim to elucidate and characterise the entire complement of sugar chains in a cell that is called the 'glycome'. This is likely to be most complicated branch on the 'omics' tree as every protein has numerous glycosylation sites, each of which may have different sugar groups attached, and these sugar chains can themselves be modified depending on the cell's state of activity (Perkel, 2002). The glycome is a complex system with each cell, tissue, organ and organism having different ever-changing glycosylation patterns. Carbohydrate-binding proteins and antibodies may play a crucial role in future high-throughput elucidation of characterising glycosylation patterns. At this moment the best way to determine glycan structures is mass spectrometry (Dell and Morris, 2001). Whether elucidation of the glycome will give relevant information for the purpose of characterisation and/or evaluation of new plant varieties, remains to be seen.

	<i>transcriptomics</i>	<i>Proteomics</i>	<i>Metabolomics</i>	<i>glycomics</i>
pro	<ul style="list-style-type: none"> -complete or large portion of the transcriptome can be investigated -high-throughput of samples possible -large worldwide effort to come to consensus on experiment design and data interpretation 	<ul style="list-style-type: none"> -more direct relation with compositional changes in plants -direct screening of allergens possible 	<ul style="list-style-type: none"> -direct relation with compositional changes in plants -availability of non-invasive techniques 	<ul style="list-style-type: none"> Relevance in relation to allergenicity
con	<ul style="list-style-type: none"> -indirect approach to changes in plant's physiology -data analysis time consuming -vast amounts of data lead to statistical difficulties -validity of technique as 'stand-alone' research tool still questionable 	<ul style="list-style-type: none"> -only small portion of proteome can be investigated -2D gel electrophoresis time consuming -protein microarrays still very much in developmental stage 	<ul style="list-style-type: none"> -very limited portion of metabolome can be investigated 	<ul style="list-style-type: none"> - glycome analysis is very complex due to the constant changes

Table 5 Pros and cons of different 'omics' technologies for safety assessment in plants

Chapter 7 General discussion.

The main attribute of the microarray technology is the possibility to scale up tremendously the rate of data acquisition. This, together, with the enormously increased capacity to sequence entire genomes is now resulting in a rapid increase of our insight into the physiology of plants. However, this increasing insight has already made clear that a number of earlier assumptions on relations between the different levels of DNA, mRNA, proteins, and secondary metabolites are not as straightforward as was anticipated. The number of genes of the human genome, for instance, turned out to be much smaller (in introduction: app. 35.000, Arabidopsis 29.000) than expected. In order to explain at the same time the complexity of the human body this probably means that the interactions between the different molecules may be more complex than anticipated. This is likely to be at least as complex in plants.

The composition of plants has not been a subject of major interest to the scientific community and even plant breeders focused mainly on other phenotypic effects such as yield or pest resistance rather than on compositional parameters. As a result, our knowledge of even the economically most prominent plant varieties is not very well documented. This lack of interest has changed markedly with the generation of the first genetically modified plant varieties. It was soon generally acknowledged that the best way to assess GM plants would be on the basis of a thorough compositional comparison, usually with the direct parent line. Data on individual plant constituents are now gathered, for instance by the ILSI and OECD, and stored in informative databases and consensus documents. At the same time it was acknowledged that the 'traditional' dataset of individual plant components provides a somewhat biased view on the plant's composition, as data on specific nutrients may be readily available whereas data on anti-nutrients, and especially natural toxins, may be limited for a number of crop species. But even if data are sufficient, it remains uncertain to what degree unintended side effects of a genetic modification can eventually be detected using this targeted compositional analysis.

Therefore international advisory bodies have urged the need for the development and evaluation of unbiased profiling methods that may be able to give additional information on the changes in the plant's physiology. As a result of this and in

accordance with the rapid development of high-throughput methodologies to investigate different types of molecules of interest, systems have been set up that may be useful in screening for differences in the new plant varieties compared to their traditional counterpart(s). The microarray technology is one of the most important developments in this respect. The advantage of this technology is that rapid expansion of the array is feasible the moment additional genomic information or materials become available. It is likely that the first crop-specific whole genome gene expression arrays will soon become commercially available. In this way tissue samples of novel plant varieties can rapidly be screened for differences in gene expression. This may become increasingly of interest now that a new generation of GMO products is being developed that have more profound physiological differences compared to the traditional parent lines. Preferably metabolic routes representing both edible as well as non-edible plant parts should be selected for the construction of the array in order to trace the activation of pathways that are usually silenced in the edible parts. Likewise the analysis of the non-edible plant parts may reveal differences in gene expression that may affect the safety of the edible parts, for instance as a result of the transportation of toxic substances to the parts to be consumed.

Different types of microarray systems are now available, differing largely in state of the art of the underlying technology and related protocols, and also in costs of equipment and materials. This differentiation will continue with new developments that are even better equipped to answer specific questions of the (scientific) community.

For the purpose of the food safety evaluation of GM plant products it is possible to use either whole transcriptome arrays, when available, or more concise, focused arrays. Focused arrays should consist of probes that are related to key nutrient and antinutrient metabolic routes, as well as to more basal cell metabolism- en stress-related pathways. Focused arrays may be more sensitive, but evidently there is a bias as to the range of unintended effects that can be detected in this way.

A number of factors will be decisive for the potential outcome of any microarray experiment, the most important ones being the nature of the probes that are spotted on the array, the quality of the mRNA to be hybridised to the array, experimental set-up and the selected data analysis procedure. It is therefore crucial that all of these

factors are given ample thought prior to any experiment. Due to the fact that the quality of the easily degradable mRNA is so important, it does not seem feasible to develop gene expression profiling strategies to compare processed food products: minor differences in processing conditions that are difficult to control for will be reflected in the resulting gene expression pattern. It is therefore recommended that gene expression profiling should only be performed on the original plant/plant parts or raw materials and not on the products derived thereof.

Another crucial factor is the sampling of mRNA populations to be investigated in relation to the characteristics of the array used. Usually the plant parts of interest will contain a number of different tissue types. Equal representation of all different tissue types in the experimental samples is very important. Furthermore, all key metabolic routes that require analysis need to be represented on the array. In the future the use of whole transcriptome arrays or well-characterised focused arrays will make it easy to check for this requirement, but nowadays careful analysis of the libraries used on the array will be necessary.

An average DNA microarray experiment results a huge number of gene expression data which need to be analysed. It is important to include quality checks at all stages of the experiment to avoid inclusion of inferior data in the analysis. The subsequent steps are generally quantification of the signal, background subtraction, normalisation, and visualisation. All of these steps can be done in numerous ways. The necessity to carefully document the followed route from raw data to conclusions in an experiment is therefore evident, including the necessity to have full insight into the subsequent data analysis steps. Current microarray data analysis and visualisation software are very user-friendly, but in practice it still remains the challenge to find the optimal route of analysis for every individual experiment.

Also, it is important to note that the analysis of DNA microarray data can be expected only to result in indications of possibly altered gene expression in the novel plant variety compared to the traditional counterpart. Additional targeted investigations will usually be required to confirm the results. These additional experiments may be based on more focused molecular biological methodologies, for instance real-time PCR or

Northern blotting, or on the targeted assessment of the related effects at the level of proteins or metabolites.

It may prove necessary to establish more exact (regulatory) criteria to determine differential gene expression in the GM plant versus the parent line. The practical approach is a two-step approach: first gene expression patterns of the GM plant are compared to the parent line or, if the parent line is not available, any other near comparator. If differences are detected, the levels are compared to the knowledge on the natural variation of the specific compound, mRNA, in other traditional varieties that are already on the market. It can, however, be imagined that changes in the GM plant versus the parent line that fall within the range of the natural variation of traditional varieties may, however, be of a nature that should nevertheless require further study.

The resulting data of microarray experiments may be informative not only to the scientists performing the experiment, but also to the broader scientific community. Therefore initiatives are ongoing to set up large databases with strict quality control criteria for microarray gene expression data. MIAME-compliant data reporting of gene expression experiments is now already becoming standard in the scientific literature. The objective of the MIAME initiative is to standardise the information related to any individual experiment in order to prevent loss of information in the future due to the incompatibility of different systems. In addition it is recommended to start considering the possibility of linking MIAME-type databases with data from other sources representing other 'omics'-methodologies or techniques such as real-time PCR, Northern blotting or individual protein analysis to avoid future incompatibilities between the different systems. For the purpose of the food safety analysis of new plant varieties it is important to analyse different integration levels in order to obtain optimal information on potential unintended side effects as a result of the selected breeding strategy.

Other 'omics' technologies, such as proteomics, metabolomics and glycomics are also becoming increasingly advanced and informative therewith. Gene expression profiling now has the advantage that it can basically cover the whole transcriptome of individual species, tissues or cell systems. The moment, however, similar informative methodologies have been developed for the proteome and metabolome the added

value of gene expression profiling may be reduced. In the meantime we can learn a lot from the potential of microarray experiments to elucidate metabolic networks and the documentation of the natural variation in gene expression in traditional crop plant varieties as well as of observed differences in novel (GM) varieties versus their parent lines.

Chapter 8 Conclusions

The information in the previous chapters about the DNA microarray technique and the discussions in the working group lead to conclusions about the usefulness of this technique in safety assessment of food from GM-plants.

Some of the conclusions are based on the fact that this technique is new and experiences are limited. These conclusions may therefore be of a general nature for new techniques. Others are linked to the use of DNA microarray technology with mRNA as a target for analysis. The analysis of RNA has the weakness that difference in mRNA level between samples is not a health issue in itself and they are not directly predictive for the level of proteins formed or quantities of metabolic product and therefore any final health issue.

When a comparison is done between plants in the same experiment and especially between two experiments, the observed differences can easily be due to small unintended variations in the experimental environment such as soil and temperature, damaged by insects, fungi etc. This implies that much knowledge is needed about how these factors may influence the results of the analysis and how this interference can be overcome before any conclusions can be made.

The most important conclusions that emerge from the present report are:

The quality of the mRNA populations to be hybridised in microarray experiments is crucial for the reliability of the resulting data. The mRNA has to be of high quality in order to compare two subsequent samples. Therefore standardisation is essential for the different steps in the microarray analysis including procedures for sampling, mRNA isolation and quality control.

The quantitative correlation between mRNA, proteins and secondary metabolites is very complex which hinders the interpretation of mRNA microarray results and particularly their consequences. This barrier for interpretations can be expected to diminish as the knowledge about this complex system increase.

The benefit from analyzing for thousands of genes/mRNA might be limited because of their minor impact on health-related aspects or the lack of knowledge about the correlation between the mRNA levels and safety issues. However the microarray has the potential, unlike proteomics or metabolomics, to screen multiple networks for unintended alterations as a result of modifications in a plant.

Gene expression profiling seems only applicable on the original plant/plant parts or raw materials and not on the products derived thereof as slightly different conditions in the food or food product will generally have large effects on the quality of the isolated mRNA.

Result from the RNA microarray analysis should be considered only as suggestive due to e.g. the incomplete knowledge about the natural variation.

Although the RNA microarray technique has the great potential in unbiased screening for unintended effect, the technique has not yet proved its relevance in risk assessment.

For the purpose of the food safety analysis of new plant varieties, including GM plants, it is important to analyse on other integrative levels, such as the protein or metabolite level in order to obtain optimal information on potential side effects as a result of the breeding strategy.

As the other “omics” are under development, such as the proteome microarrays and metabolomics, this may lead in future to ‘whole proteome’ or “whole metabolome” approaches that may reduce the need for gene expression profiling at the mRNA level.

The value of microarray will probably be highest in the developmental phase of food crops by suggesting metabolic reactions or pathways that should be explored more closely by more targeted methods.

The potential for mRNA microarray technology is considerable in specific applications as it miniaturises mRNA analysis, thereby enabling the establishment of gene expression profiles of individual mRNA samples on the basis of thousands of different genes, representing numerous different metabolic routes.

For the safety assessment of novel plant varieties, including GM plants, relevant metabolic routes may be 1) nutrient-related pathways, 2) metabolic pathways that can be used to screen for differences in the basal physiology of the plants and 3) stress-related metabolic pathways as it is known that upregulation of these pathways may lead to an increased production of anti-nutrients, including natural toxins. If the microarray is to be used for safety assessment, arrays comprising these three different categories of probes might be more applicable for safety assessment as the probes have been chosen for their relevance to safety issues.

For GM-plant evaluation, the microarray methods and data documentation need to be standardised.

Availability of “standard microarrays” for the individual plants, representing relevant plant part such as leaves, fruits and roots will allow cooperation in this field. The same apply to standards for data storage, data retrieval and availability of those data in the public domain.

Glossary

Abscisic acid	Plant hormone
ANN	Artificial Neural Network, computer algorithm for pattern recognition in large datasets, either unsupervised (SOM) or supervised.
Annotation	Recording of a gene's DNA sequences and eventually their chromosomal position and function
ANOVA	Analysis of variance, statistical method for validation of potential difference in a variable in multiple samples.
Anthesis	Anther formation, <i>i.e.</i> reproductive organ of plant flower
cDNA	Complementary DNA, in vitro made DNA molecule, contains the information in a mRNA molecule transferred to a DNA backbone.
Dendrogram	Schematic overview of the relationships between various variables or samples in an experiment, result of hierarchical cluster analysis.
DFA	Discriminant Function Analysis, supervised method for pattern recognition in large datasets.
DNA	Deoxyribonucleic acid, contains genetic information in living cells.
Down regulation	Process that causes a decrease of gene expression
DTL	Decision Tree Learning, supervised method for pattern recognition in large datasets.
EBI	European Bioinformatics Institute, participates in MGED.
Enzyme	Protein that catalyses a specific process in living cells or in vitro.
EST	Expressed Sequence Tag, contains the sequence of part of a gene that has not yet been annotated, known to be expressed in certain circumstances in a particular tissue.
Ethylene	Plant hormone
Evolutionary computing	Supervised method for pattern recognition in large datasets.
FSA	Food Standards Agency (UK).
Glycome	The total of sugar molecules per cell.
Hierarchical clustering	Unsupervised method for pattern recognition in large datasets.

Housekeeping genes	Genes that encode for enzymes, RNA, or other functional biomolecules with elementary roles in the maintenance and growth of cells
Jasmonic acid	Plant hormone
MAGE	Microarray gene expression, MGED working group for the establishment of a uniform data exchange format for microarray experiments.
Metabolome	The total of metabolites per cell.
Methyl jasmonate	Plant hormone
MGED	Microarray gene expression society, international group of scientists aiming towards a consensus on microarray experiments.
MIAME	Minimal information about a microarray experiment, guidelines for the information that should be provided in order to peer-review microarray experiments, formulated by the MGED.
Microarray	Collection of large amounts of different molecules, organised in spots, immobilised on various types of surfaces.
mRNA	messenger RNA, transcript of the DNA, encoding a specific gene.
Mutagenesis	Induction of nucleotide sequence mutations, such as with the aid of chemicals or radiation
MS	Mass Spectrometry, technique for identification of molecules based on their difference in mass.
NCBI	National Centre for Bioinformatics, American organisation for data storage and analysis of sequence databases and microarray experiments.
NMR	Nuclear Magnetic Resonance, technique for identification of molecules, based on the different magnetic behaviour of hydrogen (or other) nuclei in different molecules.
OWG	Ontology working group, MGED working group for the development of a common vocabulary for microarray experiment description.
PCA	Principal Component Analysis, unsupervised method for pattern recognition in large datasets.
PCR	Polymerase Chain Reaction, method for the amplification of a specific DNA fragment.

Probe	'Known' DNA sequence that is used in molecular biological analysis
promoter	DNA sequence in front of a gene, regulating the expression of that gene.
Proteome	The total of proteins per cell
Q-RT-PCR	Quantitative reverse transcriptase PCR, process of quantifying mRNA molecules, after converting them to cDNA.
Regulon	Genes with common regulation patterns
Responsive gene	Gene of which the expression is altered in response to a specific stimulus/effector
RNA	Ribonucleic acid, molecules in living cells, with different functions involved in transferring the information present in the DNA to functional proteins.
RT	Reverse transcriptase, enzyme used for the conversion of mRNA to cDNA.
Salicylic acid	Plant hormone
SAR	Systemic acquired resistance, <i>i.e.</i> non specific defense reactions, such as plant tissue destruction, in response to systemically delivered effector molecules
Stress	Physiological condition that demands an above normal turnover of energy and/or intrinsic substances, such as during infection by a pathogen
Target DNA	(Pool of) labelled DNA fragments that are hybridised to the array
TIGR	The Institute for Genomic Research
Transcription	Process in the cell when information is transferred from DNA to mRNA.
Transcriptional factor	Protein that controls the transcription of (a set of) specific genes by binding regulatory DNA sequences
Transcriptome	The total of mRNAs per cell
Translation	Process in the cell when information is transferred from mRNA to proteins.
Upregulation	Process that causes a increase of gene expression

References

Adler K., Broadbent J., Garlick R., Joseph R., Khimani A., Mikulskis A., Rapiejko P., and Kilian J. MICROMAX: a highly sensitive system for differential gene expression on microarrays. *Microarray Biochip Technology* (2000), p. 221- 230. Biotechniques Books, Natick, MA. Mark Schena (ed.).

Aharoni A., and Vorst O. (2001) DNA microarrays for functional plant genomics. *Plant Molecular Biology* 48:99-118.

Aharoni, A., Keizer, L.C.P., Van Den Broeck, H.C., Blanco-Portales, R., Munoz-Blanco, J., Bois, G., Smit, P., De Vos, R.C.H., O'Connell, A.P. (2002) Novel insight into vascular, stress, and auxin-dependent and -independent gene expression programs in strawberry, a non-climacteric fruit. *Plant Physiology* 129: 1019-1031.

Alwine J.C., Kemp D.J., and Stark G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridisation with DNA probes. *Proc. Natl. Acad. Sci. USA* 74, 5350-5354.

Alizadeh A.A., Eisen M.B., Davis R.E., Ma C., Lossos I.S., Rosenwald A., Bodrick J.G., Sabet H., Tran T., Yu X., Powell J.I., Yang L.M., Marti G.E., Moore T., Hudson J., Lu L.S., Lewis D.B., Tibshirani R., Sherlock G., Chan W.C., Greiner T.C., Weisenburger D.D., Armitage J.O., Warnke R., Levy R., Wilson W., Grever M.R., Byrd J.C., Botstein D., Brown P.O., and Staudt L.M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503-511.

Anderson J.S., and Mann M. (2000) Functional genomics by mass spectrometry. *FEBS Letters* 480:25-31.

Becker K.G. (2001) The sharing of cDNA microarray data; *Nature reviews neuroscience* 2, 438-440

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.

Brazma A. (2001). On the importance of standardisation in life sciences. *Bioinformatics* 17, 113-4.

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. *Nat Genet.* 29, 365-71.

Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D.H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S.R., Moon K., Burcham T., Pallas

M., DuBridg e R.B., Kirchner J., Fearon K., Mao J., and Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18 (2000) 630-634.

Chen, W., Provart, N.J., Glazebrook, J., Katagiri, F., Chang, H.S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S.A., Budworth, P.R., Tao, Y., Xie, Z., Chen, X., Lam, S., Kreps, J.A., Harper, J.F., Si-Ammour, A., Mauch-Mani, B., Heinlein, M., Kobayashi, K., Hohn, T., Dangl, J.L., Wang, X., Zhu, T. (2002) Expression profile matrix of Arabidopsis transcription factor genes suggest their putative functions in response to environmental stress. *Plant Cell* 14: 559-574.

Constans A. A new approach to gene expression analysis. *The Scientist* 16 (2002) 8: 44.

Defernez, M., Colqhoun, I.J. (2003) Factors affecting the robustness of metabolite fingerprinting using 1H NMR spectra. *Phytochemistry* 62: 1009-1017.

Dell and H.R. Morris. Glycoprotein structure determination by mass spectrometry. *Science* 291 (2001) 2351-2356

Edelstein R.L., Tamanaha C.R., Sheehan P.E., Miller M.M., Baselt D.R., Whitman L.J., Colton R.J. (2000) The BARC biosensor applied to the detection of biological warfare agents. *Biosensors and Bioelectronics* 14,10-11:805-813.

EU Regulation (EC) No 258/97 of the European Parliament and of the Council of 27 January 1997 concerning novel foods and novel food ingredients. (1997a) *Off. J. Eur. Commun.* L43, 1-7. http://europa.eu.int/eur-lex/en/lif/dat/1997/en_397R0258.html.

EU Regulation 97/618/EC: Commission Recommendation of 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) 258/97 of the European Parliament and of the Council. (1997b) *Off. J. Eur. Commun.* L253, 1-36. http://europa.eu.int/eur-lex/en/lif/dat/1997/en_397X0618.html.

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, 29 May- 2 June 2000. Rome: Food and Agriculture Organisation of the United Nations. <http://www.fao.org/es/esn/gm/biotece.htm>.

Fiehn O., Kopka J., Dormann P., Altmann T., Trethewey RN., Willmitzer L. (2000) Metabolite profiling for plant functional genomics. *Nature Biotechnology* 18:1157-61

Franssen - van Hal N.LW., Vorst O., Kramer E., Hall R.D., and Keijer J. (2002) Factors influencing cDNA microarray hybridization on silylated glass slides. *Analytical Chemistry* 308:5-17.

Gallie, D.R. (2002) Protein-protein interactions required during translation. *Plant Molecular Biology* 50: 949-970.

- Gry J., S. Holm, M. Morgan, A. Møller, R. Preece, G. Speijers and I. Søbørg (2002) Bioactive Substances in Food Plants Information System (BASIS). BASIS FINAL REPORT. Danish Veterinary and Food Administration, Søborg, Denmark.
- Gygi S.P., Corthals G.L., Zhang Y., Rochon Y., and Aebersold R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci USA* 97:9390-9395.
- Hadacek, F. (2002) Secondary metabolites as plant traits: current assessment and future perspectives. *Crit. Rev. Plant Sci.* 21: 273-322.
- Heath, L.S., Ramakrishnan, N., Sederoff, R.R., Whetten, R.W., Chevone, B.I., Struble, C.A., Jouenne, V.Y., Chen, D., Van Zyl, L., Grene, R. (2002) Studying the functional genomics of stress responses in loblolly pine with the Espresso microarray experiment management system. *Comparative and Functional Genomics* 3: 226-243.
- Heazlewood, J.L., Millar A.H. (2003) Integrated plant proteomics - putting green genomics to work. *Functional Plant Biology* 30, 471-482.
- Hess K.R., Zhang W., Baggerly K.A., Stivers D.N. and Coombes K.R. (2001) Microarrays: handling the deluge of data and extracting reliable information; *Trends in Biotechnology* 19, 463-468.
- Jordan, B.R. (2002) Molecular response of plant cells to UV-B stress. *Functional Plant Biology* 29: 909-916.
- Kane M.D., Jatkoe T.A., Stumpf C.R., Lu J., Thomas J.D. and Madore S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res.* 28(22):4552-4557.
- Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K., Galbraith D., and Bohnert H.J. (2001) Gene expression profiles during the initial phase of salt stress in rice. *The Plant Cell* 13:889-905.
- Kehoe D.M., Volland P., and Somerville S. (1999) DNA microarrays for studies of higher plants and other photosynthetic organisms. *Trends in Plant Science* 4:38-41.
- Kok E.J., Keijer J., Van Hoef A.M.A., Van der Wal - Winnubst E.N.W., Henkens M.H.C., and Kuiper H.A. (1998) mRNA fingerprinting of transgenic food crops. Report of the demonstration programme on Food Safety Evaluation of Genetically Modified Foods as a Basis for Market Introduction, Ministry of Economic Affairs, The Hague, the Netherlands, pp. 37-49.
- Kok E.J., Van der Wal - Winnubst E.N.W., Van Hoef A.M.A. and Keijer J. (2001) Differential display of mRNA. *Molecular Microbial Ecology Manual* 8.1.1.:1-10.
- 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. *The Plant Journal* 27(6)503-528.

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.

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

Le Gall G., Colquhoun I.J., Davis A.L., Collins G.J., Verhoeyen M.E. (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using ¹H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *Journal of Agricultural and Food Chemistry* 51:2447-2456 44: 1967-1974.

Liang P. and Pardee A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257,967-971.

Lockhart D.J. and Winzeler E.A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405 (6788)827-836.

MacBeath, G., and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760-1763

Maleck K., Levine A., Eulgem T., Morgan A., Schmid J., Lawton K.A., Dangl J.L., and Dietrich R. A. (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature genetics* 26:403-410.

Meijer, H.A., Thomas, A.A.M. (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochemical Journal* 367: 1-11.

Meinke D.W., Cherry J.M., Dean C., Rounsley S.D., Koornneef M. (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282:679-682.

Mitra, M., Shah, N., Mueller, L., Pin, S., Fedoroff, N. (2002) StressDB: a locally installable web-based relational microarray database designed for small-user communities. *Comparative and Functional Genomics* 3: 91-96.

Moore S., Vrebalov J., Payton P., Giovannoni J. (2002) Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *Journal of Experimental Botany* 53:2023-2030

Nadon R., and Schoemaker J. (2002) Statistical issues with microarrays: processing and analysis; *Trends in genetics* 18, 265-271.

Nair, R.S., Fuchs, R.L., Schuette, S.A. (2002) Current methods for assessing safety of genetically modified crops as exemplified by data on Roundup Ready soybeans. *Toxicologic Pathology* 20: 117-125.

Nakamura Y., Gojobori T., and Ikemura T. (2000) Codon usage tabulated from international DNA sequence database: status for the year 2000. *Nucleic Acids Research* 28:1, 292.

Negishi, T., Nakanishi, H., Yazaki, J., Kishimoto, N., Fujii, F., Shimbo, K., Yamamoto, K., Sakata, K., Sasaki, T., Kikuchi, S., Mori, S., Nishizawa, N.K. (2002) cDNA microarray analysis of gene expression during Fe-deficiency stress in barley suggests that polar transport of vesicles is implicated in phytosiderophore secretion in Fe-deficient barley roots. *Plant Journal* 30: 83-94.

NETTOX 1998: Inherent Food Plant Toxicants. Gry J., A. Kovatsis, W. Jongen, A. Møller, M. Rhodes, E. Rosa, H. Rosner, G. Speijers, I. Søborg and A. Walker (eds.) vol 1-9. Danish Veterinary and Food Administration, Søborg, Denmark. [The report is expected on the internet in 2004. For further information contact J. Gry jg@dfvf.dk]

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

Nolan J.P., and Sklar L.A. (2002) Suspension array technology: evolution of the flat-array paradigm. *Trends in Biotechnology* 20,1:9-12

Nordic Council (1998) Safety assessment of novel food plants - chemical analytical approaches to the establishment of substantial equivalence. *TemaNord* 1998:591.

Noteborn H.P., Lommen A., van der Jagt R.C., Weseman J.M. (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *Journal of Biotechnology* 77:103-114

OECD (1993) Safety evaluation of Foods Derived by Modern Biotechnology. Paris: Organisation for Economic Cooperation and Development. [Http://www.oecd.org/dsti/sti/st/biotech/prod/modern.htm](http://www.oecd.org/dsti/sti/st/biotech/prod/modern.htm).

OECD (1996) Food Safety Evaluation. Paris: Organisation for Economic Cooperation and Development.

OECD (1998) Report of the OECD workshop on the toxicological and nutritional testing of novel foods, Aussois, France, 5-8 March 1997. Paris: Organisation for Economic Cooperation and Development. [Http://www.oecd.org/ehs/ehsmono/aussoidrEN.pdf](http://www.oecd.org/ehs/ehsmono/aussoidrEN.pdf).

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/document/9/0,2340,en_2649_34385_1812041_1_1_1_1,00.html

Ott K.H., Aranibar N., Singh B., Stockton G.W. (2003) Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* 62:971-985

- Ouellet, F. (2002) Out of the cold: unveiling the elements required for low temperature induction of gene expression in plants. *In Vitro Cellular and Developmental Biology - Plant* 38: 396-403.
- Ozturk, Z.N., Talamé, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R., Bohnert, H.J. (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Molecular Biology* 48: 551-573.
- Parinov S., and Sundaresan V. (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. *Current Opinion in Biotechnology* 11:157-161.
- Park S-J., Taton T.A., and Mirkin C.A. (2002) Array-based electrical detection of DNA with nanoparticle probes. *Science* 295:1503-1506
- Pereira A. (2000) A transgenic perspective on plant functional genomics. *Transgenic Research* 9:245-260
- Perkel J.M. (2002) *The Scientist* 16(9): 32.
- Phelps T.J., Palumbo A.V., Beliaev A.S. (2002) Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Current Opinion in Biotechnology* 13:20-24
- Pradet-Balade B., Boulmé F., Beug H., Müllner E.W. and Garcia-Sanz A. (2001) Translation control: bridging the gap between genomics and proteomics? *Trends in Biochemical Science* 26:225-229
- Regierer B., Fernie A.R., Springer F., Perez-Melis A., Leisse A., Koehl K., Willmitzer L., Geigenberger P., and Kossmann J. 2002 Starch content and yield increase as a result of altering adenylate pools in transgenic plants. *Nature Biotechnology* 20:1256-1260
- Reymond, P. (2001) DNA microarrays and plant defence. *Plant Physiology and Metabolism* 39: 313-321.
- Richmond T., and Somerville S. (2000) Chasing the dream: plant EST microarrays. *Current Opinion in Biotechnology* 3:108-116.
- Ridley W.P., Sidhu R.S., Pyla P.D., Nemeth M.A., Breeze M.L., and Astwood J.D. (2002) Comparison of the Nutritional Profile of Glyphosate-Tolerant Corn Event NK603 with That of Conventional Corn (*Zea mays* L.) *Journal of Agricultural and Food Chemistry* 50:7235-7243.
- Roessner U., Luedemann A., Brust D., Fiehn O., Linke T., Willmitzer L., Fernie A. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13:11-29
- Rogan, G.J., Bookout, J.T., Duncan, D.R., Fuchs, R.L., Lavrik, P.B., Love, S.L., Mueth, M., Olson, T., Owens, E.D., Raymond, P.J., Zalewski, J. (2000) Compositional analysis of tubers from insect and

virus resistant potato plants. *Journal of Agricultural and Food Chemistry* 48: 5936-5945.

Ronning C.M., Stegalkina S.S., Ascenzi R.A., Bougri O., Hart A.L., Utterbach T.R., Vanaken S.E., Riedmuller S.B., White J.A., Cho J., Perteau G.M., Lee Y., Karamycheva S., Sultana R., Tsai J., Quackenbush J., Griffiths H.M., Restrepo S., Smart C.D., Fry W.E., Van der Hoeven R., Tanksley S., Zhang P., Jin H., Yamamoto M.L., Baker B.J., and Buell C.R. Comparative analyses of potato Expressed Sequence Tag libraries. *Plant Physiology* 131 (2003): 419-429.

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.

Schaffer R., Landgraf J., Perez-Amador M., and Wisman E. (2000) Monitoring genome-wide expression in plants. *Current Opinion in Biotechnology* 11:162-167.

Regierer B., Fernie A.R., Springer F., Perez-Melis A., Leisse A., Koehl K., Willmitzer L., Geigenberger P., Kossmann J. (2002) Starch content and yield increase as a result of altering adenylate pools in transgenic plants. *Nature Biotechnology* 20:1256-1260

Rudd S., Mewes H.-W., Mayer K.F.X. (2003) Sputnik: a database platform for comparative plant genomics. *Nucleic Acids Research* 31,1:128-132.

Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C., Manners, J.M. (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proceedings of the National Academy of Sciences USA* 97: 11655-11660.

Schwab W. (2003) Metabolome diversity: too few genes, too many metabolites? *Phytochemistry* 62:837-849

Serikawa, K.A., Xu, X.L., MacKay, V.L., Law, G.L., Zong, Q., Zhao, L.P., Bumgarner, R., Morris, D.R. (2003) The transcriptome and its translation during recovery from cell cycle arrest in *Saccharomyces cerevisiae*. *Molecular and Cellular Proteomics* 2: 191-204.

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

Sterrenburg E., Turk R., Boer J.M., Van Ommen G.B., and Den Dunnen J.T. A common reference for cDNA microarray hybridisations. *Nucleic Acids Res* (2002) 30(21)e116.

Stich N., Gandhum A., Matushin V., Mayer Ch., Bauer G., and Schalkhammer T. (2001) Nanofilms and nanoclusters: energy sources driving fluorophores of biochip bound labels. *Journal of Nanoscience and nanotechnology*, 1:1:1-9.

Sumner L.W., Mendes P., Dixon R.A. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era.

Phytochemistry 62:817-836

Swidzinski, J.A., Sweetlove, L.J., Leaver, C.J. (2002) A custom microarray analysis of gene expression during programmed cell death in *Arabidopsis thaliana*. *Plant Journal* 30: 431-446.

Templin M.F., Stoll D., Schrenk M., Traub P.C., Voehringer C.F., and Joos T.O. (2002) Protein microarray technology. *TIBTECH* 20, 160-166

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

Thomashow, M.F., Gilmour, S.J., Stockinger, E.J., Jaglo-Ottosen, K.R., Zarka, D.G. (2001) Role of the Arabidopsis CBF transcriptional activators in cold acclimation. *Physiologia Plantarum* 112: 171-175.

Van Hal N.L.W., Vorst O., van Houwelingen A.M.M.L., Kok E.J., Peijnenburg A., Aharoni A., Van Tunen A.J., Keijzer J. (2000) The application of DNA microarrays in gene expression analysis. *J. Biotechnol.* 78: 271-280.

Velculescu V.E., Zhang L., Vogelstein B., and Kinzler K.W. (1995) Serial analysis of gene expression. *Science* 270, 484-487.

Ward J.L., Harris C., Lewis J., Beale M.H. (2003) Assessment of 1H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* 62:949-57

Washburn M.P., Koller A., Oshiro G., Ulaszek R.R., Plouffe D., Deciu C., Winzeler E. and Yates III J.R. (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* 100: 3107-3112.

Welsh J. and McClelland M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18, 7213-7218.

Welsh J., Chada K., Dalal S.S., Ralph D., Cheng L., and McClelland M. (1992) Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res.* 20, 4965-4970.

Workman Ch., Jensen L.J., Jarmer H., Berka R., Gautier L., Nielsen H.B., Saxild H.-H., Nielsen C., Brunak S., and Knudsen S. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology* 3,9:0048.1-0048.16.

Xu W., Bak S., Decker A., Paquette S.M., Feyereisen R. and Galbraith D.W. (2001). Microarray-based analysis of gene expression in very large gene families: the Cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene* 272:61-74.

Yang Y.H., Dudoit S., Luu P., Lin D.M., Peng V., Ngai J., and Speed T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systemic variation. *Nucleic Acids Research* 30,4:15, e15.

Young, N.D. The genetic architecture of resistance. *Current Opinion in Plant Biology* 3 (2000) 285-290.

Yuan Q., Ouyang S., Liu J., Suh B., Cheung F., Sultana R., Lee D., Quackenbush J., and Buell C.R. The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic Acids Research* 31(2003)1:229-233.

Ziauddin J. and Sabatini D.M. Microarrays of cells expressing defined cDNAs. *Nature* 411 (2001):107-110.

Zhu T., and Wang X. (2000) Large-scale profiling of the *Arabidopsis* genome. *Plant Physiology* 124:1472-1476.

Zhu, J.K. (2002) Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology* 53: 247-273.

Zinselmeier, C., Sun, Y., Helentjaris, T., Beatty, M., Yang, S., Smith, H., Habben, J. (2002) The use of gene expression profiling to dissect the stress sensitivity of reproductive development in maize. *Field Crops Research* 75: 111-121.

Obtain your Nord and TemaNord reports from:

BELGIUM & LUXEMBOURG

Jean de Lannoy
Avenue du Roi, 202, 1190 Brussels
Tel +32 (0)2 538 5169 Fax +32 (0)2 538 0841
jean.de.lannoy@euronet.be

CANADA

Renouf Publishing Company Ltd
5369 Canotek Road, Ottawa, Ontario K1J 9J3
Tel + 1 (613) 745 2665 Fax + 1 (613) 745 7660
order.dept@renoufbooks.com
www.renoufbooks.com

CHINA

C N P I E C
Europe Division 16 Gongti East Road, P.O. Box 88, Beijing
Tel +86 10 50 66 688-8 Fax +86 10 50 63 101

DENMARK

Svensk-Norsk Bogimport A/S
Esplanaden 8 B, 1263 København K
Tel +45 33 14 26 66 Fax +45 33 14 35 88
snb@bog.dk
www.snbog.dk

ESTLAND

Astro Raamatud AS
Pärnu mnt 142, 11317 Tallinn
Tel +372 654 8485 Fax +372 654 8475
book@astro.ee

FAROE ISLANDS

H.N. Jacobsens Bókahandil
Postboks 55, 110 Tórshavn
Tel +298 31 10 36 Fax +298 31 78 73
hnj-bokh@post.olivant.fo

FINLAND

Akademiska Bokhandeln
PB 128, Centralgatan 1, 00101 Helsingfors
Tel +358 9 12141
akatilaus@akateeminen.com
www.akateeminen.com

FRANCE

Librairie LAVOISIER
14, rue de Provigny, 94236 Cachan Cedex
Tel +33 (1) 4740 6700 Fax +33 (1) 4740 6702
group@lavoisier.fr
www.lavoisier.fr

GERMANY

UNO-Verlag GmbH
Am Hofgarten 10, 53113 Bonn
Tel +49 (0)228 949020 Fax +49 (0)228 9490 222
info@uno-verlag.de
www.uno-verlag.de

HUNGARY

Euro Info Service
PO Box 1039, 1245 Budapest
Tel +36 (1) 329 2487 Fax +36 (1) 349 2053
euroinfo@euroinfo.hu

ICELAND

Mál og Menning

Laugavegi 18, 101 Reykjavik
Tel +354 (9)515 2500 Fax +354 (9)515 2505
verslun@mm.is

LATVIA

Jana Rozes Gramātnica
Kr. Barona iela 5, 1011 Riga
Tel +371 (0)2 284288 Fax +371 7 370 922

LITHUANIA

Penki Kontinentai
A. Stulginskio 5, 2001 Vilnius
Tel +370 (5) 2664540 Fax +370 (5) 2664565
books@5ci.lt
www.books.lt

NORWAY

Akademika A/S
Postboks 84 Blindern, 0314 Oslo
Tel +47 22 85 30 30 Fax +47 22 85 30 80
bloken@sio.uio.no
www.akademika.no

ROMANIA

Euromedia s.r.l.
Str Dionisie Lupu nr 65, 70184 Bucuresti
Tel + 40 1 614 06 64 Fax + 40 1 312 96 46

SWEDEN

Fritzes
Kundservice, 106 47 Stockholm
Tel +46 (0)8 690 9190 Fax +46 (0)8 690 9191
order.fritzes@nj.se
www.fritzes.se

THE NETHERLANDS

De Lindeboom Internationale Publikaties b.v.
Postbus 202
7480 AE Haaksbergen
Tel +31 (053) 57 40004 Fax +31 (053) 57 29296
books@delindeboom.com
www.delindeboom.com

UNITED KINGDOM

The Stationery Office
P.O. Box 276, London SW8 5DT
Tel +44 870 600 5522 Fax +44 870 600 5533
customer.services@tso.co.uk
www.tso.co.uk/bookshop

USA

Bernan
4611-F Assembly Drive, Lanham MD 20706-4391
Tel +1 (301) 459 7666 Fax +1 (301) 459 0056
query@bernan.com
www.bernan.com

ÅLAND

Lisco bok- och pappershandel
Skarpansvägen 25, Box 8, 22101 Mariehamn
Tel +358 (0)18 17 177 Fax +358 (0)18 19 771
info@lisco.fi