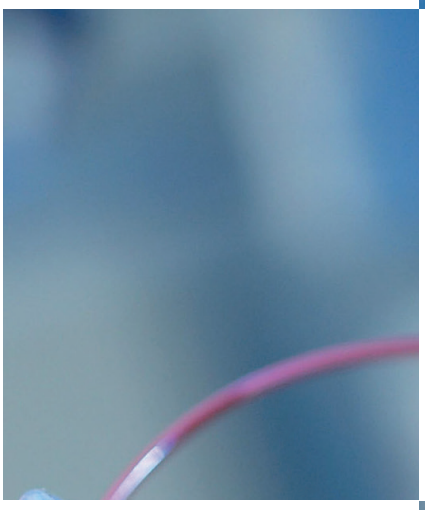
The contribution of metabolomics research to the environmental risk assessment of genetically modified plants



CGM 2013-04

ONDERZOEKSRAPPORT

PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

The contribution of metabolomics research to the environmental risk assessment of genetically modified plants

A literature review

Ruud A. de Maagd & Robert D. Hall

Commissioned by

The Netherlands Commission on Genetic Modification (COGEM) and the GMO office of the National Institute for Public Health and the Environment (RIVM)

Clusters Plant Development Systems and Metabolic Regulation, BU Bioscience, Plant Research International, Wageningen University and Research Centre, Wageningen, The Netherlands

Plant Research International, part of Wageningen UR
Business Unit Bioscience
September 2013

Report 528

Plant Research International, part of Wageningen UR
Business Unit Bioscience

Address : Droevendaalsesteeg 1, Wageningen, The Netherlands
: P.O. Box 619, 6700 AP Wageningen, The Netherlands
Tel. : +31 317 48 60 01
Fax : +31 317 41 80 94
E-mail : info.pri@wur.nl
Internet : www.pri.wur.nl

PREFACE

Assigned by the Netherlands Commission on Genetic Modification (COGEM) and the National Institute for Public Health and the Environment (RIVM), an exploratory desk study was performed concerning "The contribution of metabolomics research to the environmental risk assessment of genetically modified plants."

This report describes the state of the art in metabolomics research and the expected developments in this research field in the future. Metabolomics aims to give a wide overview of all metabolites present in a specific sample of a plant at a specific time. In the interaction with the environment changes in metabolite composition play an important role. Such composition might be influenced by genetic modification, both intentionally and unintentionally. This raised the question whether metabolomics could be used in the risk analysis of transgenic plants and in this way improve the existing Environmental Risk Assessment (ERA) procedures. Such studies on metabolomics are not yet required for the assessment.

The main conclusion of the report is that metabolomics is a fast developing discipline and might in the future contribute to ERA of transgenic crops; however, the potential contribution to ERA is limited for the time being. This is mainly due to the fact that there is limited knowledge of the role of any particular metabolite in the interaction of a particular plant with a particular component of the environment.

The study was performed by Dr. Ruud de Maagd en Dr. Robert Hall of Plant Research International, part of Wageningen UR, and was supervised by a committee consisting of experts in the field of metabolomics and GM regulations.

Chairman Advisory Committee
Prof. dr. J.J.M. Dons

ADVISORY COMMITTEE
Prof. Dr. J.J.M. Dons, BioSeeds B.V., member of COGEM
Prof. Dr. R. Verpoorte, Leiden University
Dr. J.M. Kooter, Vrije Universiteit Amsterdam, member of COGEM
Prof. Dr. Ir. Nicole van Dam, department of Ecogenomics, Radboud University Nijmegen
Dr. D.C.M. Glandorf, GMO office, National Institute of Public Health and the Environment
B. Erkamp MSc, COGEM Secretariat

Table of contents

DISCLAIMER

This report was commissioned by the COGEM and the GMO Office of the National Institute for Public Health and the Environment (RIVM). The contents of this publication are the sole responsibility of the author and do not necessarily reflect the views of COGEM or the GMO office

Dit rapport is in opdracht van de Commissie Genetische Modificatie en het Bureau GGO samengesteld. De meningen die in het rapport worden weergegeven zijn die van de auteurs en weerspiegelen niet noodzakelijkerwijs de mening van de COGEM of Bureau GGO.

ACKNOWLEDGEMENTS

We gratefully acknowledge the members of the advisory committee for the valuable discussions.

Samenvatting

Doel van de studie
De snelle ontwikkeling van “omics” technologieën roept de vraag op of, en zo ja in welke mate, deze zouden kunnen bijdragen aan verbetering van de milieurisicobeoordelingprocedures voor de toelating van genetisch gemodificeerde planten. Dit rapport beschrijft de resultaten van een literatuurstudie gericht op de vraag “Wat is de potentiele bijdrage van “metabolomics” aan de milieurisicobeoordeling van genetisch gemodificeerde (GM) planten?” Metabolomics is de Engelse term voor de studie van kleine organische moleculen (metabolieten) in planten, dieren of micro-organismen.

Mogelijke scenario’s met relevantie voor milieurisicobeoordeling
In een gepubliceerde leidraad heeft de Europese voedselveiligheidautoriteit (EFSA) de te volgen procedure voor milieurisicobeoordeling geschetst. Hierin worden zeven aandachtsterreinen geïdentificeerd, waarvan de meeste relevant zijn voor deze studie. “Vergelijkende veiligheidsbeoordeling” van GM planten dient plaats te vinden met passend vergelijkingsmateriaal (niet-GM planten). Vier methoden om onbedoelde effecten van GM planten te kunnen detecteren worden genoemd, van welke “Compositionele Analyse” de meest relevante is en met behulp van metabolomics kan worden uitgevoerd. Deze analyse is op dit moment nog niet verplicht in de Europese milieurisicobeoordeling. Transgene gewassen kunnen bedoelde (waarvoor de modificatie was ontworpen) en ook onbedoelde veranderingen hebben, die ook effect kunnen hebben op metabolieten. De onbedoelde veranderingen kunnen verder onderverdeeld worden in “voorspelbare” (konden voorspeld worden op basis van onze huidige kennis) en “onvoorspelbare” veranderingen.
In dit rapport worden zes mogelijke scenario’s beschreven die leiden tot onbedoelde veranderingen in het plantenmetaboloom als gevolg van genetische modificaties. Waar mogelijk worden ook voorbeelden gegeven. Hieruit blijkt dat van deze zes, alleen het scenario “metabolic pathway connectivity” heeft geleid tot gepubliceerde voorbeelden van onbedoelde effecten van genetische modificatie. In dit scenario kunnen veranderingen in een metabolische route leiden tot veranderingen in andere synthese routes, welke afhankelijk zijn van dezelfde substraten, producten of enzymen.

Huidige en toekomstige stand van zaken van metabolomics
Een uitgebreid overzicht van metabolomics technieken en hun mogelijkheden en beperkingen wordt gegeven . Metabolomics heeft tot doel een breed overzicht van alle metabolieten in een specifiek monster op een specifiek tijdstip te geven. De analyses zijn doorgaans niet gericht. Er wordt een beperkt aantal verschillende extractie-, scheidings-, en detectiemethoden gebruikt die goed werken voor sommige klassen metabolieten, en minder goed of niet voor andere klassen. Daarom geven deze analyses nooit een compleet beeld, en dient men van tevoren te weten welke klassen van belang geacht worden teneinde de juiste methoden te kiezen. “Metabolic fingerprinting” en “-profiling” zijn beiden niet-gerichte methoden. De eerste methode is bedoeld voor snelle verwerking van grote aantallen monsters waarbij het onderscheid tussen de monsters belangrijker is dan exacte identificatie en kwantificatie van metabolieten. De tweede methode, “metabolic fingerprinting” richt zich op het identificeren en kwantificeren van op zijn minst een deel van de verbindingen, vooral dat deel dat voor het specifieke onderzoek belangrijk zijn. Beperkingen van metabolomics voor de milieurisicobeoordeling zijn de specificiteit van de extractiemethoden en de noodzaak van zeer strikte monsterbehandeling om te voorkomen dat er technische variatie wordt geïntroduceerd. Dit maakt het vergelijken van monsters van verschillende laboratoria en zelfs binnen een laboratorium een mogelijke bron van fouten, hetgeen standaardisatie van methoden moeilijk maakt. De waargenomen grote variatie in concentraties van verbindingen betekent ook dat een bepaalde gekozen methode lage concentraties van nochtans belangrijke metabolieten kan missen. Verder overstijgt het aantal detecteerbare metabolieten vele malen het aantal geïdentificeerde verbindingen, terwijl juist identificatie belangrijk is bij de milieurisicobeoordeling. Ten slotte is de technologie nog relatief duur en is de inzet van gespecialiseerde onderzoekers voor de correcte uitvoering van de bepalingen en interpretatie van de uitkomsten noodzakelijk. Het is te verwachten dat deze beperkingen in de toekomst minder belangrijk worden omdat de technologie zich snel ontwikkelt.

Literatuur over het gebruik van metabolomics voor risicoanalyse van GM planten
Aangezien metabolomics tot nu toe nog geen deel heeft uitgemaakt van de milieurisicobeoordeling, is er geen specifieke literatuur over die toepassing beschikbaar. Vergelijkende veiligheidstesten voor het bepalen van "substantial equivalence" voor de toepassing van GM planten in (voeder)gewassen zijn al lange tijd praktijk. De gebruikte methoden en gepubliceerde uitkomsten van zulk onderzoek zijn informatief en kunnen als leidraad voor het gebruik van metabolomics in milieurisicobeoordelingen dienen. Op basis van deze studies is het aan te bevelen rekening te houden met natuurlijke variatie in metaboliet profielen op verschillende groeilocaties en onder verschillende groeicondities, evenals het meenemen van een reeks van commerciële variëteiten in de vergelijking. Gepubliceerde studies concluderen meestal dat er sprake is van 'substantial equivalence' met niet-GM controle planten, maar dat kan komen doordat studies met andere uitkomsten niet worden gepubliceerd. Het rapport bespreekt ook de verschillende statistische methoden voor het bepalen van de aanwezigheid en betrouwbaarheid van waargenomen verschillen die ook afkomstig zijn uit dit werk.

GM planten met doelbewuste modificatie van metabole routes
Genetische modificatie voor het verbeteren van de voedingswaarde of verwerkbaarheid van gewassen is een actief onderzoeksveld, waarvan de publicaties waarbij metabolomics is gebruikt enigszins een beeld kunnen geven van onbedoelde veranderingen die voor kunnen komen. In het algemeen worden onbedoelde veranderingen, als ze worden waargenomen, veroorzaakt door de verbondenheid van metabole routes en de activatie van "stille" routes. Zo kan de succesvolle, geplande verhoging van een metaboliet gepaard gaan met de afname van metabolieten in een andere route. Anderzijds kan een geplande toename ook gepaard gaan met een toename in metabolieten van een soortgelijke route die hieraan gekoppeld is. Beide kunnen nadelige effecten voor de plant hebben. Naarmate de kennis van metabole routes en hun connecties toeneemt, zullen zulke ongewenste veranderingen in toenemende mate gedetecteerd kunnen worden met een gerichte metabolomics benadering. Het voorbeeld van verandering van het carotenoidgehalte van diverse gewassen wordt in deze context in meer detail besproken.

Mogelijke milieueffecten van veranderingen in het metaboloom
Uit de beschikbare literatuur komt duidelijk naar voren dat de samenstelling van de plant een grote invloed heeft op zijn interactie met de omgeving, zowel boven- als ondergronds. Genetische modificatie die bedoeld of onbedoeld deze samenstelling verandert, kan dus effecten op deze interacties hebben. Deze effecten kunnen inhouden, maar zijn niet beperkt tot, de aantrekkelijkheid of voedingswaarde voor planteneters en bestuivers, hun predators en parasieten, evenals bodem- en rhizosfeerorganismen. Twee veelal onderzochte gevallen, de productie en rol van glucosinolaten in koolachtigen, en de rol van metabolieten die de respons tegen pathogenen in tabak bepalen, worden in meer detail belicht. Hoewel de hoeveelheid literatuur over de rol van metabolieten in plant-omgevingsinteracties erg groot is, is de bruikbaarheid hiervan voor geteelde gewassen en voor de voorspelling van effecten van specifieke metabolieten op deze interacties beperkt omdat er vaak modelorganismen gebruikt worden.

Conclusies
Metabolomics is een zich snel ontwikkelende discipline en kan in de toekomst mogelijk bijdragen aan de milieurisicoanalyse van genetisch gemodificeerde gewassen. Hoewel inmiddels een groot aantal metabolieten al kan worden gedetecteerd, beperkt het veel kleinere aantal van identificeerbare metabolieten de waarde voor risicoanalyse. Het begrip van het functioneren van het plantengenoom en de beschikbaarheid van andere "omics" technologieën vergroot het vermogen om onbedoelde effecten op het metaboloom te voorspellen, vaak in een fase voorafgaand aan introductie in het milieu van de gewassen. Deze kennis kan aanleiding geven tot gerichte metabolomicstechnieken om te bepalen of zulke onbedoelde effecten inderdaad optreden. De grootste hindernis voor de toepassing van metabolomics in de milieurisicoanalyse is echter de beperkte kennis van de rol van specifieke metabolieten in de interacties van de plant met zijn omgeving. Het zal misschien mogelijk zijn om significante onbedoelde veranderingen als gevolg van genetische modificatie waar te nemen, maar door gebrek aan kennis is het doorgaans niet mogelijk aan te geven of een dergelijke verandering als mogelijk schadelijk moet worden beoordeeld. Als metabolomics deel gaat uitmaken van het protocol voor milieurisicoanalyse van GM planten, dan is het opzetten van een database over de milieueffecten van specifieke plantenmetabolieten aan te raden. Het rapport bekijkt in de conclusies een aantal scenario's

voor de wijze waarop metabolomics in de milieurisicoanalyse zou kunnen worden ingezet. De conclusie is dat, het gebrek aan kennis over de rol van metabolieten in plant/omgeving-interacties in aanmerking genomen, op dit moment de mogelijke bijdrage van metabolomics aan milieurisicoanalyse echter nog beperkt.

Summary

Objectives of the study
With the rapid development of "-omics" technologies, the question arises whether these technologies, and to what extent, could contribute to an improvement of the Environmental Risk Assessment (ERA) of transgenic plants. This report describes the results of a literature search focussing on the question "What is the potential contribution of metabolomics for the environmental risk assessment of genetically modified crops?" Metabolomics is the science of measuring small organic molecules (metabolites) in plants, animals or microorganisms.

Possible scenario's relevant for ERA
The guidance document published by the European Food Safety Agency, outlining the procedure for ERA, identifies seven areas of concern of which most are relevant in the context of this study. It also states that "Comparative safety assessment" is to be applied using appropriate comparators (non-GM plants) and it identifies four data sources that may reveal unintended effects of GM plants, of which "Compositional analysis" is the most relevant to be addressed by metabolomics. This analysis is currently not a requirement in European ERA protocols. Transgenic crops may show intended (for which the modification was designed) as well as unintended effects, also on metabolites. The latter category may be further divided into "predictable" (could be expected based on our current knowledge) and "unpredictable" effects.
Six possible scenarios for genetic modification leading to unintended effects on the plant metabolome are described and where available, examples are given. It appears that of these six, "metabolic pathway connectivity" is the only scenario leading to –published- unintended effects of genetic modification. In this scenario, modifications targeting a particular pathway may affect other pathways that depend on the same substrates, products, or enzymes.

Current and expected state of the art in metabolomics
A comprehensive overview of metabolomics technologies and their abilities and limitations is presented. Metabolomics intends to give a wide view of all metabolites in a specific sample at a specific time. Metabolomics analyses are generally untargeted, but since they employ a limited number of different extraction, separation, and detection methods that may work well for certain classes of compounds and less well – or not - for others, they are not unbiased. Therefore these analyses never give a complete picture, and one needs to know beforehand what class of compounds is considered relevant in order to choose the proper methodology. Metabolic fingerprinting and metabolic profiling are both untargeted approaches. The former is designed for high throughput screening of large numbers of samples where discrimination between samples is more important than exact identification and quantification of all compounds. The latter approach aims at identifying and quantifying at least some of the compounds, especially those deemed important for the study at hand. Identified limitations for the use of metabolomics in ERA are: biased extraction methods and the requirement for very strict sample handling to avoid introduction of technical variation. This makes comparison between samples of different labs or even within one lab error-prone and thus limits standardisation of methods. The wide range of concentrations of compounds (dynamic range) experienced in metabolomic analyses also carries the risk of a particular approach missing low concentration compounds that are nevertheless important for ERA. Furthermore, the number of detectable compounds far exceeds the number of identified compounds. Yet, establishing the identity of an altered metabolite is important for use in ERA. Finally, the technology is relatively expensive and requires specialized researchers for correct application as well as for interpretation of the data. All these limitations are expected to decrease in the near future as the technology continues to develop fast.

Literature on the use of metabolomics for risk analysis of GM plants
Since metabolomics is not part of ERA anywhere so far, no literature focussing on this subject is available. However, comparative safety testing for the establishment of substantial equivalence for GM food or feed is a long-standing practice. Hence, the protocols used and published data are informative as they provide a primer

of problems and questions to address for the use of metabolomics in ERA. This means addressing natural variability by including different growing locations and conditions and using a range of commercially available varieties as comparators. Published results of such comparative safety studies usually establish substantial equivalence, but there may be a publication bias for such results. The report also discusses different statistical methods, each with their own advantages and drawbacks for determining which of the observed differences are statistically significant.

GMO's for targeted engineering of metabolic pathways

Genetic modification for improvement of nutritional content or processability is an active field from which the published results can give some measure of the extent of unintended effects that can occur, if they were assessed by any kind of metabolomics approach. Generally, unintended effects, if detected, were caused by pathway connectivity and the activation of silent metabolic pathways. Thus the increased production of a particular metabolite might be successful but can be accompanied by a decrease in compounds from other pathways, or by an increase in compounds from the same pathway and derived from the targeted metabolite. Both may have severe effects on the plant. As knowledge of metabolic pathways and their connectivity grows, such unintended effects might be increasingly detected by targeted metabolomics. The case of carotenoid content engineering in different crop species is in this regard discussed in more detail.

Possible environmental effects of metabolome changes

From the available literature it is evident that a plant's metabolite composition plays a large role in its interaction with the environment, both below as well as above ground. Thus genetic modification affecting this composition intentionally or unintentionally may well affect such interactions. These may include (but are not limited to) attractiveness or nutritional value for herbivores and pollinators, their predators or parasitoids, as well as soil and rhizosphere organisms. Two well documented cases, those of glucosinolates in Brassica or Arabidopsis, and that of metabolites affecting the pathogen response in tobacco are discussed in more detail. Although the amount of literature on the role of metabolites in plant-environment interactions is very large, the usefulness of these for cultivated crop sand for predicting effects of particular metabolites on the interaction is very limited, because in many studies model plant species have been used.

Conclusions

Metabolomics is a fast developing discipline and may contribute to ERA of transgenic crops in the future. Although a large and growing number of metabolites can be detected, the much smaller number of identified metabolites lowers its value for risk assessment. Our understanding of plant genome function and the availability of alternative "omics" technologies is increasing our ability to predict unintended effects on the metabolome, often in a stage far before introduction of the environment of the crop. Alternatively, this knowledge can prompt targeted metabolomics approaches to check if such effects actually occur. The main problem for the use of metabolomics in ERA is our limited knowledge of the role of any particular metabolite in the interaction of a particular plant with a particular component of the environment. Thus we might be able to detect significant (unintended) changes in the metabolome as a result of genetic modification, but because of this knowledge gap one would normally not be able to identify or exclude such a change as harmful. If metabolomics is going to be part of ERA protocols for GM plants, the establishment of a database for such knowledge is recommended. In the conclusions section the report reviews a number of potential scenarios for the application of metabolomics in ERA. It is concluded that, considering the knowledge gap with regard to the role of plant metabolites in plant/environment-interactions, the potential contribution of metabolomics in Environmental Risk Assessment is limited for the time being.

1 Introduction

1.1 Objectives of the research for this report

Submissions for approval for the cultivation of new genetically modified plant varieties in the European Union need to be accompanied by data files describing the molecular characterization, data from laboratory and greenhouse experiments as well as results of field testing experiments, in order to support Environmental Risk Assessment. The currently approved varieties in the EU contain one or both of only two input traits, insect resistance and herbicide resistance. However, several crops with altered metabolite content for improved quality as food, feed, or biofuel, as well as for the production of raw chemicals are entering the approval process or are expected to do so in the near future. These crops may, potentially, have intended or unintended alterations in their metabolite content that could potentially affect the interaction with their environment and have undesirable effects on that environment in terms of nature conservation goals or agroecosystem functions.

Metabolomics, a rapidly developing subdiscipline of biology, is the systematic characterization of small molecules and their concentrations in cells, tissues, or organisms, in ever more detail. While (targeted) compositional analysis of harvested products from new varieties, for establishing substantial equivalence is already part of the data required for approval of genetically modified foods and feeds, it is not a regular part of the data used for Environmental Risk Assessment. The combination of a rapidly developing technology with the prospect of genetically modified plant varieties with altered or new metabolic pathways has prompted the sponsors of this research to ask the question:

What is the potential contribution of metabolomics for the environmental risk assessment of genetically modified crops?

In order to answer this question satisfactorily, the following objectives were set for the literature search and for the report describing the outcomes of this research:

- To describe possible scenarios for the occurrence of unintended effects of genetic modification on metabolite content, where relevant for ERA. As part of this:
    - To deliver relevant examples for each of the scenario's from the literature where available, or alternatively, examples that demonstrate a part of such a scenario even when no genetic modification with the aim of eventual commercialization had taken place
    - To demonstrate the usefulness, or the lack thereof, of metabolomics applications for detecting such unintended effects
- To describe the metabolomics technology, its limitations, and the advantages and disadvantages of the various methods relevant for ERA
- To review the literature that includes metabolomics as part of the characterization of genetically modified plants, such as for substantial equivalence testing as well as for non-risk assessment studies
- To search the literature for examples of environmental effects of plant metabolites or effects that might occur when metabolite content is altered

The focus of these objectives is on the intended or unintended effects on the plant metabolome and, as far it is known, how these might affect the environment. Whether such effects on the environment, if they can be identified, are deemed adverse or acceptable, or the definition of what constitutes adverse effects prior to the ERA, is determined by policy makers and is therefore no part of the scope of this report.

1.2 Possible scenario's relevant for ERA

1.2.1 Introduction

Environmental risk assessment (ERA) of genetically modified (GM) plants is required according to European Commission regulations for commercial cultivation of crops for food or feed use (Regulation (EC) no. 1829/2003) or for cultivation or crops without a food/feed use (Directive 2001/18/EC). The European Food Safety Agency (EFSA) Panel on Genetically Modified organisms has published a guidance document, outlining the procedure and overall steps that such an ERA should comprise.

EFSA, in its guidance document, identifies 7 "areas of concern" to be addressed in ERA, as listed in Box 1.

Box 1. Seven "areas of concern" for ERA as listed in the EFSA Guidance Document (EFSA Panel on Genetically Modified Organisms, 2010)

1. Persistence and invasiveness, including plant-to-plant gene flow
2. Potential for plant to micro-organisms gene transfer
3. Interaction of the GM plant with target organism
4. Interactions of the GM plant with non-target organisms
5. Impacts of the specific cultivation, management and harvesting techniques
6. Effects on biogeochemical processes
7. Effects on human and animal health

Relevant to the issue investigated in this report are the following "areas of concern": (1) persistence and invasiveness as far as regulated by metabolites, such as in allelopathy or other selective advantages for the plant (3) interaction of the GM plant with target organisms and (4) interaction of the GM plant with non-target organisms, including criteria for selection of appropriate species and relevant functional groups for risk assessment; (5) impact of the specific cultivation, management and harvesting techniques; including consideration of the production systems and the receiving environment(s); (6) effects on biogeochemical processes; and (7) effects on human and animal health (as far as this is not covered by food and feed safety assessments). Area (2) potential for plant to micro-organisms gene transfer is considered to be outside the scope of this research as it cannot likely be addressed by metabolomics research. The guidance document indicates that "Comparative safety assessment" is to be applied, i.e. the use of appropriate methods to compare the GM plant and its products with their appropriate comparators (non-GM plants). While the intended alterations in GM plants, in the context of metabolomics, may be identified and confirmed by measurements of single compounds, unintended effects of genetic modification may be detected by analysis of data from different sources in a four-pillar approach as depicted in Box 2.

.

Box 2. The four data sources that may reveal unintended effects of GM plants (EFSA Panel on Genetically Modified Organisms, 2010)

1. Molecular characterization
2. Compositional Analysis
3. Agronomic and phenotypic characterization
4. GM plant-environment interactions

A more elaborate discussion of these items can be found in the EFSA guidance document (EFSA Panel on Genetically Modified Organisms, 2010). It is evident from the above list that all four data types may be able to predict or to detect changes in the metabolome of a GM plant compared to its comparator indirectly, through its effects. However, only appropriate Compositional Analysis is able to confirm intended changes in the metabolome or to detect unintended changes in the metabolome, while the other types can merely hint at such changes. The outcome of the Compositional Analysis is the identification of both intended as well as unintended effects on the GM plant metabolome that constitute potential "hazards", i.e. have the potential to do harm in the environment. A general definition of intended and unintended effects is shown in Table 1. Currently Compositional Analysis is not a requirement in European ERA protocols for applications for cultivation.

In this report we shall more specifically identify mechanisms or scenarios through which unintended effects on metabolome composition of plants as a result of genetic modification may occur, and where metabolomics for Compositional Analysis may help to identify such differences. The ability to detect such differences will rely mostly on the technical options available (and selected for use) and their detection limits, as well as on the experimental setup. A more contentious issue occurs after the detection of differences, namely determining the statistical and biological significance of observed differences. The former relies on the availability of robust statistical techniques to determine the significance of observed changes, and on the application of an experimental set-up that fits the statistical analysis used later on. The latter relies on the available knowledge on the interaction of plants with their environment and the role of its metabolome in these interactions, in order to identify potential harm to the environment. This "hazard characterization" step is then to be followed by, for those hazards that are deemed relevant, a consideration of the possible exposure pathways through which the GM plant may adversely affect the environment. Following this, "Risk characterization" combines the magnitude of the consequences of the hazard with the likelihood of these consequences occurring (EFSA Panel on Genetically Modified Organisms, 2010). In other words, Risk equals Hazard (magnitude of the consequences) x Exposure (the likelihood that these consequences occur).

Table 1. Definition of intended and unintended effects of transgenic crops. From (Rischer and Oksman-Caldentey, 2006)

Intended effects	• effects that are targeted to occur after gene transfer has been successfully accomplished • lead to improvements in phenotype, composition or agronomical traits • metabolites of interest can be analysed by targeted quantitative methods
Unintended effects	• effects which represent a statistically significant difference (e.g. in chemical composition of the GM plant compared with a suitable non-GM plant grown under the same conditions) • the evaluation should take the biological variation of the comparator into account and, therefore, an appropriate number of background samples has to be analysed • changes might have an impact on potential agronomic performance but they do not necessarily pose safety threats for human health or the environment
Predictable	• effects which are unintended but could be expected based on our current knowledge on plant metabolite pathways, gene-to-gene interactions and general plant biology

	• the class of influenced metabolites is predictable, thus targeted quantitative methods can be applied • allergenicity is often discussed in this context
Unpredictable	• effects that represent a statistically significant difference (e.g. in the chemical composition of the GM plant compared with the non-GM plant), and could not be expected in advance based on our current understanding • metabolic profiling methods are suitable to provide information, even concerning small but potentially important changes

Table 2 is shows a probably non-exhaustive list of possible ways in which unintended or unexpected effects of transgenesis on the metabolome might occur.

Table 2. Possible (hypothetical) types of unintended and/or unexpected effects of genetic modification on the plant metabolome

Transgenic construct-specific effects regardless of insertion position in the host genome	• Promiscuity of the enzyme encoded by the inserted gene(s): substrate and/or product specificity of the enzyme are less strict than expected or required (section 1.2.2) • Changes in metabolomics pathways other than the one(s) targeted by the inserted gene caused by the interconnectivity of the host's metabolic pathways or by the activation of "silent" metabolic pathways (section 1.2.3)
Host genome insertion site-specific effects (mostly) regardless of the insert type	• Inactivation or specific activity change of a host gene due to the insertion event, affecting host metabolic pathways (section 1.2.4)
Insert type and insertion site-combination specific events	• Insertion of a specific transgene in a specific location causing, for example through transcriptional and/ or translational read-through, the formation of a new recombinant protein with a unique enzymatic activity (section 1.2.5)
Genetic or epigenetic changes in metabolism as a result of the regeneration of transgenic plants	• The process of regeneration of transgenic plants may cause genetic or epigenetic changes to the host plant that affect its metabolism, even though this is not a consequence of the transgenic nature of the plant. When a transgenic line is compared to a non-transgenic isogenic line in isolation, such effect would not be readily distinguishable from effects of the transformation itself (section 1.2.6)

1.2.2 Enzyme promiscuity

Enzymes functioning in metabolic pathways are usually depicted as performing a single conversion from a specific substrate to a specific product (and *vice versa*). In reality, many enzymes are much more generalist, being able to convert multiple (often related) substrates into just as many products. In an analysis of a genome-scale model of *Escherichia coli* metabolism, it was found that 37% of the enzymes act on a variety of substrates and catalyse 65% of the known metabolic reactions. Specialist enzymes, catalysing a single reaction on one particular substrate tend to be frequently essential for survival of the organism , have a higher metabolic flux, and require more regulation of activity than do generalist enzymes (Nam *et al.*, 2012). Conversely, generalist enzymes are usually less or not essential for survival and are promiscuous, i.e. able to

catalyse multiple reactions on various substrates. Many examples of these exist and their (lack of) specificity has been studied *in vitro* (Hult and Berglund, 2007; Humble and Berglund, 2011; Kapoor and Gupta, 2012; Khersonsky *et al.*, 2006; Khersonsky and Tawfik, 2010; Mohamed and Hollfelder, 2013). Terpene synthases in plants are known to be able to handle a variety of substrates and produce multiple products, and these may be further developed *in vitro* to become more specific and increase in activity (Yoshikuni *et al.*, 2006). Although many of these properties may be studied and modified *in vitro* and thus be predictable to some extent and be able to guide targeted analysis of the metabolome for hazard identification when deemed necessary, it is possible that some of the alternative reactions would be overlooked by *in vitro* testing or by the availability of alternative substrates in the host.

1.2.3 Metabolic pathway connectivity

Most biochemical reactions do not occur in isolation but are connected, through their substrates and products, to other metabolic pathways. Thus, it is conceivable that targeted changes in one particular pathway, with the goal of changing the level of a particular metabolite or the production of a novel metabolite, may have knock-on effect in other pathways that depend on or use some of the same substrates, products, or enzymes. One example of such an effect of connectivity is a result of the modification of lignin production in transgenic plants on its flavonoid content. Plants that were modified to decrease the expression of hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT), a lignin biosynthetic gene, were severely stunted and had dark-green/purple leaves (Hoffmann *et al.*, 2004). Metabolite analysis showed that those plants have increased flavonoid accumulation affecting plant growth through the misregulation of auxin transport. Flavonoid synthesis and lignin biosynthesis are both branches of the phenylpropanoid pathway and both utilize a common precursor, the substrate of HCT. Apparently the blocking of HCT activity redirected the flux in the pathway towards flavonoids. Simultaneous blocking of the flavonoid pathway reverted the effects on plant growth (Besseau *et al.*, 2007).
A special case is that of activation of "silent metabolic pathways", i.e. metabolic pathways of the host plant, which are not active, and possibly unknown, because of the lack of an initial substrate while one or several components of that pathway are already expressed. A supply of substrate produced as results of the genetic modification, whether it is the intended product or not, could lead to further conversion of that product into metabolites that were not detectable in the host plant without the modification. An example of this is found in rice modified for carotenoid production (see section 3.3.1 of this report).
As our knowledge of metabolic pathways and their connectivity increases, it will be increasingly possible to predict potential unintended effects of changes to existing pathways or of the introduction of new pathways. However, until the host plant's metabolic pathways have been completely characterized there is room for unpredicted metabolome changes that could be detected by some form of metabolic profiling or fingerprinting.

1.2.4 Host gene activity change due to the insertion

Although much research into targeted and site-specific modification of host genomes is on-going, currently genetic modification of plants still occurs principally through two techniques: by T-DNA insertion into the plant genome following co-culture of plants cells or tissues with *Agrobacterium tumefaciens* harbouring a Ti-plasmid (or variations thereof) and by ballistic or membrane permeabilization techniques allowing DNA transfer into the plant cell followed by integration of the introduced DNA into the host genome.

As a result of the mostly random nature of introduced DNA insertion into the host genome, transgenes may "land" in or near actively expressed host genes and disrupt or alter their expression, as well as produce novel "hybrid" transcripts and/or, in rare cases, truncated or hybrid proteins consisting of part transgenic and part host-derived sequences, as well as more large-scale genomic rearrangements. The types and frequencies of such insertion effects are described in more detail in several reviews (Filipecki and Malepszy, 2006; Wilson *et al.*, 2006). In general, *A. tumefaciens* appears to more often give simple single insertions, with a bias for gene-rich chromosomal locations and few large scale rearrangements, while particle bombardment leads more often to complex insertion events with substantial rearrangements. This indicates that a thorough

characterisation of the insertion event(s) in transgenic lines subject to ERA may help identify potential undesired effects of plant transformation that might affect metabolite content.

Obviously, a transgene construct landing in a host gene that encodes an enzyme, which is actively involved in a metabolic pathway can have large unintended effects on the metabolome if it thereby disrupts the function of the gene. An obvious example is the utilization, for functional genomics purposes, of T-DNA insertion mutants or T-DNA tagging, which provides for the different scenario's described here (Walden, 2002). The simplest use of such mutants is knock-out insertions, where the insertion of T-DNA results in the disruption and inactivation of the host gene. Alternatively, in "*Activation tagging*", the T-DNA contains an active promoter directing transcription outward of the T-DNA, thereby regulating transcription of any sequence nearby its integration site. This strategy is used to activate otherwise transcriptionally silenced genes in order to study the resulting phenotype and to infer gene function from this. Telling examples of such insertions altering the host's metabolome are, among many others, the activation of a phenylpropanoid pathway-regulatory MYB transcription factor encoding gene leading to purple *Arabidopsis thaliana* plants due to anthocyanin accumulation (Borevitz *et al.*, 2000) and the activation of *YUCCA*, encoding a rate-limiting enzyme in tryptophan-dependent auxin biosynthesis (Zhao *et al.*, 2001).

1.2.5 Formation of hybrid transcripts or proteins due to the specific insertion type and location

Landing of the transgene construct in or upstream of a host gene may lead to the formation of a truncated version of the host enzyme, the formation of a hybrid transcript and potentially resulting in the formation of a hybrid protein. While most transgenic constructs are designed to include a promoter and terminator to drive initiation and termination of the transgene's transcription, respectively, transgene constructs or T-DNA insertion are known to be frequently incomplete, leading to insertion of truncated versions of the construct. One example is event MON810, a Cry1Ab expressing transgene widely used for insect-resistance in maize cultivation, which contains a 3' truncation of the *cry1Ab* gene leading to a 91 kDa protein instead of the intended 131 kDa protein as originally encoded by the construct used for transformation. In other cases, transcriptional read-through when using the *nos* transcriptional terminator in the transgene construct has been observed. If a similar truncated insertion would form a hybrid protein with (part of) a host-encoded open reading frame at the insertion site, this might lead to altered or novel enzyme activity or specificity, although it is more likely that the produced protein would be altogether inactive due to improper folding as a result of incompatibility of the parts. We have found no examples of such novel enzyme activity as a result of transgene insertion in the literature surveyed by us, nor in critical reviews by others (Wilson *et al.*, 2006) (http://natureinstitute.org/nontarget/). While such a mechanism is conceivable, it has to be noted that with our current knowledge of gene identification in plant genomes, the increasing number of fully sequenced genomes and with the routine characterization and sequencing of transgene insertion sites, the potential for such an event would be easily spotted in an early stage of product development. Characterization, including sequencing of the insertion site(s), where necessary supplemented with gene expression analysis and transcript sequencing would indicate whether the formation of such a hybrid protein is likely. In the case that this question were answered positively this might prompt further metabolomics analysis, but more likely it would lead to the selection of an alternative transgenic event with fewer complications.

1.2.6 Effects of *in vitro* tissue culture and transgenic plant regeneration

There is a multitude of data as well as anecdotal experience indicating that the processes of tissue culture and plant regeneration, that are part of most protocols for transgenic plant production, can affect the plant phenotype in unpredictable ways. Such phenotypic variation can vary from light to severe depending on the species and variety used in the transformation process (so-called "somaclonal variation"). Van de Brink *et al.* (2010) mention two examples of unintended effects that may have been caused by somaclonal variation in potato, without details. Although such events are not strictly an effect of the transgenesis they would be indistinguishable from any unintended effects caused by the transgene and as matter of course would be subject to ERA when presented in a variety for approval. It has to be noted that during product development

transgenic plants from a variety that may have been chosen for its ease of transformation and regeneration are often back-crossed with elite varieties for several generations in order to move the transgene to varieties that are better suited for commercialization. Such a strategy would usually separate any somaclonal variation effects from the transgene effect and not further be addressed by ERA. Only where the eventual product would be genetically close to the originally transformed plant (such as in clonally propagated crops), somaclonal variation effects may find their way into the eventual product subject to ERA.

1.3 The usefulness of metabolomics for detecting or predicting unintended effects relevant for ERA

The "unintended effects" mentioned in Table 1 and further exemplified in sections 1.2.2 to 1.2.6 may constitute effects on the metabolome of small molecules (see Chapter 2, section 1), which could be detectable by metabolomics technologies. These may be predictable or unpredictable. Chapter 2 describes the current state of the art in metabolomics, and what different approaches are available such as "fingerprinting" and "profiling" for characterization of samples. As noted in Table 1 for "unintended effects", there is nowadays a clear realization among researchers as well as regulators, that an "effect" should be defined as a statistically significant difference, taking into consideration the natural variation in the non-GM comparators, caused by differences in genotypes or varieties and differences in growing conditions (season, geographical location, temperature, etc.). For this one could draw from the considerable experience accumulated from substantial equivalence testing for food and feed (Chapter 3, section 2), including the consideration of natural variation, the proper choice of statistical testing including an experimental design with sufficient statistical power. Examples of metabolomics studies on GM plants for targeted engineering of metabolic pathways, and some of the unintended effects observed, are discussed in Chapter 3, section 3. Detection of unintended effects on the metabolome of GM plants would be achieved by "Compositional Analysis" in Box 2 on page 9, which is currently not required for approval for cultivation in the EU. As will be shown further on in this report, for ERA "Compositional Analysis" may have to be extended to other parts of the plant than those used as food, or in some cases to the root exudate and released volatiles if all interactions with the environment are considered. Our current knowledge of the role of metabolites in plant-environment interactions, which could be relevant for ERA, is reviewed in Chapter 3, section 4. Based on this knowledge, observed unintended changes in the metabolome will have to be characterized, which may or may not result in the identification of a "hazard", i.e. determined to be potentially harmful to the environment and contrary to the protection goals set for that environment. Thus, metabolomics analysis as currently practiced may be well suited for detecting many possible unintended effects on plant metabolite composition. However, knowledge of the role of metabolites in the plant's interaction with its environment as well as a clear a priori definition of what constitutes a hazard for the environment and what the protection goals are, is equally important for the usefulness of metabolomics in ERA.

2 State of the art in metabolomics and expected future developments

2.1 Current state of the art

2.1.1 Metabolomics is the science of measuring small molecules

The term was coined at the end of the 1990s (Oliver *et al.*, 1998) (Table 3) and the technology aims to be fully complementarily to the other, potentially unbiased or non-targeted '-omics' approaches. The specific focus on the smaller metabolites entails that metabolomics excludes the larger organic polymers. These low molecular weight metabolites are the (end) products of a huge network of metabolic pathways and represent the activities of cell regulatory processes (Fiehn *et al.*, 2011). They advertise the response of biological systems to a variety of genetical and environmental perturbations (Fiehn, 2002). Metabolomics methods are untargeted and directed to global sets of compounds involved both in primary and secondary metabolism. As such, compounds of fundamental importance to plant survival as well as others, perhaps still with unknown function, are included.

2.1.2 Metabolomics provides a momentary snap-shot of plant metabolic composition

The primary goal of any plant metabolomics approach is to gain a 'helicopter view' of metabolism at a specific point in time, in a chosen tissue, obtained either under control or experimental conditions. The plant metabolome is however constantly changing – also throughout the day – and hence, for example, all samples should be taken at the same time of day and under the same conditions. As such, every metabolomics 'snapshot' needs to be assessed in accordance with these restrictions when assessing any potential risk associated with observed differences. The plant metabolome is not in a steady state but rather, is inherently highly amenable to variation. Under normal growing conditions, as experienced by plants in nature, basic processes such as growth, development, transition to flowering, etc. are accompanied or even determined by a global reprogramming of plant metabolism. Such reprogramming may be predominantly under genetic control in relation to normal plant development or may be consequential to externally imposed environmental perturbation. This in turn, is also predominantly under genetic control. All analyses need to be performed and take account of natural changes which routinely occur in plants under a specific set of environmental conditions; the relevance of any changes observed are weighed against these changes. A case in point is e.g. the phytoalexins (Ahuja *et al.*, 2012). This is a group of compounds which routinely appear in plants and crops in nature but which are generally not detectable as they are only present when plant parts are exposed to serious stress conditions. Contemporary research approaches are progressing in their sophistication but nevertheless, we still have poor knowledge of the complexity of such changes and responses.

2.1.3 Metabolomics analyses are untargeted but are not unbiased

Metabolomics experiments entail the employment of a number of extraction, separation and detection conditions (see below). The view gained is broad (untargeted). However, the approach will never be truly holistic as some element of bias will always be involved. This results from a failure to extract or detect certain compounds that perhaps are unstable or that have chemical or chemo-physical properties unsuited to the methodologies chosen. Nevertheless, educated choices as to the best approaches to use and optimization of data collection and mining strategies have greatly enhanced our capacity to expand our biochemical knowledge of plant materials. More specific methodologies are also being developed for untargeted analyses of specific compound groups. One example is lipidomics which analyses lipid and lipid – associated compounds such as the sterols, phospholipids, glycerides, waxes etc. (Lessire *et al.*, 2009). Other compound sub-groups relevant for plants may also be developed in the future targeting additional major groups of plant

compounds such as the terpenoids, alkaloids and polyphenols, each of which are already known to contain many thousands of different chemical structures, often with identical accurate masses and elemental formulae.

2.1.4 Metabolomics approaches – 'fingerprinting' and 'profiling'

Two terms regularly used in the literature are metabolic 'fingerprinting' and metabolic 'profiling'. Both types are of potential relevance for an ERA type approach but the latter has greatest power in risk assessment. The former is generally taken to refer to the use of machine output as a potentially recognizable chemical pattern which is specific to an individual sample. These unique fingerprints can be the starting point for comparative metabolomics where the researcher wishes to compare and contrast hundreds of extracts in order to assess quickly the degree of variation and often, select the most divergent samples or genotypes for more detailed study (Hall, 2006; Saito, 2006). Specific software tools have been developed to speed up and semi-automate this process and to optimize the output. Metabolic profiling is a term referring to a deeper form of analysis where one proceeds to perform a metabolite structural elucidation strategy. Known compounds can be annotated using commercial or in house databases. For example, Fernie can unambiguously identify 150 small polar molecules using GC-ToF-MS (Fernie, 2007). Metabolite identification, and particularly of so-called secondary plant metabolites remains however, a significant challenge, or indeed, a major bottleneck (see below).

Table 3. A short list of some of the most commonly encountered technological terms in the metabolomics literature

Plant metabolomics	An analytical approach focused on generating the least biased and most comprehensive qualitative and quantitative overview of the metabolites present in a tissue, organ or whole plant.
The plant metabolome	The complete complement of low molecular weight molecules present in a specific plant. Generally, 'low molecular weight' usually refers to those molecules smaller than 1500 Da.
Metabolic (metabolomic) fingerprinting	Screening of the metabolic composition of an organism, usually in a high-throughput approach involving large sample numbers. Quantification is usually relative to other (control/treated) samples and the initial goal is generally sample comparison and discrimination analysis. Identification of the metabolites present is not usually performed and differences are based upon contrasts in spectral pattern. Fingerprinting allows the most contrasting samples to be discriminated and selected further for more detailed analysis such as metabolic profiling.
Metabolic (metabolomic) profiling	In contrast to fingerprinting, metabolic profiling aims to identify and quantify at least some of the metabolites present, often focusing on those which have been identified through multivariate analyses as being discriminatory between samples/treatments /genotypes, etc. of the metabolites present in an organism. Such compounds may then form the basis of hypotheses linking them to genetic or phenotypic differences.
Targeted analysis	Generally refers to more traditional methods of chemical analysis where the extraction, separation and detection approach chosen has been focused on and optimized for a specific, chosen group

	of metabolites that have similar properties (e.g. amino acids) or that share a common biosynthetic pathway (e.g. flavonoids). Methods are usually fully quantitative and comprehensive.
Lipidomics	Specific metabolomic characterization of lipid species.

2.2 A brief overview of the technologies used that are relevant to ERA

Most metabolomics experiments involve a combination of *separation* and *detection* technologies that can be used in serial or in parallel combinations, often referred to as 'hyphenated approaches'. The terminology can be quite daunting (Appendix Table X1). LC-MS or GC-MS are most common but there are 'extreme' examples – such as HPLC-PDA-SPE-NMR-ESI-(ToF) MS (Moco and Vervoort, 2012). In this case, one separation technique (HPLC) has been combined with three complementary detection techniques (PDA, NMR, MS). The hyphenated code essentially refers to the analytical workflow used for that particular analysis. Many reviews have been written on the separation and detection technologies available for plant metabolomics. For more detailed information the reader is referred to seminal volumes such as: (Hardy and Hall, 2012; Saito, 2006; Weckwerth, 2007; Weckwerth and Kahl, 2013). More compact summaries have also been provided by Beale and Sussman (2011) and Browne *et al.* (2011).

2.2.1 Extraction

From an ERA perspective, the method of extraction used will ultimately determine those compounds that will be detected and consequently, a range of methods may be required in order to gain a sufficiently broad overview of biochemical composition. However, 'intended' effects can help guide which approaches might be the most appropriate. The major challenge for complementary, untargeted metabolomics approaches relates to the huge diversity of molecules present as well as their concentration (Cevallos-Cevallos *et al.*, 2009; Fernie and Keurentjes, 2011; Makkar *et al.*, 2007). This complexity, involving both the basic chemical structures as well as their different combinations of functional groups such as hydroxyls, carboxyls, amines, etc. has been well documented (Fernie and Keurentjes, 2011; Saito, 2006). The choice and optimization of sample preparation and extraction procedures are therefore critical (Hall, 2006; Hardy and Hall, 2012; Weckwerth, 2007). For volatile components, organic solvents or Solid Phase Extraction approaches can be employed (Tikunov *et al.*, 2007; Verhoeven *et al.*, 2011). An element of bias is inevitably introduced at the extraction stage as few compounds will be extracted to 100%, irrespective of the solvent or method used. Any risk assessment should therefore involve a combination of both polar and semi/non-polar extraction methods to cover adequately both hydrophobic and hydrophilic metabolites. Concentration differences is an additional technical challenge as important compounds can vary in plant tissues from molar to sub-nanomolar levels (<9 orders of magnitude!). Measuring such range differences is still hugely challenging in a single metabolomics run. However, as there is never a clear link between concentration and potential risk (low levels do not necessarily mean low risk – especially for highly toxic compounds), any method(s) chosen do need to have the potential to cover all difficulties associated with dynamic range restrictions.

2.2.2 Separation

Chromatographic separation using GC-MS of (semi)polar primary metabolite extracts is one of the most widely used metabolomics approaches currently employed (Fiehn *et al.*, 2011). LC methods often involving high pressure (HPLC) or Ultra Performance (UPLC) are also particularly popular with plant scientists (Verhoeven *et al.*, 2006). An approach not yet widely applied but which is becoming increasingly popular is Capillary Electrophoresis (CE) (Soga, 2007). CE-MS in certain circumstances can have particular advantages relating to sensitivity, rapidity and resolving power (Timischl *et al.*, 2008). For a full review of current CE developments, please see the special issue of Electrophoresis, 2009, volume 30, issue 10.

2.2.3 Detection

For metabolite detection, there are basically two key players – NMR and MS. From an ERA perspective, each has its own advantages and disadvantages (see Appendix Table X2). NMR requires relatively minor sample preparation, is non-destructive and inherently quantitative. Its use is also not restricted to specific chemical groups and NMR can give unambiguous metabolite identification together with being fully quantitative. Its greatest drawback relates to its lower sensitivity compared to MS-based approaches and its requirement for relatively large samples. However, recent improvements are making this less of an issue. Components at the lower concentration ranges will likely be missed using NMR. MS has a much wider dynamic range and high sensitivity but does require molecules to be ionized (charged) in order for them to enter the instrument. Irrespective of concentration, compounds which are not charged will be invisible to the MS and are undetectable. Significant ion suppression/matrix effects can also potentially mask molecule detection in complex extracts. This, together with variable ionization frequencies, makes MS-based quantification more difficult and totally dependent on reference standards. For less common metabolites, these are often not available and hence only relative quantification (i.e. value relative to the control value) is possible.

2.2.4 Data generation, storage, processing and mining

Metabolomics is a data-rich and also data-driven technology and this has generated considerable technical and intellectual challenges. Successful metabolomics research is perhaps more dependent upon the 'dry' side of science than on the 'wet' side. Effective, sophisticated data management tools and infrastructure are essential for storing both raw and processed data. Dedicated bioinformatics tools are crucial for subsequent data analysis, integration, visualization and translation into biological knowledge (Fiehn *et al.*, 2011; Redestig *et al.*, 2011). All aspects of data management are at the core of metabolomics technologies (Sumner *et al.*, 2008) and generally discriminate metabolomics from analytical chemistry. Reliably and robustly extracting qualitative and quantitative information from metabolomics datasets requires a high level of training. With the number of metabolites being detected within a sample always greatly exceeding the available number of samples advanced statistical input is essential in order to identify potential differences between wild type / GM samples and predict their likely relevance.

2.3 Bottlenecks, developments and expectations

Metabolomics is hugely valuable as an untargeted screening tool and is already giving us the deepest possible insights into plant metabolic composition. However, every technology has its limitations. No more so than for metabolomics which, as a recently developed approach, can still be considered to be in a phase of rapid development.

2.3.1 Bottlenecks:

From the point of view of risk assessment, assessing current limitations has yielded the following list of possible bottlenecks which are of particular relevance:

- In contrast to proteomics, transcriptomics and genomics, metabolomics cannot be limited to a single standardised method. **No extraction or detection method is unbiased** and hence no single or small number of methods will give a truly holistic overview of all metabolites. Hence, while intended changes can be exploited to choose specific approaches, unintended changes may still be missed by unfortunate methodology choices or technological limitations. There are therefore no guarantees of detecting perturbation of all potentially-relevant metabolites.
- By providing a **'snapshot'** of metabolic composition, metabolomics analyses require strict sampling and sample handling. This is particularly challenging in a field-type situation, which is required for ERA. Furthermore, the natural dynamics of plant metabolism, its sensitivity to daily and seasonal environmental change and tissue – specific differences entail that non only multiple platforms but also multiple sampling times under multiple temporal / environmental regimes are required to obtain

a true global picture of metabolite composition and the assessment of potential risk associated with observed deviations.

- Also related to the above point, **sample comparison** both between labs (even when using the same machine) and also within a lab (i.e. comparing materials different seasons) is highly prone to error. This is especially the case for MS-based technologies where e.g. when a lab wishes to compare results from last year with those of samples from the next year, it is essential to re-run the previous year's samples together with the new ones.
- Metabolomics approaches are always, by definition, **methods of compromise**. Plant extracts will always comprise mixtures of compounds crossing a huge dynamic range (concentration). Choosing a generic approach risks missing components at the lower concentrations. These compounds nevertheless may have considerable significance from a toxicity / risk assessment viewpoint - as is known for example, with proteins and food related allergies.
- Current state of the art means that with metabolomics we can detect ca. 10x more metabolites than we can identify. Identification implies the availability of standards and hence, unidentified metabolites detected by metabolomics can only be quantified in relative terms. This is generally not a problem in comparative analyses (identifying deviant lines; identifying relative range changes in metabolite concentration). However, from a risk assessment viewpoint knowing the **identification of altered metabolites** (and hence their true concentration) is an important aspect of the risk assessment process. Major changes in metabolite levels do not in any way imply increased risk.
- From a technology viewpoint, metabolomics remains an **expensive approach**, both in terms of infrastructure and also in terms of manpower. Metabolomics is a multidisciplinary science requiring a range of divergent expertise's not usually found in one individual. The overall level of training required is also high.
- Metabolomics entails a good grasp of both the wet lab technologies as well as the statistics behind the data analysis procedure. Both the scale and the richness of metabolomics datasets carry with them an inherent risk of incorrect interpretation and the generation of false conclusions. A new generation of biologists is required who are confident in multi-variate statistical approaches for complex, large-scale data management and manipulation. Such broad-minded and broadly disciplined individuals are not easy to find.
- Perhaps the overall key limitation to any metabolomics experiment is relating any observed changes to their true biological relevance (in this case likelihood of risk).

2.3.2 Developments and expectations

Metabolite identification: Tools for systematic metabolite annotation are being worked and databases continue to be populated every day with new information. However, it will likely be many years before we are in a position to automatically reel-off lists of named metabolites with full confidence. Limitations incurred through having a lack of authentic standards for the majority of compounds that can now be detected are also considered significant. However, a number of groups have established mechanisms for easy exchange of commercially-available and also in-house synthesised compounds to advance database construction. Reference standards are essential for unambiguous formula assignment and the confirmation of metabolite identity and need to be available to the different platform managers. Without these, metabolic pathway elucidation, biological interpretation and proper assessment of metabolite-associate risk will remain a limitation.

Detection sensitivity: While it is clear that we are unlikely ever to be able to perform fully holistic metabolite analyses for a range of different reasons, one major current limitation of metabolomics approaches from a risk assessment viewpoint is the inability to include routinely some groups of very low abundance molecules which remain generally undetectable with current state of the art hardware (Draper *et al.*, 2011). However, machine sensitivity continues to increase and remains a key goal of instrument makers. Furthermore, with improved methodologies (NMR cryoprobes etc) it is anticipated that the lower levels of reliable detection will continue to decrease in the foreseeable future. However, from a risk point of view, the importance of such very minor compounds remains a point of discussion.

Robust methodology: Metabolomics is still best considered as a reliable research tool more than a robust standardized / automated screening tool. The general robustness of methods and the inter- exchangeability of data between labs is an issue which has not fully been resolved. Some approaches and instrumentation are known to be more reliable than others (Browne *et al.*, 2011), but there are still concerns regarding inter-laboratory reproducibility and hence, comparability of data. As a start, recommendations for the standardization of methodologies for data generation and collection – 'minimum reporting standards'– have been made in recent publications (Fernie *et al.*, 2011; Fiehn *et al.*, 2007; Sumner *et al.*, 2007). Furthermore, the metabolomics community – being led e.g. by EBI-Cambridge and the International Metabolomics Society is now uniting to tackle these specific issues. This should also help to facilitate future desires for data exchange and long-term data validity. However, the ultimate desire to create common databases of experimental data and to have a system that enables direct comparison of data from different experiments and different laboratories is considerably more challenging. Poorly defined experimental goals may result in different parameter choices for data collection as might also the precise instrumentation available in different laboratories. Furthermore, even after data generation, different labs have different experiences or preferences regarding data pre-processing. Regarding this issue, a number of inter-laboratory comparisons have recently been performed with the view to estimate both technical and non-technical (human) sources of variation in data generation. Data for both NMR (Viant *et al.*, 2009; Ward *et al.*, 2010) and GC-ESI-ToF-MS (Biais *et al.*, 2009) experiments, generated from up to five laboratories using identical, split samples and standard operating procedures, have been reported and early results are encouraging. Nevertheless, the strict requirement to follow precisely defined procedures at all steps in the data generation and data manipulation processes is clearly evident. Additional comparisons are required in order to help define better fully standardized procedures that will enable robust and reliable cross-laboratory data exchange and the creation of common metabolite databases from various reliable sources.

3 Literature review on the use of metabolomics for risk analysis of GM plants

3.1 Introduction

In this chapter we give an overview of the literature on metabolomics, which for our purpose is limited to, and can be broadly divided into three classes: 1) Literature on the use of metabolomics for risk assessment with regards to use of GM crops in food or feed; 2) An overview of metabolomics research in connection to GM plants specifically produced for the modification of the metabolic content, including some that are expected to be submitted for approval and 3) Areas of interest for ERA of GM crops, and the state-of-the-art of metabolomics in these respective areas.

3.2 Metabolomics in general GMO research – lessons from substantial equivalence testing

3.2.1 Metabolomics applications

Applications of metabolomics in the plant sciences are already extensive and broad. Information on the biochemical composition of plants can help identify biochemical changes as a potential basis of causality for phenomena of interest such as visible (phenotype) or invisible (chemical / physiological) differences. GMOs as well as random mutations may lead to (unexpected) chemical differences only detectable by destructive technologies (Fernie and Keurentjes, 2011; Hall, 2006). The progress made in the last (or rather, first) 10 years of plant metabolomics research is impressive. Publication numbers continue to increase exponentially (Hall & Hardy, 2011) and the field is gaining growing interest in a wide range of research areas. The diversity of plant species targeted, from algae to orchids, and the biological questions tackled, clearly emphasize the broad applicability of the technologies and demonstrate their added value and complementarity to other, already better established, '-omics' approaches. This has widely been covered in an extensive set of literature and the reader is referred to publications in broad reference works such as those of Hall (2006), Hardy and Hall (2012), Fernie, (2007), Saito *et al.* (2006), and Weckwerth (2013).

3.2.2 Substantial equivalence testing and natural variation

Most of the published experience in metabolomics testing of genetically modified plants comes from the comparative safety assessments of commercialized products for food and feed. These are reviewed in (Kok *et al.*, 2008) and in a recent book chapter (Harrigan and Harrison, 2012). OECD consensus documents for the safety assessment of new crops (OECD (Organisation for Economic Cooperation and Development), 2006) describe compositional considerations for these crops. For each relevant crop (alfalfa, canola, maize, cotton, soybean, potato, wheat, rice and sugar beet) a number of nutrients, vitamins and minerals, amino and/or fatty acids, and specific antinutrients or secondary metabolites that should be considered in a compositional analysis for comparative safety assessment. The assessment for a transgenic crop should involve a control with a history of safe use and commercial acceptance, which is (near-) isogenic but not containing the transgene. Natural variability in composition is addressed by including samples from multiple replicated field trials in diverse geographic regions, allowing assessing reproducibility of compositional differences in different locations. Secondly, natural variability is addressed by the inclusion of a set of different commercially available varieties from the same species (reference material), grown concurrently at the same locations as the test and control varieties. Next, to account for effects of different (climatological) growing conditions, the comparison over several growing seasons is advocated. Finally literature data from compositional studies on the same crop may be incorporated to describe inherent compositional variation. Finding statistically significant

differences of sufficient magnitude (hazard identification) would then trigger subsequent steps of hazard characterization. Thresholds ranges were suggested for different components, taking into account their nutritional relevance and the precision of their measurement (Hothorn and Oberdoerfer, 2006).

Several published examples of compositional analysis in safety assessments of GM crops exist. These were generally targeted at demonstrating substantial equivalence at the metabolome level, as a method of hazard identification, the first step of ERA. These include studies for potato (Catchpole *et al.*, 2005; Rogan *et al.*, 2000; Shepherd *et al.*, 2006), soybean (Berman *et al.*, 2011; Berman *et al.*, 2010; Berman *et al.*, 2009; Harrigan *et al.*, 2007; Lundry *et al.*, 2008; Zhou *et al.*, 2011a; Zhou *et al.*, 2011b), maize (Harrigan *et al.*, 2009; McCann *et al.*, 2007; Ridley *et al.*, 2002; Sidhu *et al.*, 2000; Skogerson *et al.*, 2010; Zhou *et al.*, 2011b), rice (Keymanesh *et al.*, 2009; Oberdoerfer *et al.*, 2005; Takahashi *et al.*, 2005), and wheat (Baker *et al.*, 2006; Obert *et al.*, 2004). Overall, the conclusion from these studies was that there were, if any, statistically significant differences in only a small number of components, that only a small part of these were reproducible over different locations and growing seasons, and that the latter were small in magnitude and all fell within the range of natural variation found among different varieties. Thus, substantial equivalence was concluded for all these cases. However, this observation does not allow the conclusion that substantial equivalence is the expected outcome in all studies. The available publications may be biased towards those cases were substantial equivalence was indeed found, while studies with different outcomes have not been published. Without access to all studies, it is not possible to estimate the true ratio between these two outcomes. An important observation from these studies, particularly on corn modified for drought tolerance (Harrigan *et al.*, 2009), for insect protection (Drury *et al.*, 2008; McCann *et al.*, 2007), or for herbicide tolerance (Ridley *et al.*, 2002) and soybean for insect protection (Berman *et al.*, 2011; Berman *et al.*, 2010; Berman *et al.*, 2009) or herbicide tolerance (Harrigan *et al.*, 2007; Lundry *et al.*, 2008; Zhou *et al.*, 2011a; Zhou *et al.*, 2011b), was that there is a high degree of compositional variation between samples grown at different locations or in different seasons, most notably for region-specific controls with regard to fatty acids and isoflavones in soybean. These studies underline the importance of understanding and including natural variation in providing biological context to observed differences between transgenic and control plants. What should also be noted is that since in all these instances substantial equivalence was concluded at the hazard identification stage, no further steps with regards to hazard characterization were deemed necessary and indeed were not taken. There is therefore no real example of a more extended risk assessment process as food or feed for these crops that could be taken as an example or as a primer for specific environmental risk assessment. However, one could expect that the same considerations (substantial equivalence, inclusion of natural variation in the form of multiple sites, multiple seasons and reference varieties) would apply there.

3.2.3 Statistical methods for substantial equivalence testing

Substantial equivalence for novel food or feed is established if there is no statistically significant difference from its conventional counterpart. In general, non-significance is determined by a p-value being lower than a pre-determined cut-off, usually 0.05. In **univariate** analysis, all samples are tested for the same variables (in metabolomics, compound concentrations or peak intensities) and all variables are tested one by one for statistical significance. Since metabolomics analysis usually yields a very large number of variables and contains also the effect of natural variation due to (non-transgenic) genotype, culture conditions and experimental variation, it can be advantageous to be able to globally analyse differences for all available variable data. This **multivariate** analysis can take different shapes, of which Principal Component Analysis (PCA) is particularly popular. In PCA a set of rotated axes is generated using linear combinations of the original axes, in this way reducing the number of variables needed to describe the variance in the dataset to a manageable number. Thus a PCA plot generally shows the 2 or 3 most important sources of variance (principle components) and the placement of the samples, and their differences for these components. The accompanying loading plot identifies the contribution of each compound or peak to these differences. Since there is no prior information about sample classes (for example, transgenic and non-transgenic) included, this is an unsupervised method of classification. **Discriminant factor analysis, linear discriminant analysis** and **decision tree analysis** are supervised methods that do take known classes as input and build a model that best describes the differences between these classes and thereby identify which variables (compounds or

peaks) differ the most strongly between the classes (Colquhoun *et al.*, 2006). The use of p-values for proof of safety has been criticized for its wide range of component-specific false negative error rates, and the use of defined safety thresholds has been propagated (Hothorn and Oberdoerfer, 2006). Another principle of sound statistical analysis that has been emerging is the concept of statistical power, i.e. the ability of a particular experimental set-up to identify differences of a particular size with statistical significance. This emphasizes the *a priori* inclusion of sufficient statistical power in the experimental design (Perry *et al.*, 2009).
The inherent difficulty of interpreting the results of multivariate analysis, due to their multi-dimensionality that is difficult to grasp by the human mind, has spurred the development of new methods for analysis. Recently, "metabolic distance" was proposed as a measure of similarity between samples and genotypes for 9 *Arabidopsis thaliana* ecotypes analysed for shoot and root metabolites by three different profiling methods. In this method, a limited number of components from a PCA are used and the sample scores of groups (replicates of the ecotypes) are used to calculate the Euclidian distance between the samples. An inter-sample distance matrix is used to assign an R-value (ranging from -1 to +1, R=0 for completely overlapping samples) for the distance between two accessions, the "metabolic distance". The use of more than two PC's in this calculation can result in a lower metabolic distance than apparent from inspecting just the first two PC's in a conventional PCA analysis (Houshyani *et al.*, 2012). This method constitutes an improvement in the analysis of the overall similarity of samples in a complex dataset. Any multivariate analysis, once significant differences or metabolic distance between samples (transgenic versus control) is detected, would have to be followed by a supervised multivariate or an univariate analysis to determine which metabolite(s) are responsible for the difference, and whether the identity and qualitative differences of the component are considered to be biologically significant. It is not immediately obvious how sensitive these multivariate analyses are for minor but biologically-relevant changes in only one 'minor' compound.

3.3 GMO's for targeted engineering of metabolic pathways

A large variety of literature on the subject of metabolic engineering in plants exists, which comprise varying depths of analysis with regards to the metabolome (reviewed by Bohnert *et al.* (2008) and Newell-McGloughlin (2008). Far fewer are crops that have been approved or submitted for approval and thus have undergone some kind of risk assessment (Center for Environmental Risk Assessment (CERA), 2012). Most of the examples listed here were not approved for cultivation in Europe and thus have not been subject to ERA according to the EFSA guidelines. These examples, with their references, are listed in Table 4.

Table 4. Summary of observed effects of targeted engineering of metabolic pathways on the metabolome

Species	Gene	Intended effect	Unintended effect	Reference
Maize LY038	*cordapA*	Higher lysine content	saccharopine and α-aminoadipic acid[1]	(Frizzi *et al.*, 2008)[2]
Argentine canola	thioesterase	High laurate and myristate in oil	None reported	(Voelker *et al.*, 1992
Soybean	*FAD2-1*	High oleic acid	None reported[3]	
Potato	*GBSS*	Starch composition	None reported[4]	
Maize	*Alpha amylase*	Starch hydrolysis	None reported	
Tree species	*various*	Lignin composition	Flavonoid accumulation, plant growth;	(Besseau *et al.*, 2007; Halpin, 2004; Halpin *et al.*, 2007; Vanholme *et al.*, 2008)

[1] Although unintended, these effects were not unexpected as the intermediates are from the same pathway.

[2] Although this reference uses a combination of a *cordapA* gene with a LKR/SDH inverted repeat sequence. LY038 contains only the *cordapA* gene.
[3] No literature reference is available, but high oleic-acid soybean was marked as substantially equivalent by Health Canada (http://www.hc-sc.gc.ca/fn-an/gmf-agm/appro/oleic_soybean-soja-eng.php)
[4] No literature reference is available, but high amylopectin potato was marked as substantially equivalent by the European Commission (http://ec.europa.eu/food/fs/sc/scp/out24_en.html)

- Monsanto's genetically modified LY038 maize expresses the *cordapA* gene from *Corynebacterium glutamicum*. This gene encodes for a lysine-feedback inhibition insensitive dihydropicolinate synthase (cDHDPS) enzyme, a regulatory enzyme in the lysine biosynthetic pathway. The decrease of feedback inhibition allows accumulation of higher levels of lysine, a desired trait particularly for feed production (Frizzi *et al.*, 2008; Huang *et al.*, 2005; Lucas *et al.*, 2007). The composition of forage and grain from maize LY038 was analysed and compared with a negative segregant LY038(-), which did not contain the *cordapA* gene. LY038 and LY038(-) maize lines were grown at five different sites. In addition, 20 conventional maize hybrids (reference material) were grown, four per site, to determine the amount of variation in nutrient composition that might be expected at each site. Forage and grain samples were collected from all plots and analysed for nutritional components, anti-nutrients and secondary metabolites. In addition, six lysine-related secondary metabolites from lysine biosynthetic and catabolic pathways, as well as free (not incorporated into protein) and total lysine, were analysed in LY038, LY038(-), and the conventional maize hybrid grain samples (CFSAN/Office of Food Additive Safety, 2005). Thus, this would constitute a targeted analysis. The compositional analyses of grain and forage of Lysine maize was shown to be compositionally equivalent to that of the negative segregant and to that of conventional maize in general, except for the intended increase in grain lysine content and an associated increase in the lysine-related catabolites, saccharopine and α-aminoadipic acid.

- Nutritional improvement by modifying oil or lipid content has been achieved in many crop and non-crop species (Cahoon and Schmid, 2008). Both Argentine canola (*Brassica napus*) as well as soybean has been engineered for modified oil content, by several companies. In canola, high laurate (12:0) and myristate (14:0) levels were achieved by Monsanto by inserting a thioesterase encoding gene from the California bay laurel (*Umbellularia californica*) (Voelker *et al.*, 1992). Other oil content modifications have been achieved by chemical mutagenesis and breeding. The analysis of levels of erucic acid and glucosinolates were within acceptable levels for canola quality. Similarly, the quality and quantity of seed protein were as expected. The fatty acid profile of high laurate canola was different from all control varieties, with increased levels of lauric acid (up to 40%) and myristic acid, and lower levels of oleic acid and linoleic acid, and less palmitic acid, less stearic, linoleic, arachidic, gadoleic, and slightly more behenic acid. No statistical differences in crude protein, crude fibre, gross energy content, and amino acids were noted between the processed meal of high laurate and control *B. napus* cultivars. It was determined that canola meal derived from transgenic lines was equivalent to meal from traditional *B. napus* varieties in terms of nutritional composition and safety.

- BASF developed a potato with altered starch composition (increased amylopectin to amylose ratio) for cultivation in Europe. The altered starch is achieved through the down-regulation of the expression of the granule-bound starch synthase encoding gene (*GBSS*). The observed difference in starch composition was from 85% amylopectin and 15 % amylose to 98% amylopectin and 2% amylose in transgenic line EH92-527-1. The amylopectin production serves the non-food industry, such as for paper production. By-products of the starch extraction process (e.g., pulp) can be used for other purposes, including animal feed or for other conventional, non-food purposes. These potatoes will be kept strictly separated from those for human consumption.

- Trees and other plants with altered lignin content, improving their quality for paper processing, biofuel production, or as feed are being investigated on a wide scale and some risk assessment studies have been performed on these (Halpin *et al.*, 2007). Here, increased insights in the metabolic pathways involved in lignin biosynthesis (reviewed in (Vanholme *et al.*, 2008) have revealed interactions between lignin metabolism and the biosynthesis of other cell wall polymers. Furthermore, systems biology approaches applied to 20 Arabidopsis mutants each affected in a single step of the lignin biosynthesis pathway revealed interactions

with global metabolism, such as primary metabolism and stress pathways even where there was no visible phenotype (Vanholme *et al.*, 2012). Although this study did not use transgenic plants, it highlights the connection of the lignin biosynthesis pathway with other pathways and could indicate research directions when looking for unintended effects of genetic modification of lignin production. As one of the few examples of transgenic crops in development, field trial results have been published for some of the projects. In a 4 year study at two sites, it was shown that poplar trees down-regulated for two lignin biosynthesis pathway genes in order to improve pulping properties, had low ecological impact. Of all parameters measured, only initial transgenic trunk and root decomposition was significantly faster than in controls. This was also observed in transgenic tobacco (Halpin *et al.*, 2007).

Other crops to be expected to be approved in the future are those with improved micronutrient (zinc, iron, provitamin A) levels for human consumption, such as those developed in the HarvestPlus program (http://www.harvestplus.org) and Golden Rice with increased carotenoid levels (see next paragraph).

3.3.1 Case study: Carotenoids and Vitamin A

Carotenoids are a class of isoprenoid metabolites that are produced by all photosynthetic organisms and provide plant organs such as fruits and flowers, but also leaves, with colours and in some cases are precursors of volatile compounds involved in aroma formation. They are important for plants as means of attraction for pollinators, photoprotectants, and as precursors for several hormones and signalling molecules, such as abscisic acid, strigolactones, and mycorradicin. In animal and human nutrition they have important (and supposedly health-promoting) roles as antioxidants, as feed additives and most importantly, as precursor (in the form of β-carotene, provitamin A) of the essential vitamin A (reviewed in(Cazzonelli, 2011).

The carotenoid biosynthesis pathway in plants and its links with other biosynthetic pathways have been thoroughly studied in many species and are reviewed elsewhere (Cazzonelli, 2011; Giuliano *et al.*, 2008). In summary, biosynthesis of carotenoids in plants requires an isoprenoid substrate, geranylgeranyl diphosphate (GGPP) derived from the plastidic MEP pathway in plants. GGPP and its precursor isopentyl phosphate (IPP) are also precursor for other plant metabolites with important functions, such as (tri- and sesqui-) terpenes, cytokinins, tocopherols, chlorophylls and gibberellins. The first dedicated step in carotenoid biosynthesis is the formation of phytoene from GGPP by phytoene synthase (PSY), followed by a desaturation and several isomerization steps leading to all-trans-lycopene, the red pigment of ripe tomato fruit. Cyclization of the linear lycopene molecule leads to carotene (the vitamin A precursor, and putative precursor of strigolactones), which itself is precursor of the xantophylls, among which capsorubin, the red pigment of pepper and lutein, the principal photoprotectant carotenoid in green tissues. Violaxanthin and neoxanthin are precursors of the plant hormone abscisic acid (ABA).

Many studies aimed at carotenoid metabolic engineering have been reported in literature and were reviewed elsewhere (Giuliano *et al.*, 2008). These studies can be broadly divided according to the approach and the utilized genes, into studies that directly increase expression of carotenoid biosynthetic genes by inserting gene copies either from plants or from bacteria, and studies that modify expression of regulatory genes or of genes encoding carotenoid-associated proteins, both increasing the sink strength of the plastid pool and hence not or only indirectly affecting the expression of the biosynthetic genes. Most of the reported studies were aimed at increasing lycopene content in tomato fruits or carotenoid production in potato tubers. The other well-known example is the biofortification of rice with β-carotene in order to raise vitamin A consumption in countries where rice is a staple food, the “Golden Rice” project (http://www.goldenrice.org). A more recent development is the synthesis of ketocarotenoids derived from xantophylls, in crop plants (tomato, potato, carrot, tobacco, canola and maize (Zhu *et al.*, 2009). This class of carotenoids, such as astaxanthin and canthaxanthin, is normally produced bacteria, fungi and green algae. Astaxanthin from marine algae gives pink flesh colour to salmon, trout, shrimp and lobster, and hence is an important food additive in aquaculture.

Tomato

The authors of one review (Giuliano *et al.*, 2008) identify as an area were advances in knowledge are needed, (quote) " the extent of metabolic cross-talk between the carotenoid pathway and other pathways influencing plant development and nutritional value". Tomato fruit is probably the best studied system with regards to metabolomics, and specifically so for studies focusing on raising lycopene or β-carotene content. A variety of approaches has been used towards this goal (reviewed in (Fraser *et al.*, 2009)), although not all of these also contain a thorough metabolomics characterization of expected and unexpected effects of the modification. The constitutive overexpression of fruit phytoene synthase under control of the CaMV 35S promoter resulted in plants with ectopic pigmentation, reduced leaf chlorophyll and reduced height. The dwarfing effect was attributed to an up to 30-fold reduction in gibberellins (GAs) and could be complemented with exogenous gibberellin (Fray *et al.*, 1995). Apparently in these plants, through the above described pathway connectivity, GGPP is diverted from GA to carotenoid synthesis. A smaller number of plants had no apparent vegetative phenotype and slightly raised β-carotene levels in ripe fruits. Both lycopene and β-carotene were produced in earlier fruit stages (from immature green fruit onwards) compared to wild type fruits. This was accompanied by extensive perturbations in the fruit metabolome, resembling those normally occurring during ripening. In another case, the production of β-carotene tomato through constitutive over-expression of a bacterial phytoene desaturase was achieved. At the same time, tocopherol (a vitamin E) levels in fruit were upregulated through an as yet unknown mechanism (Römer *et al.*, 2000).

Several approaches have been reported utilizing regulatory genes for raising carotenoid content in tomato and other crops. In tomato, several mutations affecting the light signal transduction pathway (High Pigment, *hp* mutations) affect carotenoid levels as well as anthocyanins and phenylpropanoids in plants, at least partially through increase of plastid sink strength (Azari *et al.*, 2010; Levin *et al.*, 2006). Many of these mutations have pleiotropic effects on plant vigour or fruit yield as well. Fruit-specific RNAi-mediated down-regulation of the underlying gene for the *hp2* mutation, *DE-ETIOLATED1*, resulted in similar changes in fruit secondary metabolite content without these agronomically negative effects (Davuluri *et al.*, 2005; Enfissi *et al.*, 2010). This observation highlights a potential difficulty when choosing reference material for comparative safety assessment if such plants were to be commercialized. Although the spontaneous *hp* mutants have not been commercialized so far, their inclusion as reference material in a comparative safety assessment of transgenic *DET1* knock-down varieties would probably greatly alter the acceptable range of many metabolite concentrations for substantial equivalence. Expression of a fibrillin protein, involved in sequestration of carotenoids in plastids, resulted in a 2-fold increase in carotenoids, as well in an increase of carotenoid-derived flavour volatiles (Simkin *et al.*, 2007). Although carotenoid-derived volatiles were not reported in the other studies mentioned above, this increase may well be a general consequence of increased carotenoid production in tomato. Carotenoid-derived volatiles in tomato are involved in flavour formation (Simkin *et al.*, 2004), and in many other plants are an important component of floral scent.

Potato

Similar to in tomato fruit, expression of a bacterial phytoene synthase in potato tubers increased β-carotene, lutein, and violaxanthin (Ducreux *et al.*, 2005). Overexpression of a bacterial *DXS* resulted in increased phytoene levels and early sprouting of the tubers. This could be likely attributed to the observed increase in the level of trans-zeatin riboside (a cytokinin) in tubers, which was correlated with *DXS* expression (Morris *et al.*, 2006). Since cytokinins are produced from isopentyl pyrophosphate, the product of the MEP pathway and precursor of carotenoids, this is probably an unexpected by-product of the increased flux through the MEP pathway caused by *DXS* overexpression.

Golden Rice

In the "Golden Rice" project, the aim is to raise provitamin A (β-carotene) content in rice endosperm to levels at which rice can deliver the recommended daily uptake in developing societies where rice is the main food source. Similar initiatives are being undertaken for maize, potato, canola and tomato (see above). These studies are reviewed in (Farré *et al.*, 2011). Briefly, the first Golden Rice (GR1) contained a *PSY* gene from daffodil and the carotene desaturase (*CRTI*) gene from a bacterium, the former under control of an endosperm-specific promoter, and the latter constitutively expressed under control of the CaMV 35S promoter (Ye and Beyer, 2000). Endosperm from transgenic plants contained higher levels of carotenes as well as of

zeaxanthin and lutein, even in the absence of an exogenous source of lycopene cyclase. This indicates that in rice, endosperm lycopene cyclase activity is constitutively expressed or induced upon lycopene production in transgenic seeds. This is an example of a "silent metabolic pathway" as mentioned in Chapter 1.2.3 of this report, that is activated. While lycopene was the expected end product of CRTI, the unsuspected presence of lycopene cyclase activity in rice endosperm converted the product further, although in this case mostly to the desired end product, carotene. In Golden Rice 2, the *PSY* gene originated from maize, leading to 23 times higher β-carotene levels than in GR1 (Paine *et al.*, 2005). In both cases the reported metabolomics analysis does not extend beyond the analysis of the major carotenoids, so no assessment of unintended changes can be made in these cases.

The above examples show that unintended effects from a genetic modification aimed at altering a single metabolite can result in unintended effects with respect to compounds from the same pathway or from connected pathways. This would imply that, if sufficient knowledge of the pathway network is available, targeted metabolomics would detect such unintended effects.

3.4 Possible environmental effects of metabolome changes

Plants do not function in isolation but are part of an (agro-) ecosystem, in which they interact with their environment during their lifetime and beyond. Herbivores feed on the plant and are affected by the (anti-) nutritional quality of the plant. Effects of this feeding extend further into the food web as predators or parasites feed on the herbivores (tritrophic interactions) and may be directly or indirectly affected by plant metabolites. Flowering plants are visited by pollinators, for their pollen or nectar. Furthermore, plants emit volatiles that have functions as attractant or repellent for pollinators, herbivores or their natural enemies. Plant roots exude metabolites into the rhizosphere, which may affect rhizobacteria, soil fungi and mycorrhizae, nematodes as well as macrofauna. Finally, unharvested parts of the plants are left in the field to naturally decompose, sometimes after tilling or burning. Plant parts will be further degraded by above ground or underground detritus feeders and finally by microorganisms. During degradation, plant metabolites permeate into the soil by leaching. Clearly, all these processes are to some extent affected by the plant's composition, including its metabolites. In this context it has to be noted that in many commercialized transgenic crops the modification is to some extent, or very specifically, targeted to the edible, harvested part of the plant (for example, oil content in seed) and thus may have little effect on other plant parts. On the other hand, this also means that available data on compositional equivalence for food/feed use are limited to those parts and do not include analysis of, for example, roots and leaves that do play a role in the interaction with the environment. The only other source of information on unexpected environmental effects of GM crops comes from the experience of monitoring of environmental effects in already commercialized crops, mainly from countries outside Europe and specifically the US. van den Brink *et al.*, (2010) have produced an inventory of all observed unexpected environmental effects in maize, sugar beet, potato, oilseed rape, alfalfa, soybean, and cotton. They did not identify any unexpected effects that might be attributed to a significant change in the metabolome of these plants.

3.4.1 Effects on target and non-target organisms

Target organisms in this context are defined as plant pests or pathogens of the modified crop plants, which are targeted by the specific modification, in order to make the plant resistant or to deter insect pests from feeding on the plant tissues. A clear example is the group of Bt crops that have been commercialized since 1996. Although there is a lot of evidence that secondary metabolites play a role in the interaction between host plant and pest or pathogen both in attracting as well as in repelling or killing the pest, there are no plants with intended modified secondary metabolites which are currently under commercialization.

Glucosinolates

A relatively well characterized type of secondary metabolite that plays a role in plant/insect-interactions is the glucosinolate complex of compounds in the Brassicaceae, for which some 120 different structures have been identified. These compounds affect important quality and flavour characteristics as well as agronomic characteristics such as resistance to pathogen and insect pests. These have been characterised in Arabidopsis as well as in crucifer crops such as oilseed rape, cabbage and broccoli. The diversity and function of glucosinolates in both above-ground tissues as well as in roots has been comprehensively reviewed by van Dam *et al.* (2009). Overall, roots have a higher diversity and higher concentrations of glucosinolates than shoots, and the predominant compounds are different for roots and shoots, probably reflecting their different functions. Glucosinolates are constitutively produced as well as being induced upon herbivore feeding.

From the accumulated data on the role of glucosinolates on plant interactions with other organisms, it is clear that both targeted modification of the glucosinolate production in crucifer crops, as well as unintended changes in this production due to other modifications may very well alter the interaction of such crops with their environment. Some of these effects have been explored in studies performed in the context of the Dutch "ERGO" programme. More specifically, Houshyani et al. have modified glucosinolate content in Arabidopsis by overexpression of a regulatory MYB transcription factor, HAG1/MYB28, in 4 different ecotypes. Effects of the modification on specific glucosinolates was highly dependent on the background genotype and had differing effects on both general as well as specialist herbivores (Houshyani, 2012).
The effect of an introduced indirect defence mechanism in Arabidopsis - the production of volatile compounds in order to attract natural enemies of an herbivore – has also been investigated. In two out of three transformed ecotypes this resulted in production of the volatile nerolidol and a higher emission of two other volatiles. The overall effect of the HAG1/MYB28 modification, the nerolidol modification, and an unrelated overexpression of a Bt gene (SN19) on the metabolome of Arabidopsis was also investigated. A comparison was made for four wild type ecotypes, and 18 lines with the three traits in 3 backgrounds. The earlier mentioned "metabolic distance" method was used to determine the divergence at the metabolome level of the transgenic lines from the baseline as was defined by the four wild type ecotypes. Both roots and shoots were analysed using two complementary metabolomics platforms. Overall, the effect of modifications was larger in the root metabolome than it was in the shoot metabolome. Several of the transgenic lines exhibited substantial equivalence when compared to the baseline, while other lines, sometimes with the same modification, were statistically distinguishable from that baseline. It was suggested that these significant distances might disappear when the effect of varying (natural) environmental conditions on the baseline is included. However, this was not tested in this particular study.

Tobacco and altered resistance to pathogens
Salicylic acid (SA) plays a key role in the induction of increased resistance of plants to pathogens after an initial pathogen attack, so-called Systemic Acquired Resistance (SAR) (Fu and Dong, 2013). Expression of two bacterial SA biosynthetic genes producing SA from chorismate in transgenic tobacco raised SA and SA glycoside levels up to one thousand times. This rendered the plants resistant to bacterial and viral pathogens (Verberne *et al.*, 2000). Non-targeted (2D-) NMR analysis combined with multivariate data analysis of wild type tobacco leaves after TMV infection, or leaves with systemically acquired resistance from the same plants showed a large number of changes in metabolites, associated with infection, SAR, or both. Increased 5-caffeoylquinic acid, R-linolenic acid analogues, and sesqui- and diterpenoids were found in infected plants (Choi *et al.*, 2004). Metabolic fingerprinting using NMR of wild type and transgenic plants combined with multivariate data analysis allowed the discrimination of the two genotypes, with differences in chlorogenic acid, malic acid, and sugars (Choi *et al.*, 2004). Flavonoids and chlorogenic acid were suppressed in the transgenic plants compared to controls (Nugroho *et al.*, 2002). These studies show that non-targeted metabolite analysis is capable of capturing the complex and difficult to predict changes occurring during TMV infection and SAR in tobacco.

Effect on plant-associated organisms from altered composition of the plant
Plant-herbivore interactions
Herbivores depend on plant cues such as colour and smell for finding host plants and on metabolite contents for attractiveness and usefulness as food, an ultimately for their fitness. There is a vast amount of literature on

host preference and the role of metabolites in insect/herbivore-interactions, which are reviewed elsewhere. Although several different plant species and insects have been studied, only a few were on relevant crops (Åhman *et al.*, 2010; Clavijo McCormick *et al.*, 2012; Farré-Armengol *et al.*, 2013; Irwin *et al.*, 2004; Riffell, 2011; Rodríguez *et al.*, 2013; Schiestl, 2010; Turlings *et al.*, 2012). Plants also react to herbivore damage by induced synthesis of defence compounds or defend themselves by the constitutive production of defence compounds or antinutritional metabolites. Probably one of the best studied is the interaction of tobacco with its pathogens and herbivores, in which induced nicotine production plays an important role (Baldwin, 2001). Various studies describe the complexity of plant defence mechanisms against pathogens and herbivores for tobacco, with several describing how interference with metabolite production affects herbivore performance on the plant (Gaquerel *et al.*, 2012; Jassbi *et al.*, 2008; Kajikawa *et al.*, 2011; Steppuhn and Baldwin, 2007; Stitz *et al.*, 2011; Wünsche *et al.*, 2011). Genetic modification with intended or unintended effects on such pathways may have similar effects. For example, tobacco plants with a reduced nicotine content, due to a genetic modification, are more attractive to and sustain considerable more damage from various herbivores both in the laboratory and in natural settings (Steppuhn *et al.*, 2004).

Tritrophic interactions

Due to the introduction of insect-resistant Bt crops (since 1996), a lot of attention has been given to research into the effect of genetic modification on the predators and parasitoids of the primary herbivores that are targets or non-targets of the modification. This plant-herbivore-predator/parasitoid interaction is termed a 'tritrophic interaction'. Modification leading to mortality of the target herbivore may have direct effects on the next trophic level if, as in the case of highly specialized parasitoids, the predator or parasitoid is completely dependent on the targeted herbivore. Herbivores feeding on Bt crops may accumulate plant metabolites and transfer those to their predators and parasitoids, where they may have an effect. The accumulation might also affect prey preferences for the predator or parasitoids. Additionally, it has been shown that herbivory causes plants to produce and/or release volatiles that attract natural enemies of the herbivore. Changes in the metabolome as a result of genetic modification may, intended or unintended, change the released volatiles in type or quantity and thereby alter this mechanism. Also here, natural variation is a factor to consider. Volatile emissions upon insect feeding or mimicking of the wounding response by jasmonate treatment showed marked quantitative differences in 9 ecotypes of Arabidopsis. Moreover, the variation in volatile production after jasmonate treatment was reflected in the behaviour of a parasitoid (Snoeren et al. 2010). Increasing green leaf volatile production in transgenic Arabidopsis increased parasitation of a leaf-feeding herbivore by parasitoids, while decreasing volatile production decreased parasitation of the same herbivores (Shiojiri *et al.*, 2006). These results indicate how (unintentionally) modifying volatile production in transgenic crops could affect tritrophic interactions as well.

Effects on soil and rhizosphere organisms

The rhizosphere contains plant root exudates, and secreted metabolites from bacteria, fungi, and other soil organisms. Those compounds may affect growth and replication of microorganisms, while the microorganisms themselves may have important functions in plant growth promotion, or are pathogens to the plant. Metabolomics studies on the rhizosphere have been reviewed by Chakravarthy et al. (2012). Root exudates contain mineral ions, inorganic acids as well as organic compounds such as amino acids, organic acids, sugars, phenolics and many secondary metabolites. These will likely, to a large extent, influence the microbiome and determine which specific microorganisms can grow around the root.

Particularly dependent on certain exuded or secreted root secondary metabolites, are symbiotic microorganisms, such as root-nodulating *Rhizobium* and *Frankia* species or mycorrhizae. These organisms depend on very specific plant metabolites for the establishment of a successful symbiosis. Genetic modification resulting in quantitative or qualitative changes in those compounds may affect those symbioses.

Also here, glucosinolate production in *Brassica* is one of the most well studied systems. One of the most comprehensive studies on the possible effects of genetic modification on soil and rhizosphere organisms related to changes in the metabolome is that by Kabouw, who studied intraspecific variation and its effects on above and below-ground organisms in white cabbage (Kabouw *et al.*, 2010). It was shown that both above and

below-ground parts of the plants show intra-specific variation in glucosinolate content, and the latter (glucosinolate in roots) is reflected in intra-specific variation in root exudate glucosinolate concentration, albeit that the profiles of glucosinolates were not correlated between roots and their exudates. Variation in glucosinolate concentration had effects on an aboveground non-target insect directly associated with the plant (aphid, *Brevicoryne brassicae*) as well as in the next trophic level, on the development of a parasitoid of that aphid (*Diaeretiella rapae*). Below ground, intraspecific variation did affect a root-associated nematode pest (*Pratylenchus sp.*), but not non-associated soil organisms. Moreover, below-ground organisms could affect aboveground organisms through alteration of the host plant. However, intra-specific variation in metabolome content did not alter these below ground/above ground interactions significantly in this study. Similar below ground/above ground-interactions were also found on two grass species (Bezemer *et al.*, 2005) and two different oat cultivars (Sell and Kuo-Sell, 1990).

Known effects of genetic modification on rhizosphere processes are scarce and have been reviewed by Kabouw *et al.* (2012). However, also here no specific examples of altered root or root exudate metabolites affecting rhizosphere organisms were found. Altered rhizosphere microorganism diversity as a result of genetic modification of tobacco was reported, but the description of the modifications used was too poor to correlate it with the altered metabolome, nor was an attempt made to propose a mechanism involving altered metabolite content (Andreote *et al.*, 2008)

In conclusion, there is a large amount of literature or published data on the effects of plant metabolites on plant-environment interactions, but these studies are very diverse and usually are not specifically addressing cultivated crop species. Examples of genetic modification altering metabolite content with altered plant-environment interactions as a consequence are limited to examples where the modification was intended to increase resistance to insects or pathogens. Only in a few cases, the effects of these modifications on non-target organisms, i.c. non-herbivorous species associated with plants, were assessed.

4 Conclusions

What is the potential contribution of metabolomics for the environmental risk assessment of genetically modified crops?

Unintentional effects

Metabolomics is a relatively new but fast developing discipline in biology. In the future it may contribute to the Environmental Risk Assessment of transgenic crops, as it already does in the Comparative Safety Assessment for food and feed uses. The possible scenario's through which genetic modification of a crop plant may lead to an altered metabolome, which in turn may or may not have undesirable effects on the environment, are described. Such metabolome alterations may be intentional (the goal of the modification) or unintentional, and in the latter case predictable to a certain extent, or not predictable at all. The mechanisms leading to such unintended effects may be due to the specific insertion site of the transgene in the genome, or be a result of an incomplete understanding of the metabolic pathways of the host plant.

The use of knowledge on metabolic pathway regulation and connectivity

Our understanding of plant genome function is currently at a level at which, in combination with sequencing of the transgene insert region (molecular characterization in ERA), the potential of a particular event to have unintended effects on the metabolome may often be recognized at an early stage of development. It may trigger the selection of another event for further development or when it is subject to ERA, it may trigger additional queries. In the case of suspected interference with host gene function or the formation of hybrid transcripts, transcriptomics would be able to detect these changes most of the time. It is shown that although in theory there are many ways in which transgene insertion can alter enzyme specificity or completely knock out activity of a gene, there are no published examples of events where this occurred unintentionally and was discovered later on to be causing undesired phenotypes. There are however, several examples where modification of a metabolic pathway affected more than one compound within that pathway or even in pathways connected to it (such as in the examples for modification of carotenoid content). These examples suggest that such unintended effects would be relatively easily discovered with a limited experimental approach targeted at intermediates of the pathway or those from connected pathways. This may prevent the requirement of performing a holistic, all-inclusive metabolic profiling exercise which would be costly and demanding. Additionally, without clear endpoints, based on our knowledge of environmental effects, such risk assessment would likely be inconclusive (see below).

A clear limitation of the current metabolomics technologies is that still a relatively small number (~ 10%) of detectable compounds can actually be identified. Additionally, the remaining unidentified compounds can only be quantified in relative terms. Without the identification, risk assessment for compounds with concentrations that significantly deviate from those in the reference samples is not possible. Another limitation, as in all compositional characterizations, is that metabolomics delivers a snapshot of the content at a certain time point. Therefore, metabolomics cannot replace testing or monitoring in field trials, where the (accumulated) effects on populations of species in realistic conditions can be measured.

A lack of knowledge of metabolite effects on the interaction with the environment

The biggest problem for the predictive power of metabolomics in ERA is likely to be the knowledge gap that exists for the role or function of most plant metabolites in their interaction with the environment. Apart from a few well studied systems, such as glucosinolate production in Brassicaceae (discussed in this report) and volatiles in insect resistance and signalling, little is known of the effect of most plant metabolites on herbivores and their predators or parasitoids, on soil microorganisms, etc. Also it is difficult to predict how a change in one or a few metabolites would affect the interaction of the plant with its environment as a whole. Thus, even if alterations in metabolite content would be detected and the contributing metabolites could be identified, this

would hardly contribute to a more targeted ERA. In those cases where prior knowledge exists about metabolite effects on the environment, such as for glucosinolates, targeted metabolomics may identify significant changes in the metabolome of transgenic plants. These data could help to further optimize experiments such as laboratory tests on selected organisms or define organisms to focus on in field trials on which potential adverse effects could be expected. One observation made during the study was that data on metabolite-environment interactions are highly diverse and fragmented, with relatively few data for actual crop species, including those regularly cultivated in the Netherlands. Thus ERA, should it include metabolomics analysis in the future, would highly benefit from a centralized database that collects data from all studies involving intended and unintended effects on the metabolome from genetic modification or otherwise. Additionally, such a database should collect data on the effects of individual plant metabolites or metabolite classes on the environment, such as on species with an import role in ecosystem functioning.

Possible scenarios for the use of metabolomics in ERA

Compositional analysis for ERA as part of the approval process for commercial cultivation (not including use for food or feed) is as far as we have been able to determine, not a standard requirement component in any country. Untargeted metabolomics in order to establish "substantial equivalence" for non-harvested plant parts (both above and below ground) and even volatiles in the emitted by the plant may be introduced as part of the compositional analysis step in ERA. Statistical analysis comparable to that used for substantial equivalence in food or feed and described in this report could be applied to accept or reject substantial equivalence. This could even be done in a two-tiered approach: 1). initial testing of plant parts (roots, above-ground organs) using metabolomics analyses and proceed to the next step only if significant differences are found. A confounding factor here is that the plant metabolome may be affected by developmental stage, wounding, herbivory and other biotic or abiotic stresses, which would rapidly increase the number of different conditions under which to perform comparisons. 2). testing of emitted volatiles and of root exudates to further characterize "exposure" of the environment. Compared to the testing for substantial equivalence in food or feed, an additional complication may be the choice of comparators. In food or feed tests, these are non-transgenic plants representing a range of commercial varieties of the same species. Regardless of the comparators chosen, in the event that substantial equivalence is rejected, and considering our knowledge gap with regard to the role of metabolites in plant/environment-interactions, metabolomics could contribute to the ERA only in isolated cases. In all other cases other data sources such as agronomic and phenotypic characterisation, as well study of plant-environment interactions in the laboratory or in field trials would be more informative. Alternatively, the decision to require metabolomics data for the approval process, and if so what type of data, could be required on a case-by-case basis, e.g. dependent on the intention of the genetic modification. Where no metabolic changes are intended, such as in Bt crops, no metabolomics data would have to be required. Where particular metabolic changes are intended, one could require targeted metabolomics, the extent of which should be made depending on our knowledge of the intentionally changed metabolite pathway and its connectivity with other pathways. Only if significant unintentional changes are observed, more extensive further tests, such as tests on non-target organism in the laboratory or the field, should be required.

In conclusion, the field of metabolomics is undergoing a fast development and is already contributing to establishing substantial equivalence in GM plant-derived food and feed. In the future, it may contribute more and more to compositional analysis in ERA for the detection of unintended effects. However, there is a knowledge gap with regard to the role of plant metabolites in plant/environment-interactions. Apart from a few specific cases where more pre-existing knowledge is available, the contribution of metabolomics in Environmental Risk Assessment is considered limited for the time being.

Literature

Åhman, I., R. Glinwood, and V. Ninkovic. 2010.
The potential for modifying plant volatile composition to enhance resistance to arthropod pests. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 5.

Ahuja, I., R. Kissen, and A. M. Bones. 2012.
Phytoalexins in defense against pathogens. Trends in Plant Science 17:73-90.

Andreote, F. D., R. Mendes, F. Dini-Andreote, P. B. Rossetto, C. A. Labate, A. A. Pizzirani-Kleiner, J. D. van Elsas, J. L. Azevedo, and W. L. Araúo. 2008.
Transgenic tobacco revealing altered bacterial diversity in the rhizosphere during early plant development. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology 93:415-424.

Azari, R., Y. Tadmor, A. Meir, M. Reuveni, D. Evenor, S. Nahon, H. Shlomo, L. Chen, and I. Levin. 2010.
Light signaling genes and their manipulation towards modulation of phytonutrient content in tomato fruits. Biotechnology Advances 28:108-118.

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

Baldwin, I. T. 2001.
An ecologically motivated analysis of plant-herbivore interactions in native tobacco. Plant Physiology 127:1449-1458.

Beale, M. H., and M. R. sussman. 2011.
Metabolomics of *Arabidopsis thaliana*. p. 157-180. *In* R. D. Hall (ed.), Biology of Plant Metabolomics., Annual Plant Reviews. Blackwell Publishing Ltd., Oxford, UK.

Berman, K. H., G. G. Harrigan, M. A. Nemeth, W. S. Oliveira, G. U. Berger, and F. S. Tagliaferro. 2011.
Compositional equivalence of insect-protected glyphosate-tolerant soybean MON 87701 × MON 89788 to conventional soybean extends across different world regions and multiple growing seasons. Journal of Agricultural and Food Chemistry 59:11643-11651.

Berman, K. H., G. G. Harrigan, S. G. Riordan, M. A. Nemeth, C. Hanson, M. Smith, R. Sorbet, E. Zhu, and W. P. Ridley. 2010.
Compositions of forage and seed from second-generation glyphosate-tolerant soybean MON 89788 and insect-protected soybean MON 87701 from Brazil are equivalent to those of conventional soybean (glycine max). Journal of Agricultural and Food Chemistry 58:6270-6276.

Berman, K. H., G. G. Harrigan, S. G. Riordan, M. A. Nemeth, C. Hanson, M. Smith, R. Sorbet, E. Zhu, and W. P. Ridley. 2009.
Compositions of seed, forage, and processed fractions from insect-protected soybean MON 87701 are equivalent to those of conventional soybean. Journal of Agricultural and Food Chemistry 57:11360-11369.

Besseau, S., L. Hoffmann, P. Geoffroy, C. Lapierre, B. Pollet, and M. Legrand. 2007.
Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. Plant Cell 19:148-162.

Bezemer, T. M., G. B. De Deyn, T. M. Bossinga, N. M. Van Dam, J. A. Harvey, and W. H. Van Der Putten. 2005.
Soil community composition drives aboveground plant-herbivore-parasitoid interactions. Ecol Lett 8:652-661.

Biais, B., J. W. Allwood, C. Deborde, Y. Xu, M. Maucourt, B. Beauvoit, W. B. Dunn, D. Jacob, R. Goodacre, D. Rolin, and A. Moing. 2009.
1H NMR, GC-EI-TOFMS, and data set correlation for fruit metabolomics: Application to spatial metabolite analysis in melon. Analytical Chemistry 81:2884-2894.

Bohnert, H. J., H. T. Nguyen, and N. G. Lewis. 2008.
Bioengineering and Molecular Biology of Plant Pathways. Advances in Plant Biochemistry and Molecular Biology. vol. 1. Elsevier Ltd.

Borevitz, J. O., Y. Xia, J. Blount, R. A. Dixon, and C. Lamb. 2000.

Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12:2383-2393.

Browne, J., A. Bacic, M. Tester, and U. Roessner. 2011.
Abiotic stress and metabolomics. p. 61-86. *In* R. D. Hall (ed.), Biology of Plant Metabolomics., Annual Plant Reviews, . Blackwell Publishing Ltd., Oxford, UK.

Cahoon, E. B., and K. M. Schmid. 2008.
Metabolic Engineering of the Content and Fatty Acid Composition of Vegetable Oils. p. 161-200, Advances in Plant Biochemistry and Molecular Biology, vol. 1.

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

Cazzonelli, C. I. 2011.
Carotenoids in nature: Insights from plants and beyond. Functional Plant Biology 38:833-847.

Center for Environmental Risk Assessment (CERA). 2012.
GM Crop Database.(Center for Environmental Risk Assessment (CERA), ILSI Research Foundation,);

Cevallos-Cevallos, J. M., J. I. Reyes-De-Corcuera, E. Etxeberria, M. D. Danyluk, and G. E. Rodrick. 2009.
Metabolomic analysis in food science: a review. Trends in Food Science and Technology 20:557-566.

CFSAN/Office of Food Additive Safety. 2005.
Biotechnology Consultation Note to the File BNF No. 000087.

Choi, H. K., Y. H. Choi, M. Verberne, A. W. M. Lefeber, C. Erkelens, and R. Verpoorte. 2004.
Metabolic fingerprinting of wild type and transgenic tobacco plants by 1H NMR and multivariate analysis technique. Phytochemistry 65:857-864.

Clavijo McCormick, A., S. B. Unsicker, and J. Gershenzon. 2012.
The specificity of herbivore-induced plant volatiles in attracting herbivore enemies. Trends in Plant Science 17:303-310.

Colquhoun, I. J., G. Le Gall, K. A. Elliott, F. A. Mellon, and A. J. Michael. 2006.
Shall I compare thee to a GM potato? Trends in Genetics 22:525-528.

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

Draper, J. H., S. Rasmussen, and H. Zubair. 2011.
Metabolite analysis and metabolomics in the study of biotrophic interactions between plants and microbes., p. 25–60. *In* R. D. Hall (ed.), Biology of Plant Metabolomics., Annual Plant Reviews. Blackwell Publishing Ltd., Oxford, UK.

Drury, S. M., T. L. Reynolds, W. P. Ridley, N. Bogdanova, S. Riordan, M. A. Nemeth, R. Sorbet, W. A. Trujillo, and M. L. Breeze. 2008.
Composition of forage and grain from second-generation insect-protected corn MON 89034 is equivalent to that of conventional corn (*Zea mays* L.). Journal of Agricultural and Food Chemistry 56:4623-4630.

Ducreux, L. J. M., W. L. Morris, P. E. Hedley, T. Shepherd, H. V. Davies, S. Millam, and M. A. Taylor. 2005.
Metabolic engineering of high carotenoid potato tubers containing enhanced levels of β-carotene and lutein. Journal of Experimental Botany 56:81-89.

EFSA Panel on Genetically Modified Organisms. 2010.
Guidance on the environmental risk assessment of genetially modified plants. EFSA Journal 8:1879-1990.

Enfissi, E. M. A., F. Barneche, I. Ahmed, C. Lichtlé, C. Gerrish, R. P. McQuinn, J. J. Giovannoni, E. Lopez-Juez, C. Bowler, P. M. Bramley, and P. D. Fraser. 2010.
Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. Plant Cell 22:1190-1215.

Farré-Armengol, G., I. Filella, J. Llusia, and J. Peñuelas. 2013.
Floral volatile organic compounds: Between attraction and deterrence of visitors under global change. Perspectives in Plant Ecology, Evolution and Systematics 15:56-67.

Farré, G., C. Bai, R. M. Twyman, T. Capell, P. Christou, and C. Zhu. 2011.

Nutritious crops producing multiple carotenoids - a metabolic balancing act. Trends in Plant Science 16:532-540.

Fernie, A. R. 2007.
The future of metabolic phytochemistry: Larger numbers of metabolites, higher resolution, greater understanding. Phytochemistry 68:2861-2880.

Fernie, A. R., A. Aharoni, L. Willmitzer, M. Stitt, T. Tohge, J. Kopka, A. J. Carroll, K. Saito, P. D. Fraser, and V. de Luca. 2011.
Recommendations for reporting metabolite data. Plant Cell 23:2477-2482.

Fernie, A. R., and J. J. B. Keurentjes. 2011.
Genetics, genomics and metabolomics. p. 219–260. *In* R. D. Hall (ed.), Biology of Plant Metabolomics., Annual Plant Reviews. Blackwell Publishing Ltd., Oxford, UK.

Fiehn, O. 2002.
Metabolomics – the link between genotypes and phenotypes. . Plant Molecular Biology 48:155–171.

Fiehn, O., T. Kind, and D. Kumar Barupal. 2011.
Data processing, metabolomic databases and pathway analysis., p. 367–406. *In* R. D. Hall (ed.), Biology of Plant Metabolomics., Annual Plant Reviews. Blackwell Publishing Ltd., Oxford, UK.

Fiehn, O., L. W. Sumner, S. Y. Rhee, J. Ward, J. Dickerson, B. M. Lange, G. Lane, U. Roessner, R. Last, and B. Nikolau. 2007.
Minimum reporting standards for plant biology context information in metabolomic studies. Metabolomics 3:195-201.

Filipecki, M., and S. Malepszy. 2006.
Unintended consequences of plant transformation: A molecular insight. Journal of Applied Genetics 47:277-286.

Fraser, P. D., E. M. A. Enfissi, and P. M. Bramley. 2009.
Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. Archives of Biochemistry and Biophysics 483:196-204.

Fray, R. G., A. Wallace, P. D. Fraser, D. Valero, P. Hedden, P. M. Bramley, and D. Grierson. 1995.
Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. Plant Journal 8:693-701.

Frizzi, A., S. Huang, L. A. Gilbertson, T. A. Armstrong, M. H. Luethy, and T. M. Malvar. 2008.
Modifying lysine biosynthesis and catabolism in corn with a single bifunctional expression/silencing transgene cassette. Plant Biotechnology Journal 6:13-21.

Fu, Z. Q., and X. Dong. 2013.
Systemic acquired resistance: Turning local infection into global defense. p. 839-863, Annual Reviews in Plant Biology, vol. 64.

Gaquerel, E., A. Steppuhn, and I. T. Baldwin. 2012.
Nicotiana attenuata α-dioxygenase1 through its production of 2-hydroxylinolenic acid is required for intact plant defense expression against attack from *Manduca sexta* larvae. New Phytologist 196:574-585.

Giuliano, G., R. Tavazza, G. Diretto, P. Beyer, and M. A. Taylor. 2008.
Metabolic engineering of carotenoid biosynthesis in plants. Trends in Biotechnology 26:139-145.

Hall, R. D. 2006.
Plant metabolomics: From holistic hope, to hype, to hot topic. New Phytologist 169:453-468.

Halpin, C. 2004.
Re-designing lignin for industry and agriculture. Biotechnology and Genetic Engineering Reviews 21:229-245.

Halpin, C., S. C. Thain, E. L. Tilston, E. Guiney, C. Lapierre, and D. W. Hopkins. 2007.
Ecological impacts of trees with modified lignin. Tree Genetics and Genomes 3:101-110.

Hardy, N. W., and R. D. Hall. 2012.
Plant Metabolomics Methods and Protocols. Methods in Molecular Biology vol. 860. Humana Press, Otawa, USA.

Harrigan, G. G., and J. M. Harrison. 2012.
Assessing compositional variability through graphical analysis and Bayesian statistical approaches: Case studies on transgenic crops. Biotechnology and Genetic Engineering Reviews 28:15-32.

Harrigan, G. G., W. P. Ridley, K. D. Miller, R. Sorbet, S. G. Riordan, M. A. Nemeth, W. Reeves, and T. A. Pester. 2009.
The forage and grain of MON 87460, a drought-tolerant corn hybrid, are compositionally equivalent to that of conventional corn. Journal of Agricultural and Food Chemistry 57:9754-9763.

Harrigan, G. G., W. P. Ridley, S. G. Riordan, M. A. Nemeth, R. Sorbet, W. A. Trujillo, M. L. Breeze, and R. W. Schneider. 2007.
Chemical composition of glyphosate-tolerant soybean 40-3-2 grown in Europe remains equivalent with that of conventional soybean (*Glycine max* L.). Journal of Agricultural and Food Chemistry 55:6160-6168.

Hoffmann, L., S. Besseau, P. Geoffroy, C. Ritzenthaler, D. Meyer, C. Lapierre, B. Pollet, and M. Legrand. 2004.
Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell 16:1446-1465.

Hothorn, L. A., and R. Oberdoerfer. 2006.
Statistical analysis used in the nutritional assessment of novel food using the proof of safety. Regulatory Toxicology and Pharmacology 44:125-135.

Houshyani, B. 2012.
PhD thesis. Application of omics technologies for environmental risk assessment of genetically modified plants. Wageningen University.

Houshyani, B., P. Kabouw, D. Muth, R. C. H. de Vos, R. J. Bino, and H. J. Bouwmeester. 2012.
Characterization of the natural variation in *Arabidopsis thaliana* metabolome by the analysis of metabolic distance. Metabolomics 8:131-145.

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

Hult, K., and P. Berglund. 2007.
Enzyme promiscuity: mechanism and applications. Trends in Biotechnology 25:231-238.

Humble, M. S., and P. Berglund. 2011.
Biocatalytic promiscuity. European Journal of Organic Chemistry:3391-3401.

Irwin, R. E., L. S. Adler, and A. K. Brody. 2004.
The dual role of floral traits: Pollinator attraction and plant defense. Ecology 85:1503-1511.

Jassbi, A. R., K. Gase, C. Hettenhausen, A. Schmidt, and I. T. Baldwin. 2008.
Silencing geranylgeranyl diphosphate synthase in *Nicotiana attenuata* dramatically impairs resistance to tobacco hornworm. Plant Physiology 146:974-986.

Kabouw, P., N. M. Van Dam, W. H. Van Der Putten, and A. Biere. 2012.
How genetic modification of roots affects rhizosphere processes and plant performance. Journal of Experimental Botany 63:3475-3483.

Kabouw, P., W. H. van der Putten, N. M. van Dam, and A. Biere. 2010.
Effects of intraspecific variation in white cabbage (*Brassica oleracea* var. *capitata*) on soil organisms. Plant and Soil 336:509-518.

Kajikawa, M., T. Shoji, A. Kato, and T. Hashimoto. 2011.
Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant physiology 155:2010-2022.

Kapoor, M., and M. N. Gupta. 2012.
Lipase promiscuity and its biochemical applications. Process Biochemistry 47:555-569.

Keymanesh, K., M. H. Darvishi, and S. Sardari. 2009.
Metabolome comparison of transgenic and non-transgenic rice by statistical analysis of FTIR and NMR Spectra. Rice Science 16:119-123.

Khersonsky, O., C. Roodveldt, and D. S. Tawfik. 2006.
Enzyme promiscuity: evolutionary and mechanistic aspects. Current Opinion in Chemical Biology 10:498-508.

Khersonsky, O., and D. S. Tawfik. 2010.
Enzyme promiscuity: A mechanistic and evolutionary perspective. Annu Rev Biochem 79:471-505.

Kok, E. J., J. Keijer, G. A. Kleter, and H. A. Kuiper. 2008.
Comparative safety assessment of plant-derived foods. Regulatory Toxicology and Pharmacology 50:98-113.

Lessire, R., E. Cahoon, K. Chapman, J. Dyer, P. Eastmond, and E. Heinz. 2009.
Highlights of recent progress in plant lipid research. Plant Physiology and Biochemistry 47:443-447.

Levin, I., R. C. H. de Vos, Y. Tadmor, A. Bovy, M. Lieberman, M. Oren-Shamir, O. Segev, I. Kolotilin, M. Keller, R. Ovadia, A. Meir, and R. J. Bino. 2006.

High pigment tomato mutants - More than just lycopene (a review). Israel Journal of Plant Sciences 54:179-190.

Lucas, D. M., M. L. Taylor, G. F. Hartnell, M. A. Nemeth, K. C. Glenn, and S. W. Davis. 2007.
Broiler performance and carcass characteristics when fed diets containing lysine maize (LY038 or LY038 x MON 810), control, or conventional reference maize. Poultry Science 86:2152-2161.

Lundry, D. R., W. P. Ridley, J. J. Meyer, S. G. Riordan, M. A. Nemeth, W. A. Trujillo, M. L. Breeze, and R. Sorbet. 2008.
Composition of grain, forage, and processed fractions from second-generation glyphosate-tolerant soybean, MON 89788, is equivalent to that of conventional soybean (*Glycine max* L.). Journal of Agricultural and Food Chemistry 56:4611-4622.

Makkar, H. P. S., P. Siddhuraju, and K. Becker. 2007.
Plant secondary metabolites., Methods Mol Biol. vol. 393. Humana Press, Totowa, USA.

McCann, M. C., W. A. Trujillo, S. G. Riordan, R. Sorbet, N. N. Bogdanova, and R. S. Sidhu. 2007.
Comparison of the forage and grain composition from insect-protected and glyphosate-tolerant MON 88017 corn to conventional corn (*Zea mays* L.). Journal of Agricultural and Food Chemistry 55:4034-4042.

Moco, S., and J. Vervoort. 2012.
Chemical identification strategies by liquid chromatography (LC)-photo diode array (PDA)-solid phase extraction (SPE)-nuclear magnetic resonance (NMR)/mass spectrometry (MS). p. 287-316. *In* N. W. Hardy, and R. D. Hall (eds), Plant Metabolomics Methods and Protocols, Methods in Molecular Biology. vol. 860. Humana Press, Otawa, USA.

Mohamed, M. F., and F. Hollfelder. 2013.
Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer. Biochimica et Biophysica Acta - Proteins and Proteomics 1834:417-424.

Morris, W. L., L. J. M. Ducreux, P. Hedden, S. Millam, and M. A. Taylor. 2006.
Overexpression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: Implications for the control of the tuber life cycle. Journal of Experimental Botany 57:3007-3018.

Nam, H., N. E. Lewis, J. A. Lerman, D. H. Lee, R. L. Chang, D. Kim, and B. O. Palsson. 2012.
Network context and selection in the evolution to enzyme specificity. Science 337:1101-1104.

Newell-McGloughlin, M. 2008.
Nutritionally improved agricultural crops. Plant physiology 147:939-953.

Nugroho, L. H., M. C. Verberne, and R. Verpoorte. 2002.
Activities of enzymes involved in the phenylpropanoid pathway in constitutively salicylic acid-producing tobacco plants. Plant Physiology and Biochemistry 40:755-760.

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

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

OECD (Organisation for Economic Cooperation and Development). 2006.
An introduction to the food/feed safety consensus documents of the Task Force. Series on the Safety of Novel Foods and Feeds. no 14, ENV/JM/MONO(2006)10.

Oliver, S. G., M. K. Winson, D. B. Kell, and F. Baganz. 1998.
Systematic functional analysis of the yeast genome. Trends in Biotechnology 16:373-378.

Paine, J. A., C. A. Shipton, S. Chaggar, R. M. Howells, M. J. Kennedy, G. Vernon, S. Y. Wright, E. Hinchliffe, J. L. Adams, A. L. Silverstone, and R. Drake. 2005.
Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nature Biotechnology 23:482-487.

Perry, J. N., C. J. F. T. Braak, P. M. Dixon, J. J. Duan, R. S. Hails, A. Huesken, M. Lavielle, M. Marvier, M. Scardi, K. Schmidt, B. Tothmeresz, F. Schaarschmidt, and H. V. D. Voet. 2009.
Statistical aspects of environmental risk assessment of GM plants for effects on non-target organisms. Environmental Biosafety Research 8:65-78.

Redestig, H., J. Szymanski, M. Y. Hirai, J. Selbig, L. Willmitzer, Z. Nikoloski, and K. Saito. 2011.

Data integration, metabolic networks and systems biology., p. 261–316. *In* R. D. Hall (ed.), Biology of Plant Metabolomics., Annual Plant Reviews. Blackwell Publishing Ltd., Oxford, UK.

Ridley, W. P., R. S. Sidhu, P. D. Pyla, M. A. Nemeth, M. L. Breeze, and J. D. Astwood. 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.

Riffell, J. A. 2011.
The neuroecology of a pollinator's buffet: Olfactory preferences and learning in insect pollinators. Integrative and Comparative Biology 51:781-793.

Rischer, H., and K.-M. Oksman-Caldentey. 2006.
Unintended effects in genetically modified crops: revealed by metabolomics? Trends in Biotechnology 24:102-104.

Rodríguez, A., B. Alquézar, and L. Peña. 2013.
Fruit aromas in mature fleshy fruits as signals of readiness for predation and seed dispersal. New Phytologist 197:36-48.

Rogan, G. J., J. T. Bookout, D. R. Duncan, R. L. Fuchs, P. B. Lavrik, S. L. Love, M. Mueth, T. Olson, E. D. Owens, P. J. Raymond, and J. Zalewski. 2000.
Compositional analysis of tubers from insect and virus resistant potato plants. Journal of Agricultural and Food Chemistry 48:5936-5945.

Römer, S., P. D. Fraser, J. W. Kiano, C. A. Shipton, N. Misawa, W. Schuch, and P. M. Bramley. 2000.
Elevation of the provitamin A content of transgenic tomato plants. Nature Biotechnology 18:666-669.

Saito, K., Dixon, R., Willmitzer, L. 2006.
Plant Metabolomics. Springer-Verlag, Berlin.

Schiestl, F. P. 2010.
The evolution of floral scent and insect chemical communication. Ecol Lett 13:643-656.

Sell, P., and H. L. Kuo-Sell. 1990.
The influence of infestation of oats by root-knot nematodes (Meloidogyne sp.) on the performance of the cereal aphid, Metopolophium dirhodum (Walk.) (Hom., Aphididae). J Appl Entomol 109:37-43.

Shepherd, L. V. T., J. W. McNicol, R. Razzo, M. A. Taylor, and H. V. Davies. 2006.
Assessing the potential for unintended effects in genetically modified potatoes perturbed in metabolic and developmental processes. Targeted analysis of key nutrients and anti-nutrients. Transgenic Research 15:409-425.

Shiojiri, K., K. Kishimoto, R. Ozawa, S. Kugimiya, S. Urashimo, G. Arimura, J. Horiuchi, T. Nishioka, K. Matsui, and J. Takabayashi. 2006.
Changing green leaf volatile biosynthesis in plants: An approach for improving plant resistance against both herbivores and pathogens. Proceedings of the National Academy of Sciences of the United States of America 103:16672-16676.

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

Simkin, A. J., J. Gaffé, J. P. Alcaraz, J. P. Carde, P. M. Bramley, P. D. Fraser, and M. Kuntz. 2007.
Fibrillin influence on plastid ultrastructure and pigment content in tomato fruit. Phytochemistry 68:1545-1556.

Simkin, A. J., S. H. Schwartz, M. Auldridge, M. G. Taylor, and H. J. Klee. 2004.
The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles β-ionone, pseudoionone, and geranylacetone. Plant Journal 40:882-892.

Skogerson, K., G. G. Harrigan, T. L. Reynolds, S. C. Halls, M. Ruebelt, A. Iandolino, A. Pandravada, K. C. Glenn, and O. Fiehn. 2010.
Impact of genetics and environment on the metabolite composition of maize grain. Journal of Agricultural and Food Chemistry 58:3600-3610.

Soga, T. 2007.
Capillary electrophoresis – Mass spectrometry for metabolomics. p. 287-316. *In* W. Weckwerth (ed.), Plant Metabolomics Methods and Protocols, Methods in Molecular Biology. vol. 358. Humana Press, Otawa, USA.

Steppuhn, A., and I. T. Baldwin. 2007.
Resistance management in a native plant: Nicotine prevents herbivores from compensating for plant protease inhibitors. Ecol Lett 10:499-511.

Steppuhn, A., K. Gase, B. Krock, R. Halitschke, and I. T. Baldwin. 2004.

Nicotine's defensive function in nature. PLoS biology 2.

Stitz, M., I. T. Baldwin, and E. Gaquerel. 2011.
Diverting the flux of the JA pathway in *Nicotiana attenuata* compromises the plant's defense metabolism and fitness in nature and glasshouse. PLoS ONE 6.

Sumner, L. W., A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden, and M. R. Viant. 2007.
Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211-221.

Sumner, L. W., E. Urbanczyk-Wochniak, and C. D. Broeckling. 2008.
Metabolomics data analysis, visualization and integration. p. 409–436. *In* D. Edwards (ed.), Methods in Molecular Biology. vol. 406. Humana Press, Totowa, USA.

Takahashi, H., Y. Hotta, M. Hayashi, M. Kawai-Yamada, S. Komatsu, and H. Uchimiya. 2005.
High throughput metabolome and proteome analysis of transgenic rice plants (*Oryza sativa* L.). Plant Biotechnology 22:47-50.

Tikunov, Y., F. W. A. Verstappen, and R. D. Hall. 2007.
Metabolic profiling of natural volatiles: headspace trapping GC-MS. . p. 39-54. *In* W. Weckwerth (ed.), Metabolomics: methods and protocols, Methods in Molecular Biology. vol. 358. Humana Press, Totowa, USA.

Timischl, B., K. Dettmer, H. Kaspar, M. Thieme, and P. J. Oefner. 2008.
Development of a quantitative, validated Capillary electrophoresis-time of flight-mass spectrometry method with integrated high-confidence analyte identification for metabolomics. Electrophoresis 29:2203-2214.

Turlings, T. C. J., I. Hiltpold, and S. Rasmann. 2012.
The importance of root-produced volatiles as foraging cues for entomopathogenic nematodes. Plant and Soil 358:51-60.

van den Brink, L., C. B. Bus, A. C. Franke, J. A. M. Groten, L. A. P. Lotz, R. D. Timmer, and C. C. M. van de Wiel. 2010.
Inventory of observed unexpected environmental effects of genetically modified crops. Applied Plant Research.

Vanholme, R., K. Morreel, J. Ralph, and W. Boerjan. 2008.
Lignin engineering. Current Opinion in Plant Biology 11:278-285.

Vanholme, R., V. Storme, B. Vanholme, L. Sundin, J. H. Christensen, G. Goeminne, C. Halpin, A. Rohde, K. Morreel, and W. Boerjana. 2012.
A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. Plant Cell 24:3506-3529.

Verberne, M. C., R. Verpoorte, J. F. Bol, J. Mercado-Blanco, and H. J. M. Linthorst. 2000.
Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. Nature Biotechnology 18:779-783.

Verhoeven, H. A., C. H. R. de Vos, and R. J. Bino. 2006.
Plant metabolomics strategies based upon Quadrupole Time-Of-Flight Mass Spectrometry (QTOF-MS). p. 34–47. *In* S. K., R. A. Dixon, and L. Willmitzer (eds), Plant Metabolomics. Springer-Verlag., Berlin.

Verhoeven, H. A., H. Jonker, R. C. H. de Vos, and R. D. Hall. 2011.
Chemical identification strategies by liquid chromatography (LC)-photo diode array (PDA)-solid phase extraction (SPE)-nuclear magnetic resonance (NMR)/mass spectrometry (MS). p. 85-100. *In* N. W. Hardy, and R. D. Hall (eds), Plant Metabolomics Methods and Protocols, Methods in Molecular Biology. vol. 860. Humana Press, Otawa, USA.

Viant, M. R., D. W. Bearden, J. G. Bundy, I. W. Burton, T. W. Collette, D. R. Ekman, V. Ezernieks, T. K. Karakach, C. Y. U. Lin, S. Rochfort, J. S. De Ropp, Q. Teng, R. S. Tjeerdema, J. A. Walter, and W. D. Huifeng. 2009.
International NMR-based environmental metabolomics intercomparison exercise. Environmental Science and Technology 43:219-225.

Voelker, T. A., A. C. Worrell, L. Anderson, J. Bleibaum, C. Fan, D. J. Hawkins, S. E. Radke, and H. M. Davies. 1992.
Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257:72-74.

Walden, R. 2002.
T-DNA tagging in a genomics era. Crit Rev Plant Sci 21:143-165.

Ward, J. L., J. M. Baker, S. J. Miller, C. Deborde, M. Maucourt, B. Biais, D. Rolin, A. Moing, S. Moco, J. Vervoort, A. Lommen, H. Schäfer, E. Humpfer, and M. H. Beale. 2010.
An inter-laboratory comparison demonstrates that [1H]-NMR metabolite fingerprinting is a robust technique for collaborative plant metabolomic data collection. Metabolomics:1-11.

Weckwerth, W. 2007.
Metabolomics: methods and protocols., Methods Mol Biol. vol. 358. Humana Press, Otowa, USA.

Weckwerth, W., and G. Kahl. 2013.
The Handbook of Plant Metabolomics. Molecular Plant Biology. Wiley-Blackwell.

Wilson, A. K., J. R. Latham, and R. A. Steinbrecher. 2006.
Transformation-induced mutations in transgenic plants: Analysis and biosafety implications. Biotechnology and Genetic Engineering Reviews 23:209-237.

Wünsche, H., I. T. Baldwin, and J. Wu. 2011.
Silencing NOA1 elevates herbivory-induced jasmonic acid accumulation and compromises most of the carbon-based defense metabolites in *Nicotiana attenuata*. Journal of Integrative Plant Biology 53:619-631.

Ye, X., and P. Beyer. 2000.
Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303-305.

Yoshikuni, Y., T. E. Ferrin, and J. D. Keasling. 2006.
Designed divergent evolution of enzyme function. Nature 440:1078-1082.

Zhao, Y., S. K. Christensen, C. Fankhauser, J. R. Cashman, J. D. Cohen, D. Weigel, and J. Chory. 2001.
A role for flavin monooxygenase-like enzymes in auxin biosynthesis. Science 291:306-309.

Zhou, J., K. H. Berman, M. L. Breeze, M. A. Nemeth, W. S. Oliveira, D. P. V. Braga, G. U. Berger, and G. G. Harrigan. 2011a.
Compositional variability in conventional and glyphosate-tolerant soybean (*Glycine max* L.) varieties grown in different regions in Brazil. Journal of Agricultural and Food Chemistry 59:11652-11656.

Zhou, J., G. G. Harrigan, K. H. Berman, E. G. Webb, T. H. Klusmeyer, and M. A. Nemeth. 2011b.
Stability in the composition equivalence of grain from insect-protected maize and seed from glyphosate-tolerant soybean to conventional counterparts over multiple seasons, locations, and breeding germplasms. Journal of Agricultural and Food Chemistry 59:8822-8828.

Zhu, C., S. Naqvi, T. Capell, and P. Christou. 2009.
Metabolic engineering of ketocarotenoid biosynthesis in higher plants. Archives of Biochemistry and Biophysics 483:182-190.

Appendix

Table X1. A non-exhaustive list of the most common technological abbreviations likely to be encountered in the metabolomics literature

APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photo-ionization
CAS	Chemical abstracts service
CE	Capillary electrophoresis
CID	Collision-induced dissociation
DI-MS	Direct infusion mass spectrometry
ESI	Electrospray ionization
FIA/DIA	Flow injection analysis/direct infusion analysis (also FI-MS or DI-MS)
FT-ICR-MS	Fourier transform – ion cyclotron resonance – MS (or FTMS)
HCD	High-energy collision-induced dissociation
HILIC	Hydrophilic interaction chromatography
HPLC	High performance (pressure) liquid chromatography
HTP	High throughput
ICP-MS	Inductively coupled plasma MS
LC/GC	Liquid/gas chromatography
LTQ	Linear trap quadrupole
MALDI-MS	Matrix assisted laser desorption ionization – MS
MS	Mass spectrometry
MS/MS; MS*n*	Double (MS/MS) or multiple levels (MS*n*) of molecular fragmentation/ re-fragmentation with MS detection (also Tandem MS)
m/z	Mass/charge ratio
netCDF	Network common data form
NIST	National Institute of Standards and Technology (metabolite database) (http://www.nist.gov/srd/analy.htm)
NMR	Nuclear magnetic resonance
PDA (DAD)	Photodiode array detection (diode array detection)
PI	Photo-ionization
SEC	Size exclusion chromatography
SPE/SPME	Solid phase extraction or solid phase micro-extraction
ToF	Time of flight (also ToF)
UPLC	Ultra performance liquid chromatography

Table X2. The three main metabolomic approaches most commonly used with their contrasting advantages and disadvantages

Technology	Advantages	Disadvantages
NMR	• Non-destructive • Quantitative • Robust, long-established technology • Good software and metabolite database support • Short measurement time • Almost no sample preparation needed • No separation required • No derivatization required • Detects all organic compound classes • Structural identification of unknown compounds relatively straightforward • Compatible with liquids and solids	• Relatively low sensitivity • Expensive instrumentation • Limited to protonated compounds • Cannot detect salts and inorganic ions • Relatively large (0.5 mL) samples required
GC-MS	• Relatively inexpensive • Quantitative (when standards available) • Modest sample size required • Good sensitivity • Good software and metabolite database support • Detects most organic and some inorganic molecules • Excellent separation reproducibility	• Destructive • Non-volatiles need derivatization • Unsuitable for heat-labile components • Extensive sample preparation • procedures • Separation step means longer measurement times (20–50 min) • Identification of unknown compounds is difficult
LC-MS	• Flexible technology • Good sensitivity • Detects most organic and some inorganic molecules • Small sample size (mg) requirement • Direct injection can be used for very rapid analysis (1–2 min) • Has potential for detecting largest portion of metabolome	• Destructive • Quantification limited • Relatively expensive instrumentation • Extensive sample preparation procedures • Separation step means longer measurement times (20–50 min) • Resolution and reproducibility poorer than GC • Instrumentation less robust than NMR or GC-MS • Software and metabolite database support poor • Identification of unknowns is difficult

COMMISSIE
COGEM
GENETISCHE
MODIFICATIE

POSTBUS 578
3720 AN BILTHOVEN
TEL.: 030 274 2777
FAX: 030 274 4476
INFO@COGEM.NET
WWW.COGEM.NET