TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Allgemeine Lebensmitteltechnologie

Metabolite Profiling of Maize Grain: Differentiation due to Genetics, Environment and Input System

Richard M. Röhlig

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. W. Schwab
Prüfer der Dissertation:	1. Univ.-Prof. Dr. K.-H. Engel
	2. Univ.-Prof. Dr. A. Gierl

Die Dissertation wurde am 14.10.2010 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 18.01.2011 angenommen.

DANKSAGUNG

Die vorliegende Arbeit wurde unter Betreuung von Prof. Dr. Karl-Heinz Engel, Ordinarius des Lehrstuhls für Allgemeine Lebensmitteltechnologie, Technische Universität München, angefertigt. Ich möchte Herrn Prof. Engel für die Gelegenheit danken auf dem spannenden Gebiet der Mais Metabolomics zu arbeiten. Außerdem möchte ich mich für seine wissenschaftliche Unterstützung, viele fruchtbare Diskussionen, sowie die Möglichkeit bedanken, das tun zu dürfen, was ich gut kann und gerne mache, und dabei noch besser zu werden.

Bei Herrn Prof. Dr. Alfons Gierl bedanke ich mich für die Übernahme des Koreferats. Herrn Prof. Dr. Wilfried Schwab sei für die Bereitschaft gedankt, den Prüfungsvorsitz zu übernehmen.

Dem EU Projekt SAFEFOODS und allen Beteiligten bin ich für den interessanten wissenschaftlichen Austausch und die Erweiterung meines europäischen Verständnisses verbunden.

Besonderer Dank gilt Herrn Dr. Eder und Frau Dr. Schwarzfischer von der Bayerischen Landesanstalt für Landwirtschaft (LfL), sowie Herrn Gerl vom Versuchsgut Scheyern, Helmholtz Zentrum München, für die Bereitstellung wertvollen Pflanzenmaterials. Auch die Unterstützung in der NIR Spektroskopie von Herrn Dieter Nast, sowie die Einführung in die Agrar-Statistik von Herrn Rudof Graf waren mir eine große Hilfe.

Ich bin dankbar für die angenehme Zusammenarbeit am Lehrstuhl mit meinen Kollegen Andreas Barnsteiner, Dr. Ron Baudler, Catherine Delaporte, Anja Demmel, Dr. Marti Dregus, Oxana Fastovskaya, Dr. Alexandra Hahn, Katrin Hempfling, Tim Lubinus, Dr. Andreas Miller, Dr. Francisco Moreano, Tobias Müller, Kriskamol “Tar” Na Jom, Anne-Marie Orth, Iulia Poplacean, Steffi Speiser, Dr. Hedwig Strohalm, Dr. Ludwig Adam, Dr. Bernhard Meier, Dr. Walter Weiss und Dr. Ludwig Ziegler. Extra Dank geht an Martina Denk für die zuverlässige Aufarbeitung von 6930 Proben und Dr. Thomas Frank für seine geduldigen Ohren und den vielfältigen Austausch, sowie Frau Röske für das tägliche Lächeln. Veronika Ziegmann und Elisabeth Hertle danke ich für ihre tatkräftike Unterstützung im Rahmen von Bachelorarbeiten, Peter Rauscher und der Werkstatt für die Hilfsbereitschaft.

Dank geht auch an meinen früh verstorbenen Freund und Kollegen Bertrand Seumo Meuleye, an dessen Offenheit und Lebensfreude ich oft denke.

Zuletzt gilt großer Dank meinen Eltern, meiner Schwester und meinen Schwiegereltern für den Rückhalt und das Vertrauen. Meiner lieben wundervollen Frau danke ich ganz besonders.

Table of Contents

1 INTRODUCTION AND OBJECTIVES

In recent years, a trend towards the development of pathway-based breeding approaches can be observed, which aim at the understanding of metabolic interaction, and its application to breeding driven metabolic engineering (Schauer and Fernie, 2006). In this context, the investigation of metabolites is of particular interest, as they are the end products of cellular processes, that can be considered as the ultimate response of an organism to genetic and environmental influence factors (Fiehn, 2002). To challenge the complex interplay of different impact factors, the assessment of a broad spectrum of plant constituents is essential. Therefore, metabolite profiling techniques have been developed for the parallel assessment of a comprehensive set of metabolites (Allwood *et al.*, 2008; Kopka *et al.*, 2004), complementing other untargeted approaches of molecular profiling, such as transcriptomics (Saito *et al.*, 2008) and proteomics (Alvarez *et al.*, 2008). These so-called *omics*-techniques aim at extracting, detecting, identifying and quantifying a broad spectrum of plant molecules to provide a deeper insight into complex biological systems (Fiehn, 2001).

In the field of plant metabolomics gas-chromatography coupled to mass-spectrometry (GC/MS) has been perceived as the best understood and exemplary profiling approach for the simultaneous and non-biased detection of metabolites from biological samples (Steinhauser and Kopka, 2007) and has been proven to be a powerful tool for the analysis of complex plant matrices (Castro and Manetti, 2007; Fiehn *et al.*, 2000; Lozovaya *et al.*, 2006; Roessner *et al.*, 2000). For example, GC/MS-based profiling methods have been applied to the investigation of phenotypic diversity in rice (Kusano *et al.*, 2007), for the assessment of compositional variability due to environmental impact (Semel *et al.*, 2007), different input systems, e.g. conventional vs. organic (Davies, 2007; Zörb *et al.*, 2006), or due to genetic background (Harrigan *et al.*, 2007a), as well as to monitoring and investigation of plant developmental systems (Fait *et al.*, 2008; Mounet *et al.*, 2007; Shu *et al.*, 2008; Tarpley *et al.*, 2005).

Objectives. Maize (*Zea mays*) is the most important cereal food and feed crop worldwide (FAO, 2008). However, very few reports are available on the natural variation among maize varieties based on metabolite profiling techniques (Hazebroek et al., 2007; Zörb et al., 2006).

One objective of this work was the establishment and application of a GC/MS-based metabolite profiling approach to the identification and the assessment of a broad spectrum of maize constituents from different chemical classes. An increased understanding of metabolic variation should be achieved by investigation of a representative sample set considering potential impact factors on maize kernel composition, such as cultivar, farming location, growing season, and fertilization practice. The evaluation of results by sound statistics should help to demonstrate the potential of untargeted metabolite profiling to evaluate the variations in maize grain metabolite pools resulting from the interplay of environment, season and genotype, and to put the results into the context of natural variability. This information will be of relevance to our basic understanding of the regulation of crop composition, and will be of assistance to breeders, farmers and downstream industries where consistency in crop composition is important for product quality.

The second objective was the application of near infrared spectroscopy (NIRS) profiling for the qualitative and quantitative screening of maize samples. Substantiation of potential differences should be evaluated based on NIRS calibration models developed by use of GC/MS-metabolite profiling data. The suitability of the applied NIRS approach as tool for the pre-assessment of large sample-sets should be tested, thus demonstrating that NIRS could complement existing GC/MS-metabolite profiling methods to the investigation of complex crop sample sets with potential for use in metabolomic studies and breeding programs.

2 BACKGROUND

2.1 Maize (*Zea mays*)

2.1.1 Maize as an Agricultural Product

Field maize has been cultivated for 8000 years in Mexico and Central America and for 500 years in Europe. It is the most important cereal crop worldwide followed by wheat and rice (FAO, 2008). The world's production of maize is more than 800.000.000 tons, with the US and China accounting for 37% and 20%, respectively, as the major producers. In the EU 11% of the total amount of field maize is grown (FAO, 2008). Maize is used as a food and feed crop, as well as a source of energy. For human nutrition maize and cornmeal is a staple food in many regions of the world. The starchy grains are also the source of fermented products such as bourbon whiskey or maize beer. However, the major part of maize is produced for feeding of livestock (Pingali, 2001). In addition, maize is increasingly used as the basis for biomass fuels, such as ethanol, and as feedstock for biomass gas-plants.

Maize is naturally cross-pollinated for reproduction, and until the 1920's farmers were growing only open-pollinated varieties. Today, mainly hybrid maize is cultivated, owing to its more vigorous growth and higher yields (heterosis-effect). Hybrid seeds are produced by crossing two homozygous parental lines in removing the tassels from the plants before pollen shedding, so that only one sort of pollen will be received by the silks (OECD, 2002). Hybridization and backcrossing are the integral processes of breeding programs, where the genetic variation within species and between related species is used as a major source for crop improvement. Traditionally, researchers focused on a few number of single target traits with importance for industrial or nutritive value. One of the longest continuous experiments in this field is the Illinois long-term selection experiment for protein and oil content, starting in 1896. With more than 100 cycles of selection, nine related maize populations have evolved, that contain the known phenotypic extremes for maize kernel composition (Moose *et al.*, 2004). These single trait approaches were greatly enhanced, when researchers began in the 1980s to develop molecular-marker techniques, that both revolutionized plant breeding, and assisted basic research by

facilitating the introgression of defined genes or genomic regions from wild species or landraces (Fernie and Schauer, 2008).

2.1.2 Maize Composition

Maize kernels consist of endosperm (containing starch) and germ (containing oil), wrapped in the pericarp, a cellulose layer. The major constituents of mature kernels are starch (65–80%), protein (5–15%), moisture (5–20%), fat (3–9%) and dietary fiber (5–15%). Minor constituents are minerals and vitamins (1–2%) (Kirchhoff, 2008; OECD, 2002). Due to the balanced amino acid composition of its protein, maize has a high nutritional value as a food and feed crop, except for the two limiting essential amino acids lysine and tryptophan. Several maize variants have been developed with specific improvements in composition, such as Quality Protein Maize (QPM) variants, which exhibit improved levels of lysine and tryptophan. Other specialty types of maize are characterized by higher oil content, higher amylose content or higher amylopectin content (waxy maize) (Jugenheimer, 1976).

2.2 Metabolomics

2.2.1 Targeted Analysis vs. Unbiased Profiling

Targeted analysis of single compounds has been the analytical standard for a long period owing to a number of advantages. Highly specialized analytical protocols allow the accurate detection and quantification of constituents from different matrices with an excellent methodological reproducibility (Matissek and Steiner, 2005). For most of the analytes of interest in the food and feed industry accepted and approved standard methods allow the equal measurement at different laboratories with the same analytical quality. However, these methods reach their limits, if a high number of analytes should be detected in parallel. In addition, a targeted approach will be able to cover only analytes that were known *a priori*. These limitations have led to the development of unbiased analytical approaches, the so-called fingerprinting and profiling methods.

The unbiased assessment of a broad spectrum of analytes with a single method enlarges the capabilities of analytical techniques. One example is the demonstration

of substiantial equivalence in the course of the authorization procedure of novel foods. The concept of substantial equivalence is a key element in the safety assessment of novel foods and has been introduced by the OECD in the beginning of the 1990s (OECD, 1993). The composition of a novel food is compared to a food that is characterized by a history of safe use. If the novel food is found to be substantially equivalent, it is considered as safe as the conventional counterpart. The OECD has suggested a number of constituents to be investigated, which need several different techniques for analysis. If these compounds, or at least a great part of them, could be analyzed by a single approach, this would provide the possiblity to scan even more samples to allow a truly comprehensive comparison of the novel food with its conventional counterpart. Another aspect of the use of unbiased profiling methods is the possibility to simultaneously investigate further compounds, that so far have not been in the focus of the OECD, but may also impact the decision about substantial equivalence. This could assist in the detection of potential unintended effects that are one of the major concerns associated with crop modification based on genetic engineering. The random integration of the transgene may cause gene disruptions that can lead to sequence changes, production of new proteins or formation of either new metabolites or altered levels of exisiting metabolites that could compromise safety (Kuiper *et al.*, 2001). Other unintended effects, related to the genetic modification, may be secondary effects of the introduced sequences. However, unintended effects are not limited to breeding techniques comprising genetic engineering, but may also occur during conventional breeding as a result of mutagenesis, as well as hybridization and backcrossing that are integral processes of breeding programs where the genetic variation within species and between related species is used as a major source for crop improvement. Unintended effects can be subdivided into predictable and unpredictable unintended effects. A predictable unintended effect goes beyond the the expected effect of the newly introduced trait, but may be explained through the current knowledge of plant biology and metabolic pathways. Unpredictable unintended effects cannot be explained by the present knowledge, and therefore may not be present in the predefined set of target analytes, whereas an unbiased molecular profiling may close this gap.

2.2.2 *Omics*-Techniques

Fingerprinting is an uncomplex approach without a major pre-treatment of the raw material that is sufficient for the detection of major effects (Fiehn, 2002). It is mainly applied if the aim is to look for compositional similarities or the overall natural variability should be explored in a large sample set, and if it might not be necessary to determine the individual levels of all constituents. Exemplary techniques for fingerprinting approaches are nuclear magnetic resonance (NMR), direct infusion mass spectrometry (DI-MS) or near infrared spectroscopy (NIRS) (Davies, 2007; Manetti *et al.*, 2006; Osborne, 2008).

Based on the results of a rapid screening by fingerprinting techniques, pre-selected samples could be subjected to further thorough analysis by a comprehensive molecular profiling. Metabolite profiling techniques aim at the parallel assessment of a comprehensive set of metabolites (Allwood *et al.*, 2008; Kopka *et al.*, 2004), complementing other untargeted approaches of molecular profiling, such as transcriptomics (Saito *et al.*, 2008) and proteomics (Alvarez *et al.*, 2008).

For example the impact of different amounts and forms (organic, inorganic) of nitrogen supply on the gene expression level in the wheat endosperm have been investigated (Lu *et al.*, 2005). Many of the genes showing differential expression in this study are known to participate in nitrogen metabolism and storage protein synthesis. Another study employing a proteomics approach compared the protein compositions of potato tubers subjected to organic and mineral-based fertility management practices, respectively. The results suggested an increased stress response in organic farming (Lehesranta *et al.*, 2007). In wheat 16 “diagnostic” proteins with potential to afford a signature to prove authenticity of organic wheat were proposed (Zörb *et al.*, 2009). In addition to transcriptomics and proteomics, metabolomics-based approaches should also be suitable to reflect the impact of different input systems on crops, as metabolites can be considered as the ultimate response of organisms to processes regulating metabolism (Fiehn, 2002). The combination of profiling approaches gives a comprehensive picture of the whole organism. Systems biology aspires for the integration of information from genes, transcripts, proteins and metabolites to illuminate the connections and intercorrelations between different biochemical pathways. This combined approach

may also help in establishing a molecular signature to prove authenticity of natural products, which may for example be of importance for crops grown under distinct agricultural practices, such as organic and conventional farming.

2.2.3 GC/MS-Metabolite Profiling

Metablomics can be described as the approach to assess the totality of all metabolites—the metabolome. In the field of plant metabolomics gas-chromatography coupled to mass-spectrometry (GC/MS) has been perceived as the best understood and exemplary profiling approach for the simultaneous and non-biased detection of metabolites from biological samples (Steinhauser and Kopka, 2007). GC/MS-based profiling methods can rely on well established GC instrumentary, which is characterized by high resolution and reliability. The use of different capillary materials allows the detection of compounds from various chemical classes. Also non-volatile compounds can be captured after derivatization. For identification of the metabolites different types of mass spectrometers can be coupled to the GC instruments, e.g. Quadrupole or time-of-flight (TOF) mass spectrometers are most commonly used for profiling purposes. Another great advantage of GC/MS is the availability of still growing mass spectral libraries for assignment of metabolites (Ausloos *et al.*, 1999; Kopka *et al.*, 2005).

GC/MS based profiling methods have been applied to the investigation of phenotypic diversity in plants (Kusano *et al.*, 2007) and to the assessment of compositional variability due to environmental impact (Semel *et al.*, 2007), different input systems, e.g. conventional vs. organic, (Davies, 2007; Zörb *et al.*, 2006) or due to genetic background (Harrigan *et al.*, 2007a). Metabolomics-based approaches have also been suggested for the use in safety assessment, e.g. for the evaluation of substantial equivalence and the detection of potential unintended effects (Cellini *et al.*, 2004; Kuiper *et al.*, 2003), and demonstrated its application for breeding-driven metabolic engineering (Schauer and Fernie, 2006), as well as in monitoring and investigation of plant developmental systems (Fait *et al.*, 2008; Mounet *et al.*, 2007; Shu *et al.*, 2008; Tarpley *et al.*, 2005). Recently, statistical assessment of the metabolite profiling data from mung beans (*Vigna radiata*) via principal component analysis demonstrated that the metabolic changes during the sprouting of mung beans are reflected by

time-dependent shifts of the scores which were comparable for two sprouting processes independently conducted under the same conditions (Na Jom *et al.*, 2010). In a similar study concerning malting barley, the capability of GC/MS profiling was proven to reflect the dynamic changes of the metabolites in the course of the different malting stages for compounds ranging from lipophilic to hydrophilic (Frank *et al.*, 2011).

2.3 Analysis of Profiling Data—Chemometrics

2.3.1 *Omics*-Data

What is special about data obtained by a profiling approach in comparison to results from targeted analysis? The peculiarity of profiling data originates from its unbiased character. Targeted methods detect values for each analyte individually, whereas unbiased approaches intend to measure *everything* in a single run. Of course, this is an objective out of reach; however, profiling techniques still aim at capturing at least *everything they can*. This approach has a number of implications on the type of data obtained.

The most obvious consequence is the amount of data. A typical GC/MS-based metabolite profiling technique has coverage of a few hundred metabolites. More sophisticated methods, for example multidimensional methods, such as GC x GC-TOFMS, are capable to detect up to 1400 peaks in one chromatogram (Pierce *et al.*, 2006). Microarray chips for maize transcriptomics with a few 10.000 genes are available (Ma *et al.*, 2006), and 2D gel electrophoresis-based proteomics covers around 500–2000 different proteins per gel (Govorun and Archakov, 2002). Owing to the high number of simultaneously detected compounds, metabolic crosslinks in an organism will be reflected. This intercorrelation dependency is also called *collinearity*, and represents another characteristic of *Omics*-data with importance for data analysis.

2.3.2 Identifying the Major Sources of Variation (Multivariate)

Traditional data analysis will compare each analyte value by value. For assessment of profiling data this is not a satisfactory approach. Where to look if there are a few hundred or thousand values? In a first step multivariate statistics should be applied for the investigation of the major sources of variation in a data set. This can be achieved by Principal Component Analysis (PCA). PCA enables the rapid differentiation of samples based on their whole profile by visualizing the data as dots in a two-dimensional plot. For example, the metabolite profiles of two relatively similar samples will result in a close grouping, whereas samples varying in their composition will be differentiated (Figure 1). Owing to the use of a linear

combination algorithm for calculation of the plot axes—the principal components—redundancy in the dataset in terms of collinearity is eliminated, because constituents exhibiting a similar variation will be recognized in the same principal components.

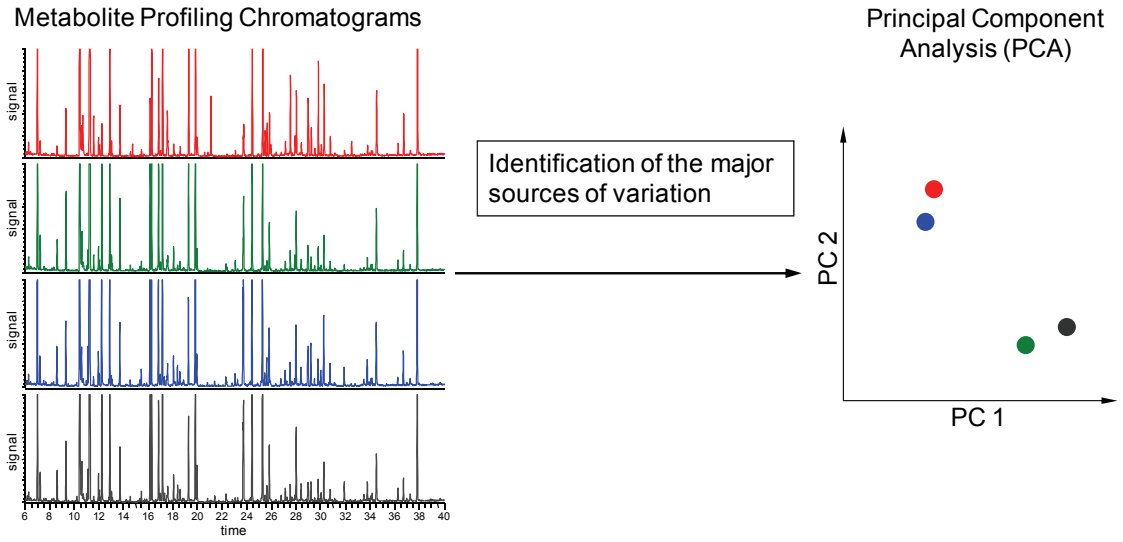

Figure 1: Differentiation of visually similar metabolite profiling chromatograms obtained by GC/MS via Principal Component Analysis (PCA)

The principal components are determined by linear combination of the individual peak signals *s*. The simplest way would be to summarize all values (Equation 1).

$$Sum = s_1 + s_2 + \ldots + s_n \qquad (1)$$

As a result, one principal component would represent the total content of each sample as determined by the chosen analytical technique, and potential variation within each profile would be lost in the total content parameter. As the aim of a PCA is to determine the major sources of variation, different *sums*, i.e. principal components (PC), are calculated by multiplying the signals *s* with an individual factor *f*, which is determined according to the variance of each analyte (Equation 2).

$$\begin{aligned} PC1 &= s_1 * f_{1,PC1} + s_2 * f_{2,PC1} + \ldots + s_{n,PC1} \\ PC2 &= s_1 * f_{1,PC2} + s_2 * f_{2,PC2} + \ldots + s_{n,PC2} \qquad (2) \\ &\vdots \\ PC(n-1) &= s_1 * f_{1,PC(n-1)} + s_2 * f_{2,PC(n-1)} + \ldots + s_{n,PC(n-1)} \end{aligned}$$

By definition, the first principal component (PC1) will include the greatest differences in the data set. For calculation of PC1 the analytes with the strongest variation will get the highest loading factors *f*. The next principal component explains as much as

possible of the residual variation not covered in the first PC. Therefore, in case of PC2, the highest loadings will be given to the compounds with the greatest variances, which have not been highly weighted in PC1. The total number of principal components is defined by the number of analyzed samples; for *n* samples, *n* – 1 principal components can be calculated. However, usually the first 5 to 10 components will already explain 80–90% of the total variation in a data set.

2.3.3 Substantiation and Quantification of Differences (Univariate)

The investigation of the major sources of variation is followed by a substantiation of differences. A quick way for the identification of the drivers of variation is to examine the factor loadings as determined by PCA. For example, if a clear separation of samples is revealed on the first principal component, the main contributors to this effect will exhibit high absolute loading factors in PC1 (Figure 2). In a next step, the signals of these compounds can be statistically evaluated by univariate comparisons. For investigation of two groups of samples, e.g. cultivar A and cultivar B, Student's *t*-test seems to be adequate. If a sample set contains more than two groups (e.g. 3 cultivars), or more than one factor (e.g. factor cultivar and factor year) an Analysis of Variance (ANOVA) is more appropriate. In the Student's *t*-test mean values are compared, whereas in an ANOVA the variance within a group is compared to the variance between groups. Because the information about different factors is incorporated into the calculation, the results will be more accurate.

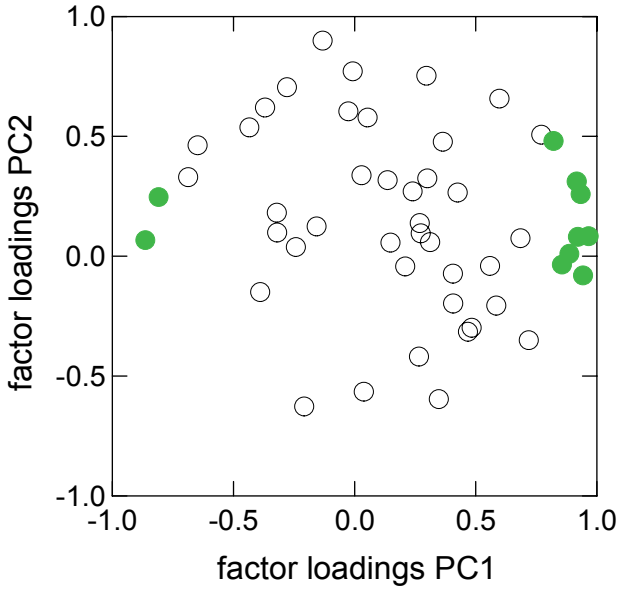

Figure 2: Factor loading scores plot of a PCA of metabolite profiling data. The metabolites with the 10 highest loadings on PC1 are highlighted (green)

As a small example, maize kernels from three cultivars should have been collected from field trials in two different years. In the case of only a small influence of the growing season, samples will differ mainly due to cultivar. If data from both seasons

are analyzed by Student's *t*-test, the standard deviation will become smaller due to a higher number of degrees of freedom, because cultivar data from both years were averaged. Therefore, the comparison of cultivars will result in more differences, than compared to a single-year sample set. On the other hand, if the influence of the year would be very strong, the standard deviation for each cultivar will become very large, due to the big differences of samples from different years. The result will be no observable difference as determined by Student's *t*-test. In contrast, ANOVA *knows* about the second factor year and will create sub-sets for each year before comparing the variance within cultivars with the variance between cultivars. Therefore, in the first case, ANOVA will not overinterprete differences between cultivars; in the second case ANOVA will still find differences between cultivars, if there are any.

2.3.4 Visualizing *Omics*-Data

2.3.4.1 Box Plots

There are numerous possibilities for the visual comparison of the results obtained by an analytical technique. In the following, two methods with use for interpretation of complex *omics*-data will be explained, that were also applied in this work.

The comparison of mean values, as obtained by averaging the results of multiple analyses for each individual sample, may be sufficiently performed by plotting the means as a column diagram. If data are obtained from large studies, mean values will usually include more than analytical replicates of a single sample, but rather include samples combining a variety of impact factors, such as cultivar, location or growing season. In this case more information about the density and the distribution of the data around the mean value would be desirable. Box plots, or whisker plots, show robust statistics (median, quartiles, etc.) rather than comparing sample values to the normal distribution (mean, standard deviation, etc.). In a box plot, the center horizontal line marks the median of the sample (Figure 3).

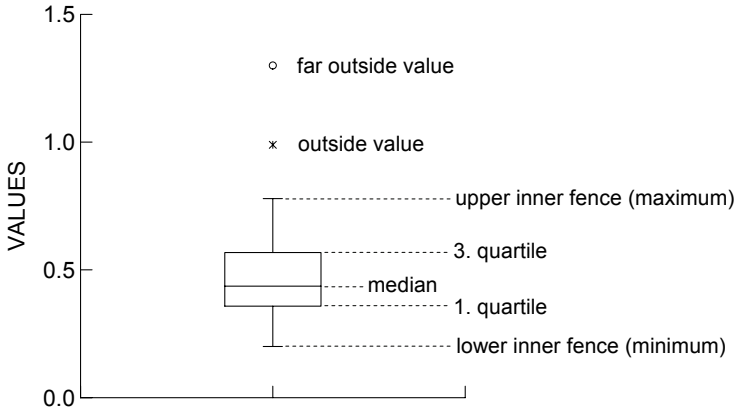

Figure 3: Box plot. Values between the inner and outer fences are plotted by asterisks. Far outside values are plotted by empty circles.

The difference between the top and the bottom of the box can be called the interquartile or the midrange with the box edges at the first and third quartiles, representing the central 50% of all values. Fences define outside and far outside values that are defined as follows:

lower inner fence = bottom of box − (1.5 * midrange)

upper inner fence = top of box + (1.5 * midrange)

lower outer fence = bottom of box − (3 * midrange)

upper outer fence = top of box + (3 * midrange)

The whiskers show the range of observed values that fall within the inner fences. In other words, they show the range of values that fall within 1.5 * range from bottom to top of the box. Because the whiskers extend to observed values and the fences need not correspond to observed values, the whiskers do not necessarily extend all the way to the inner fences. Values beyond the outer fences are called far outside values.

2.3.4.2 Histograms and Logit (p)

One of the most familiar displays for a distribution of values of a high number of samples is the histogram. A histogram is a pictorial display of vertically standing bars. It is a crude density estimator because the shape of a histogram depends upon the choice of the number of bars. In addition to the description of sample properties, histograms can also be used for visualization of statistical results. One application is

the investigation of *p*-values obtained by ANOVA of profiling data. In addition to the absolute *p*-values themselves, also their distribution among the constituents is of interest. However, the *p*-values of an ANOVA based, for example, on a broad range of metabolites are not Gaussian variables. Usually, the major part of *p*-values will exhibit only small levels, and the decision on which treatment has the main impact (e.g. factor 1 vs. factor 2) will be only weak (Figure 4A). Therefore, the data should be transformed for enabling a better evaluation of the data.

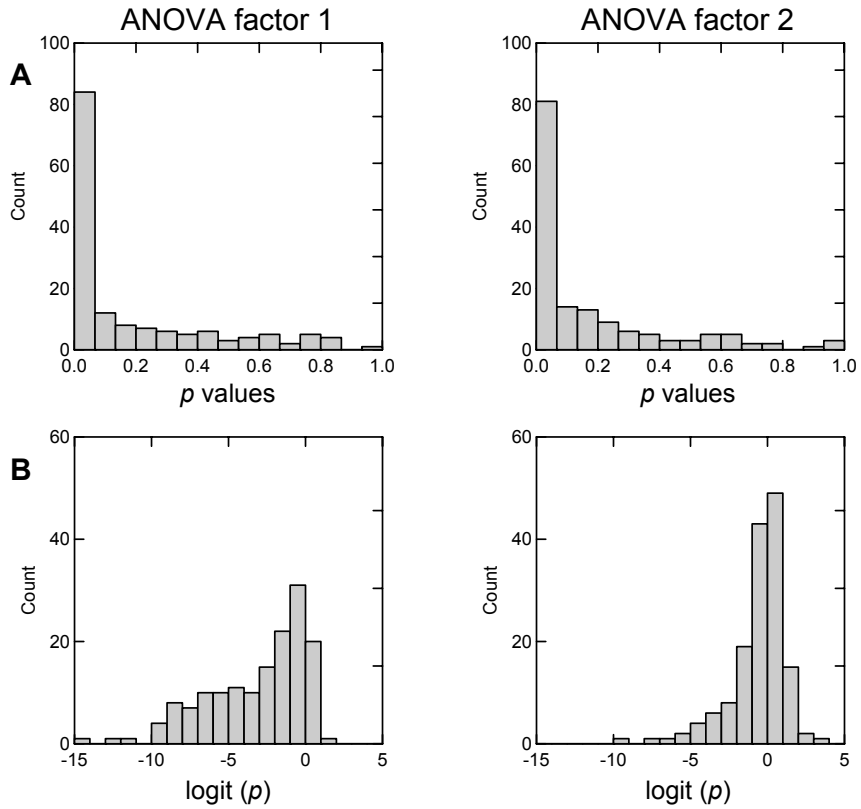

Figure 4: Histograms of *p*-values obtained by a two-way ANOVA (factors 1 and 2) of metabolite data for 152 compounds. (A) comparison of the original *p*-values distribution and (B) distribution of *p*-values transformed by use of logit (*p*) = log *(p*) − log (1−*p*)

The Logit-function will modify the distribution of *p*-values from $[0 \leq p \leq 1]$ to $[-\infty < \text{logit}\ (p) < +\infty]$, e.g. logit ($p$=0.5) = 0, logit ($p$=0.1) ≈ −1, logit ($p$=0.01) ≈ −2, logit ($p$=0.001) ≈ −3 (Ashton, 1972; Greiff et al., 2002; Ter Braak and Gremmen, 1987). Histograms showing logit (*p*) can be interpreted as follows: the more equal the distribution of logit (*p*)-values, the more significant is a treatment; the more logit (*p*)-values are around zero, the less significant is a treatment. In Figure 4B this will allow the conclusion that factor 1 has a stronger effect on the variation between samples than factor 2.

2.4 Near Infrared Spectroscopy (NIRS)

> *NIR spectroscopy is the art and science of interactions of NIR energy with matter. A spectroscopist is a person who studies spectroscopy. NIR spectrometry, on the other hand is the art and science of building NIR instrumentation that a spectroscopist would use to pursue spectroscopy. NIR spectroscopy, requiring human interaction and interpretation, results in a report or paper. NIR spectrometry, also requiring human interaction and interpretation, culminates in an instrument—a spectrometer.*
>
> *(McClure and Tsuchikawa, 2007)*

2.4.1 Short History

The existence of near infrared light (800–2500 nm) was discovered not until 1800, when William Herschel put a thermometer beyond the red end of a light spectrum generated by use of a prism (Herschel, 1800a). He noticed an increase in temperature, which he explained by the existence of a so far unknown type of invisible energy in this spectral range (Herschel, 1800b).

NIRS as a potential analytical tool was firstly described in the middle of the 20th century by Gordy and others (Gordy and Martin, 1937; Plyler and Williams, 1936; Williams, 1936), who *replaced* the thermometer by mica absorption layers (Williams and Rogers, 1937). Further improvement of spectroscopic techniques was achieved from the 1960s on by Goddu and Delker (1958), Kaye *et al.* (1951), Wheeler (1959), Whetsel (1968), and especially Norris (1963), who was the first to demonstrate that NIR spectrometry could be calibrated with multiple linear regression. It was already at this time, that the potential of NIRS was recognized by the food industry. The major efforts made in the years until now were the introduction of new detectors and more compact in-line measuring NIRS units. The use of computers for calculation of regression models and thus the ability to employ more sophisticated algorithms, such as neural networks, introduced additional verve into the field of near infrared spectroscopy (Wang and Paliwal, 2007).

2.4.2 NIRS as Analytical Tool

The absorption of near infrared light corresponds to overtones and combinations involving C-H, O-H or N-H chemical bonds, resulting in a characteristic spectral profile for each sample based on its chemical composition. Near infrared spectroscopy, as an instrumental method for acquiring spectra of foods and other materials, is widely used for the determination of both qualitative and quantitative characteristics. The popularity of NIRS methods is owed to five main advantages, which are (1) speed, (2) little or no sample preparation, (3) multiple analyses from a single scan, and (4) a non-destructive measurement process which is (5) highly reproducible. These come along with three shortcomings: (1) NIRS is a technology that must be trained/calibrated. (2) Determination of constituents at known levels often involves expensive and complicated reference methods closely tied to wet chemistry that demand the input of highly skilled personnel. (3) Modern-day calibration methods rely on rather sophisticated chemometric techniques, thus calling into play the assistance of personnel who are highly trained in chemometrics or statistics in chemistry (McClure and Tsuchikawa, 2007).

2.4.3 Investigation of Food and Feed

Together with multivariate analysis, such as principal component or linear discriminant analysis, NIRS showed great promise as a screening tool for monitoring biochemical changes in crop developing systems, such as malting barley (Allison and Maule, 1991), or for both discriminating between yeast strains and grouping strains with deletions in genes that disturb similar metabolic pathways (Cozzolino *et al.*, 2006). NIRS is commonly used for targeted proximate analyses (Osborne, 2008; Woodcock *et al.*, 2008) and detection of physical properties, such as hardness or solubility (Blanco *et al.*, 2006; Miralbés, 2004). These applications are based on calibration models for each analyte that have to be developed based on multivariate statistics. By mathematically correlating spectral data with the data obtained by the currently accepted laboratory procedures for reference analyses, the content of a respective constituent can be predicted by NIRS (Shenk *et al.*, 2008).

Also non-targeted qualitative applications of NIRS were reported. For example, comparison of raw NIR spectra from GM maize and non-GM maize was used to

differentiate between GM maize and its parental line (Bertrand *et al.*, 1985). In a similar experiment six cultivars were differentiated by Principal Component Analyses of wavelengths selected from NIR spectra (Delwiche and Norris, 1993). In another study, a combined approach of both qualitative and quantitative data evaluation was applied. Classification of hard wheat was obtained by calibration/prediction of protein content and degree of hardness, and in addition by Principal Component Analysis of full spectrum data (Sandorfy *et al.*, 2007).

Applications of NIRS in the food industry for routine analysis range from reception inspection to process monitoring (Ferreira *et al.*, 2005; Zhou *et al.*, 2006) to quality control of finished products (Esteban-Díez *et al.*, 2007). Examples are the detection of the ratio between Arabica and Robusta in coffee powder, that is the determining factor for coffee quality (Quilitzsch *et al.*, 2005); in agriculture NIRS is applied to the assessment of α- and β-carotene in produce (Tian *et al.*, 2007), and in pomiculture for the analysis of the solid fraction (Miralbés, 2004). In the milling industry rheological and chemical properties of flours are determined by use of NIRS (Baye *et al.*, 2006). Also proximate analyses in crop plants, such as protein, oil or starch in maize, are performed based on NIRS methods (Emura *et al.*, 2006).

2.4.4 Instrumentational Set Up

A spectrometer consists of an energy source, a dispersive element to enable the measurement at different wavelengths, a cell where the interaction with matter will take place, and a detector (Figure 5). In analytical NIR spectrometers usually a quartz halogen bulb is used as source, owing to its broad band emission of NIR light (800–2500 nm). Discrete wavelengths can be generated by monochromator systems or, as in state of the art instruments, by Fourier-Transform-Elements. The core piece of these so-called FT-NIR spectrometers is the interferometer, where the light is split into two beams and subsequently reunited by use of a mirror. The distance between the beam splitter and the mirror can be altered periodically to generate interference patterns. Depending on the optical path length intensity maxima will be obtained for discrete wavelengths, which can be used for the absorbance measurement at one wavelength after another. These spectrometers exhibit an excellent signal/noise ratio, short analysis time and a high resolution of less than 0.1 cm^{-1}. NIR light that is

not absorbed by the sample, can be measured in reflectance or transmission mode and results in a characteristic spectral profile for each sample based on its chemical composition (Osborne, 2000). Owing to strong scattering of the not absorbed light, an integration sphere can optionally be used (Figure 5) for forwarding to the detector. Several types of NIR detectors, such as photoinductive or photovoltaic elements, are in use, of which InGaAs or PbS detectors are mostly used for laboratory equipment. Absorption of the sample is calculated from the reflected light, which is measured by the detector.

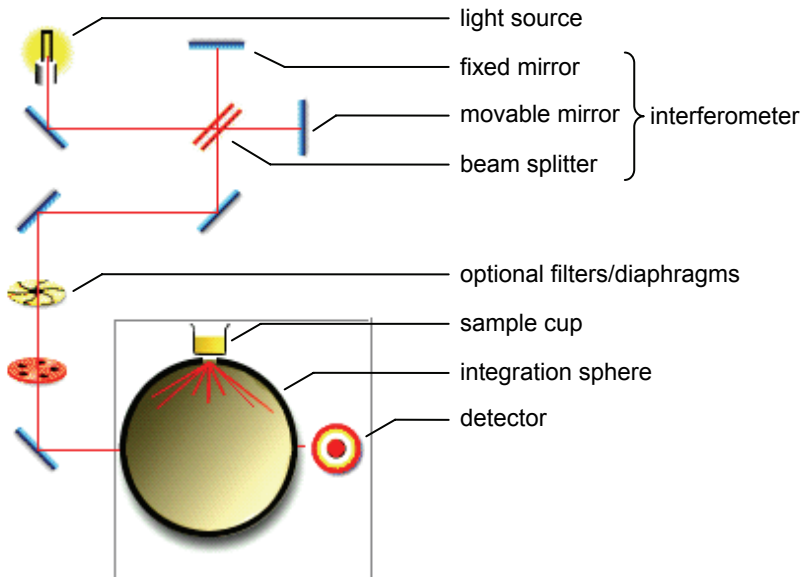

Figure 5: Principle of typical a NIR spectrometer using the example of a FT-NIR spectrometer (Spectrum One NTS, Perkin Elmer, Rodgau-Jügesheim, Germany)

2.4.5 Calibrating a NIR Spectrometer for Quantification

2.4.5.1 Principles of NIRS Calibrations

Absorption of near infrared light is based on overtone and combination vibrations resulting in broad bands which overlap to a large extent. Each absorption band contains the information of several constituents. Therefore, for quantification of an analyte, it will be not sufficient to assess the absorption at one distinct wavelength. This is the major difference between NIR spectral data and chromatographic data, in which the peak signal at a distinct retention time will correspond quantitatively to the analyte. For quantification based on NIR spectral data, the absorptions at different

wavelengths are usually mathematically combined by linear combination algorithms. The challenge is to locate the appropriate wavelengths and to define proper weighing factors to reflect the proportion of the analyte's absorption to the total absorption at a distinct wavelength. This step is performed by correlation of quantitative data obtained from reference samples by analysis with accepted methods and the spectroscopic data of these samples. The correlation of NIR spectral data with quantitative reference analytical data is defined as *calibration* resulting in a calibration *model*, that can be used to predict the content of an analyte in unknown samples based only on NIR spectral data (Perez-Marin *et al.*, 2007).

Typically, so-called global calibration models are developed, that will be able to properly predict the content of the target analyte in 90–95% of all samples of a given matrix (e.g. maize kernels) (Shenk *et al.*, 2001). These global calibrations are based on the Lambert-Beer law, which follows the assumption that the relation between absorption and concentration of an analyte is linear (Hruschka, 2001). To enable the accurate prediction by a NIRS calibration model, the reference samples selected for establishment of a global model have to reflect all potential variation in the samples that may be investigated by NIRS. In addition, the content of the analyte itself plays an important role for building a calibration model, and the reference samples should equally cover all possible concentrations from low to high values, to ensure the proper prediction by the model. Therefore, possible impact factors on absorption, composition as well as technical influences such as water content or milling quality, have to be considered.

To ensure the accurate prediction of contents, a NIRS calibration model has to be validated by use of samples that exhibit the same characteristics as the reference samples, but have not been incorporated for building the model. Validation means to predict the content of the target compounds by use of the calibration model in additional samples, and to compare these results with reference analytical values. Two major strategies can be applied for selection of test-samples. One option is to divide the reference sample into two groups; one group which is used for calibration, and another group that is used as test-set for validation of the calibration equation. This procedure is dependent on a large number of samples, which all have to be analyzed by reference analytical methods. In addition, the half of the samples cannot

be included in building the calibration model, and additional information of these samples may be lost. Therefore, as a second possibility the test-sets will be replaced by a cross-validation step. Cross-validation incorporates all available samples into the calibration and simultaneously will perform a validation during calculation of the model by use of a leave-one-out strategy. A leave-one-out strategy means to successively remove each individual sample from the calibration set and to analyze them on the basis of models build with the remaining samples. The great advantage of this approach is the possibility to use all available samples for both calibration and validation, and enables the establishment of calibration models also for small sets of reference samples.

2.4.5.2 Calibration Algorithms for NIRS

Three multivariate, linear calibration algorithms are mainly used for the development of NIRS calibration models: Multiple Linear Regression (MLR), Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR) (Wang and Paliwal, 2007).

MLR is based on the correlation of reference data with absorption values at a small number of wavelengths that were selected by the user. This selection may be based on literature data or on results from calibrations performed earlier (Perez-Marin *et al.*, 2007). In contrast, PCR and PLSR will not pre-select a number of wavelengths, but will use the whole spectrum for calculation of the regression equation. As a typical NIR spectrum contains redundant information, i.e. one constituent will cause absorption at different wavelengths, this procedure would cause overfitting of the model due to collinearity. Overfitting means, that the model will very precisely describe the levels of an analyte for the samples that were included in the calibration set, but the content in additional samples will be predicted very poorly, because the model is too specialized to the reference samples. Therefore, PCR and PLSR will reduce the number of variables/wavelengths (i.e. dimensions) to avoid collinearity. PCR will calculate principal components as new variables that contain the greatest proportion of variance between spectra. By nature, these principal components will contain no collinearity (see 2.3.2). The values of the principal components will then be analyzed for correlation with the reference data. However, the principal

components selected by PCR may not primarily contain the absorption differences of the target compound in the samples, but may be determined by other influences, which yields in a bad calibration. Therefore, PLSR will calculate principal components considering only wavelength bands, where a correlation between absorption differences and differences in the content of the target compound is observed (Roggo *et al.*, 2007; Rosipal and Krämer, 2006).

There are also other approaches for establishing a correlation between NIR spectral data and reference values, most notably neural networks (El-Sanhoty *et al.*, 2006). These methods follow an iterative process to create nested functions that describe the absorption at different wavelengths in relation to the content of a target analyte. The main advantage of these techniques is the incorporation of non-linear behavior, that can be observed e.g. for very broad concentration ranges, that will extend the range which is covered by the Lambert-Beer law, or for very low concentrations in complex matrices. However, these techniques are very sophisticated in implementation with benefits only in a few special cases, and have not been commercially established yet.

2.4.5.3 Evaluation of NIRS Calibration Models

The quality of a calibration model depends on a number of factors. In general an excellent calibration will be achieved considering the following factors: (a) good reproducibility of the reference method, which should be less than 5% of the (b) wide concentration range of reference samples, (c) an equal distribution of the reference values, (d) a large number of reference samples, that (e) will comprehensively represent the sample population, that will be analyzed by the model. A weak reproducibility of the reference method can be partially compensated by increasing the number of samples. However, too many reference samples that contain similar information may also decrease the quality of a model. Several statistical parameters can be considered for evaluation of a calibration model:

Coefficient of Correlation (r)

The coefficient of correlation is telling to which extent a statistical, linear coherence is given between two characteristics. It will be calculated for both calibration (r_{cal}) and

validation (r_{val}). It is ranging from −1 to 1; the algebraic sign will correspond to negative or positive correlation, respectively. The closer the value is to 1, the stronger is the correlation (Figure 6C, D). In case of zero, both characteristics are either independent from each other, or the correlation is non-linear, which will not be able to detect by the coefficient of correlation (Figure 6A). The coefficient of correlation can be used for determination of a linear relation, which is quantitatively, but not qualitatively (Sachs and Hedderich, 2006).

Slope of Regression Line

Another quality criterion for the correlation of values predicted by the calibration model and reference data is the slope of the regression line. In the best of cases its value is m = 1, representing the slope of the line through the origin (Figure 6B, D). A successful calibration therefore will be determined by a value close to the line through origin; the higher the deviation, the less accurate are the predicted values (Figure 6A, C) (Sachs and Hedderich, 2006).

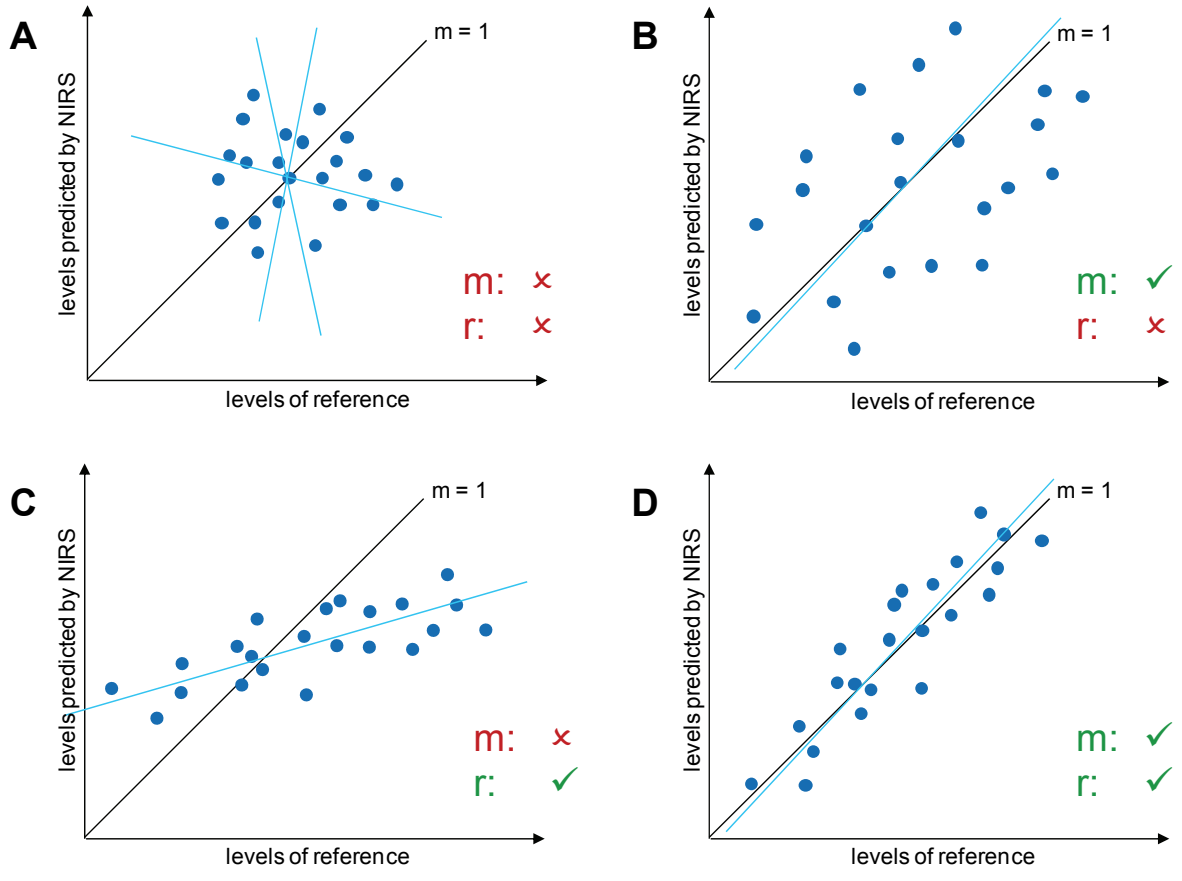

Figure 6: Illustration of statistical parameters for the evaluation of NIRS calibration models: m = slope, r = coefficient of correlation.

Standard Error of Model (SEC und SECV)

Performance of a calibration model can be evaluated on the basis of the Root Mean Squared Error of Calibration (SEC) and Root Mean Squared Error of Prediction (SECV) as determined by crossvalidation (leave-one-out strategy). These parameters are determined as the square root of the sum of the squared differences between predicted values and reference values for calibration and validation, respectively. They correspond to the standard error of the regression and serve as an indicator for the error of the prediction model (Martens and Naes, 1992).

Performance Criteria

To enable the comparison of different calibration models, the statistical parameters described above have to be standardized to remove the influence of units and absolute scales. The following performance criteria are described in literature for evaluation of NIRS calibration models:

Criterion 1 = (max − min) / SECV, minimum requirement > 10

max (maximum level observed in the calibration sample set)

min (minimum level observed in the calibration sample set)

(De la Roza *et al.*, 1998; Fontaine *et al.*, 2001)

Criterion 2 = s / SECV, minimum requirement > 3

s (standard deviation of levels in the calibration sample set)

(De la Roza *et al.*, 1998; Williams and Sobering, 1996)

Criterion 3 = SECV / $s_{reference}$, minimum requirement ≤ 2

$s_{reference}$ (standard deviation of the reference method)

(Fontaine *et al.*, 2001)

Criteria 1 and 2 describe the relation of the prediction error of the model to the distribution (range and variance) of reference values. Criterion 1 takes into account the concentration range of the reference data, i.e. the ratio of the range in reference values (minimum to maximum) to the standard error of the calibration model (SECV); criterion 2 will consider the standard deviation SD of the sample population in

relation to the standard error of the prediction by the model SECV. The minimum requirements of 10 and 3 for criterion 1 and 2, respectively, allow the differentiation of samples low, medium and high in the target constituent (Williams and Sobering, 1996). Criterion 3 is deduced according to requirements described in literature (Fontaine *et al.*, 2001), that the SECV should not be higher than twice the standard deviation of the reference method.

3 MATERIALS AND METHODS

3.1 Maize Seed Materials

3.1.1 Sample Sets

Investigation of the Influence of Genetics and Environment (section 4.2)

Four maize (*Zea mays*) cultivars (*Flavi,* Caussade Semences, Caussade, France; *Lukas,* Limagrain GmbH Edemissen, Germany; *Pontos*, Limagrain GmbH Edemissen, Germany; *ES Shorty*, Euralis Semences, Lescar, France) were grown at one location (Frankendorf) in Bavaria (Germany). One cultivar (*Amade*o, KWS Mais GmbH, Germany) was grown at four different locations (Mittich, Reith, Strassmoos, Thann) in Bavaria (Germany). Samples were obtained from field trials in the growing seasons 2004, 2005 and 2006. Growing periods were as follows: Frankendorf: 21.04.–19.10.2004, 03.05.–26.10.2005, 02.05.–12.10.2006; Mittich: 17.04.–05.10.2004, 14.04.–12.10.2005, 24.04.–11.10.2006; Reith: 19.04.–20.10.2004, 15.04.–18.10.2005, 24.04.–12.10.2006; Strassmoos: 21.04.–28.10.2004, 02.05.–20.10.2005, 24.04.–10.10.2006; Thann: 21.04.–22.10.2004, 02.05.–25.10.2005, 21.04.–17.10.2006.

Investigation of the Influence of Input System (section 4.3)

Three maize (*Zea mays*) cultivars (*Amadeo*, KWS Mais GmbH, Germany; *Lukas*, Limagrain GmbH, Edemissen, Germany; *Flavi*, Caussade Semences, Caussade, France) were grown in the season 2004 at two locations with different input regimes. At location Frankendorf (Bavaria, Germany) the crops were grown conventionally, at location Schönbrunn (Bavaria, Germany) organic farming was employed (experiment Ia). The same procedure was repeated in the season 2005 (experiment Ib). For experiment II samples from two plots with different input systems (conventional/organic) were obtained from one location Scheyern (Bavaria, Germany). The distance between the field plots was approximately 400 m. Growing periods were as follows: Frankendorf: 21.04.–19.10.2004, 03.05.–26.10.2005; Schönbrunn: 22.04.–19.09.2004, 12.05.–27.09.2005; Scheyern: 12.05.–25.10.2006. Additional Metadata on agronomy are provided in Table 1.

Table 1: Metadata on agronomy (2004–2006)

										fertilization			plant-protection		
	Sowing date	harvest date	temp.[a] [°C]	rain[b] [mm]	soil[c] index	N[d] [kg/ha]	P_2O_5[d] [mg/kg]	K_2O[d] [mg/kg]	pH[d]	fertilizer	[kg/ha][e]	date	herbizide/ treatment	[l/ha]	date
							season 2004								
Frankendorf (conv)	Apr 21	Oct 19	14.0	501	80	104	90	160	7.1	NK NK	30 80	Apr 21 Jun 06	Artett Spectrum	2.8 1.4	May 12 May 12
Thann (conv)	Apr 21	Oct 22	14.4	447	41	76	36	24	6.2	NP NK NK	30 70 50	Apr 21 May 19 Jun 14	Gardo Gold Callisto	4.0 1.0	May 19 May 19
Reith (conv)	Apr 19	Oct 20	14.6	471	42	78	10	19	6.3	NP NK	30 130	Apr 17 May 11	Gardo Gold Callisto	3.0 0.75	May 19
Mittich (conv)	Apr 17	Oct 05	15.1	591	46	100	16	16	6.4	NP NK	30 130	Apr 17 May 11	Gardo Gold Callisto	3.0 0.75	May 19
Strassmoos (conv)	Apr 21	Oct 28	14.2	336	41	59	23	24	6.4	NP Alzon 47 N	36 150	Apr 21 Apr 26	Gardo Gold Callisto Gardo Gold Callisto Callisto Certrol B	3.0 0.75 1.0 0.25 1.0 0.3	May 19 May 19 Jun 02 Jun 02 Jun 17 Jun 17
Schönbrunn (org)	Apr 22	Sep 19	14.7	403	52	53	220	310	7.5	-	-	-	currycomb roller hoe		May 13 May 25
							season 2005								
Frankendorf (conv)	May 03	Oct 25	14.3	570	80	82	200	230	6.8	NK NK	30 100	May 03 Jun 13	Artett Spectrum	2.8 1.4	Jun 02 Jun 02
Thann (conv)	May 02	Oct 25	14.6	570	47	50	33	25	6.4	NP NK NK	30 70 50	May 02 May 30 Jun 20	Gardo Gold Callisto	4.0 1.0	Jun 02
Reith (conv)	Apr 15	Oct 18	14.8	583	70	66	22	21	-	NP NK	40 115	Apr 14 May 19	Callisto	1.0	May 25
Mittich (conv)	Apr 14	Oct 12	15.5	535	46	93	28	19	-	farmyard manure NP NK	120 40 35	Apr 01 Apr 14 May 19	Gardo Gold Callisto	4.0 1.0	May 25
Strassmoos (conv)	May 02	Oct 20	14.6	490	40	50	18	20	6.4	NPK NP Alzon 47 N	34 27 110	Apr 12 May 02 May 12	Gardo Gold Callisto Gardo Gold Callisto	3.0 0.75 1.0 0.25	May 25 May 25 Jun 06 Jun 06
Schönbrunn (org)	May 12	Sep 27	14.8	482	52	-[f]	170	190	7.6	-	-	-	currycomb currycomb roller hoe roller hoe		May 27 Jun 01 May 06 Jun 21
							season 2006								
Frankendorf (conv)	May 02	Oct 12	15.3	420	80	73	23	27	7.2	NK NK	30 100	May 02 Jun 13	Artett Spectrum	2.8 1.4	Jun 10
Thann (conv)	Apr 21	Oct 17	15.7	438	60	40	31	44	6.7	NP NK NK	30 70 50	Apr 21 May 15 Jun 12	Gardo Gold Callisto	4.0 1.0	May 22
Reith (conv)	Apr 24	Oct 12	15.8	444	65	82	33	18	7.1	NP NK NK	35 60 45	Apr 24 May 08 Jun 06	Gardo Gold Callisto	3.0 0.75	May 22
Mittich (conv)	Apr 24	Oct 11	16.4	387	73	65	9	6	6.8	NP NK NK	35 60 60	Apr 24 May 08 Jun 06	Gardo Gold Callisto	3.0 0.75	May 22
Strassmoos (conv)	Apr 24	Oct 10	15.5	309	36	60	12	14	5.8	NPK NP ENTEC	39 27 110	Apr 20 Apr 24 May 04	Gardo Gold Callisto Gardo Gold Callisto	2.0 0.5 2.0 0.5	May 15 May 15 Jun 08 Jun 08
Scheyern (conv)	May 12	Oct 25	15.2	387	63	79	110-200	110-200	6.1	NP N	60 130	May 12 May 15	Motivell Certrol	0.9 0.8	Jun 12 Jun 12
Scheyern (org)	May 12	Oct 25	15.2	387	64	70	110-200	110-200	6.0	-[g]	-[g]	-[g]	roller hoe roller hoe roller hoe hoe		May 22 Jun 16 Jul 01 Jul 02

a Mean temperatures May–October; b Sum of precipitation May–October; c Productivity indicator "Ackerzahl": Relative yield in comparison to the best German site (Ackerzahl = 100) (BewRL, 1967); d Soil samples (0 to −90 cm) taken in spring before application of fertilizer; e Total amount of fertilizer; f No data available; g Mustard grown in autumn of the pre-season was ploughed in to the ground as green manure

Investigation of the Influence of Genetic Engineering (section 4.4)

South African Maize. White maize samples were derived from the transgenic Bt hybrid variety DKC78-15B (hybrid of event MON 810 from Monsanto), from the transgenic glyphosate-tolerant Roundup Ready variety DKC78-35R (hybrid of event NK603 from Monsanto), and the near-isogenic non-GM hybrid variety CRN 3505 (Monsanto). The plants were grown at two different sites, namely, Petit and Lichtenburg (South Africa), under high-input system; the varieties were planted in Petit over three growing seasons (2004, 2005 and 2006) and in Lichtenburg over one growing season (2004). At planting, the plants were fertilized with 300 kg ⁄ Ha 4 : 3 : 4 (33), topdressing 300 kg ⁄ Ha KAN (28) and treated with herbicide 1.8 L ⁄ Ha Guardian + 200 mL ⁄ Ha Sumi Alpha. Two months after planting, the plants were treated with herbicide, 2.2 L ⁄ Ha A-maizing + 1L ⁄ Ha Harness + 220 mL ⁄ Ha alphacypermytrin. Three months after planting, the material was treated with pesticide, 750 mL ⁄ Ha Endosulfan against stalkborer. The DKC78-15B and the control variety CRN3505 were also grown in Potchefstroom (South Africa) under low-input system, which means that no fertilizer, no fungicide and no herbicide were applied throughout the growth of the plants. The plant material was harvested 8 months after planting.

For the field trial performed at location Petit in 2005, three replicate samples were available and the results were averaged prior to further analysis for all techniques. For all other field trials, one sample was analyzed.

German Maize. The transgenic Bt hybrid variety TXP 138-F (hybrid of event MON 810 from Monsanto), and its isogenic counterpart DKC3420 (Monsanto) were harvested in season 2004. The samples were grown under conventional farming practice in Bavaria (Germany). At location Neuhof four field replicates, at Pfaffenhofen three field replicates were collected.

NIRS Screening of Maize (section 4.5)

Three maize (Zea mays) cultivars (*Gomera EU*, Euralis Saaten GmbH, Norderstedt, Germany; *Amadeo*, KWS Mais GmbH, Germany; *Lukas*, Limagrain GmbH, Edemissen, Germany) were grown at location Strassmoos (Bavaria, Germany) in three consecutive seasons. Growing periods were as follows: 21.04.–28.10.2004, 02.05.–20.10.2005, 24.04.–10.10.2006. Metadata on agronomy are provided in Table 1.

The samples for investigation of the influence of input system (experiment Ia, Ib and II) were also included for analysis by both GC/MS and NIRS.

For the development of NIRS calibration models the sample sets described above were extended to a total of 383 maize samples, that were collected over three growing seasons (2004–2006) from 17 different locations in Bayern (Germany), and 3 locations in South Africa, covering 32 different genotypes.

3.1.2 Growing Parameters and Sampling

All samples were obtained from field trials with totally randomized field plot design. For each cultivar/location/season three field replicates were available, if not mentioned otherwise. Ten cobs were harvested from the two mid rows of each plot and a sub-sample of 100 g kernels was taken for further processing. *Dry matter* was determined by drying the kernels at 105°C to constant weight. *Plant height* was measured after flowering from ground to top of tassel. *Kernel yield* was determined for the two mid rows of each plot (9 m^2) and extrapolated to one hectare. Calculation of yield was based on dry matter standardized to a moisture content of 16%.

The locations employing organic farming had been managed for at least 3 years according to the provisions laid down in Council Regulation (EEC) 2092/1991.

3.1.3 Sample Processing

Maize collected from the German sites was air-dried (30–40°C) for 3 days. Samples from South Africa were dried already in the field before harvest. The dried maize kernels (10–15% moisture) were frozen in liquid nitrogen and immediately ground with a cyclone mill (Cyclotec, Foss, Germany) equipped with a 500 µm sieve. The flour was freeze-dried (ALPHA 1-4 LSC, Christ, Osterode, Germany) for 48 hours. The moisture content of the resulting material (<2%) was determined as loss of weight by drying at 105 °C for 3 hours. Freeze-dried flour samples were stored at −18 °C in tightly closed LDPE bottles (Kautex Textron, Bonn, Germany).

3.1 GC/MS-Metabolite Profiling of Maize

3.1.1 Sample Extraction

Four hundred milligrams of freeze-dried maize flour were weighed into a 3 mL cartridge (Merck, Darmstadt, Germany) which was sealed with PTFE frits at the bottom and top of the flour layer. The cartridge was connected to a vacuum manifold (Supelco, Taufkirchen, Germany). For disintegration of the matrix, the maize flour was pre-soaked in 200 µL of methanol (Merck, Darmstadt, Germany) for 20 min at ambient temperature with vents of the manifold closed (Frenzel *et al.*, 2002). The methanol was removed by application of vacuum (30–20 mbar max) on the top of the cartridge for 30 min. Lipids were eluted with 4 mL of dichloromethane (Riedel de Haën, Seelze, Germany) into 11 mL vials (lipid extract) by gravity flow. Residual dichloromethane was removed from the flour by application of vacuum on the bottom of the cartridge. Polar compounds were eluted with a total of 10 mL of methanol/water (80+20, v+v) within 40 min into 11 mL-vials by application of weak vacuum at the bottom.

3.1.2 Preparation of Standard Solutions

Reference compounds were obtained from Merck KgaA (Darmstadt, Germany), Fluka (Buchs, Switzerland), Riedel de Haën (Seelze, Germany) and Oxeno (Marl, Germany).

Retention time standard mix I: Solutions of undecane (1.5 mL, 1 mg/mL), hexadecane (2.5 mL, 1 mg/mL), tetracosane (4 mL, 1 mg/mL) and triacontane (4 mL, 1 mg/mL) in *n*-hexane (Merck, Darmstadt, Germany) were added to 10 mg of octatriacontane. Hydrocarbons were purchased from Fluka (Buchs, Switzerland).

Retention time standard mix II: 1.5 mL of *n*-hexane (Merck, Darmstadt, Germany) and solutions of hexadecane (2.5 mL, 1 mg/mL), tetracosane (4 mL, 1 mg/mL) and triacontane (4 mL, 1 mg/mL) in *n*-hexane were added to 10 mg of octatriacontane. Hydrocarbons were purchased from Fluka (Buchs, Switzerland). Alanine was used as retention time standard in place of undecane for fraction IV.

Internal standard solution for fraction I: Identical to Retention time standard mix I. Tetracosane was used as internal standard for quantification of major lipids.

Internal standard solution for fraction II: 6 mg of 5α-cholestan-3β-ol (Fluka, Buchs, Switzerland) were dissolved in 10 mL of dichloromethane (Riedel de Haën, Seelze, Germany).

Internal standard solution for fraction III: 40 mg of phenyl-β-D-glucopyranoside (Fluka, Buchs, Switzerland) were dissolved in 25 mL of distilled water.

Internal standard solution for fraction IV: 20 mg of *p*-chloro-L-phenylalanine (Fluka, Buchs, Switzerland) were dissolved in 25 mL of distilled water.

3.1.3 Fractionation and Analysis of Lipids

100 µL of internal standard solution for fraction I and 100 µL of internal standard solution for fraction II were added to the lipid extract. The solution was evaporated in 4 mL vials to dryness by rotary evaporation (ACTEVap Evaporator, Activotec, Cambridge, United Kingdom). Residual solvents were removed by application of nitrogen. The lipids were re-dissolved in 500 µL of dry methyl *tert*-butyl ether (MTBE, Oxeno, Marl, Germany) and 250 µL dry methanol 50 µL of sodium methylate, 5.4 M in methanol (Fluka, Buchs, Switzerland) were added. After reaction for 90 min at room temperature in the dark, 1 mL of dichloromethane and 2 mL of aqueous 0.35 M HCl were added. The solution was shaken vigorously and the upper phase was discarded. After re-extraction of the lower phase containing the transmethylated lipids with another 2 mL of aqueous 0.35 M HCl the solution was evaporated to dryness by rotary evaporation.

The dry transmethylated lipid extract was sub-fractionated by solid phase extraction (SPE). After placing 200 - 300 mg of sodium sulfate on top of the cartridge one column volume (CV, 2.5 mL) of *n*-hexane was used for conditioning the SPE column. The *n*-hexane was removed by application of weak vacuum on the bottom. Transmethylated lipids were re-dissolved in 250 µL of dichloromethane and transferred to the SPE cartridge.

The methyl ester fraction (fraction I) was eluted with 3x2 mL of *n*-hexane and MTBE (100:2, v+v). The eluate was evaporated to dryness by rotary evaporation (160 mbar min) and re-dissolved in 300 µL of *n*-hexane and transferred into a TPX plastic auto-sampler vial with an integrated 0.2 mL glass micro-insert and closed using a 6 mm silicone/PTFE red screw cap. 1 µL was injected into to GC/MS.

The minor polar lipid fraction (fraction II) was eluted with 3x2 mL of *n*-hexane and MTBE (70:30, v+v). After addition of 100 µL of retention time standard mix I the eluate was evaporated to dryness by rotary evaporation (160 mbar min). Residual solvents were removed by application of nitrogen. Fraction II was re-dissolved in 250 µL of dry pyridine (Fluka, Buchs, Switzerland) and 50 µL of N-methyl-N-trimethylsilyl-trifluoracetamide (MSTFA, Merck, Darmstadt, Germany). After flushing with argon, the vial was tightly sealed with PTFE-sealings and silylated for 15 min at 70°C. After transfer into a TPX plastic auto-sampler vial with an integrated 0.2 mL glass micro-insert and closed using a 6 mm silicone/PTFE red screw cap 1 µL was injected into the GC/MS.

3.1.4 Fractionation of Polar Extract

150 µL of internal standard solution for fraction III and 150 µL of internal standard solution for fraction IV were added to the polar extract. 1 mL of this solution was concentrated in 4 mL vials by rotary evaporation and dried over phosphorus pentoxide. After re-dissolving in 200 µL of dry pyridine and 100 µL of dry trimethylsilylimidazole (TMSIM, Fluka, Buchs, Switzerland) were added, the sample was silylated for 20 min at 70°C in a tightly sealed vial. For differential hydrolysis of the silylated derivatives, 200 µL of *n*-hexane and 400 µL of water were added. After slightly shaking at room temperature and subsequent phase separation (5 min) 150 µL of the upper phase (fraction III) were transferred into a TPX plastic auto-sampler vial with an integrated 0.2 mL glass micro-insert and 75 µL of retention time standard mix I was added and closed using a 6 mm silicone/PTFE red screw cap. 1 µL was injected into the GC/MS.

2 mL of polar extract were concentrated by rotary evaporation and dried over phosphorus pentoxide. After re-dissolving in 250 µL of dry hydroxylammonium-chloride (Merck, Darmstadt, Germany) the sample was oximated for 30 min at 70°C. 100 µL of MSTFA were added. After flushing with argon the tightly sealed vials were allowed to stand for 20 min at 70°C. 500 µl of *n*-hexane and 300 µL of water were added. After vortexing and phase separation the upper phase was removed and the lower phase was re-extraced with 2 x 500µL of *n*-hexane. The lower phase, containing acids, amino acids and amines (fraction IV) was concentrated to dryness by rotary

evaporation and dried over phosphorus pentoxide. 100 µL of retention time standard mix II were added and the solvent was removed by application of nitrogen. The dry extract was redissolved in 250 µL of dry acetonitrile (Merck, Darmstadt, Germany) and 50 µL of MSTFA were added. After flushing with argon, the sample was resilylated for 60 min at 70°C. After transfer into a TPX plastic auto-sampler vial with an integrated 0.2 mL glass micro-insert and closed using a 6 mm silicone/PTFE red screw cap 1 µL was injected into the GC/MS.

3.1.5 GC/MS Analysis

Gas chromatography was performed on a Finnigan TraceGC Ultra (Thermo Electron Corp., Austin, TX) with split/splitless injector combined with a Finnigan Trace DSQ mass spectrometer (Thermo Electron Corp., Austin, TX) with electron ionization (EI) ion source. The column used was a factorFOUR VF-1ms, 60 m x 0.32 mm internal diameter (ID), coated with a 0.25 µm film of 100% polydimethylsiloxane (Varian, Darmstadt, Germany). Injection was performed in split mode (split flow 15 mL/min) at an injection temperature of 280°C. Helium as carrier gas was used at a constant flow of 1 mL/min. Column temperature was programmed from 100°C to 320°C (10 min hold) at a 4°C/min. The MS interface temperature was set to 320°C. After a solvent delay of 6 min full scan mass spectra were recorded within a scan range of 40-700 mu at an electron energy of 70 eV and a source temperature of 250°C. Identification of maize constituents was achieved by comparison of retention times and mass spectra with those of silylated and methylated reference compounds or by comparing mass spectra of the NIST02 MS database (Ausloos *et al.*, 1999). Peak heights were normalized by the heights of the internal standards in the respective fraction.

3.1.6 Metabolite Identification

Metabolites were identified according to mass spectral data from custom (A: mass spectral data and retention times of reference compounds), public (B: mass spectral data and retention index of Golm Metabolome Database (Kopka *et al.*, 2005)), commercial (C: mass spectral data of NIST02 mass spectral library (Ausloos *et al.*,

1999)) mass spectral libraries and from literature (E: Xu and Godber (1999), F: Kamal-Eldin *et al.* (1992), and G: Meyna (2005), Miller (1982)).

For A, B and C metabolites were denoted as identified if the similarity index was >750 on a scale of 0 to 1,000; in addition, for A a maximum relative retention time deviation of 0.1% and for B a maximum relative retention index deviation of 1.0% were required.

3.1.7 Validation of GC/MS Methodology

Repeatability

Repeatability was determined by triplicate analysis of a maize sample. Relative standard deviations (RSD) of peak heights normalized to the internal standard of the respective fraction were calculated.

Recovery

Recovery rates of selected compounds were determined by analyzing three aliquots of both lipophilic and polar liquid extracts of the same maize flour in triplicate. Extraction, fractionation, derivatization and GC/MS investigation were performed as described above. The first aliquot of the extract was spiked with the standard compounds at the beginning of the analytical procedure. The second aliquot was spiked at the end of the work up prior to GC/MS investigation. The third aliquot of the extract was analyzed to calculate peak heights of the standard compounds naturally observed in the unspiked flour. Peak heights of the first extract aliquot were compared to peak heights of the second sample, taking the peak heights of extract aliquot three into account. Recoveries were calculated according equation 3:

$$W = \frac{H_{P1} - H_{P3}}{H_{P2} - H_{P3}} * 100\% \qquad (3)$$

Equation 3: calculation of recovery rates: W recovery rate, H_{P1} peak height in the aliquot of the extract spiked in the beginning of the analytical procedure, H_{P2} peak height in the aliquot of the extract spiked prior to GC/MS investigation, H_{P3} peak height in the unspiked aliquot of the extract

Preparation of standard compounds for recovery rate

Fraction I: 20 µL of *n*-hexane solution containing 1.5 mg/mL triarachin, 8.5 mg/mL tripalmitin, 7.5 mg/mL tristearin and 0.75 mg/mL squalene was added to the flour at the start of analysis. 20 µL of *n*-hexane solution containing 1.5 mg/mL methyl arachidate, 8.5 mg/mL methyl palmitate, 7.5 mg/mL methyl stearate and 0.75 mg/mL squalene was added to the sample prior to GC/MS investigation.

Fraction II: 20 µL of dichloromethane solution containing 3.0 mg/mL 5α-cholestan-3β-ol, 0.25 mg/mL δ-tocopherol, 0.5 mg/mL stigmasterol, 5.0 mg/mL β-sitosterol, 7.5 mg/mL palmitic acid, 5.0 mg/mL stearic acid and 0.25 mg/mL octadecanol was added to the maize flour and to fraction II prior to silylation, respectively.

Fraction III: 20 mg sucrose and 100 µL of hexane solution containing 5 mg/mL fructose, 5 mg/mL glucose, 5 mg/mL galactose, 5 mg/mL raffinose, 5 mg/mL sorbitol and 5 mg/mL inositol (standard compound mix for fraction III) were added to the flour at the start of analysis. 15 µL of internal standard solution for fraction III was added to 1 mL of polar extract. At the end of analysis 75 µL of a pyridine/TMSIM (50+25, v+v) solution of 6 mg sucrose, 4.5 µL dried internal standard solution for fraction III and 3 µL dried standard compound mix for fraction III was added to 150 µl of upper phase of fraction III.

Fraction IV: 50 µL of an aqueous solutions containing 2.4 mg/mL *p*-chloro-L-phenylalanine, 0.1 mg/mL alanine, 0.1 mg/mL glycine, 0.5 mg/mL malic acid, 0.5 mg/mL glutamine, 0.1 mg/mL lysine hydrochloride, 0.5 mg/mL aspartic acid, 0.5 mg/mL asparagine, 0.1 mg/mL lactic acid and 0.1 mg/mL fumaric acid was added to 2 mL of polar extract and to fraction IV prior to silylation, respectively.

3.1.8 Pre-Processing of Chromatographic Data

Retention time matching of GC/MS data was performed by use of Chrom*pare*, a self-tailored MS Excel® tool, basically based on Student's *t*-test (Frenzel *et al.* (2003); www.chrompare.com). The tool is optimized for comparison of chromatographic data, including automated retention time adjustment according to retention time standards. Metabolites were quantified by relative peak levels according to the respective internal standard. Data from triplicate analysis of each sample were averaged for further statistical analyses.

3.2 NIRS-Metabolite Profiling of Maize

3.2.1 NIRS analysis

The ground and freeze-dried maize samples were conditioned for 15 h in the freeze-dryer before measurement in the NIR spectrometer to standardize the water content of the samples. NIRS analysis was performed in diffuse reflection by means of a commercial Fourier-transform (FT) NIR spectrometer, equipped with an integrating sphere and an InGaAs detector (Spectrum One NTS, Perkin Elmer, Rodgau-Jügesheim, Germany). A total of 10 g of flour was placed into a quartz petri dish and flattened by a stamp. After closing the petri dish, it was subsequently mounted onto a continuously rotating measurement unit (NIRA sample spinner, Perkin Elmer, Rodgau-Jügesheim, Germany). Spectra were recorded at room temperature from 1000 nm to 2500 nm in 0.5 nm steps (resolution: 8 cm^{-1}). One hundred spectra were averaged to obtain a mean NIR spectrum for each sample. Spectral data (background subtracted) were stored as log (1/R) values at corresponding wavelength points.

3.2.1 Pre-Processing of Spectral Data

Smoothing (Savitzky Golay, 7 smoothing points) and multiple scatter correction (MSC) were performed by means of commercially available software (Spectrum Quant, Spectrum Quant+, Perkin Elmer, Rodgau-Jügesheim, Germany). For export, pre-processed spectral data were saved as ASCII files. The wavelength ranges 1000–1081 nm and 2448–2500 nm were excluded for further data analysis due to increased noise in the beginning and the end of the spectra.

3.3 NIRS Calibration

3.3.1 Sample Selection

Reference data for calculation of NIRS calibration models were obtained by GC/MS-metabolite profiling. Out of a pool of 383 samples, for sugars and acids/amino acids a total of 99 and 101, respectively, maize samples including various influence factors on maize composition such as genotype (32 cultivars), location (20 farming sites) or growing season (2004–2006) were selected. Also 20 samples from three field sites located in South Africa were included.

3.3.2 Establishment of Calibration Models

In order to establish a correlation between data obtained by NIRS and the total contents of sugars and of acids/amino acids in the samples, sum parameters were calculated for the respective fractions based on relative quantifications as determined by GC/MS-metabolite profiling. The calibration datasets consisting of pre-processed spectral data and reference GC/MS-metabolite profiling data were imported into the Unscrambler® (CAMO Software ASA, Oslo, Norway) software for statistical analysis. Calibration equations were calculated based on Partial Least Squares Regression algorithms (PLSR) and stepwise Multiple Linear Regression (MLR). For stepwise MLR the resolution of the spectral data was averaged to 5 nm segments to reduce the number of variables.

3.3.3 Validation of NIRS Calibration Models

Performance of calibration models was evaluated on the basis of the Root Mean Squared Error of Calibration (SEC) and Root Mean Squared Error of Prediction (SECV) as determined by crossvalidation (leave-one-out strategy). A leave-one-out strategy means to successively remove each individual sample from the calibration set and to analyze them on the basis of models build with the remaining samples. Performance criteria were calculated as follows: Criterion 1 = (max − min) / SECV, minimum requirement > 10, max (maximum level observed in the calibration sample set), min (minimum level observed in the calibration sample set) (De la Roza *et al.*, 1998; Fontaine *et al.*, 2001); criterion 2 = s / SECV, minimum requirement > 3, s (standard deviation of levels in the calibration sample set) (De la Roza *et al.*, 1998; Williams and Sobering, 1996); criterion 3 = SECV / $s_{reference}$, minimum requirement ≤ 2, $s_{reference}$ (standard deviation of the reference metabolite profiling method (10%) according to section 4.1, (Fontaine *et al.*, 2001)).

3.3.4 Application by NIRS Calibration Models

In order to determine the levels of sugars and acids/amino acids in maize by means of NIRS, sample pre-treatment, acquisition of NIR spectra and pre-processing of spectral data were performed in the same way as described for calibration and validation

samples. Subsequently, sugar and acid/amino acid contents were predicted on the basis of pre-processed spectral data by applying the respective calibration equations.

3.4 Statistical Analysis

3.4.1 Principal Component Analysis (PCA)

PCA was performed by use of Systat 11 (Systat Software Inc., CA). Profiling data were auto-scaled by the standard deviation of each variable (correlation matrix, (Jackson, 1991)) to reduce the influence of metabolites with high abundance. For PCA of NIRS data the water bond between 1858–2041 nm was excluded.

3.4.2 Peak-by-Peak Comparison (Chrompare)

Peak-by-peak comparison was performed by use of Chrom*pare*, a self-tailored MS Excel® tool, basically based on student's *t*-test ((Frenzel *et al.*, 2003); www.chrompare.com). The tool is optimized for comparison of chromatographic data, including automated retention time adjustment according to retention time standards. Peaks below noise level were discarded on the basis of a threshold of 2% relative peak height in fractions I/II and 3% relative peak height in fractions III/IV. Trace constituents for whom the confidence intervals ($p < 0.05$) were higher than their mean levels were also not included for comparison.

3.4.3 Analysis of Variance (ANOVA)

ANOVA was performed by use of Systat 11 (Systat Software Inc., CA). The significance level was set to $p < 0.01$ for all statistical comparisons. Differences were considered to be statistically significant if no interaction effect was observed and the main effect was significant by ANOVA and after post hoc testing by Tukey's HSD.

For investigation of the influence of genetics and environment (section 4.2) the ANOVA model described in equation 4 was used for each analyte:

$$v_{ij} = \mu + t_i + y_j + e_{ij} \qquad (4)$$

where v_{ij} is the response for the *i*th treatment (i.e. cultivar or location) and the *j*th year; μ is the overall mean, e_{ij} is the random error including error of field replicates ($n = 3$). For investigation of the influence of input system (section 4.3) the model was

modified according to equation 5 to include the interaction of location/input system and genotype:

$$v_{ik} = \mu + c_i + i_k + (ci)_{ik} + e_{ik} \qquad (5)$$

where v_{ik} is the response for the *i*th cultivar (n = 3) and the *k*th location/input system (n = 2); μ is the overall mean, c_i is the effect of the *i*th maize cultivar, i_k is the effect of the *k*th location/input system, $(ci)_{ik}$ is the effect of the interaction between the *i*th maize cultivar and the *k*th location/input system and e_{ik} is the random error including error of field replicates.

Logit (p) transformation

The distribution of *p*-values was transformed by use of $\text{logit}(p) = \log(p) - \log(1-p)$ (Ashton, 1972; Greiff *et al.*, 2002; Ter Braak and Gremmen, 1987). This will modify the distribution of *p*-values from $[0 \leq p \leq 1]$ to $[-\infty < \text{logit}\ (p) < +\infty]$. E.g. logit $(p=0.5) = 0$, logit $(p=0.1) \approx -1$, $\text{logit}(p=0.01) \approx -2$, $\text{logit}(p=0.001) \approx -3$. Histograms showing logit (*p*)-values can be interpreted as follows: the more equal the distribution of logit (*p*)-values, the more significant is a treatment; the more logit (*p*)-values are around zero, the less significant is a treatment.

4 RESULTS AND DISCUSSION

4.1 Adaption of the Metabolite Profiling Methodology to Maize

The metabolite profiling methodology applied in this study is based on consecutive extraction of freeze-dried flour resulting in a lipid and a polar extract (Figure 7). The fractionation method is in accordance with the procedure described for the analysis of rice grain (Frenzel *et al.*, 2002). After transesterification the lipid extract is separated by solid phase extraction (SPE) into a fraction (I) containing fatty acid methyl esters and a fraction (II) containing minor lipids such as free fatty acids and sterols. The base-catalyzed transmethylation of the triglycerides and the subsequent SPE on silica gel allow a separate analysis of the metabolically important free fatty acids that are not methylated. This procedure has been shown to proceed without

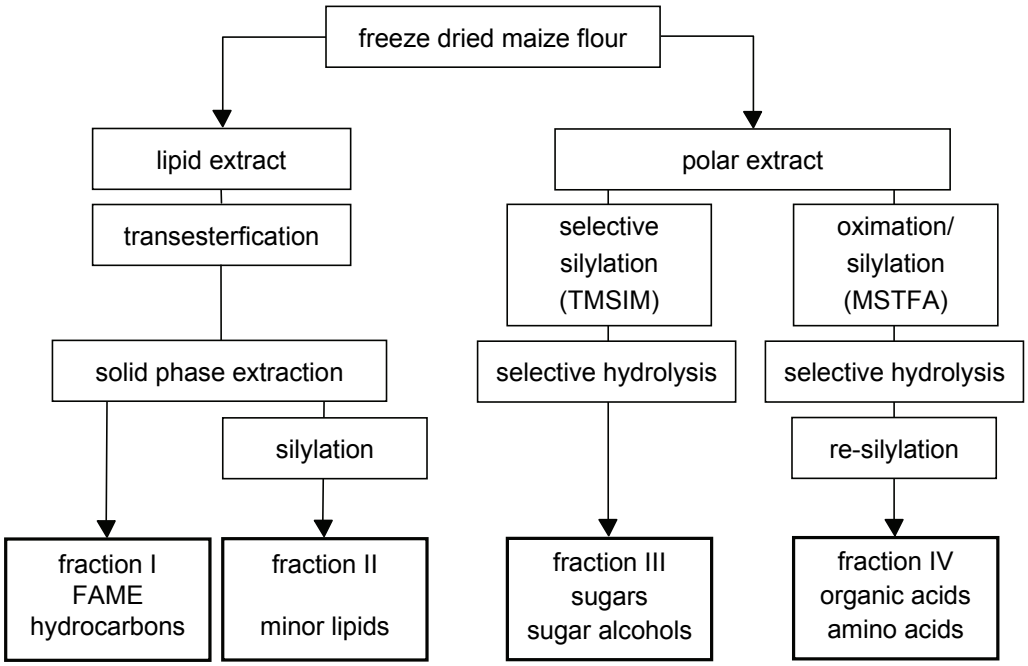

Figure 7 Fractionation scheme of the applied metabolite profiling method for maize

formation of artifacts (Frenzel *et al.*, 2002). In the polar extract sugars and sugar alcohols (fraction III) are separated from acids, amino acids and amines (fraction IV) by silylation and subsequent differential hydrolysis. This fractionation is based on the relative stability of the R-Si$(CH_3)_3$ group of sugars/polyols and amino acids/organic acids to aqueous hydrolysis (Frenzel *et al.*, 2002). An additional oximation step is

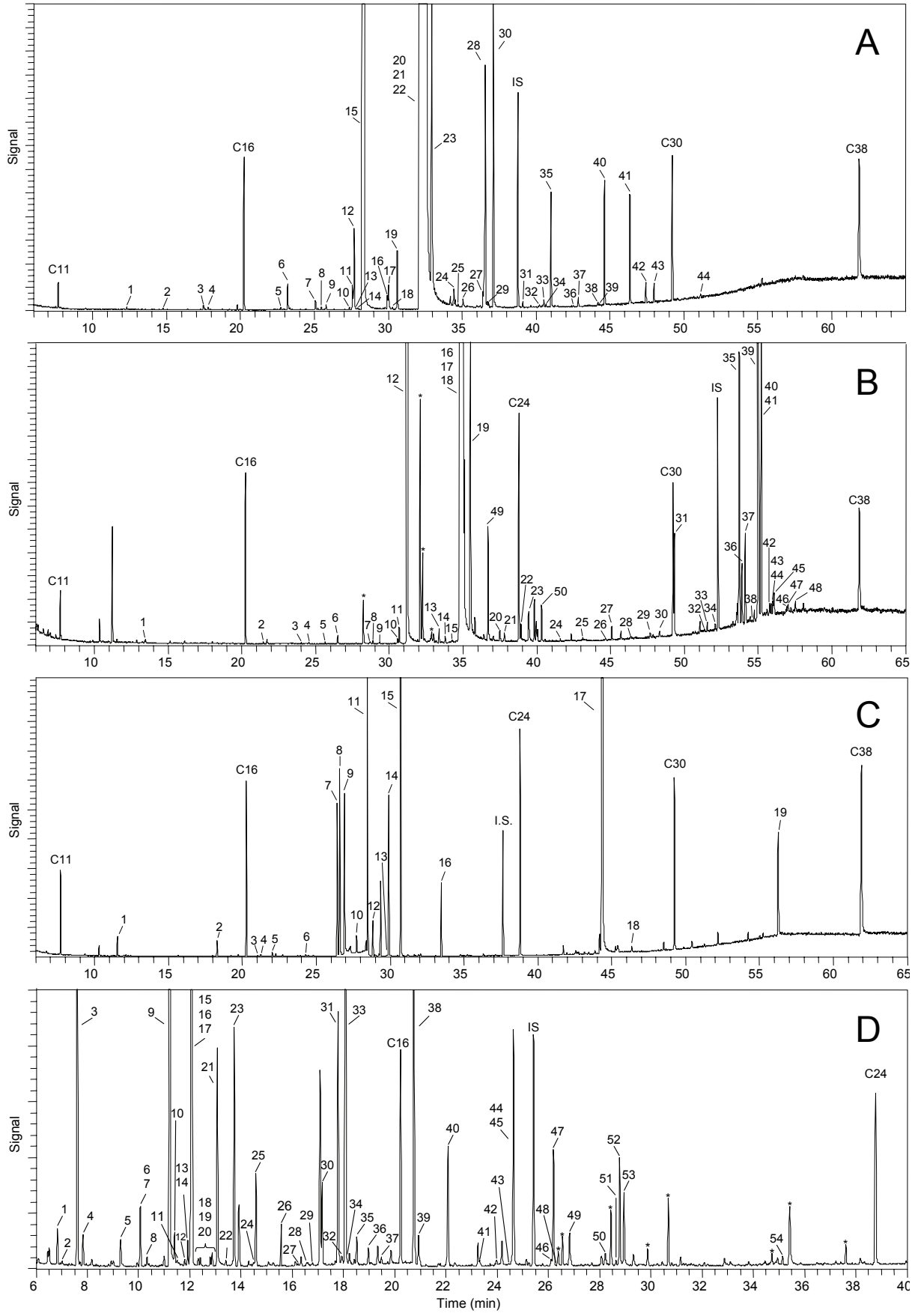

Figure 8: GC/MS total ion current chromatograms of metabolite profiling fractions (A) I (major lipids), (B) II (minor lipids), (C) III (sugars and sugar alcohols), and (D) IV (acids and amino acids). The peak numbers refer to the numbers in Tables 2 and 3.

employed to protect α-ketoacids from enolization and decarboxylation before trimethylsilylation.

The metabolite profiling methodology used in this study was originally developed for rice (Frenzel *et al.* 2002; Frank *et al.* 2007). Only small changes of the protocol were needed to adapt the metabolite profiling methodology from rice to the maize matrix. Compositional differences between rice and maize, e.g. higher levels of carbohydrates, free amino acids and lipids in maize kernels (Scherz and Senser, 2000), were taken into account by changing the amounts of reagents and solvents used in the derivatization steps or by reducing the aliquot of fraction IV employed for further sub-fractionation of the polar extract.

The four fractions are analyzed by GC/MS; Figure 8 provides an example of the respective total ion current chromatograms. A total of approximately 300 distinct analytes were detected. Comparison of mass spectral data and retention times to those of reference compounds or to literature data resulted in the identification of 167 compounds (Tables 2 and 3). The mean recovery determined for selected compounds was 89%; it ranged from 72% (β-sitosterol) to 98% (stearic acid methyl ester) for non-polar compounds and from 64% (fumaric acid) to 106% (sucrose) for polar compounds (Table 4). For selected representatives from the different chemical classes the intra-laboratory repeatability of the metabolite profiling approach was less than 10% RSD (Table 4), thus being in accordance to the data reported for rice (Frenzel et al., 2002).

In contrast to comparable metabolite profiling methodologies, additional fractionation steps for both lipophilic and polar compounds have been implemented. These sub-fractionations are time-consuming; the total work-up of a set of 12 samples required approximately 8 h. In particular, the sub-fractionation of the polar extract enables the separation of free sugars present as major low molecular weight compounds from minor constituents such as organic acids and amino acids. In addition, the fractionation allows an appropriate choice of reagents to ensure effective derivatization (Frenzel *et al.*, 2002). In the present work the sub-fractionation also facilitated the analysis of the contribution of metabolites from different chemical classes to the observed PCA separations.

Table 2: Compounds identified in fractions I (major lipids) and II (minor lipids)

no.	Compound	ident[a]	no.	compound	ident[a]	no.	compound	ident[a]
fraction I								
saturated FAME[b]			*unsaturated FAME*			*hydrocarbons*		
1	10:0	A	7	15:1 (10Z)	A	2	14	A
4	12:0	A	11	16:1	C	3	15	A
6	14:0	A	12	16:1 (9Z)	A	5	17	A
9	15:0	A	13	16:1 (9E)	A	8	18	A
15	16:0	A	17	17:1 (9Z)	A	14	19	A
19	17:0	A	22	18:1 (9Z)	A	18	20	A
23	18:0	A	24	19:1 (10Z)	C	25	22	A
26	19:0	A	28	20:1 (11Z)	A	29	23	A
30	20:0	A	33	22:1 (11Z)	A	34	25	A
31	21:0	A	38	24:1 (15Z)	A	36	26	A
35	22:0	A	10	16:2	C	39	27	A
37	23:0	A	16	17:2	D	41	squalene	A
40	24:0	A	21	18:2 (9Z, 12Z)	A	42	cholestane	C
43	26:0	A	27	20:2 (11Z, 14Z)	A			
44	28:0	A	32	22:2 (13Z, 16Z)	A			
			20	18:3 (9Z, 12Z, 15Z)	A			
fraction II								
free fatty acids[c]			*fatty alcohols*[c]			*sterols/stanols*[c]		
1	9:0	A	9	16:0	A	34	cholesterol	A
2	12:0	A	14	18:0	A	35	campesterol	A
3	13:0	A	15	phytol	A	36	campestanol	A
6	14:0	A	21	20:0	A	37	stigmasterol	A
8	15:0	A	24	22:0	A	38	Δ7-campestenol	E
10	16:1	C	27	24:0	A	39	β-sitosterol	A
11	16:1 (9Z)	A	30	26:0	A	40	sitostanol	A
12	16:0	A	32	28:0	A	41	Δ5-avenasterol	A
13	17:0	A	47	32:0	D	42	gramisterol	F
16	18:3 (9Z, 12Z, 15Z)	A				43	Δ7-stigmastenol	F
17	18:2 (9Z, 12Z)	A	*hydroxy FAME*[b,c]			44	cycloartenol	A
18	18:1 (9Z)	A	49	12-OH 18:1 (9Z)	A	45	Δ7-avenasterol	F
19	18:0	A	50	9,12-OH 18:0	G	46	24-methylene-cycloartanol	A
20	19:0	A						
22	20:1 (11Z)	A	*phenolic compounds*[c]			48	citrostadienol	F
23	20:0	A	4	methyl *p*-hydroxy-cinnamate	A			
25	22:0	A				*tocopherols*[c]		
26	23:0	A	5	methyl 3-methoxy-cinnamate	C	29	δ-tocopherol	A
28	24:0	A				31	γ-tocopherol	B,C
			7	methyl ferulate	A	33	α-tocopherol	A

[a] Identification according to
A mass spectral data and retention times of reference compounds
B mass spectral data and retention index of Golm Metabolome Database (Kopka *et al.*, 2005)
C mass spectral data of NIST02 mass spectral library (Ausloos *et al.*, 1999)
D mass spectral data
E according to (Xu and Godber, 1999)
F according to (Kamal-Eldin *et al.*, 1992)
G according to (Meyna, 2005; Miller, 1982)

[b] fatty acid methyl esters

[c] TMS derivatives of respective compound

Table 3: TMS[a] derivatives of compounds identified in fractions III (sugars and sugar alcohols) and IV (acids, amino acids and amines)

no.	compound[a]	ident[b]	no.	compound[a]	ident[b]
fraction III			*fraction IV*		
sugars and sugar alcohols			*amino acids and amines*		
1	glycerol	A	3	alanine	A
2	erythritol	A	4,18	glycine	A
3,4,5	arabinose	A	5,26	β-alanine	A
6	ribitol	A	6	valine	A
7,8,9	fructose	A	7	norvaline	A
10,12	galactose	A	10	leucine	A
11,15	glucose	A	11	ethanolamine	A
13	mannitol	A	13	alloisoleucine	A
14	sorbitol	A	15	isoleucine	A
16	*myo*-inositol	A	16	proline	A
17	sucrose	A	23	serine	A
18	trehalose	A	25	threonine	A
19	raffinose	A	27	homoserine	A
			31	pyroglutamic acid	A
fraction IV			32	methionine	A
Acids			33	aspartic acid	A
1	lactic acid	A	36	5-hydroxynorvaline	C
2	hydroxyacetic acid	A	37	threonic acid	A
8	4-hydroxybutyric acid	C	38	glutamic acid	A
9	phosphoric acid	A	39	phenylalanine	A
12	maleic acid	A	40	asparagine	A
14	4-aminobutyric acid	A	43	putrescine	A
17	succinic acid	A	44	glutamine	A
20	glyceric acid	A	46	citrulline	A
21	fumaric acid	A	48	ornithine	A
22	pyrrole-2-carboxylic acid	A	50	histidine	A
24	glutaric acid	A	52	lysine	A
28	2-piperidinecarboxylic acid	C	53	tyrosine	A
29	β-aminoisobutyric acid	A	54	tryptophan	A
30	malic acid	A			
34	cinnamic acid	A	*fraction IV*		
35	γ-aminobutyric acid	A	*others*		
41	α-aminoadipic acid	A	19	2,4-hydroxy-pyrimidine	C
42	cis-aconitic acid	B,C	49	adenine	A
45	3-glycerophosphoric acid	B,C			
47	citric acid	A			
51	*p*-cumaric acid	A			

[a] Metabolites identified as persilylated derivates

[b] Identification according to

A mass spectral data and retention times of reference compound

B mass spectral data and retention index of Golm Metabolome Database (Kopka *et al.*, 2005)

C NIST02 mass spectral library (Ausloos *et al.*, 1999)

Table 4: Repeatability of the metabolite profiling methodology and mean recoveries (n = 3) calculated for selected compounds

compound	RSD [%][a]	mean recovery [%]	compound	RSD [%][a]	mean recovery [%]
triglycerides (fraction I)[b]			*sugars (fraction III)*[c]		
16:0	7	97	fructose	4	91
18:0	3	99	glucose	3	104
20:0	2	91	galactose	4	96
			sucrose	3	106
hydrocarbons (fraction I)			raffinose	2	94
squalene	6	97			
tricosane	5	-	*sugar alcohols (fraction III)*[c]		
			sorbitol	2	96
free fatty acids (fraction II)[c]			inositol	2	97
16:0	6	94			
18:0	9	85	*amino acids (fraction IV)*[c]		
			alanine	1	78
fatty alcohols (fraction II)[c]			asparagine	11	85
octadecanol	11	82	aspartic acid	2	79
hexacosanol	5	-	glutamine	10	81
			glycine	7	79
sterols (fraction II)[c]			lysine	5	77
campesterol	1	-			
β-sitosterol	1	72	*organic acids (fraction IV)*[c]		
stigmasterol	3	98	citric acid	2	-
			fumaric acid	14	64
tocopherols (fraction II)[c]			malic acid	6	80
α-tocopherol	8	-	lactic acid	4	71
δ-tocopherol	8	92			
γ-tocopherol	7	-	*internal standards*		
			tetracosane	-	89
			5α-cholestan-3β-ol	-	92
			phenyl-β-D-	-	106
			p-chlorophenylalanine	-	94

[a] RSD: relative standard deviation (n = 3)

[b] Metabolites detected as fatty acid methyl esters

[c] Metabolites detected as persilylated derivatives

4.2 Influence of Genetics and Environment

4.2.1 Introduction

When investigating the impact of genetic background, environmental conditions or agronomic practices on maize seeds, a comparative analysis of metabolites is of particular interest as their levels can be regarded as the final response of an organism to all processes regulating metabolism (Fiehn, 2002). Recent studies on metabolic changes during growth and development (Seebauer *et al.*, 2004), the influence of environment and farming practice (Harrigan *et al.*, 2007b; Harrigan *et al.*, 2007c), and the impact of genetic background and growing seasons (Reynolds *et al.*, 2005; Ridley *et al.*, 2002) have demonstrated the importance of these factors on maize metabolite levels. In addition to these targeted approaches, unbiased metabolite profiling techniques proved to be powerful tools for the analysis of complex plant matrices (Castro and Manetti, 2007; Fiehn *et al.*, 2000; Lozovaya *et al.*, 2006; Roessner *et al.*, 2000). These techniques aim at extracting, detecting, identifying and quantifying a broad spectrum of compounds to provide a deeper insight into complex biological systems (Fiehn, 2001). Recently, investigation of low phytic acid mutants of rice and maize by means of metabolite profiling contributed to the characterization and the classification of different types of mutations (Frank *et al.*, 2007; Hazebroek *et al.*, 2007).

The aim of this study was the application of a metabolite profiling approach to the identification and the assessment of a broad spectrum of maize constituents from different chemical classes. Four maize cultivars cultivated within three growing seasons should be differentiated by comparison of metabolite profiles covering both polar and lipophilic compounds. An increased understanding of metabolic variation should be achieved by assessment of samples from four different growing locations within three years. Evaluation of results should be performed by a series of statistical approaches, including Principal Component Analysis (PCA), pairwise comparison and analysis of variance (ANOVA). This should help to demonstrate the potential of untargeted metabolite profiling to evaluate the impact of genetics and environment on maize grain composition.

4.2.2 Sample Sets

The study was based on two experimental set-ups to evaluate the impact of genetic background, growing location and growing season on the metabolite profiles of maize kernels. The first sample set comprised four cultivars which were grown at one location (Frankendorf, Bavaria). On the basis of their maturity behavior, cultivars belonging to early (ES Shorty), medium (Lukas) and late (Flavi, Pontos) maturity groups were selected (Table 5). In general, *early* and *medium* cultivars mature within a shorter period of time and are suitable for cultivation at farming sites with a short growing season. In contrast, *late* cultivars result in higher yields but require rather warm climatic conditions and longer growing periods to achieve this effect.

Table 5: Growing parameters[a] for maize cultivars (ES Shorty, Lukas, Flavi, Pontos) in seasons 2004-2006 at location Frankendorf, Bavaria.

	ES Shorty	Lukas	Flavi	Pontos
Maturity index[b]	210	240	260	270
Dry matter (%)				
2004	63.6	63.8	63.1	62.1
2005	65.8	62.6	61.8	61.5
2006	64.1	-[c]	-	63.9
Height (cm)				
2004	253	283	305	257
2005	272	305	312	265
2006	252	256	-	248
Yield (dt/ha)				
2004	138	156	146	141
2005	141	143	133	118
2006	116	125	-	122

[a] Average values calculated from single analyses of the field replicates. Accordingly, the data were not sufficient for assessment of statistical significance of differences between mean values.

[b] According to German maturity classification *Reifezahl*. A difference of 10 expressed as *Reifezahl* is equivalent to a difference of 1% expected dry matter. Lower values mean higher dry matter.

[c] no data available

If cultivars differing in maturity behavior are cultivated under the same climatic conditions—as performed in this study—their kernels will differ in dry matter content, which is expected to be higher in early and lower in late cultivars (Bundessortenamt, 2008). In addition to maturity type, the selected cultivars differed

in expected plant heights: ES Shorty and Pontos are generally short, Lukas and Flavi taller varieties.

The second sample set consisted of one cultivar (Amadeo, medium maturity index) grown at four different locations in Bavaria, Germany (Table 6). When selecting the growing sites, three locations (Thann, Reith, Mittich) with different types of loamy soil (sandy loam, slicky loam, clay loam) and one location (Strassmoos) with very sandy soil which exhibits a limited water-retaining capacity, were chosen. Regarding the climatic conditions, the location Mittich was included as it is known to have rather warm weather compared to the other three sites (IPS, 2009). Both sample sets were investigated in three consecutive growing seasons.

Table 6: Growing parameters[a] for maize cultivar Amadeo[b] in seasons 2004-2006 at four locations in Bavaria (Thann, Reith, Mittich, Strassmoos)

	Thann	Reith	Mittich	Strassmoos
Soil type[c]	sandy loam	silty loam	clay loam	loamy sand
Dry matter (%)				
2004	68.2	67.6	70.9	65.8
2005	67.4	68.2	71.7	66.5
2006	71.9	69.3	72.4	69.9
Height (cm)				
2004	303	300	300	275
2005	288	287	285	262
2006	290	276	315	295
Yield (dt/ha)				
2004	127	136	99	158
2005	123	132	114	136
2006	124	131	133	110

[a] Average values calculated from single analyses of the field replicates. Accordingly the data, were not sufficient for assessment of statistical significance of differences between mean values.

[b] Reifezahl: 230

[c] Soil types were defined according to (Wendland *et al.*, 2007)

4.2.3 Differentiation of Cultivars

Results. Metabolite profiling data obtained for the four cultivars grown at one location in 2004 were subjected to statistical analysis via Principal Component Analysis (PCA) to determine the major sources of variation within the dataset. On the basis of the data from all four fractions, a clustering of the cultivars, with Pontos being clearly separated on the first principal component score, was observed (Figure 9A). There was no obvious sub-clustering owing to field (n = 3) or analytical (n = 3) replicates.

PCAs on the basis of metabolite data from the individual fractions demonstrated that both lipid fractions I and II could be used to separate Pontos from the other cultivars (Figures 9B and 9C). The mirroring of PC1 between fraction I and II indicated that overall levels of major lipids (i.e. triglycerides) are lower and levels of minor lipids are higher in Pontos. The polar fractions III (sugars and sugar alcohols) and IV (acids, amino acids and amines) showed no or only weak differentiation between Pontos and the other cultivars (Figures 9D and 9E).

To identify the sources of variation between Pontos and the other cultivars, loading scores of the first principal component were examined (Figure 10). Subsequently, in each fraction the metabolites with the highest absolute loading scores were quantified on the basis of relative signals (Figures 11 and 12).

Levels of fatty acid methyl esters in fraction I were decreased in Pontos compared to the other three cultivars (Figure 11A). The levels of the methyl esters of palmitic, stearic, oleic and linoleic acid, reflecting more than 95% of the oil content in maize, were not among the ten highest loading scores (Figure 10A); however, they were also significantly lower (58–73%) in Pontos ($p < 0.05$). Similarly, the levels of the methyl esters of ricinoleic acid and of the tentatively identified 9,12-dihydroxy linoleic acid (which eluted in fraction II owing to their increased polarity due to the additional hydroxy group), were decreased in Pontos compared to the other cultivars (Figure 11B). In contrast, the free fatty acids detected in fraction II showed higher levels in cultivar Pontos (Figure 11B).

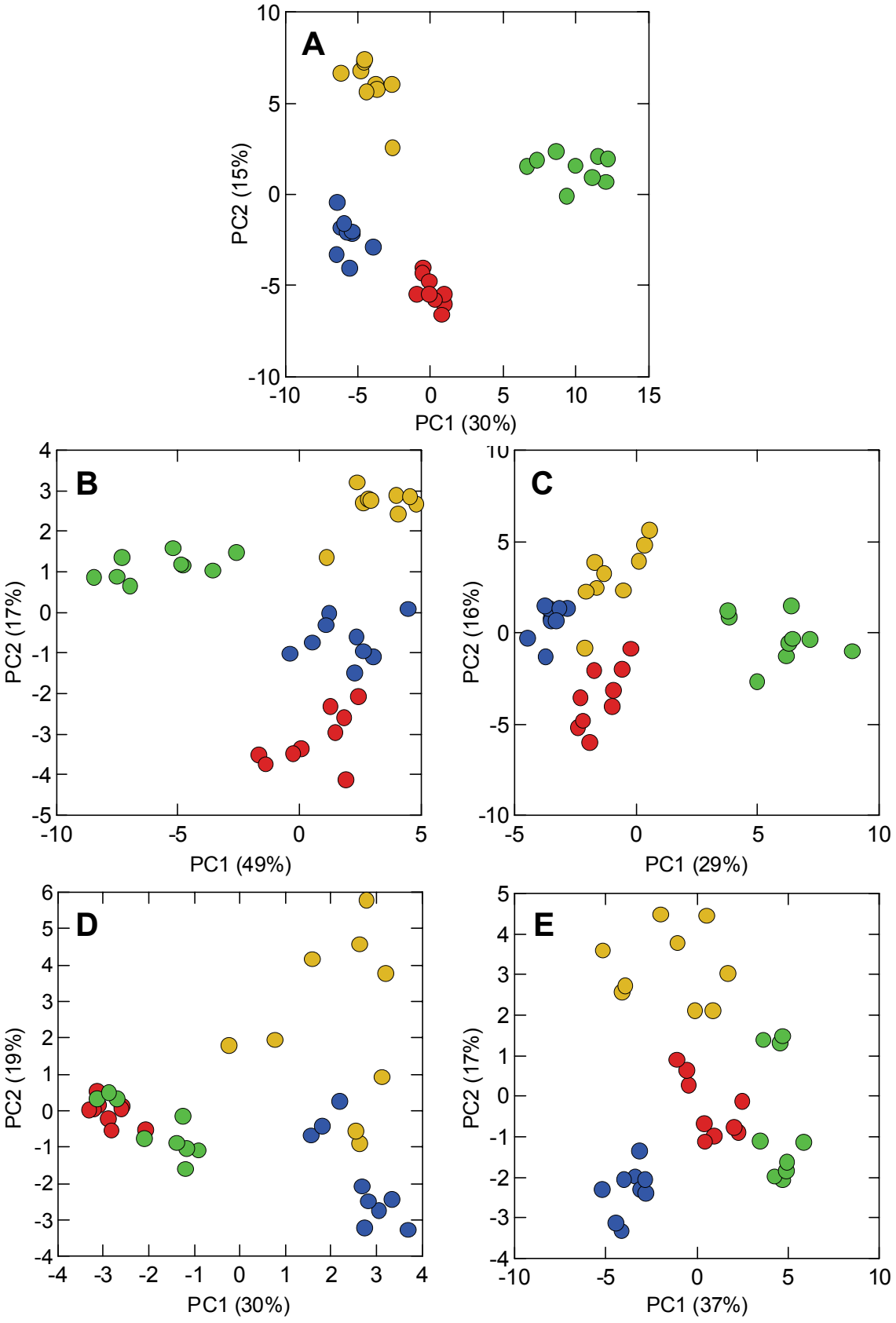

Figure 9 Principal component analysis of metabolite profiling data from (A) combined fractions I-IV (I: major lipids, II: minor lipids, III: sugars/sugar alcohols and IV: acids/amino acids/amines) and (B) individual fractions I, (C) II, (D) III and (E) IV. Triplicate analysis of four cultivars (●: Flavi, ●: Lukas, ●: Pontos, ●: ES Shorty), growing season 2004.

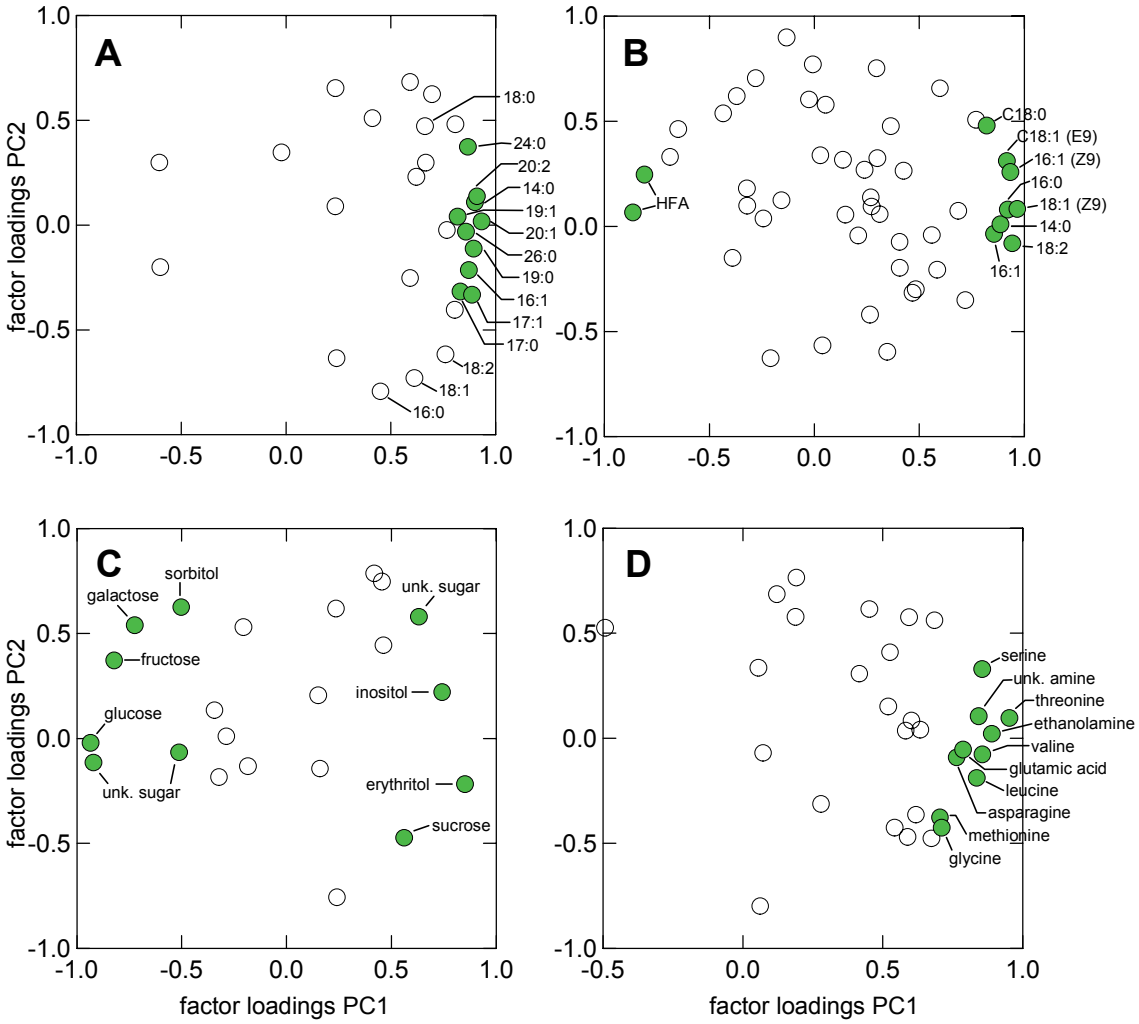

Figure 10 Factor loading scores of principal component 1 and 2 of principal component analysis of metabolite profiling data of four cultivars grown in season 2004 from (A) fraction I: major lipids, (B) fraction II: minor lipids, (C) fraction III: sugars/sugar alcohols and (D) fraction IV: acids/amino acids/amines. Black circles (●) representing compounds with 10 highest absolute loading scores on PC1.

Differences in the levels of sugars and sugar alcohols (fraction III, Figure 5A) and acids, amino acids and amines (fraction IV, Figure 12B) between Pontos and the other cultivars were less pronounced than those observed for the lipids. In fraction III none of the differences was statistically significant ($p < 0.05$); in fraction IV the amino acids serine, threonine, glycine, leucine and methionine showed significantly higher levels in Pontos.

The PCAs determined for the four cultivars in the years 2005 and 2006 (Figure 13A and 13B) were different from the results obtained in 2004. The clustering patterns varied from year to year and combining the data from all three growing seasons did not allow a separation of cultivars (Figure 13C) but rather revealed a clear clustering according to years (Figure 13D).

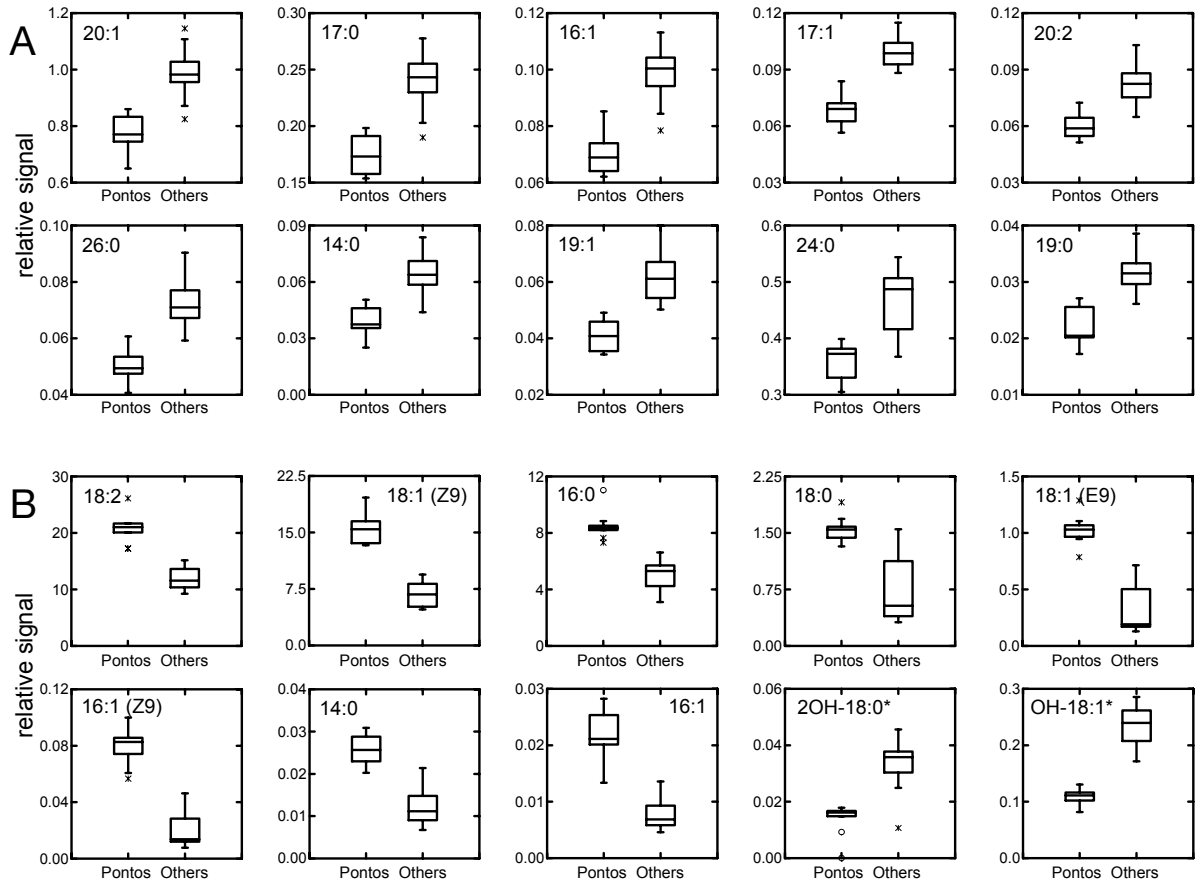

Figure 11 Comparison of cultivar Pontos with cultivars Lukas, Flavi and ES Shorty by semiquantified levels of (A) fatty acid methyl esters, fraction I and (B) free fatty acids and hydroxy fatty acid methyl esters (*), fraction II, in growing season 2004. Relative signals calculated on the basis of the respective internal standard (fraction I: tetracosane, fraction II: 5-α-cholestane-3-β-ol).

The extent of differences between cultivars was further assessed by pairwise comparisons. A total of 1958 statistical comparisons (6 pairwise cultivar comparisons x 3 growing seasons x 103–120 analytes) were performed (Table 7). In the growing season 2004 the number of statistically significant ($p < 0.05$) differences between Pontos and the other three cultivars (20–29%) was higher than the number of differences between these cultivars (12–17%). This is in agreement with the results obtained by PCA, where Pontos showed a strong separation in 2004 (Figure 9A). In the following season 2005, in which clustering of the cultivars by PCA was less pronounced (Figure 13A), the numbers of differences are more equally distributed between the comparisons with a mean value of only 15%.

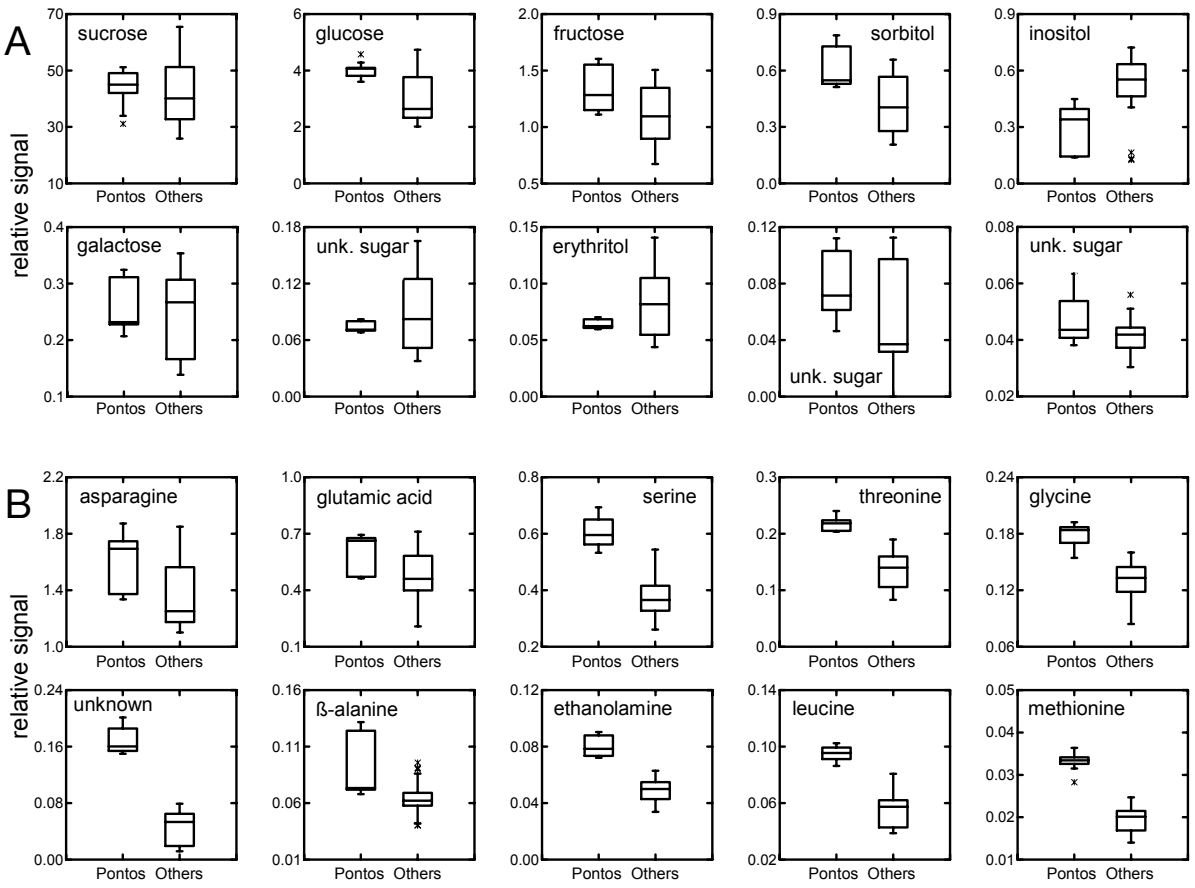

Figure 12 Comparison of cultivar Pontos with cultivars Lukas, Flavi and ES Shorty by semiquantified levels of (A) sugars and sugar alcohols, fraction III and (B) amino acids and amines, fraction IV, in growing season 2004. Relative signals calculated on the basis of the respective internal standard (fraction III: phenyl-β-D-glucopyranoside, fraction IV: p-chloro-L-phenylalanine).

In 2006 the similarity of metabolite patterns of Pontos and Flavi (Figure 13B) is reflected by a relatively small number of differences (14%), whereas the clear differentiation from the other two cultivars results in high numbers of statistically significant differences ranging from 24% to 31% (Table 7).

Pairwise comparisons on the basis of 1346 statistical comparisons (3 pairwise year comparisons x 4 cultivars x 105–123 analytes) of growing seasons within the same sample set revealed 37 % statistically significant differences between the seasons 2004 and 2006 but only 16 % differences between 2005 and 2006 (Table 7). These results are in good agreement with the separation of years by PCA (Figure 13D), where the years 2004 and 2006 could be separated on both PC1 and PC2, whereas years 2005 and 2006 were separated only on the second PC.

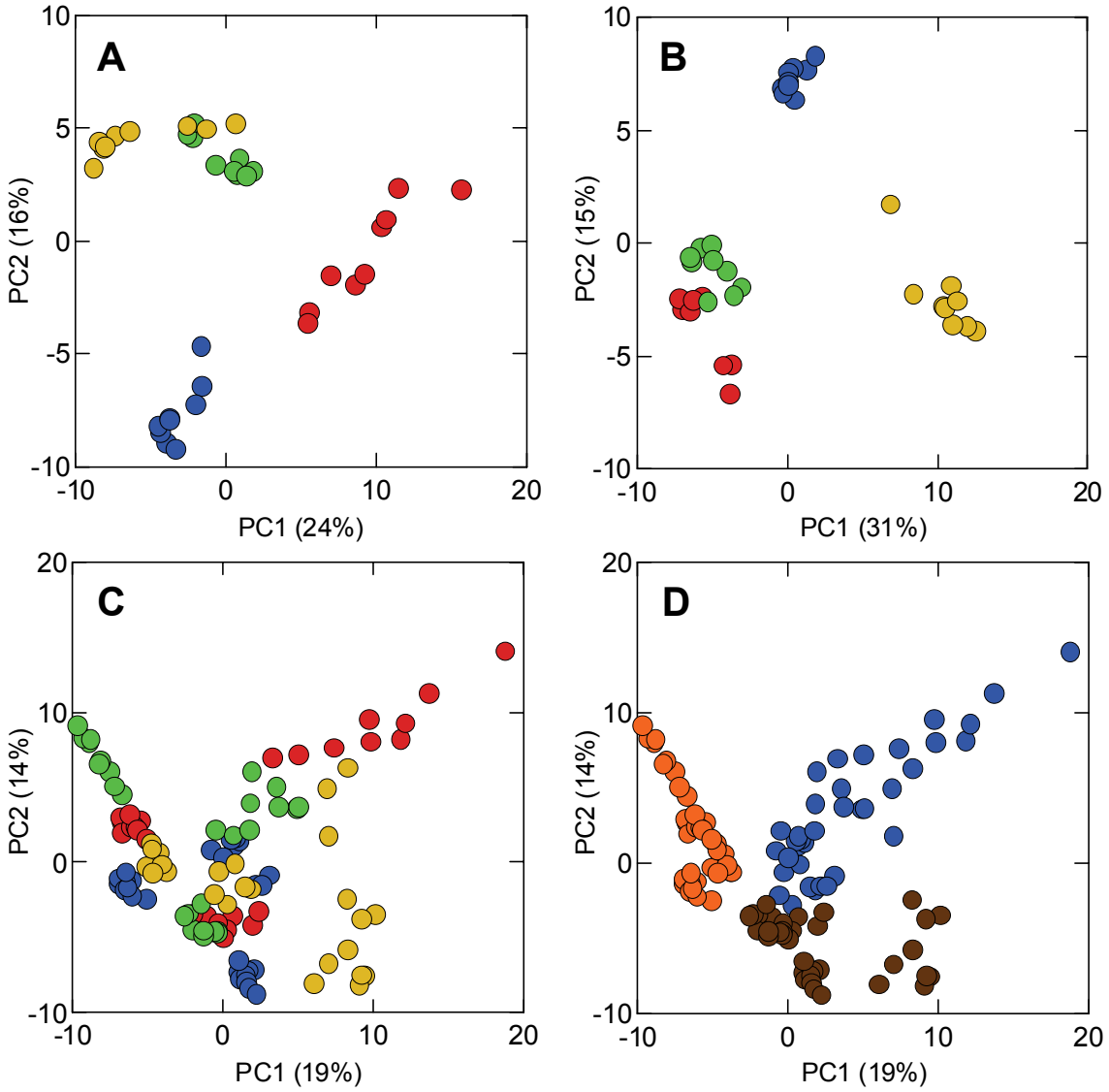

Figure 13 Principal component analysis of metabolite profiling data from fractions I-IV in growing seasons 2005 (A), 2006 (B) and 2004-2006 (C,D) at farming location Frankendorf, Bavaria. Triplicate analysis of four cultivars (A,B and C) ●: Flavi, ●: Lukas, ●: Pontos, ●: ES Shorty, (D) ●: 2004, ●: 2005, ●: 2006.

Discussion. Assessment of metabolite profiles of four different maize cultivars by PCA revealed a strong separation of one cultivar (Pontos) in the first growing season, which could be ascribed to different metabolite levels in the lipid fractions. A negative correlation between the levels of triglycerides and free fatty acids suggests an interplay between the metabolism of storage fats and their potential precursors (Voelker and Kinney, 2001), indicating that cultivar Pontos was in a less mature state at the date of harvest (Daftary and Pomeranz, 1965; Weber, 1969).

Table 7: Pairwise comparisons between cultivars, farming locations and growing seasons

	comp.[a]	differences[b] total	[%]		comp.	differences total	[%]
cultivars				**locations**			
2004				*2004*			
Pontos vs. Lukas	109	31	28	Strassmoos vs. Thann	116	25	22
Pontos vs. Shorty	111	32	29	Strassmoos vs. Reith	115	22	19
Pontos vs. Flavi	105	21	20	Strassmoos vs. Mittich	116	35	30
Lukas vs. Shorty	106	16	15	Thann vs. Reith	107	3	3
Lukas vs. Flavi	103	12	12	Thann vs. Mittich	112	14	13
Shorty vs. Flavi	105	18	17	Reith vs. Mittich	109	13	12
total	639	130	20	total	675	112	17
2005				*2005*			
Pontos vs. Shorty	104	18	17	Strassmoos vs. Thann	112	22	20
Pontos vs. Lukas	103	15	15	Strassmoos vs. Reith	105	12	11
Pontos vs. Flavi	109	14	13	Strassmoos vs. Mittich	107	16	15
Lukas vs. Shorty	106	16	15	Thann vs. Reith	107	15	14
Lukas vs. Flavi	111	16	14	Thann vs. Mittich	109	23	21
Shorty vs. Flavi	113	16	14	Reith vs. Mittich	104	7	7
total	646	95	15	total	644	95	15
2006				*2006*			
Pontos vs. Lukas	111	27	24	Strassmoos vs. Thann	110	2	2
Pontos vs. Shorty	117	35	30	Strassmoos vs. Reith	110	9	8
Pontos vs. Flavi	102	14	14	Strassmoos vs. Mittich	113	3	3
Lukas vs. Shorty	120	27	23	Thann vs. Reith	112	13	12
Lukas vs. Flavi	108	32	30	Thann vs. Mittich	113	4	4
Shorty vs. Flavi	115	36	31	Reith vs. Mittich	110	4	4
total	673	171	25	total	668	35	5
total	1958	396	20	*total*	1987	242	12
growing seasons within cultivars				**growing seasons within locations**			
2004 vs. 2005				*2004 vs. 2005*			
Shorty	107	11	10	Strassmoos	110	25	23
Lukas	112	27	24	Thann	111	18	16
Flavi	114	30	26	Reith	108	9	8
Pontos	109	29	27	Mittich	110	15	14
total	442	97	22	total	439	67	15
2004 vs. 2006				*2004 vs. 2006*			
Shorty	123	46	37	Strassmoos	113	30	27
Lukas	115	44	38	Thann	116	14	12
Flavi	107	33	31	Reith	107	4	4
Pontos	116	47	41	Mittich	115	16	14
total	461	170	37	total	451	64	14
2005 vs. 2006				*2005 vs. 2006*			
Shorty	116	12	10	Strassmoos	108	8	7
Lukas	110	17	16	Thann	117	25	21
Flavi	112	23	21	Reith	108	8	7
Pontos	105	17	16	Mittich	112	6	5
total	443	69	16	total	445	47	11
total	1346	336	25	*total*	1335	178	13

[a] Number of compounds/comparisons for each pair

[b] All statistical comparisons made by Student's *t*-test at the 5% level of significance ($p < 0.05$)

Levels of free amino acids in the endosperm of kernels have been reported to decrease significantly by the end of the vegetation period (Arruda *et al.*, 1978; Duvick, 1952; Hirel *et al.*, 2005; Miyanishi *et al.*, 1991) resulting in programmed cell death (Young and Gallie, 2000). Therefore, the higher levels of free amino acids support the argument that, in 2004, kernels of Pontos were less mature than those of the other cultivars at time of harvest. This is further supported by data on conventional agronomic performance traits shown in Table 4. Cultivar Pontos, with the highest maturity index, did not perform as well as the other cultivars due to sub-optimal growing conditions for this genotype in this particular trial region. The Pontos genotype is basically intended for farming locations with rather warm climate conditions.

The strong separation of cultivar Pontos from the other cultivars observed in 2004 was not repeated in the other years tested. In 2004 the temperature in the first two months of the growing season was 2–3 °C degrees lower than in the following seasons and in August 2004 the weather was hot and dry (IPS, 2009). These weather conditions would impact on the growth and development of cultivar Pontos in particular and are likely to be an important factor distinguishing the metabolite profile of Pontos in 2004. As Figure 6C shows, when data from all three growing seasons were combined the impact of the growing season on the metabolite profile of kernels is more pronounced than the influence of genetic background (cultivar). This crucial phenomenon may also explain differences seen when moving crop experiments from a lab environment to the field.

The number of statistically significant ($p < 0.05$) differences in metabolite levels between the four cultivars (20% in 2004, 15% in 2005 and 25% in 2006) was in the same order of magnitude as those determined for low phytic acid maize mutants. Application of a GC/MS-metabolite profiling approach similar to the one employed in the current study revealed 11–30% of the detected compounds (124 polar and lipophilic analytes) to be statistical significantly different ($p < 0.05$) between wild-type maize and low phytic acid maize mutants (Hazebroek *et al.*, 2007). Another study investigating the nutritional and metabolic profiles of different maize hybrids using targeted analysis approaches comprising 47 analytes, determined statistically significant differences ranging from 33% to 47% of total comparisons (Reynolds *et al.*,

2005). The slightly higher percentages observed in this study might be explained by the selection of maize hybrids, which were considered to be adapted to different parts of Europe (southern France, northern France, Germany, Italy).

4.2.4 Differentiation of Locations

Results. PCAs of metabolite data obtained for one maize variety (Amadeo, medium maturity index) cultivated at four locations in Bavaria are shown in Figure 14. In the growing season 2004 location Strassmoos showed a strong separation on the first principal component and location Mittich could be differentiated on the second PC (Figure 14A). In 2005 location Strassmoos showed again a clear separation on the first PC, but the clustering of the other locations changed (Figure 14B). The differentiation pattern observed in 2006 was less clear (Figure 14C). Combination of the data from all three seasons resulted in an overlap of clusters and did not result in a clear differentiation due to either location or growing season (Figure 14D). Investigations of further principal components (up to PC5) did also not reveal any clear separation.

For evaluation of differences due to farming location, a total of 1987 statistical comparisons (6 pairwise location comparisons x 3 growing seasons x 104–116 analytes) were performed. The amounts of statistically significant ($p < 0.05$) differences for locations ranged from 17% in 2004 to 5% in 2006, with a mean of 12% over three growing seasons (Table 7). The observed number of differences confirmed the clustering obtained by PCA (Figure 14). For example, in 2004 for the clearly separated location Strassmoos 19–30% statistically significant differences were determined compared to the other three locations, whereas the comparison of the locations Thann and Reith, for which the PCA clusters overlapped, resulted in only 3% statistically significant differences.

Comparison of growing seasons within locations resulted in 13% significant differences out of a total of 1335 comparisons (3 pairwise year comparisons x 4 locations x 107–117 analytes). The amounts of differences between growing seasons are in the same range as differences between farming locations. This reflects results obtained by PCA when data from all seasons were combined but no clear clustering could be observed due to either growing season or farming location (Figure 14D).

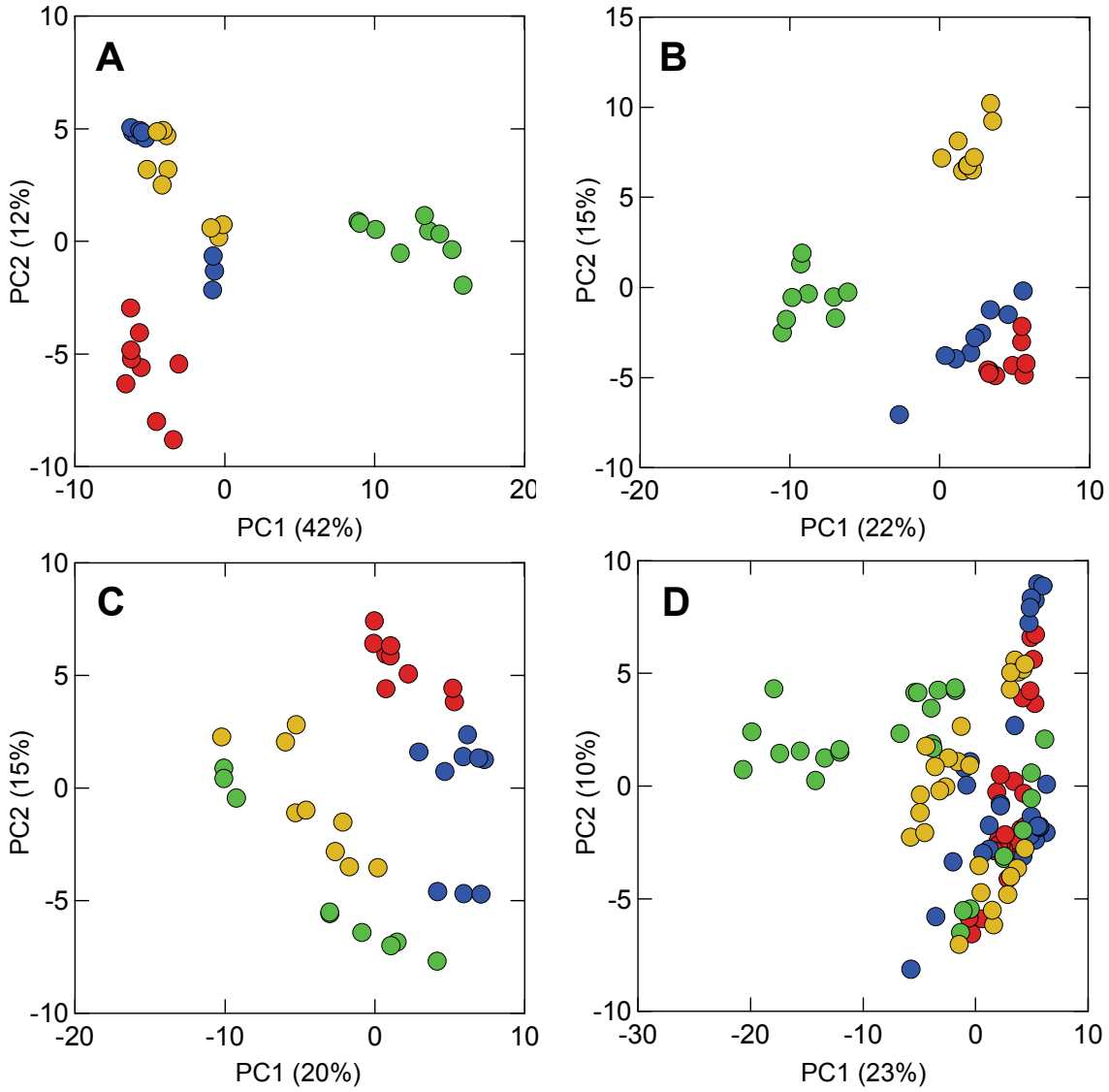

Figure 14 Principal Component Analysis of metabolite profiling data from fractions I–IV in growing seasons 2004 (A), 2005 (B), 2006 (C) and 2004-2006 (D). Triplicate analysis of cultivar Amadeo at four locations (●: Mittich, ●: Reith, ●: Strassmoos, ●: Thann).

ANOVA. To qualify the extent and distribution of significant differences found between cultivars and locations, probabilities (*p*-values) obtained by analysis of variance (ANOVA) were examined (Figure 15). To allow a better comparison of *p*-values of ANOVA for a large number of analytes the exponential distribution of *p*-values was transformed by logit (*p*). The first ANOVA was performed with data from the four cultivars grown in three consecutive seasons (2004–2006). Comparison of histograms in Figure 15A (cultivars) and 15B (growing seasons) shows more highly significant logit (*p*)-values < −3, i.e. *p*-values < 0.001 for the factor growing season, which is also the more prominent impact factor on metabolite profiles when data are

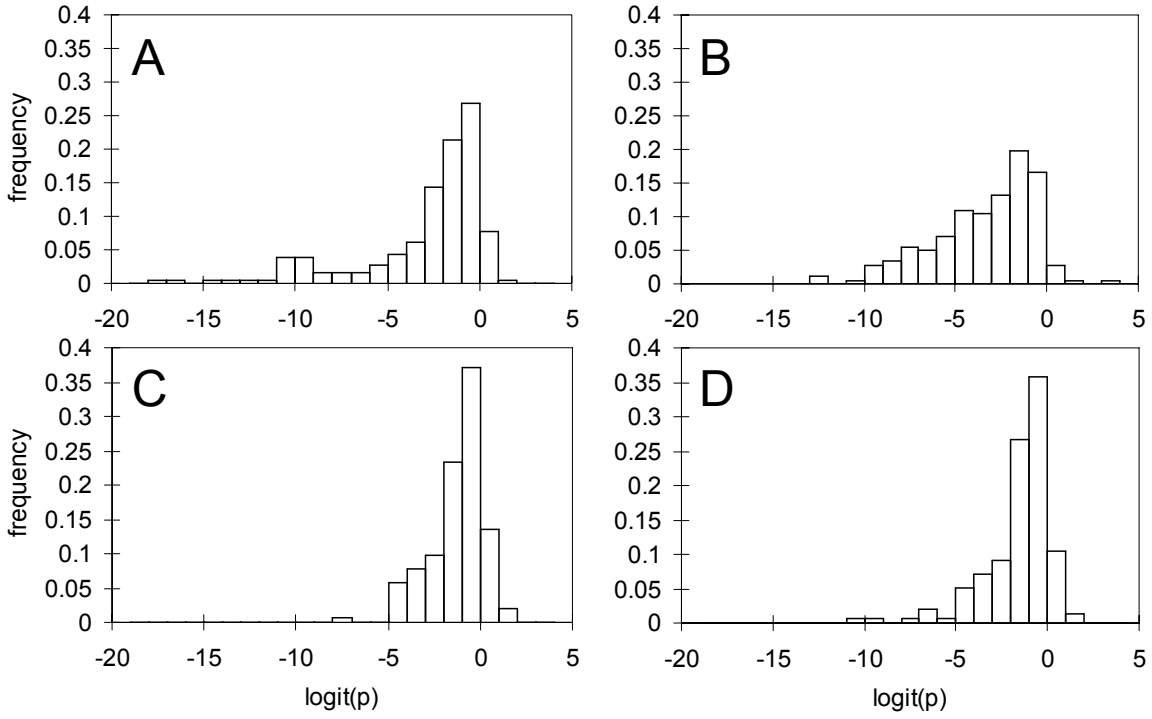

Figure 15 ANOVA of metabolite profiling data from fractions I-IV. First row: model including treatment (A) cultivar (ES Shorty, Lukas, Flavi, Pontos at location Frankendorf) and (B) growing season for cultivars; second row: model including treatments (C) location (Amadeo at locations Thann, Reith, Mittich, Strassmoos) and (D) growing season for farming locations. Histograms are showing frequency of logit (*p*) = log (*p*) − log (1−*p*); e.g. logit (*p*=0.5) = 0, logit (*p*=0.1) ≈ −1, logit (*p*=0.01) ≈ −2, logit (*p*=0.001) ≈ −3, etc.

evaluated by PCA (Figure 13D). A second ANOVA comprising data of the samples from four different locations (Figure 15C/D) revealed a more equal distribution for both factors farming site and growing season. This is in accordance with the results obtained by PCA where no clear separation was found if data from all three seasons were combined. In general, the distribution of *p*-values demonstrated that the differentiation due to farming locations (Figure 15C) was less significant than that due to cultivars (Figure 15A). The differentiation due to growing season appeared to be more pronounced within cultivars (Figure 15B).

Discussion. Metabolite profiles of samples grown at Strassmoos could be clearly separated by PCA from those grown at the other locations in two out of three growing seasons (Figure 14). This may be explained by differences in the soil compositions. Whereas Strassmoos has a very sandy soil with limited water-retaining capacity, the other locations were characterized by a more silty/clay loam soil. In

addition, differences in the local microclimate may influence the observed PCA separations. Details on these parameters have been provided in Table 1.

Pairwise comparisons and ANOVA between growing locations performed for one cultivar showed fewer statistically significant differences ($p < 0.05$) than the statistical assessment of cultivars grown at one farming location. This could lead to the conclusion that the impact of genetic background on metabolic variation is more pronounced than the influence of the environment; however, it has been shown that variation caused by environmental factors, e.g. site location, is dependent on the genotype grown (Reynolds *et al.*, 2005). The cultivar (Amadeo) grown at the four different locations has a medium maturity index, and is therefore equally suitable for all locations. This could also explain the smaller number of significant differences between years for this cultivar grown at different locations. Additional support is given by another study that highlighted the interaction of genetic and environmental background for many metabolites (Harrigan *et al.*, 2007b). In this survey a total of 58 metabolites were analyzed, of which 36% differed statistically significantly ($p < 0.05$) between maize inbreds crossed against two different testers, and 48% of statistically significant differences were due to the influence of the location.

4.2.5 Conclusions

The study demonstrates the suitability of the described extraction and fractionation approach for metabolite profiling of maize kernels. The combination with appropriate statistical tools enabled the evaluation of metabolic variation for maize cultivars differing in genetic and environmental background. The analysis of sub-fractions allowed the assessment of the contribution of metabolites from different chemical classes. The data obtained do not indicate that one of the fractions is generally more important; each of them may add information to explain genotype or environmental effects on crop compositions. Thus, the unbiased sub-fractionation as applied in this study is a technically demanding but useful approach providing additional value to metabolite profiling. The type of comparative datasets generated may serve as objective basis for crop assessment and the data confirm the potential of metabolite profiling to assist in breeding and farming approaches.

4.3 Influence of Input System (Conventional vs. Organic Farming)

4.3.1 Introduction

With the arising ecological awareness in the 1980s, farmers and consumers started to look for alternatives to conventional farming. Rather than intensively applying mineral fertilizers and relying on chemical plant protection, organic farming is based on minimal use of off-farm inputs and on ecologically friendly management practices (Lampkin *et al.*, 2000; Winter *et al.*, 2006). The share of organically farmed area has continuously increased over the last two decades in particular in Europe and North America (Willer and Kilcher, 2009). Surveys indicate that many consumers purchase organic foods because of the perceived health and nutrition benefits (Bourn and Prescott, 2002), although a recent systematic review found no evidence for a difference in nutritional quality between organically and conventionally produced foods (Dangour *et al.*, 2009).

The rising interest in this field is also reflected by an increased scientific activity; from 1993 to 2008 an 8-fold increase of scientific publications concerned with “organic farming” can be observed (SciFinder, 2009). Many of these publications deal with the impact of organic farming practice on soil parameters such as organic matter (Gong *et al.*, 2009; Herencia *et al.*, 2007), biodiversity and vitality (Mäder *et al.*, 2002) or pH (Zhang *et al.*, 2008). Others focus on the influence of input regimes and tillage systems (Fuentes *et al.*, 2009; Ghorbani *et al.*, 2008) on yield. From a food quality point of view, parameters such as protein content, nutrient levels (Jacob, 2007; Warman and Havard, 1998) and minerals (Langenkämper *et al.*, 2006) have been thoroughly investigated.

A more comprehensive approach for the assessment and evaluation of a broad spectrum of crop constituents, complementing the above-described targeted studies, is envisaged by application of the so-called *omics*-techniques. For example, the impact of different amounts and forms (organic, inorganic) of nitrogen supply on the gene expression level in the wheat endosperm have been investigated (Lu *et al.*, 2005). Many of the genes showing differential expression in this study are known to participate in nitrogen metabolism and storage protein synthesis. Other studies

involved proteomics approaches: Comparison of the protein compositions of potato tubers subjected to organic and mineral-based fertility management practices, respectively, suggested an increased stress response in organic farming (Lehesranta *et al.*, 2007). In wheat 16 *diagnostic* proteins with potential to afford a signature to prove authenticity of organic wheat were proposed (Zörb *et al.*, 2009).

In addition to transcriptomics and proteomics, metabolomics-based approaches should also be suitable to reflect the impact of different input systems on crops. Gas chromatography coupled with mass spectrometry (GC/MS) proved to be one of the most robust technologies for metabolite profiling (Kopka, 2006). At present, there is only one example for the application of this approach to organically farmed crops; the analysis of 52 polar metabolites in one wheat grain variety grown under organic and conventional farming practices, respectively, detected only moderate differences (Zörb *et al.*, 2006).

The aim of this study was to investigate the metabolite profiles of maize (*Zea mays*) grown conventionally and organically, respectively, using a methodology that recently was shown to be suitable to demonstrate variations in maize grain metabolite pools resulting from the interplay of environment, season and genotype (section 4.2). By analysis of three cultivars grown at two locations with different input systems and at a third location, where both organic and conventional farming were applied, the impact of the growing regime on the metabolite spectrum should be put into the context of natural variability.

4.3.2 Experimental Setup

Three cultivars (Amadeo, Lukas, Flavi) were used in the study to evaluate the impact of input systems on metabolite profiles of maize kernels differing in genetic background. They were grown in the seasons 2004 (experiment Ia) and 2005 (experiment Ib) at two locations in Bavaria, one (Frankendorf) with conventional and the other (Schönbrunn) with organic farming practice. The two locations were approximately 30 km apart. In order to minimize environmental influences, a further experiment (II) was designed in which the three cultivars were grown at one location (Scheyern) providing field plots for both conventional and organic farming at a distance of approximately 400 m (Figure 16).

The metabolite profiling methodology applied in this study is based on consecutive extraction of freeze-dried maize flour and subsequent sub-fractionation resulting in four fractions including (I) major lipids, (II) minor lipids such as free fatty acids and sterols, (III) sugars and sugar alcohols and (IV) organic acids, amino acids and amines. A total of approximately 300 distinct analytes were detected by GC/MS analysis. Comparison of mass spectral data and retention times to those of reference compounds or to literature data resulted in the identification of 167 compounds (Tables 2 and 3). Figure 17 provides an example of the respective total ion current chromatograms for cultivar Lukas grown at locations Frankendorf (conventional farming) and Schönbrunn (organic farming) in 2004.

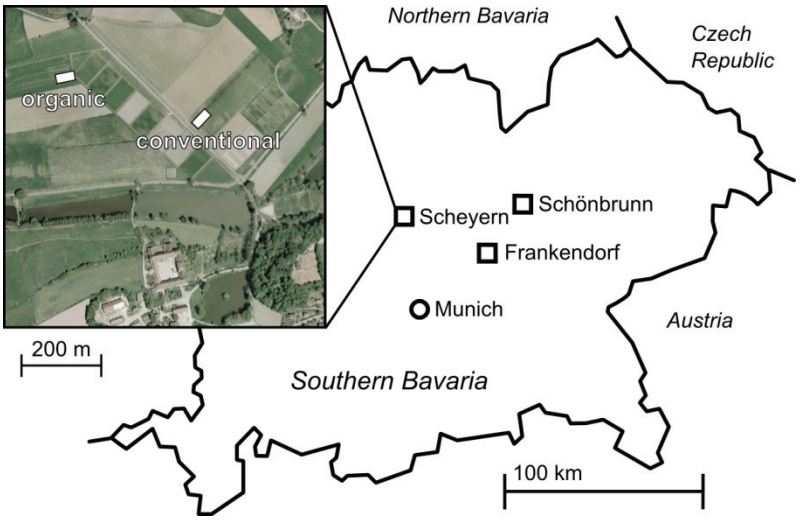

Figure 16 Geographical locations of the field trial sites Frankendorf, Schönbrunn and Scheyern. © Aerial photo: Bayerische Vermessungsverwaltung. © Cartography: Kober-Kümmerly+Frey, Köln

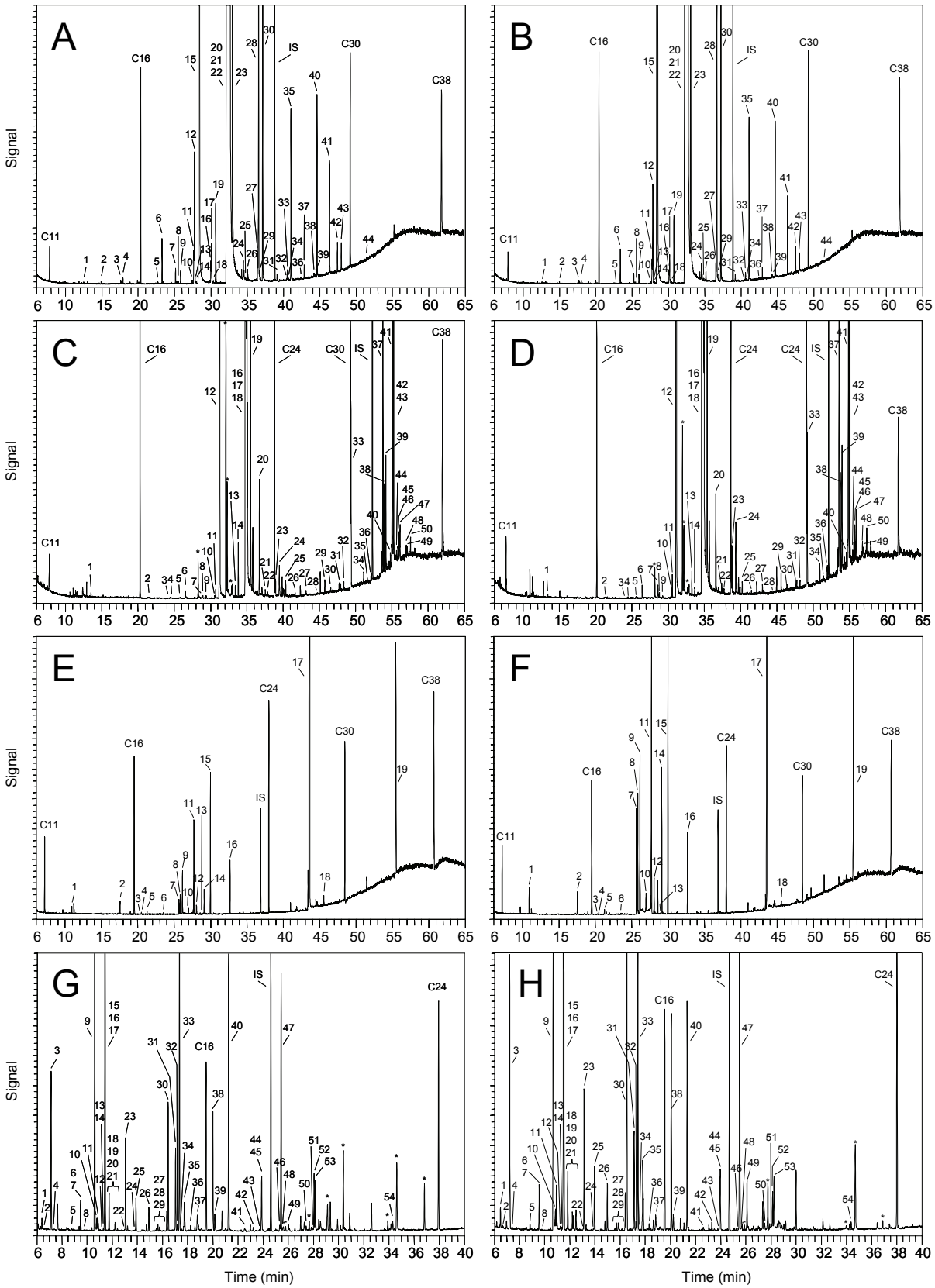

Figure 17 GC/MS total ion current chromatograms of metabolite profiling (A/B) fractions I (major lipids), (C/D) II (minor lipids), (E/F) III (sugars and sugar alcohols) and (G/H) IV (acids and amino acids) obtained by analysis of cultivar Lukas grown conventionally at location Frankendorf (left column) and organically at location Schönbrunn (right column) in season 2004. The peak numbers refer to the numbers in Tables 2 and 3.

4.3.1 Qualitative Assessment of Differences due to Farming Condition

4.3.1.1 Principal Component Analysis

Metabolite profiling data obtained for the three cultivars (Amadeo, Lukas, Flavi) grown in 2004 at the two locations with different input systems were subjected to statistical assessment via Principal Component Analysis (PCA) to determine the major sources of variation. On the basis of the data from all metabolites covered in fractions I-IV, a clear separation according to farming locations/input systems was observed on the first principal component accounting for 30% of the variation (Figure 18A). At the location Schönbrunn (organic farming) the three cultivars formed one cluster whereas at location Frankendorf (conventional farming) cultivar Lukas was differentiated on the second principal component (16% of the variation).

In the following season 2005 the clustering of the samples on the first two principal components (38% of the variation) was increasingly determined by differences between cultivars (Figure 18B). The effects of the farming location/input system were much less pronounced than in 2004.

An even clearer impact of the genetic background became obvious from the data obtained for the three cultivars grown under different input systems at the same location: As shown in Figure 18C, the cultivars showed quite distinct clusters on the first two principal components of the PCA, explaining 59% of the variation. However, only small differences were observed between the samples obtained by conventional and organic farming, respectively.

4.3.1.1 Analysis of Variance (ANOVA)

An analysis of variance (ANOVA) was performed for each of the three datasets to evaluate the number of differences due to locations/input systems and to genetic background. In 2004 the levels of a total of 125 compounds were compared; post hoc testing (Tukey's HSD, $p < 0.01$) revealed 29% to be statistically significantly different for locations/input systems and 23% different for genotype (Table 8). In agreement with the clustering seen in the PCA, in 2005 the number of the statistically significant differences for locations/input systems was much lower (13% of 127 compounds), whereas the differences due to the influence of cultivars increased to 32%.

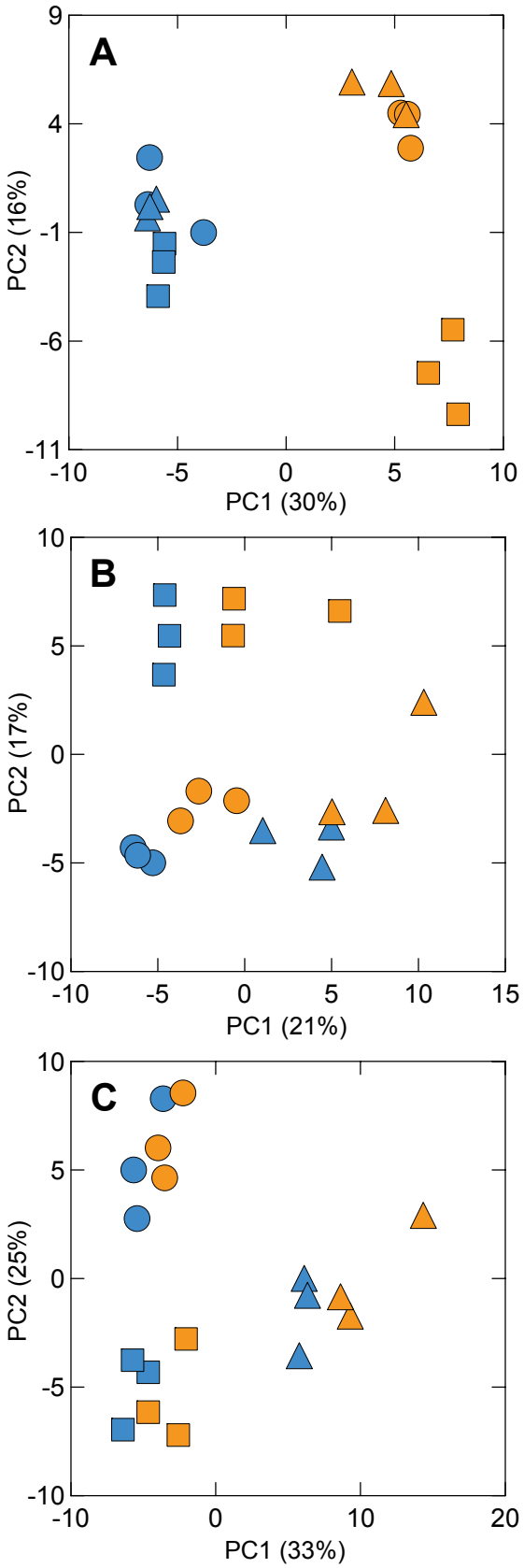

Figure 18 Principal Component Analysis of metabolite profiling data from fractions I-IV obtained by analysis of three maize cultivars (●: Amadeo, ■: Lukas, ▲: Flavi) grown at farming sites with different input regimes (●,■,▲: conventional, ●,■,▲: organic) in growing seasons (A) 2004 and (B) 2005 at locations Frankendorf (conv) and Schönbrunn (org) and (C) 2006 at location Scheyern (conv, org).

At location Scheyern, where both conventional and organic farming were applied at the same site, the number of differences due to input systems decreased further to only 11%. The clear separation of cultivars seen in this experiment is reflected by 56% of the 126 compounds being statistically significantly different due to genetic background.

Assessment of the different chemical classes revealed that the influence of locations/input systems was mainly reflected by statistically significant differences in the polar fractions III and IV, whereas differences between cultivars were found more in the lipophilic fractions I and II. In total, only two metabolites turned out to be consistently different over all three seasons (malic acid, *myo*-inositol) due to input system, but 14 due to cultivar (Table 8).

Table 8: Number of statistically significant differences obtained by ANOVA ($p < 0.01$) and Tukey's HSD ($p < 0.01$) of metabolite profiling data from fractions I (major lipids), II (minor lipids), III (sugars, sugar alcohols) and IV (organic acids, amino acids, amines) of three maize cultivars (Amadeo, Lukas, Flavi) grown in the experiments Ia, Ib and II

	experiment			
	Ia	Ib	II	consistent [a]
Location (Input System)				
compounds included	125	127	126	116
differences				
fraction I	7	3	1	0
fraction II	8	0	0	0
fraction III	10	11	3	1
fraction IV	11	3	10	1
total	36	17	14	2
differences [%]	29	13	11	2
Genetic Background (Cultivar)				
compounds included	125	127	126	116
differences				
fraction I	5	8	19	2
fraction II	12	14	17	7
fraction III	5	9	12	2
fraction IV	7	10	22	3
total	29	41	70	14
differences [%]	23	32	56	12

[a] Numbers of compounds consistently included for comparison and differences consistently detected as statistically significant in all experiments

4.3.2 Comparison of Relative Metabolite Levels

In 2004 the influence of locations/input systems was most prominent. To determine the metabolic sources of variation, loading scores of the first principal component of the PCA data were examined (Figure 19). Metabolites with the 10 highest absolute loading scores were quantified on the basis of relative signals. They all belonged to the polar fractions III and IV containing sugars, sugar alcohols, acids and amines. In addition, the levels of *myo*-inositol were determined, as this metabolite was found to be consistently different by ANOVA over all seasons (Figure 20).

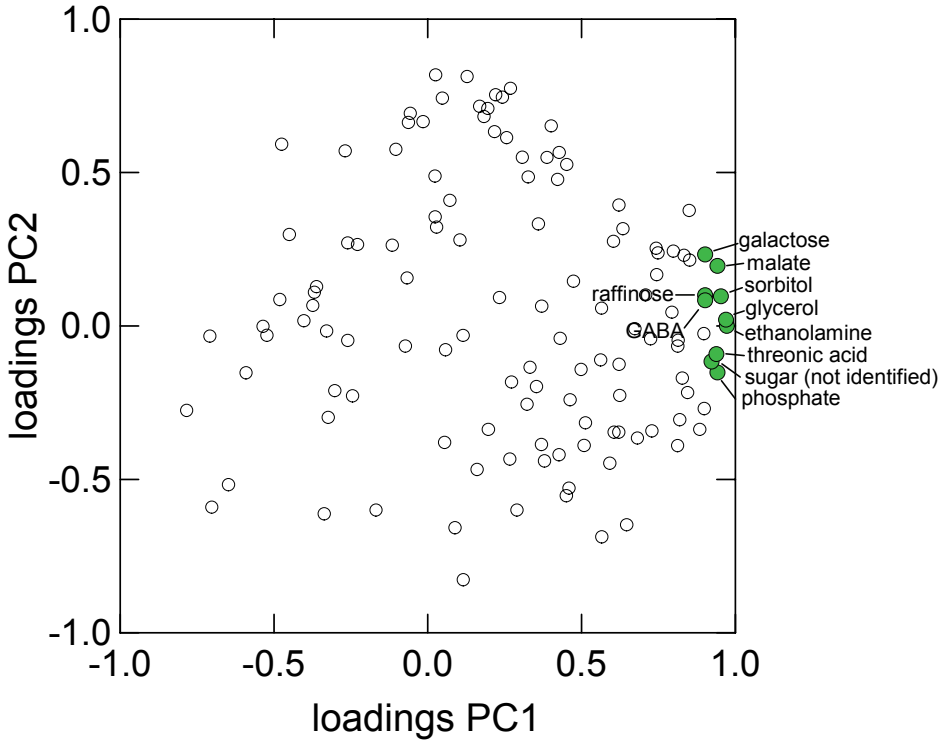

Figure 19 Factor loading scores of principal component 1 and 2 of PCA of metabolite profiling data of three cultivars grown in season 2004 at two locations (Frankendorf (conventional) / Schönbrunn (organic)) from fractions I-IV. Black circles (●) representing compounds with 10 highest absolute loading scores on PC1.

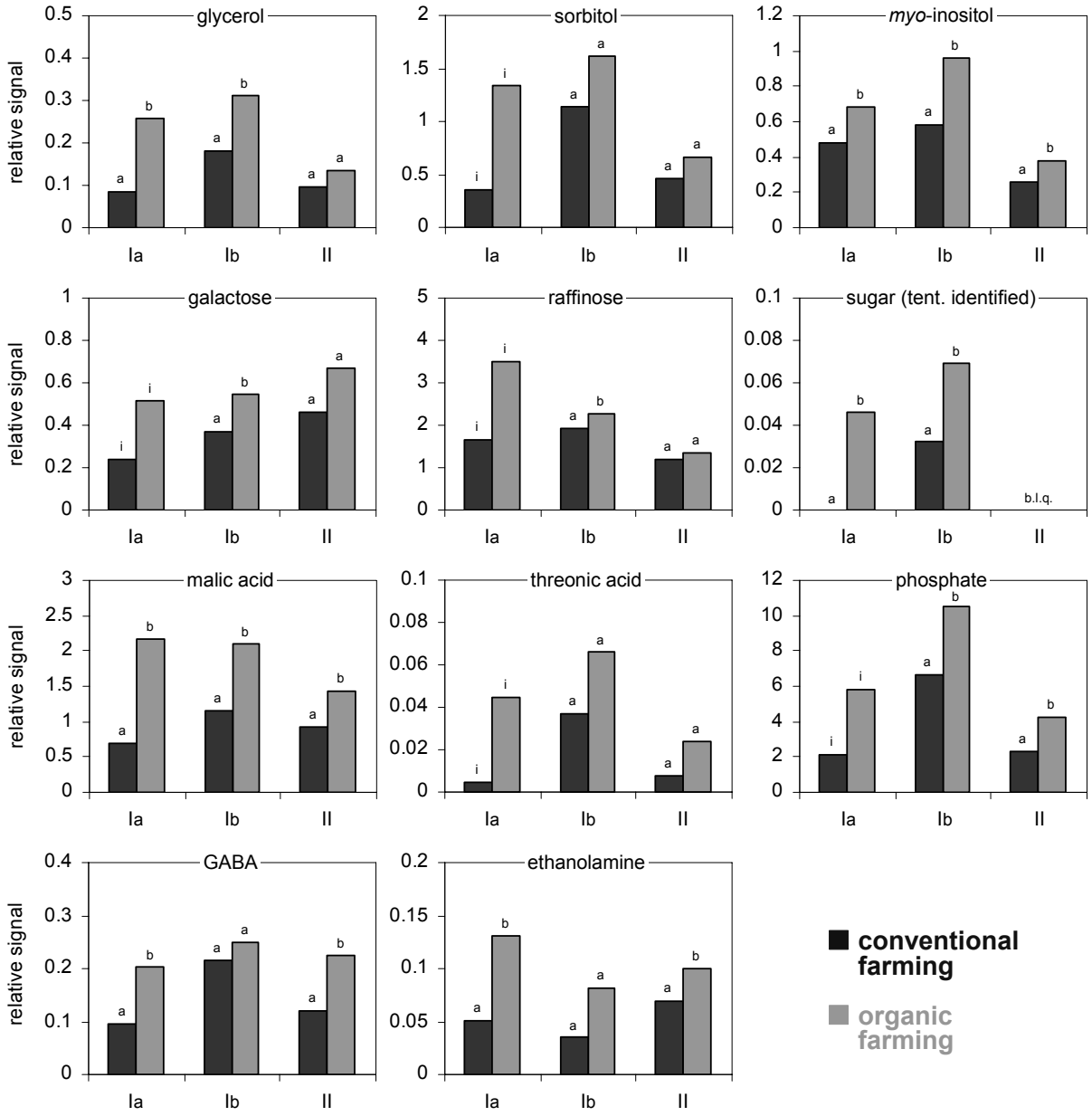

Figure 20 Comparison of locations with conventional and organic farming by semiquantified levels of compounds selected according to the ten highest loading scores on PC1 in experiment Ia and *myo*-inositol. Relative signals calculated on the basis of the respective internal standard. Statistically significant differences (Tukey's HSD, $p < 0.01$) are indicated by different characters (a,b) within each experiment (Ia, Ib, II); ^{i} significant interaction of cultivar*location ($p < 0.01$).

In the first growing season 2004 (experiment Ia) higher levels for these compounds were detected in the samples grown at the location Schönbrunn under the organic regime. All differences were statistically significant ($p < 0.01$) or the lower value was below the limit of quantification. Although, the levels of some of the metabolites showed a significant ($p < 0.01$) interaction effect of cultivar and farming location, a closer look at the interaction diagrams revealed an ordinal interaction in these cases

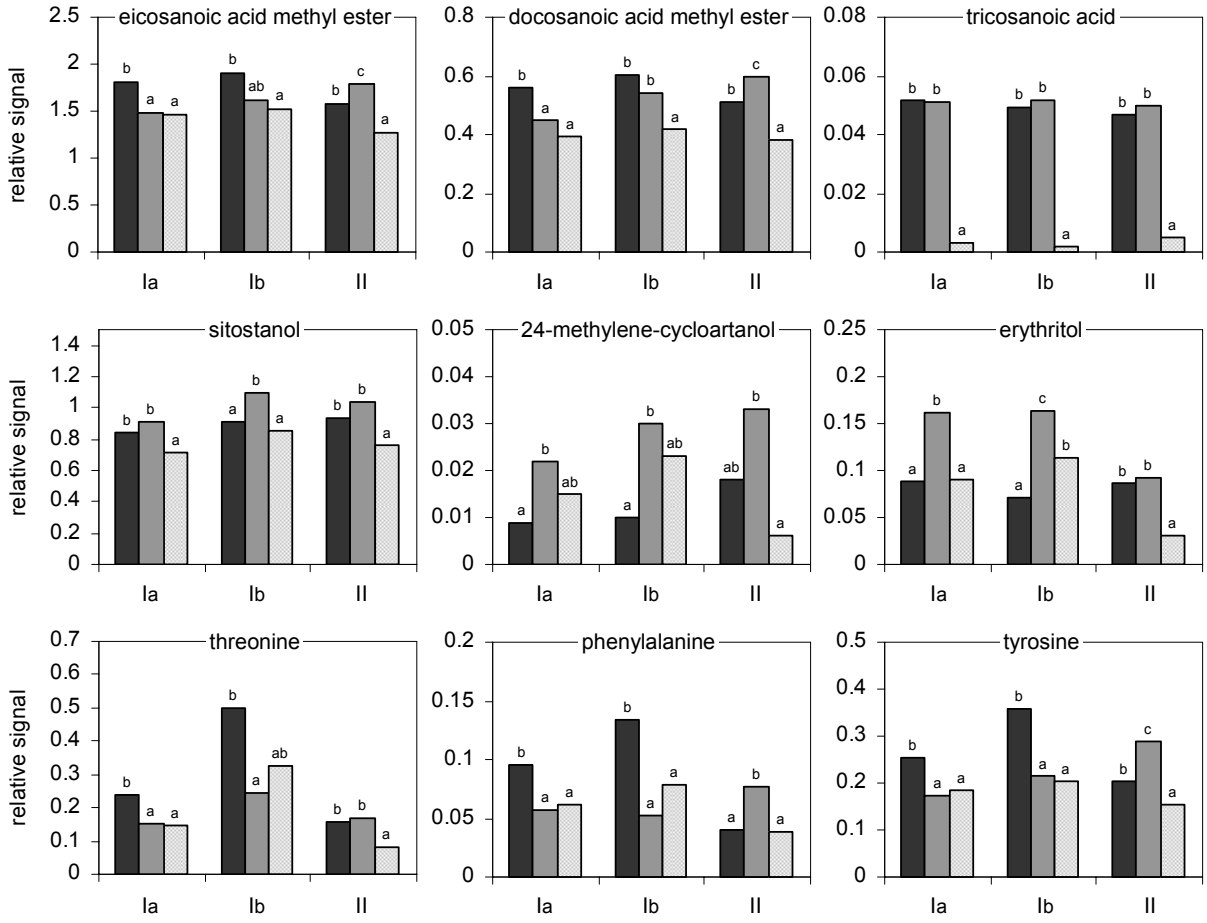

Figure 21 Comparison of three maize cultivars (■ Amadeo, ■ Lukas, ■ Flavi) by semiquantified levels of identified compounds that showed consistently different levels in three field trials (Ia, Ib, II). Relative signals calculated on the basis of the respective internal standard. Statistically significant differences (Tukey's HSD, $p < 0.01$) were indicated by different characters (a,b,c) within each experiment.

for the effect of locations/input systems, which allowed a further evaluation of these results on the basis of individual cultivars (Leigh and Kinnear, 1980). Comparison of the levels of these metabolites between conventional and organic practice by post hoc testing (Tukey's HSD, $p < 0.01$) resulted in significant differences for these compounds. Repetition of this trial in 2005 (experiment Ib) resulted in fewer statistically significant differences. The smallest differences were detected in experiment II in which both input systems were applied at one location (Scheyern). Regarding all three sample sets, only malic acid, *myo*-inositol, and after post hoc testing, phosphate turned out to be consistently different at all experiments.

Of the 14 differences consistently observed between cultivars over all three sample sets, 9 metabolites were identified and quantified (Figure 21). Differences were observed for compounds from all four metabolite profiling fractions I–IV.

4.3.3 Comparison of Major Impact Factors on Metabolite Variation

The sequence of PCA plots obtained from the experiments Ia, Ib and II (Figure 18A–C) demonstrates that the observed separations are mainly due to the genetic differences (cultivars) and to environmental influences; the different input systems (conventional/organic) only lead to minor differentiations. Figure 18A (experiment Ia) showed a strong separation of farming sites on the first principal component in 2004. At this point it remained unclear whether this effect was due to the different input systems employed or only due to the different locations. The repetition in 2005 (experiment Ib, Figure 18B) also resulted in differentiations according to locations and/or input systems; however, the genetic background (cultivars) turned out to be the dominating contributor to the observed clustering. Finally, experiment II revealed that, if environmental influences are minimized by performing the trials at one location, only a very slight differentiation according to the input system is observed, and the clustering pattern is mainly determined by the differences in cultivars (Figure 18C).

The clustering in Figure 18A may be explained by differences in nutritional supply, soil composition and influences in the local microclimates. The location Frankendorf was characterized by a silty loam soil with a rather high soil index of 80, which represents 80% of the performance capacity of an “ideal” soil (Wendland *et al.*, 2007), whereas the location Schönbrunn had a more sandy loam soil with a soil index of only 52. In addition, though temperatures were more adequate at Schönbrunn, precipitation was much more abundant at the location Frankendorf and may have contributed to a better plant growth at this location (Table 1). Additional support for the less favorable conditions in Schönbrunn is given by the clear separation of Lukas on PC2 at this location (Figure 18A). Lukas is known to be a robust cultivar under different environmental conditions from dryness to low temperature (Eder, 2009). Apparently, at location Frankendorf the growing conditions met the requirements for all cultivars, resulting in one PCA cluster. Under the conditions in Schönbrunn, the more robust cultivar Lukas behaved differently from Amadeo and Flavi. In 2005 the precipitation was higher than in 2004 at both locations (Table 1) and obviously reached a sufficient amount to ensure similar growth behavior of the maize plants at both farming sites.

However, a differentiation of locations/input systems was still observable within the clusters of each cultivar.

Evaluation of metabolite profiling data by ANOVA confirmed the observations made by PCA; most differences between locations and/or input systems were found in 2004, less in 2005 and only a few number of differences between input systems conducted at the same location in 2006. The decrease in statistically significant differences due to locations/input systems and the simultaneous increase due to genetic background (cultivar) from field trials Ia/Ib to II reflects the set-up of these experiments. The extent of changes seen due to the factors genetics and environment is in the same order of magnitude as observed for maize kernel metabolites in a previous study employing the same metabolite profiling methodology (section 4.2).

Considering the broad range of low molecular weight constituents analyzed by the applied GC/MS-metabolite profiling approach, the number of consistent differences identified owing to input system is relatively small; only for malic acid, *myo*-inositol and phosphate higher levels were determined for maize grown at organic farming sites in all three experiments. For two of these metabolites similar effects are known from other studies: A metabolite profiling approach analyzing 51 polar metabolites in wheat grown at different input practices also reported higher levels of *myo*-inositol at growing sites with organic farming compared to the respective conventional site (Zörb *et al.*, 2006). *Myo*-inositol plays important functional roles in various physiological routes involved in, for example, seed desiccation, osmo-regulation and stress response (Loewus and Murthy, 2000); at this point the data available do not allow a reasoned answer why this metabolite should be consistently changed in organically grown crops. Phosphate is one of the most important plant constituents that affect growth and metabolism (Raghothama, 1999). The increased levels of phosphate observed in the organically grown maize samples observed in this study are in agreement with higher levels of phosphate reported in various other organically grown plants (Dangour *et al.*, 2009; Winter *et al.*, 2006).

4.3.4 Conclusions

In conclusion, the application of a comprehensive metabolite profiling approach allowed the investigation of the effect of conventional and organic farming management practices on maize metabolites from different chemical classes ranging from lipophilic to polar. The assessment of impact factors on metabolic variation such as genotype, farming location and growing season, enabled the evaluation of differences in the light of natural variation. The results of this study suggest that genotype and environment are the major contributors to differentiations seen in metabolite profiles of maize kernels. The application of different input systems had only a small impact on the metabolites covered by the applied analytical approach. The few consistent differences seen between maize grown conventionally and organically, respectively, are in agreement with phenomena previously observed for organically grown crops.

4.4 Influence of Genetic Engineering

4.4.1 Introduction

Genetic engineering of agricultural crops is being employed for yield improvement, e.g. by increasing resistance to disease (Fujimoto *et al.*, 1993) and stress (Xue *et al.*, 2004) and tolerance to herbicides (Oard *et al.*, 1996), as well as for improvement of the nutritive value of crops, e.g. by increasing the availability of essential micronutrients (Lucca *et al.*, 2001; Ye *et al.*, 2000). Maize as one of the most important agricultural crops and as part of the staple diet of humans and livestock has been subjected to a variety of genetic modifications. Transgenic maize plants have been produced with different characteristics including insect-resistant Bt-maize (James, 2003) and herbicide-tolerant Roundup Ready maize (Sidhu *et al.*, 2000).

Current safety assessment procedures developed for GM crops are primarily based on a targeted compositional analysis of specific safety and nutrition-related compounds (FAO/WHO, 2000; OECD, 1993). Targeted analysis of specific key compounds, using well established and validated protocols, has provided the cornerstone for assessing the nutritional value and safety of cultivated crop species. Such a targeted approach may, however, has its limitations in detecting unintended effects in genetically modified organisms. Consequently, further assessment by non-targeted profiling technologies as unbiased analytical approaches has been suggested to overcome this drawback (Cellini *et al.*, 2004; Kuiper *et al.*, 2003).

In addition, a comparative analysis should not only focus on the GM crop itself, and its corresponding parental line, but metabolite profiles should also be assessed in the light of natural variability that is inherent in conventional crop material (EFSA, 2006). Such information would allow a more comprehensive benchmark against which the new generations of crops and advances in production systems could be evaluated. Genetic background, growing environment (geographical, seasonal) and crop management practices are major factors underpinning this variation.

In this chapter two sets of transgenic maize lines and their isogenic counterparts grown in South Africa and Bavaria (Germany) were analyzed by GC/MS-metabolite profiling. Genetically modified Bt maize expresses a crystal protein from the soil

bacterium Bacillus thuringiensis (Bt) that when ingested by insect pests causes a lethal paralysis in the digestive tract. In addition, transgenic Roundup ready (RR) maize was grown in South Africa that has been developed by genetic engineering to tolerate glyphosate, the active ingredient in the Roundup ready herbicide. The data were assessed by multivariate statistics (PCA) to allow the evaluation of the results in the light of natural variation.

4.4.2 South African Maize (GM-Bt, GM-RR, non-GM)

To assess the influence of genetic modification under different environmental conditions, a GM maize line (GM-Bt) was grown together with its near isogenic line (non-GM) at three locations in South Africa (Petit, Potchefstroom, Lichtenburg) in growing season 2004. In addition, at Petit and Lichtenburg Roundup ready-maize (GM-RR) was grown together with the Bt-maize and the isogenic line. To include the potential influence of the growing season, the maize lines grown at location Petit were also harvested in two additional years (2005 and 2006).

Statistical assessment via PCA of the GC/MS-metabolite profiling data from the samples grown at the three locations in 2004 revealed clear separations of the GM lines from the respective isogenic line at Potchefstroom and Lichtenburg (Figure 22). For the maize lines grown over three years, a distinct separation of both GM lines was observed for the location Petit in 2006; the separation of GM lines from the isogenic maize line was less pronounced for this location in 2004 and 2005. However, despite partly obvious differences between GM lines and isogenic maize determined for one location/year, no separations of the different maize lines were detectable when combining the metabolite profiling data obtained from GM lines and isogenic maize for all growing locations/years (Figure 22). This confirms that, at least in the case of the specific GMOs analyzed, the effect of environment (location, year) was more pronounced than that of the genetic background (GM, non-GM).

In addition to GC/MS-metabolite profiling, the described samples were also analyzed by cDNA microarray transcriptome profiling, two-dimensional gel electrophoresis proteome profiling and 1H-NMR metabolic fingerprinting. All techniques showed that the environmental factors caused more variation in the different transcript, protein, and metabolome profiles than the different genotypes (Barros *et al.*, 2010).

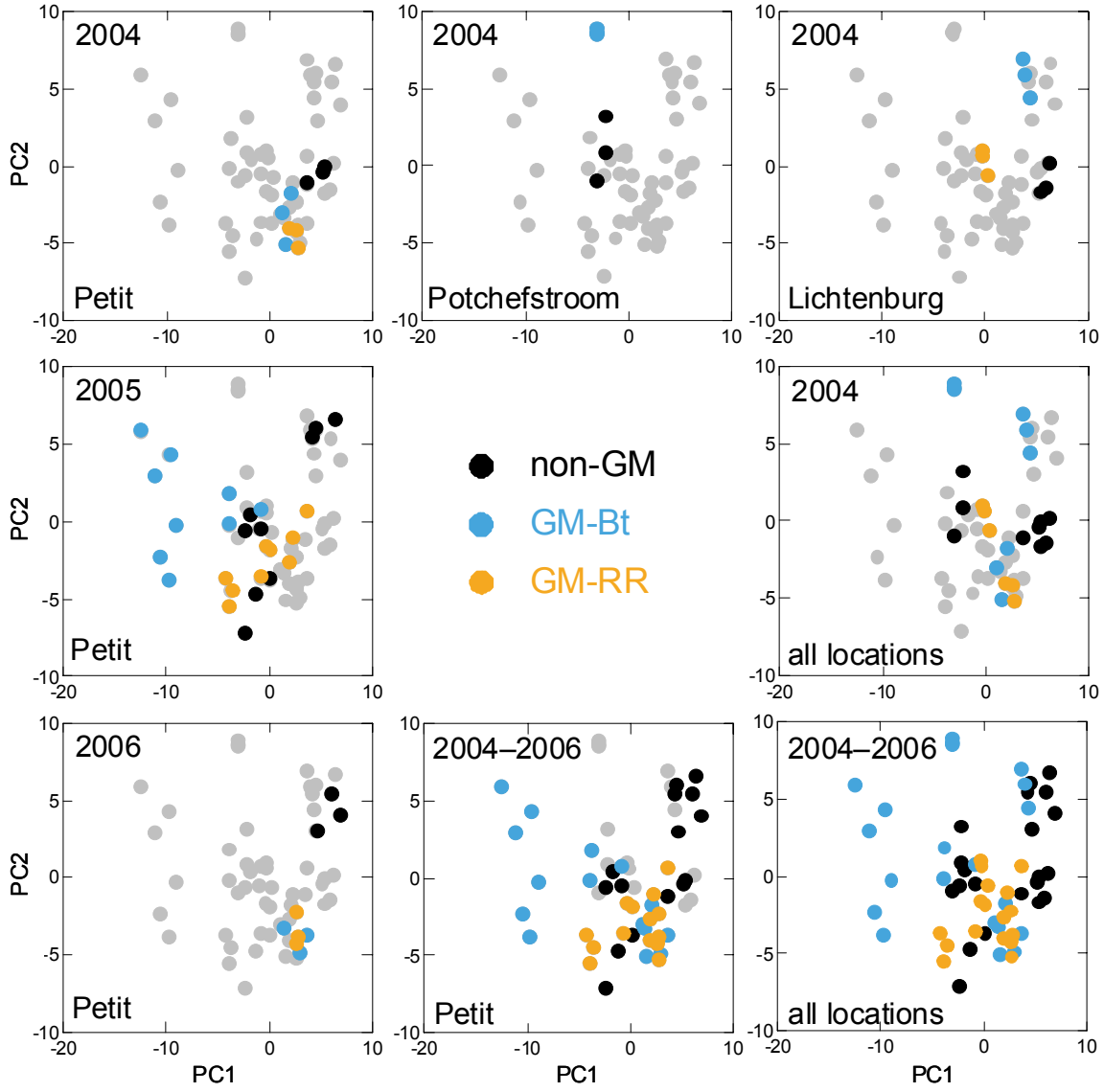

Figure 22 Principal component analysis of GC/MS-metabolite profiling data from fractions I-IV obtained by triplicate analysis of three maize lines (● non-GM, ● GM-Bt, ● GM-RR) grown in South Africa at locations Petit, Potchefstroom and Lichtenburg in growing seasons 2004, 2005 and 2006. In 2005 three technical replicates were analyzed. Each plot contains all samples, sub-sets are highlighted in color.

4.4.3 German Maize (GM-Bt, non-GM)

Two maize genotypes (GM-Bt, non-GM) were grown at two different farming locations in Bavaria, Germany, in season 2004. Field replicates (n=4 location Neuhof, n=3 location Pfaffenhofen) were collected for analysis by GC/MS-metabolite profiling. Data from triplicate analysis were subjected to Principal Component Analysis (Figure 23). PCA of pooled data from the four metabolite profiling fractions I–IV revealed a clear separation of the two farming locations on the first principal component representing 34% of the total variation in the dataset (Figure 23A).

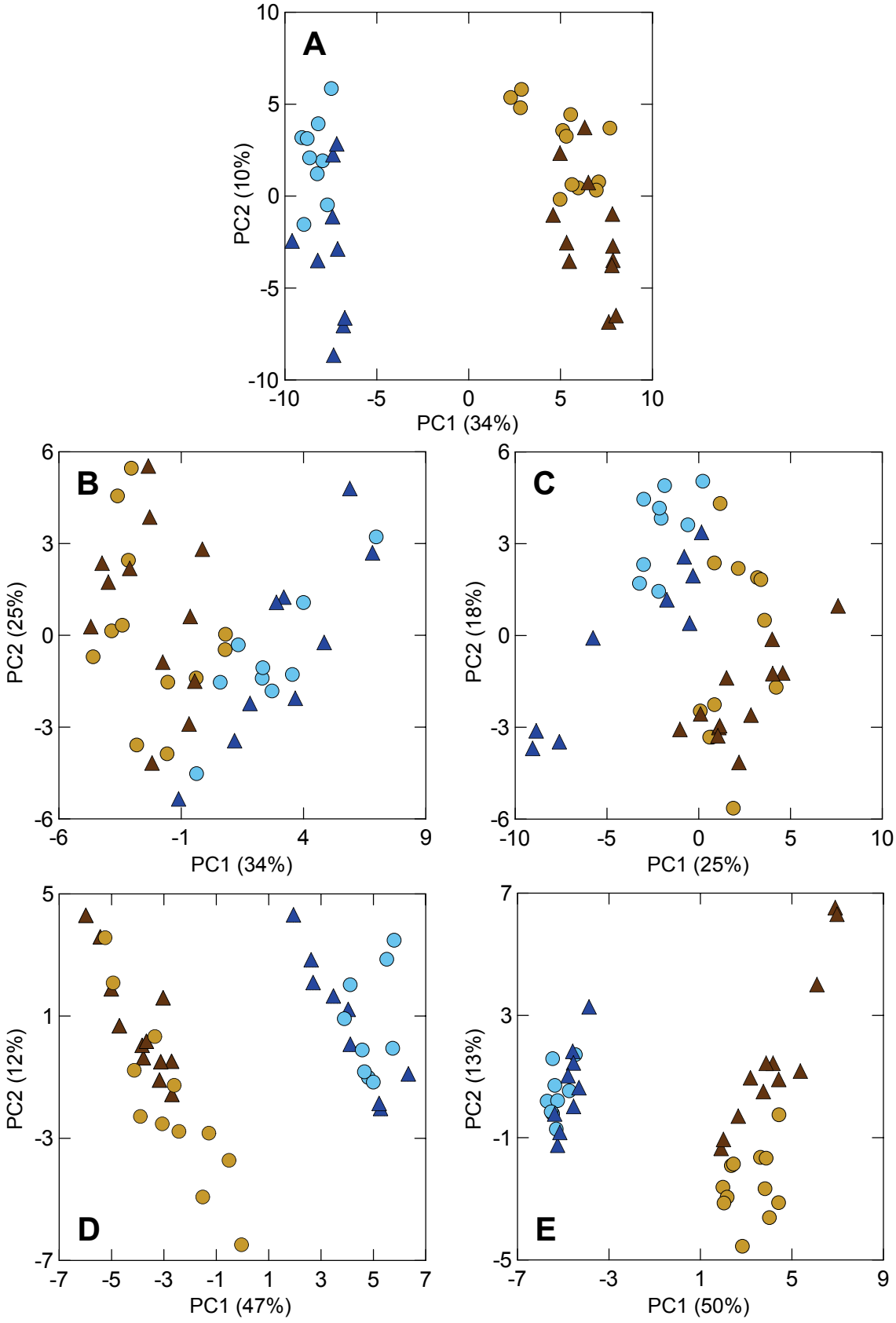

Figure 23 Principal component analysis of metabolite profiling data from (A) combined fractions I-IV (I: major lipids, II: minor lipids, III: sugar/sugar alcohols and IV: acids/amino acids/amines), and (B) individual fractions I, (C) II, (D) III and (E) IV of two maize genotypes (▲ GM-Bt, ● non-GM) grown at locations Neuhof (●, ▲) and Pfaffenhofen (●, ▲) (Bavaria, Germany) in season 2004. Field replicates for each treatment were analyzed in triplicate.

Clusters according to genotypes could be observed within farming locations only on PC2 (10% of the total variation). Tukey's post hoc testing resulted in a significant ($p < 0.001$) differentiation of the clusters.

To achieve a better understanding which maize constituents are the drivers of the observed differentiation pattern, PCAs of data from the four individual metabolite profiling fractions were investigated. PCAs of the two lipid fractions I and II (Figures 23B/C) show differentiation clusters according to locations on the first principal component. However, there is no clear separation as compared to the combined PCA (Figure 23A). In the plots based on data for minor lipids (fraction II, Figure 23C) a grouping of non-GM samples from Pfaffenhofen, but not Neuhof, was observed. In contrast to the lipid fractions, the plots obtained for data of the polar fractions III and IV, demonstrate a strong effect of the location similar to the combined PCA (Figures 23D/E). For sugars and sugar alcohols (fraction III, Figure 23D), the non-GM samples do overlap with GM-Bt at location Neuhof, whereas results for acids, amino acids and amines (fraction IV, Figure 23E) show distinct clusters for both genotypes at Neuhof, but not at location Pfaffenhofen.

The results obtained by PCA for individual fractions I–IV demonstrate unequal contributions of the maize metabolites from different chemical classes. The influence of the growing location could be detected in all plots, but this effect was much more pronounced for the polar fractions III and IV. Differences observed between the different genotypes were not consistent. For example, PCA of minor lipids data revealed a cluster for non-GM samples grown at Pfaffenhofen. For acids, amino acids and amines a similar effect was observed for samples grown in Neuhof. This indicates a more pronounced impact of environment for samples collected from different sites for GM-Bt, and non-GM maize, respectively, whereas the influence of the different genotypes on the metabolite profiles seem to play only a minor role for the investigated maize material.

The results obtained for the comparative metabolite profiling of the South African and the German GM maize, and their respective isogenic counterparts are in agreement with data obtained for wheat showing that differences observed between GM and the control lines were generally within the same range as the differences

observed between the control lines grown on different sites and in different years (Baker *et al.*, 2006). In addition, the results are in accordance with observations made for differently ripening maize cultivars over a consecutive three-year period (section 4.2). Whereas the influence of the genotype (cultivar) could be clearly shown within the single years, combination of samples from all seasons revealed the environmental impact to be the most prominent impact factor. Similar results were also reported by a research program (GO2) commissioned to assess the potential use of *omics* approaches in comparative analysis and their relevance to risk assessment (FSA, 2005). This three-year research program was focusing on the applicability and practicality of a variety of existing and emerging techniques for the safety assessment procedures for the next generation of GM foods. The program examined the use of transcriptomic, proteomic and metabolomic techniques in a number of different plant species including potato, barley, tomato and Arabidopsis. Publications arising from the FSA projects observed that the differences between conventional varieties were always significantly greater than the differences between the wild-types and their respective transgenics despite the fact that some GM lines had very distinct morphological phenotypes (Catchpole *et al.*, 2005; Defernez *et al.*, 2004). It was concluded that the vast majority of the observed changes were small (ca. 2-fold or less) with evidence provided that at least some of these changes may be due to somaclonal variation resulting from the *in vitro* manipulation of plants rather than the presence of an inserted transgene *per se*.

4.4.4 Conclusions

The influence of genetic modification on maize metabolite profiles was assessed under different environmental conditions. The data generated revealed that environmental influences (farming location, growing season) had a stronger overall effect on the metabolome of the investigated maize genotypes than the genetic modification. The data do not allow a general conclusion on the potential for unintended effects in GMO and a case-by-case approval remains pragmatic. However, for the GM plants investigated in this study, the differences in the metabolite profiles of maize genotypes grown in different environments were significantly greater than the effects of the transgene.

4.5 NIRS Screening of Maize

4.5.1 Introduction

The preceding chapters demonstrated the potential of GC/MS-based metabolite profiling for the qualitative and quantitative assessment of a broad range of maize constitutents. To achieve this, a sophisticated analytical protocol is used that employs extraction, fractionation and derivatization steps to face the different chemical characteristics inherent in a set of metabolites from a biological sample. However, if the aim is to look for compositional similarities or if the overall natural variability should be explored in a large sample set, it might not be necessary to determine the individual levels of all metabolites. In a first step, an uncomplex, rapid fingerprinting approach, without a major pre-treatment of the raw material would be sufficient for the detection of the major effects (Fiehn, 2002). Based on the results of this screening, a pre-selection of samples could be further thoroughly analyzed by comprehensive GC/MS-metabolite profiling.

A non-destructive, highly reproducible technique known for its cost effective and simple application is near infrared spectroscopy (NIRS). The absorption of infrared light (800–2500 nm) corresponding to overtones and combinations involving C-H, O-H or N-H chemical bonds results in a characteristic spectral profile for each sample based on its chemical composition. Together with multivariate analysis, such as principal component or linear discriminant analysis, NIRS showed great promise as a screening tool for monitoring biochemical changes in crop developing systems, such as malting barley (Allison and Maule, 1991), or for both discriminating between yeast strains and grouping strains with deletions in genes that disturb similar metabolic pathways (Cozzolino *et al.*, 2006).

NIRS is commonly used for targeted proximate analyses (Osborne, 2008; Woodcock *et al.*, 2008) and detection of physical properties, such as hardness or solubility (Blanco *et al.*, 2006; Miralbés, 2004). These applications are based on calibration models for each analyte that have to be developed based on multivariate statistics. By mathematically correlating spectral data with the data obtained by the currently accepted laboratory procedures for reference analyses, the content of a respective

constituent can be predicted by NIRS (Shenk *et al.*, 2008). A major challenge during the development of the calibration models are the different characteristics of spectral data, i.e. the absorption of the X-H chemical bonds, and of reference analytical data, for example the content of nitrogen in a sample based on digestion and detection by titration. As a second aspect, available calibration models are based on targeted reference analytical data from proximate analyses. For use of NIRS in a profiling context, the unbiased character of GC/MS-metabolite profiling data has to be considered, if NIRS models should be used for the non-targeted prediction of the chemical composition.

Therefore, the aim was to apply NIRS profiling for the qualitative and quantitative screening of maize samples. By multivariate comparison (principal component analysis) of the spectra with data from polar GC/MS-metabolite profiling, the ability of NIRS for the detection of the major sources of variation in a dataset should be determined. Environmental and genetic impact on maize metabolite profiles were considered by analysis of three cultivars grown in different growing seasons and at different locations. For substantiation of differences between the locations based on NIRS data, calibration models should be developed based on polar GC/MS-metabolite profiling data for the prediction of the total content of sugars and of acids/amino acids. By this, the suitability of the applied NIRS approach as tool for the pre-assessment of large sample-sets should be tested. It should be demonstrated that NIRS could complement existing GC/MS-metabolite profiling methods in the investigation of crop samples from large sample sets with potential for use in complex metabolomic studies and breeding programs.

4.5.2 Qualitative Assessment by GC/MS and NIRS

The potential of NIRS profiling for the investigation of maize kernels should be determined by analysis of cultivars grown under different environments. For comparison, the maize kernels were first analyzed by the established GC/MS-metabolite profiling approach. For NIRS profiling, near infrared spectra (1000–2500 nm) were recorded in diffuse reflection mode by means of a commercial Fourier-transform (FT) NIR spectrometer. After performing standard data pre-

processing (smoothing, scatter correction, elimination of water band), the NIR spectra were subsequently used for statistical analysis.

Influence of Farming Location / Input System

Assessment of three maize cultivars (Amadeo, Lukas, Flavi) collected from two different field trials comprising conventional (location Frankendorf) and organic (location Schönbrunn) farming practice was performed by principal component analysis (PCA) of the chromatographic GC/MS and the spectral NIRS data. PCA of GC/MS-metabolite profiling data of sugars (Figure 24A) and acids/amino acids (Figure 24B) revealed a clear separation due to locations/input systems on the first principal component. Differentiation of cultivars could be observed on the second principal component. PCA of NIRS profiling data obtained for the same sample set also resulted in clustering according to locations/input systems on the two first principal components (Figure 24C), but there was no differentiation of the three cultivars.

Influence of Growing Season

In a second experiment the influence of farming location should be investigated. Samples of three maize cultivars (Gomera EU, Amadeo, Lukas) were harvested in three consecuting growing seasons (2004–2006). Examination of the first two principal components for GC/MS data of sugars revealed a clear differentiation according to years (Figure 25A), whereas only one of the cultivars (Gomera EU) formed a more distinct cluster. GC/MS data for acids/amino acids also exhibited a strong impact of growing season, with the most prominent separation of samples obtained from 2004 (Figure 25B). Again cultivar Gomera EU could be differentiated from the other two cultivars, for which clusters overlapped. A PCA of NIR spectral data revealed a distinct clustering of all three seasons (Figure 25C). This observation was in agreement with the GC/MS results. Due to the strong impact of the growing season, the three cultivars could be distinguished only within the three years (Figure 25C).

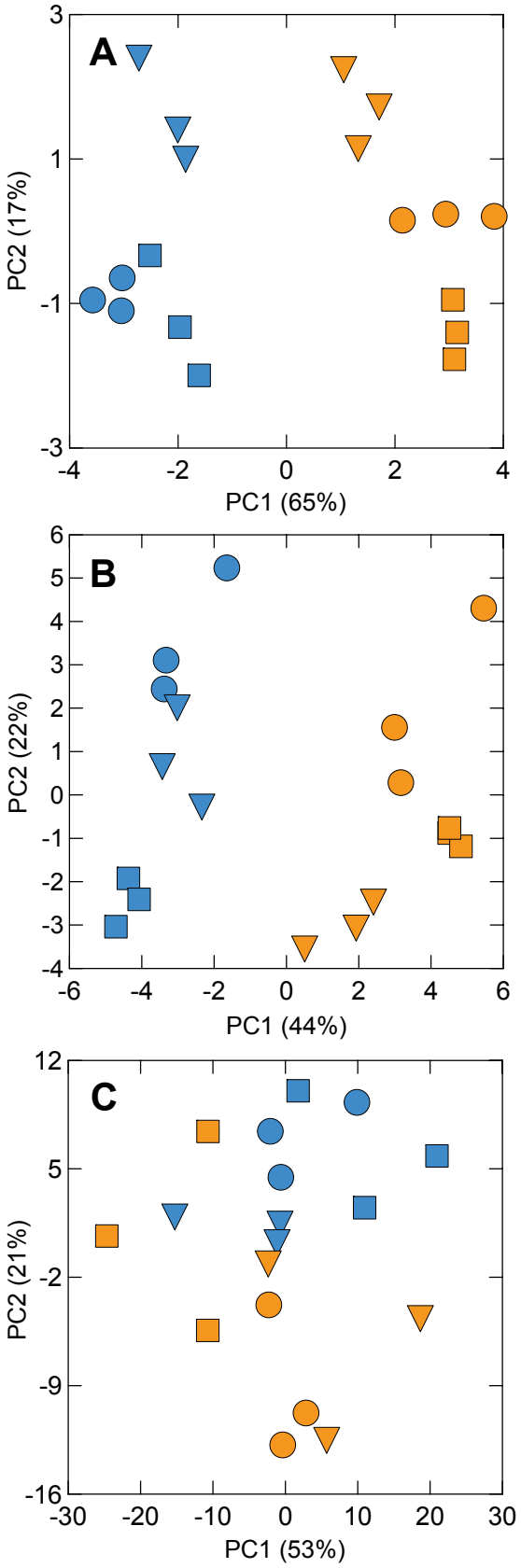

Figure 24 Principal Component Analysis of data obtained by GC/MS-metabolite profiling of (A) sugars and sugar alcohols and (B) acids/amino acids, and by (C) NIRS analysis of three maize cultivars (●: Amadeo, ■: Lukas, ▼: Flavi) grown at different locations (●,■,▼: Frankendorf, ●,■,▼: Schönbrunn).

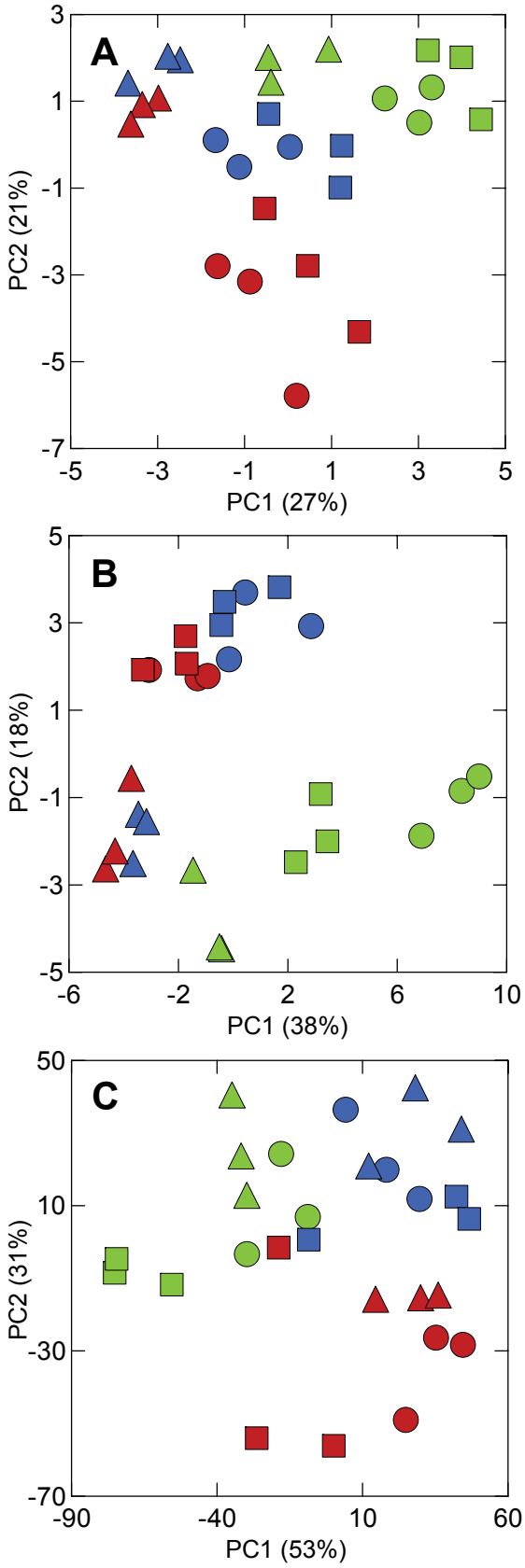

Figure 25 Principal Component Analysis of data obtained by GC/MS-metabolite profiling of (A) sugars and sugar alcohols, (B) acids/amino acids, and by (C) NIRS analysis of three maize cultivars (▲: Gomera EU, ●: Amadeo, ■: Lukas) harvested at location Strassmoos in three growing seasons (▲,●,■: 2004, ▲,●,■: 2005, ▲,●,■: 2006).

Assessment of NIR spectra via PCA revealed similar differentiation patterns as those obtained by GC/MS-metabolite profiling. The variation covered by the first two principal components was in the same order of magnitude for both approaches. NIRS as well as GC/MS were able to identify environment (farming location, growing season) as the major impact factor on maize composition. Also an effect of the genetic background (cultivar) was identified by the two profiling approaches, although PCAs based on GC/MS-metabolite profiling appeared to be more sensitive to the differentiation of cultivars on the first two principal components. This may be explained by the type of information that is contained in chromatographic GC/MS versus spectrometric NIR data. The metabolite coverage of the applied GC/MS protocol ranges from sugars and sugar alcohols to organic acids, amino acids and amines. However, the maize kernel exhibits also a broad range of non-polar metabolites such as lipids, free fatty acids, sterols and carotenoids (Kirchhoff, 2008). Profiling approaches based on GC/MS are restricted to the detection of low molecular weight constituents, selected according to the solubility of the chosen extraction solvents (Fiehn, 2008). In contrast, NIR spectra will be the result of absorption due to the total composition of a sample. That means, both low and high molecular weight constituents, including matrix compounds such as starch and cellulose, are reflected by the obtained absorption patterns (Osborne, 2008; Wang and Paliwal, 2007).

4.5.3 Establishment of NIRS Calibration Models

4.5.3.1 Calculation of Sum Parameters

A quantitative assessment should enable the elucidation of the metabolic changes underlying the observed PCA differentiation patterns. Basic quantification of GC/MS-metabolite profiling data can be performed by calculating response ratios (i.e. relative peak signals according to the internal standard). In contrast, NIR spectra do not contain quantitative data per se. A correlation has to be established between spectral information (i.e. absorptions at distinct wavelength points) and the content of the target compound. This is accomplished by building mathematical models based on reference data obtained for samples analyzed by both NIRS and an accepted analytical method. For a compositional analysis based on unbiased profiling data for

each metabolite an individual model would have to be calculated. However, GC/MS profiling approaches are also capable of detecting compounds that may not have been identified according to standard compounds, thus no reference analytical method will exist for these constituents.

Therefore, the idea of this study was to create a parameter analogous to the reference data of a single target compound, but based on GC/MS-metabolite profiling data. Accordingly, the GC/MS response ratios of sugars/sugar alcohols and of acids/amino acids were summarized representing the total metabolites contents in the respective fraction. Reference data was obtained from a broad range of maize samples investigated by GC/MS-metabolite profiling for the assessment of genetics and environment (sections 4.2, 4.3). The samples were selected according to genotype (32 cultivars), location (20 farming sites, Bavaria, Germany) and growing season (2004–2006). In addition, 20 samples from three field sites located in South Africa were included. The resulting high diversity in the reference sample set is a prerequisite for the development of a global NIRS prediction model (Shenk et al., 2008). The selected samples should be representative for the population of samples to be analyzed in the future.

Table 9: Metabolites included for calculation of sum parameter for sugars and sugar alcohols

	content [a]	SD [b]	% of total	sum %
sucrose	38.47	15.33	79.8	79.8
glucose	3.80	2.30	7.9	87.6
fructose	1.96	1.45	4.1	91.7
raffinose	1.70	0.95	3.5	95.2
sorbitol	0.71	0.52	1.5	96.7
inositol	0.50	0.25	1.0	97.7
galactose	0.27	0.19	0.6	98.3
glycerol	0.18	0.21	0.4	98.7
sugar (tent. identified)	0.09	0.05	0.2	98.9
erythritol	0.08	0.06	0.2	99.0
sugar (tent. identified)	0.07	0.03	0.1	99.2

[a] Based on relative levels according to the internal standard
[b] SD: standard deviation (n = 99)

Table 10: Metabolites included for calculation of sum parameter for acids, amino acids

	content [a]	SD [b]	% of total	sum %
phosphoric acid	3.56	2.84	23.9	23.9
proline	2.80	1.13	18.8	42.6
aspartic acid	1.29	0.76	8.7	51.3
citric acid	1.02	0.49	6.9	58.2
malic acid	0.93	0.74	6.3	64.4
asparagine	0.81	0.56	5.4	69.8
alanine	0.77	0.70	5.2	74.8
glutamic acid	0.67	0.50	4.5	79.3
pyroglutamic acid	0.46	0.24	3.1	82.4
serine	0.41	0.40	2.8	85.1
acid (tent. identified)	0.36	0.40	2.4	87.5
glutamine	0.26	0.19	1.7	89.2
tyrosine	0.21	0.09	1.4	90.7
GABA	0.20	0.14	1.3	92.0
threonine	0.16	0.15	1.1	93.1
valine	0.16	0.10	1.0	94.1
lysine	0.12	0.07	0.8	95.0
glycine	0.12	0.07	0.8	95.8
acid (tent. identified)	0.10	0.07	0.7	96.5
fumaric acid	0.09	0.23	0.6	97.0
leucine	0.07	0.07	0.5	97.5
lactic acid	0.06	0.09	0.4	97.9
phenylalanine	0.06	0.05	0.4	98.3
ethanolamine	0.06	0.05	0.4	98.7
ß-alanine	0.05	0.07	0.3	99.0
adenine	0.03	0.06	0.2	99.5
glyceric acid	0.02	0.03	0.2	99.6
acid (tent. identified)	0.02	0.03	0.1	99.7
threonic acid	0.02	0.03	0.1	99.9
tryptophan	0.01	0.02	0.1	99.9
methionine	0.01	0.02	0.1	100.0

[a] Based on relative levels according to the internal standard
[b] SD: standard deviation (n = 101)

Based on these samples the total metabolite content was calculated for sugars (a total of n = 99 samples) and for acids/amino acids (a total of n = 101 samples. The relative levels of 11 sugars and sugar alcohols were summarized, with sucrose

representing almost 80% of the total sugar content (Table 9). Within the fraction of acids/amino acids, phosphoric acid was the most abundant metabolite (24% of the total content), followed by proline (19%) and aspartic acid (9%) (Table 10). In total 31 compounds were included in the calculation of the second sum parameter. The resulting sum parameters represented more than 99% of the total content in the samples as detected by GC/MS-metabolite profiling (Tables 9, 10).

4.5.3.2 Choice of Regression Algorithm

To establish a correlation between spectral data and the content of the constituent of interest several calibration methods are in use (Wang and Paliwal, 2007). To find those wavelengths whose inclusion in the calibration model results in the best prediction of the reference value on the basis of spectral data, partial least square regression (PLS) was applied as a full-spectrum method. In addition, stepwise multiple linear regression (MLR) was used as another relatively robust methodology for comparison.

4.5.3.3 NIRS Calibration Model for Sugars

The calibration set used for development of a NIRS model for sugars exhibited 99 samples selected according to their relative sugar content (Figure 26). The histogram shows a high bar for low relative levels owing to the low contents of sugars in most of the South African samples.

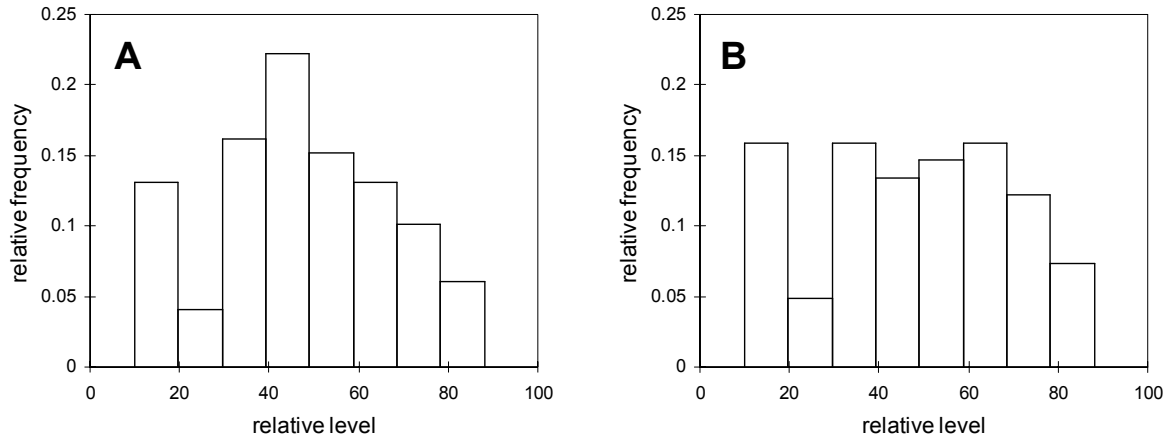

Figure 26 Histograms of the reference data obtained by metabolite profiling for the prediction of the sum parameter for sugars and sugar alcohols for (A) models PLSR, MLR1 and MLR3 (n = 99) and for (B) models MLR2 and MLR3 (n = 82). Relative signals were calculated on the basis of the respective internal standard.

For prediction of sugars by NIRS five models were evaluated. In order to assess the performance of the calibration models NIRS performance criteria were calculated and compared to literature data generally accepted as minimum requirements for NIRS methods (De la Roza *et al.*, 1998; Perez *et al.*, 2001; Williams and Sobering, 1996). The ratio of the range in reference data for the validation samples to the SECV (criterion 1) and the ratio of the standard deviation SD of the sample population to the standard error of prediction SECV (criterion 2) relate the SECV to the range and variance in the original data. A minimum requirement of 10 and 3 for criterion 1 and 2, respectively, allows the differentiation of samples low, medium and high in the target constituent (Williams and Sobering, 1996). Criterion 3 was deduced according to requirements described in literature (Fontaine *et al.*, 2001) that the SECV should not be higher than twice the standard deviation of the reference method (10% for GC/MS-metabolite profiling according to section 4.1. For prediction of sugars, a PLS regression was calculated on the basis of the whole NIR spectrum (Table 11). The correlation of the regression line of the calibration equation was > 0.9, but the model did not qualify according to criteria 1 and 2. Therefore, a MLR model was calculated including 8 wavelengths (1260 nm, 1405 nm, 1705 nm, 1720 nm, 2180 nm, 2315 nm, 2385 nm und 2395 nm) selected from bands that exhibited high loadings of the PLS regression model. MLR1 performed better than the PLSR model as reflected by the regression parameters (Table 11). The requirements 1 and 3 reported in literature for calibration models could be fulfilled (Table 11); criterion 2 was nearly fulfilled.

To pass all three performance criteria, two approaches were considered. First, for the distribution of reference values a slight Gaussian normal curve was observed, due to too many samples with medium levels of sugars. It was decided to remove 17 samples with levels between 30 and 60, resulting in a linear distribution of reference values from 82 samples in the histogram (Figure 26B). The respective model (MLR2) fulfilled all performance criteria (Table 11). In a second approach, it was strived to reach the criteria without reducing the reference sample set. It was presumed, that there was an overlap of sugar and starch bands due to structural similarities of sugars and starch. Therefore, an additional wavelength (2097 nm) from a known starch band (Osborne, 2008; Shenk *et al.*, 2008), was included into the model (MLR3). In fact, the model fulfilled criteria 1–3, although it did not perform as well as MLR2. It was

concluded, that the best approach would be to combine a reference sample distribution consisting of equal proportions of samples low, medium and high in sugar content, with the additional wavelength from the starch band. This final model (MLR4) showed the best performance, and it was used for prediction of the relative content of sugars for further experiments.

In comparison to literature data the SECV of model MLR4 may seem high (Qin *et al.*, 2007; Woodcock *et al.*, 2008). However, it has to be considered, that the standard error of the prediction model cannot be lower than the standard error of the reference method. For the applied GC/MS-metabolite profiling approach, a repeatability cut-of is set at 10%. NIRS calibrations for prediction of soluble sugar content in maize forage gave a similar SECV as compared to the results obtained in this study (Welle *et al.*, 2003). Therefore, the level of accuracy obtained in this study was considered to be adequate for the selection of samples according to low, medium and high sugar content, independently of the absolute value, as needed for a NIRS a pre-screening approach.

Table 11: Regression parameters of NIRS calibrations based on metabolite profiling data for sugars and for acids/amino acids

	r_{cal}[a]	r_{val}	m_{cal}[b]	m_{val}	SEC[c]	SECV[d]	Crit.1[e]	Crit.2[f]	Crit.3[g]
sugars									
PLSR	0.9080	0.8521	0.825	0.787	7.60	9.60	7.45	1.98	1.87
MLR1	0.9525	0.9414	0.907	0.898	5.75	6.38	11.21	2.98	1.24
MLR2	0.9626	0.9514	0.927	0.917	5.57	6.34	11.27	3.26	1.24
MLR3	0.9578	0.9462	0.917	0.908	5.42	6.11	11.70	3.11	1.19
MLR4	0.9672	0.9556	0.935	0.928	5.23	6.07	11.78	3.41	1.18
acids/amino acids									
PLSR	0.9477	0.9325	0.898	0.874	2.49	2.82	12.09	2.79	1.26
MLR	0.9615	0.9540	0.924	0.913	2.15	2.34	14.57	3.36	1.05

[cal] calibration, [val] validation

[a] coefficient of correlation

[b] slope of regression line

[c] root mean squared error of calibration

[d] root mean squared error of prediction

[e] performance criterion 1 = (max-min) / SECV > 10

[f] performance criterion 2 = s/SECV > 3

[g] performance criterion 3 = SECV / $s_{reference} \leq 2$

Details on the calculation of the performance criteria as described in the methods section

4.5.3.4 NIRS Calibration Model for Acids/Amino Acids

For prediction of the contents of acids/amino acids a reference sample set of 96 equally distributed samples was selected. In addition 5 samples with very high contents were included, resulting in a total of 101 reference samples (Figure 27). First, a PLSR was performed, fulfilling the performance criteria 1 and 2 (Table 3). As for sugars the multiple linear regression approach resulted in better regression parameters, also an MLR model was calculated. Inclusion of 8 wavelengths (1190 nm, 1215 nm, 1340 nm, 1405 nm, 1530 nm, 1650 nm, 1860 nm und 2220 nm) resulted in the fulfillment of the performance criteria 1–3 (Table 11).

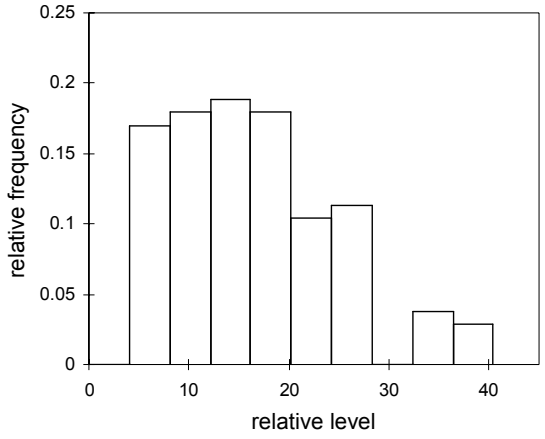

Figure 27 Histogram of the reference data obtained by metabolite profiling for the prediction of the sum parameter for acids/amino acids by NIRS models (PLSR and MLR, n = 101). Relative signals were calculated on the basis of the respective internal standard.

NIRS prediction of the total content of free amino acids as descibed in literature suggests a strong influence of the matrix to the quality of the prediction models. Whereas in tea smaller SECVs were achieved, the prediction in cheese was in the same order of magnitude as for GC/MS profiling as reference method (Ikegaya, 1990; Skeie *et al.*, 2006; Woodcock *et al.*, 2008). Taking these studies into consideration, the performance of the developed NIRS method was found to be sufficient.

4.5.4 Application of Calibration Models/Method Evaluation

Influence of Farming Locations / Input System

Applicability of the developed NIRS models was evaluated by comparison of maize grown at two different locations (Frankendorf, Schönbrunn), based on quantitative data from GC/MS and NIRS analyses. Comparison of the mean (three cultivars) levels for total sugar content as determined by GC/MS between two locations revealed higher levels at the location Schönbrunn deploying organic farming practice (Figure 28A). The relative levels for sugars as predicted by the calibration model MLR4 also exhibit higher levels at Schönbrunn, and lower sugar content at Frankendorf (Figure 28A). In addition, the relative total content for acids/amino acids was analyzed by GC/MS and NIRS confirming the results for sugars, and showing both approaches to be in excellent agreement (Figure 28B). The quantified levels of sugars and acids/amino acids shown in Figure 28 demonstrate that the separations between locations observed by PCA (Figure 24) are mainly owing to increased levels of polar metabolites.

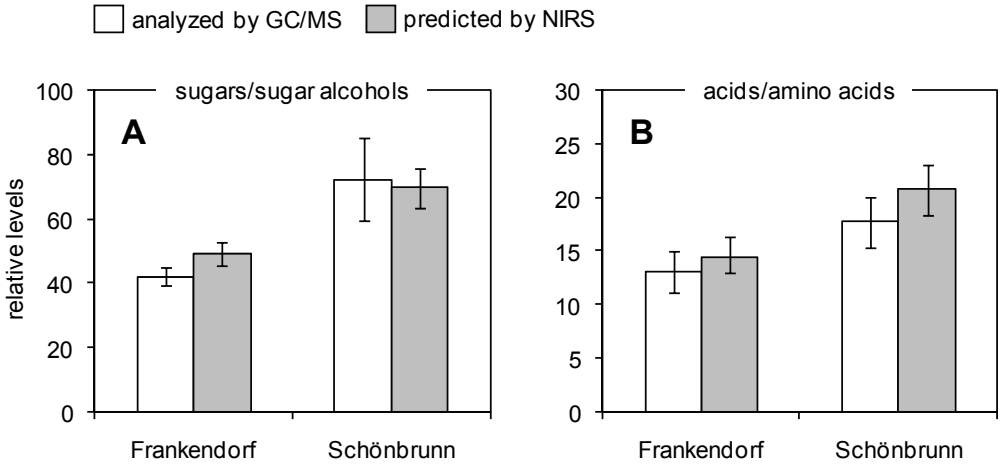

Figure 28 Comparison of maize samples grown at two locations with different input systems (conventional/organic) by relative levels of the sum parameters for sugars and sugar alcohols (A, as described in Table 1) and for acids/amino acids (B, as described in Table 3) as determined by GC/MS-metabolite profiling (light bars) and as predicted by NIRS models (grey bars) for sugars and acids/amino acids, respectively. Relative signals were calculated on the basis of the respective internal standard. Mean values and the corresponding standard errors were calculated for three maize cultivars.

Influence of Growing Season

For GC/MS data from maize cultivars grown in three consecutive growing seasons Principal Component Analysis revealed the most prominent separation of years for acids/amino acids (Figure 25B). Therefore, the respective sum parameter was compared as determined by GC/MS and as predicted by NIRS. Samples harvested in the season 2004 that were separated by PCA exhibited higher levels than samples from 2005 or 2006 (Figure 29). The results obtained by NIRS prediction were in good agreement with the levels as determined by GC/MS. The strong impact of growing season on the chemical composition of maize kernels is confirmed by results shown earlier in chapter 4.2, in which GC/MS-metabolite profiling investigated the influence of growing season on the differentiation of cultivars and farming locations.

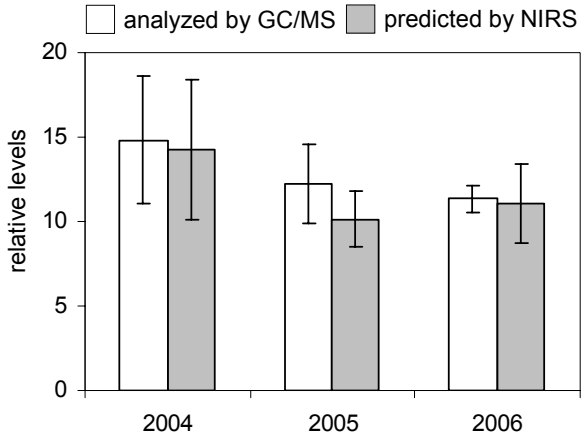

Figure 29 Comparison of maize samples grown in three consecutive seasons (2004, 2005, 2006) for three maize cultivars by relative levels of the sum parameters for acids/amino acids (Table 2) as determined by GC/MS-metabolite profiling (light bars) and as predicted by NIRS models (grey bars). Relative signals were calculated on the basis of the respective internal standard. Mean values and the corresponding standard errors were calculated for three maize cultivars.

4.5.5 Conclusions

The potential of NIRS as a profiling tool for the screening of maize samples was compared to GC/MS-based metabolite profiling. PCA evaluation of cultivars grown at different environments revealed the suitability of the full spectrum NIRS approach for the qualitative investigation of maize kernel composition. Major impact factors such as environment and genetic background could be equally determined by both GC/MS

and NIRS profiling. Multivariate statistics enabled the quantification of basic metabolic changes underlying the observed PCA clusters and allowed the differentiation of samples exhibiting low, medium and high levels of sugars or acids/amino acids. The straight forward use of GC/MS data as a basis for building NIRS prediction models demonstrated the robustness and suitability of GC/MS-metabolite profiling for both routine analysis and as reference data for spectroscopic methods. NIRS as a rapid screening approach may serve as a valuable tool for the pre-assessment of large sample sets prior to a more thorough analysis by GC/MS-metabolite profiling, thus demonstrating its potential for the investigation of crops in complex metabolomic studies and breeding programs.

5 SUMMARY

A comparative metabolite profiling approach based on gas chromatography-mass spectrometry (GC/MS) was applied to investigate the impact of genetic background, environment and farming practice on the chemical composition of maize (*Zea mays*) grain. The metabolite profiling protocol involved sub-fractionation of the metabolites and allowed the assessment of about 300 distinct analytes from different chemical classes (polar to lipophilic), of which 167 could be identified. Recoveries and intra-laboratory repeatability determined for selected representatives from the different chemical classes confirmed the suitability of the applied metabolite profiling procedure for the comprehensive assessment of maize grain composition.

The metabolite profiles of four maize cultivars that differed in their maturity classification were compared using principal component analysis (PCA) over three consecutive growing seasons. This revealed a strong separation of one cultivar in the first growing season, which could be explained by the immaturity of the kernels of this cultivar compared with the others in the field trial. Further evaluations by pairwise comparison using Student's *t*-test and analysis of variance (ANOVA) showed that the growing season was the most prominent impact factor driving variation of the metabolite pool. An increased understanding of metabolic variation was achieved by analysis of a second sample set comprising one cultivar grown for three years at four locations. The analysis of sub-fractions allowed the assessment of the contribution of metabolites from different chemical classes. The data obtained do not indicate that one of the fractions is generally more important; each of them may add information to explain genotype or environmental effects on crop compositions.

Maize kernels grown conventionally and organically, respectively, were investigated using GC/MS-based metabolite profiling. By analysis of three cultivars grown at two locations with different input systems and at a third location, where both organic and conventional farming were applied, the impact of the growing regime on the metabolite spectrum should be put into the context of natural variability. The metabolite profiling data were statistically assessed via PCA and ANOVA. The PCA demonstrated that the observed separations were mainly due to genetic differences (cultivars) and to environmental influences. The different input systems

(conventional/organic) only led to minor differentiations. ANOVA and quantification of selected constituents confirmed these observations. Only three metabolites (malic acid, *myo*-inositol and phosphate) were consistently different due to the employed input system if samples from all field trials were considered.

Two sets of transgenic maize lines and their isogenic counterparts were grown in South Africa (GM-Bt, GM-RR, non-GM) and Bavaria, Germany (GM-Bt, non-GM). Multivariate assessment by PCA of the GC/MS profiling data allowed the investigation of the potential effect of the genetic modification on the maize metabolite profiles. For evaluation of the results in the light of natural variation, samples were considered from different environments including farming locations and growing seasons. The data generated revealed that environmental influences (farming location, growing season) had a stronger overall effect on the metabolite profiles of the investigated maize genotypes than the genetic modification.

The potential of Near Infrared Spectroscopy (NIRS) as a profiling tool for the screening of maize samples was compared to GC/MS-based metabolite profiling. Qualitative assessment by PCA of three cultivars grown at different locations and in different growing seasons demonstrated the equal determination of major impact factors such as environment and genetic background on maize kernel composition by both NIRS and GC/MS profiling. Substantiation of potential differences was evaluated based on NIRS calibration models developed for the total content of sugars and acids/amino acids as determined by GC/MS-metabolite profiling. The models enabled the quantification of basic metabolic changes underlying the observed PCA clusters and allowed the differentiation of samples exhibiting low, medium and high levels of sugars or acids/amino acids. The straight forward use of GC/MS data as a basis for building NIRS prediction models demonstrated the robustness and suitability of GC/MS-metabolite profiling for both routine analysis and as reference data for spectroscopic methods. NIRS was shown to be a valuable tool for the pre-assessment of sample sets complementing existing GC/MS approaches, used subsequently for a more thorough analysis of samples of interest. The type of comparative datasets generated by both GC/MS and NIRS may serve as objective basis for crop assessment and the data confirm the potential of metabolite profiling to assist in breeding and farming approaches.

6 ZUSAMMENFASSUNG

Ein auf Gaschromatographie/Massenspektrometrie basierender Metabolite Profiling Ansatz wurde verwendet, um die Einflüsse von genetischem Hintergrund, Umwelt und Anbaupraxis auf die chemische Zusammensetzung von Mais (*Zea mays*) zu untersuchen.

Die Metabolite Profiling Methodik basiert auf Extraktion und Fraktionierung und erlaubt die Erfassung von ca. 300 individuellen Analyten (polar bis lipophil), von welchen 167 identifiziert werden konnten. Für ausgewählte Vertreter der verschiedenen chemischen Klassen zeigten Ergebnisse zur Wiederfindung und labor-internen Wiederholbarkeit die Eignung des verwendeten Metabolite Profilings zur umfassenden Untersuchung der Zusammensetzung von Mais.

Ein Vergleich – über drei aufeinanderfolgende Anbaujahre – der Metabolitenprofile von vier Maissorten, welche in ihrem Reifeverhalten variierten, wurde mittels Hauptkomponentenanalyse (PCA) durchgeführt. Die deutliche Abtrennung einer der Sorten in der ersten Anbausaison konnte auf den unreifen Zustand dieser Maiskörner zurückgeführt werden. Weiterführende Untersuchungen durch paarweisen Vergleich mittels *t*-Test und Varianzanalyse (ANOVA) zeigten, dass die Anbausaison den größten Einfluss auf die Variation innerhalb des Metabolitenpools hatte. Ein tieferes Verständnis metabolischer Schwankungen wurde durch die Betrachtung eines zweiten Probensets erreicht, welches Proben einer Sorte an vier Standorten im Zeitraum von drei Jahren enthielt. Die Analyse einzelner Fraktionen erlaubte die Erfassung des Beitrages von Metaboliten verschiedener Stoffklassen. Die erhaltenen Daten legen nahe, dass keine der Fraktionen grundsätzlich einen größeren Einfluss hat; vielmehr kann jede Fraktion Informationen enthalten, welche für die Erklärung der Effekte von Genotyp oder Umwelteinflüssen auf die Zusammensetzung von Getreide eine Rolle spielen.

Maiskörner aus konventionellem und biologischem Anbau wurden mittels GC/MS-Metabolite Profiling untersucht. Durch die Analyse dreier Sorten von zwei Anbauorten mit unterschiedlicher Anbaupraxis und einem dritten Anbauort, an welchem sowohl konventionelle als auch biologische Landwirtschaft durchgeführt werden, wurde der Einfluss der Anbaupraxis auf das Metabolitenspektrum im Licht

der natürlichen Schwankungsbreite betrachtet. Die Metabolite Profiling Daten wurden statistisch mittels PCA und ANOVA untersucht. Die PCA zeigte, dass die beobachteten Trennungen hauptsächlich auf genetische Unterschiede (Sorten) und Umwelteinflüsse zurück zu führen waren. Die Anbaupraxis (konventionell/biologisch) führte lediglich zu geringfügigen Unterschieden. ANOVA und die Quantifizierung ausgewählter Inhaltsstoffe bestätigten diese Beobachtung. Unter Berücksichtigung von Proben aller Feldversuche wiesen nur drei Metabolite (Äpfelsäure, *myo*-Inositol und Phosphat) konsistent unterschiedliche Gehalte zwischen den beiden Anbaupraktiken auf.

Zwei Sets gentechnisch veränderter Maislinien und deren entsprechenden Elternlinien wurden in Südafrika (GM-Bt, GM-RR, non-GM) und in Bayern (GM-Bt, non-GM) angebaut. Multivariate Auswertung der GC/MS Profiling Daten mittels PCA erlaubten die Untersuchung des potentiellen Einflusses der genetischen Modifikation auf die Metabolitenprofile von Mais. Zur Bewertung der Ergebnisse im Licht der natürlichen Variabilität wurden verschiedene Umgebungseinflüsse (Anbauort, Anbaujahr) berücksichtigt. Die ermittelten Daten zeigen, dass Umweltbedingungen auf die Metabolitenspektren der untersuchten Maisgenotypen einen größeren Einfluss hatten als die gentechnische Veränderung.

Nahinfrarotspektroskopie (NIRS) als potentieller Profiling Ansatz für das Screening von Maisproben wurde mit GC/MS-basiertem Metabolite Profiling verglichen. Die qualitative Untersuchung dreier Sorten aus unterschiedlichen Anbauorten und Anbaujahren mittels PCA zeigte eine vergleichbare Eignung von sowohl NIRS als auch GC/MS zur Bestimmung der Haupteinflussfaktoren auf die Zusammensetzung von Maiskörnern, wie zum Beispiel Umwelteinflüsse und genetischer Hintergrund. Zur Substantivierung möglicher Unterschiede wurden NIRS Kalibrierungsmodelle für die Gesamtgehalte an Zuckern und Säuren/Aminosäuren entwickelt, die zuvor mittels GC/MS-Metabolite Profiling bestimmt wurden. Die Modelle ermöglichten die Quantifizierung von metabolischen Veränderungen, welche die PCA Gruppierungen begründen, und erlauben die Differenzierung von Proben mit niedrigen, mittleren und hohen Gehalten an Zuckern oder Säuren/Aminosäuren. Die Möglichkeit der direkten Verwendung von GC/MS Daten als Basis für die Erstellung von NIRS Kalibrierungsmodellen belegte die Robustheit und Eignung der beschriebenen

GC/MS-Metabolite Profiling Methode sowohl zur Routineanalyse als auch als Referenzmethode für spektroskopische Anwendungen. NIRS erwies sich als wertvolles Werkzeug für die Einstufung von großen Probensets. Es ergänzt damit vorhandene GC/MS Methoden, welche im Anschluss für eine weiterführende Analytik interessanter Proben eingesetzt werden können. Die Art von umfassenden Daten, welche durch die beiden Ansätze GC/MS und NIRS gewonnen werden, können als objektive Basis für die Untersuchung von Getreide angesehen werden. Die im Rahmen dieser Arbeit erhaltenen Daten bestätigen das Potential von Metabolite Profiling zur analytischen Bewertung im Rahmen von Züchtung und Anbau.

7 REFERENCES

Allison, M.J. and Maule, A.P. (1991) Monitoring biochemical change in malting by use of reconstructed NIR spectra. *Proceedings of the Congress - European Brewery Convention*, 23rd, 133-8.

Allwood, J.W., Ellis, D.I. and Goodacre, R. (2008) Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiologia Plantarum*, 132, 117-135.

Alvarez, S., Marsh, E.L., Schroeder, S.G. and Schachtman, D.P. (2008) Metabolomic and proteomic changes in the xylem sap of maize under drought. *Plant, Cell & Environment*, 31, 325-340.

Arruda, P., da Silva, W.J. and Teixeira, J.P.F. (1978) Protein and free amino acids in a high lysine maize double mutant. *Phytochemistry*, 17, 1217-1218.

Ashton, W.D. (1972) *The Logit Transformation with special reference to its uses in Bioassay*. Charles Griffin & Company Ltd., London.

Ausloos, P., Clifton, C.L., Lias, S.G., Mikaya, A.I., Stein, S.E., Tchekhovskoi, D.V., Sparkman, O.D., Zaikin, V. and Zhu, D. (1999) The critical evaluation of a comprehensive mass spectral library. *Journal of the American Society for Mass Spectrometry*, 10, 287-299.

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

Barros, E., Lezar, S., Anttonen, M.J., Dijk, J.P.v., Röhlig, R.M., Kok, E.J. and Engel, K.-H. (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnology Journal*, 8, 1-16.

Baye, T.M., Pearson, T.C. and Settles, A.M. (2006) Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *Journal of Cereal Science*, 43, 236-243.

Bertrand, D., Robert, P. and Loisel, W. (1985) Identification of some wheat varieties by near infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, 36, 1120-1124.

BewRL (1967) Richtlinien für die Bewertung des land- und forstwirtschaftlichen Vermögens vom 17. November 1967. *Bundessteuerblatt 1967*, Teil I, 397.

Blanco, M., Alcalá, M., González, J.M. and Torras, E. (2006) A process analytical technology approach based on near infrared spectroscopy: Tablet hardness, content uniformity, and dissolution test measurements of intact tablets. *Journal of Pharmaceutical Sciences*, 95, 2137-2144.

Bourn, D. and Prescott, J. (2002) A Comparison of the Nutritional Value, Sensory Qualities, and Food Safety of Organically and Conventionally Produced Foods. *Critical Reviews in Food Science and Nutrition*, 42, 1-34.

Bundessortenamt (2008) *Beschreibende Sortenliste - Getreide, Mais, Ölfrüchte, Leguminosen (großkörnig), Hackfrüchte 2008 (außer Kartoffeln)*. Bundessortenamt, Hannover.

Castro, C. and Manetti, C. (2007) A multiway approach to analyze metabonomic data: a study of maize seeds development. *Analytical Biochemistry*, 371, 194-200.

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

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

Cozzolino, D., Flood, L., Bellon, J., Gishen, M. and Lopes, M.D.B. (2006) Combining near infrared spectroscopy and multivariate analysis as a tool to differentiate different strains of Saccharomyces cerevisiae: a metabolomic study. *Yeast*, 23, 1089-1096.

Daftary, R.D. and Pomeranz, Y. (1965) Changes in Lipid Composition in Maturing Wheat. *Journal of Food Science*, 30, 577-582.

Dangour, A.D., Dodhia, S.K., Hayter, A., Allen, E., Lock, K. and Uauy, R. (2009) Nutritional quality of organic foods: a systematic review. *American Journal of Clinical Nutrition*, 90, 680-685.

Davies, H.V. (2007) Metabolomics: applications in functional biodiversity analysis in potato. *Acta Horticulturae*, 745, 471-483.

De la Roza, B., Martinez, A., Santos, B., González, J. and Gómez, G. (1998) The estimation of crude protein and dry matter degradability of maize and grass silages by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 6, 145-152.

Defernez, M., Gunning, Y.M., Parr, A.J., Shepherd, L.V.T., Davies, H.V. and Colquhoun, I.J. (2004) NMR and HPLC-UV Profiling of Potatoes with Genetic Modifications to Metabolic Pathways. *Journal of Agricultural and Food Chemistry*, 52, 6075-6085.

Delwiche, S.R. and Norris, K.H. (1993) Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chemistry*, 70, 29-29.

Duvick, D.N. (1952) Free amino acids in the developing endosperm of maize. *American Journal of Botany*, 39, 656-61.

Eder, J. (2009) Growing Performance of Cultivar Lukas. *Bayerische Landesanstalt für Landwirtschaft (LfL)*, personal communication.

EEC (1991) Council Regulation (EEC) No 2092/91 of June 24 1991 on organic production of agricultural products and indications referring thereto on agriculturalproducts and foodstuffs.

EFSA (2006) Guidance Document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. *The EFSA Journal*, 1-100.

El-Sanhoty, R., Shahwan, T. and Ramadan, M.F. (2006) Application of artificial neural networks to develop a classification model between genetically modified maize (Bt-176) and conventional maize by applying lipid analysis data. *Journal of Food Composition and Analysis*, 19, 628-636.

Emura, K., Yamanaka, S., Isoda, H. and Watanabe, K.N. (2006) Estimation for Different Genotypes of Plants based on DNA Analysis using Near-infrared (NIR) and Fourier-transform Infrared (FT-IR) Spectroscopy. *Breeding Science*, 56, 399-403.

Esteban-Díez, I., González-Sáiz, J.M., Sáenz-González, C. and Pizarro, C. (2007) Coffee varietal differentiation based on near infrared spectroscopy. *Talanta*, 71, 221-229.

Fait, A., Hanhineva, K., Beleggia, R., Dai, N., Rogachev, I., Nikiforova, V.J., Fernie, A.R. and Aharoni, A. (2008) Reconfiguration of the Achene and Receptacle Metabolic Networks during Strawberry Fruit Development. *Plant Physiology*, 148, 730-750.

FAO (2008) ProdSTAT: Crops, Food and Agriculture Organization of the United Nations

FAO/WHO (2000) Safety aspects of genetically modified foods of plant origin.

Fernie, A.R. and Schauer, N. (2008) Metabolomics-assisted breeding: a viable option for crop improvement? *Trends in Genetics*, 25, 39-48.

Ferreira, A.F.P., Boelens, H.F.M. and Westerhuis, J.A. (2005) Inline Monitoring of Butane Isomers Adsorption on MFI Using Near-Infrared Spectroscopy: Drift Correction in Time Based Experiments. *Langmuir*, 21, 6830-6835.

Fiehn, O. (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics*, 2, 155-168.

Fiehn, O. (2002) Metabolomics - the link between genotypes and phenotypes. *Plant Molecular Biology*, 48, 155-171.

Fiehn, O. (2008) Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends in Analytical Chemistry*, 27, 261-269.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nature Biotechnology*, 18, 1157-1161.

Fontaine, J., Horr, J. and Schirmer, B. (2001) Near-Infrared Reflectance Spectroscopy Enables the Fast and Accurate Prediction of the Essential Amino Acid Contents in Soy, Rapeseed Meal, Sunflower Meal, Peas, Fishmeal, Meat Meal Products, and Poultry Meal. *Journal of Agricultural and Food Chemistry*, 49, 57-66.

Frank, T., Meuleye, B.S., Miller, A., Shu, Q.-Y. and Engel, K.-H. (2007) Metabolite Profiling of Two Low Phytic Acid (lpa) Rice Mutants. *Journal of Agricultural and Food Chemistry*, 55, 11011-11019.

Frank, T., Scholz, B., Peter, S. and Engel, K.-H. (2011) Metabolite profiling of barley: Influence of the malting process. *Food Chemistry*, 124, 948-957.

Frenzel, T., Miller, A. and Engel, K.-H. (2002) Metabolite profiling-a fractionation method for analysis of major and minor compounds in rice grains. *Cereal Chemistry*, 79, 215-221.

Frenzel, T., Miller, A. and Engel, K.-H. (2003) A methodology for automated comparative analysis of metabolite profiling data. *European Food Research and Technology*, 216, 335-342.

FSA (2005) Safety Assessment of Novel Foods - Report of GO2 Research Programme Review, UK Food Standards Agency.

Fuentes, M., Govaerts, B., De León, F., Hidalgo, C., Dendooven, L., Sayre, K.D. and Etchevers, J. (2009) Fourteen years of applying zero and conventional tillage, crop rotation and residue management systems and its effect on physical and chemical soil quality. *European Journal of Agronomy*, 30, 228-237.

Fujimoto, H., Itoh, K., Yamamoto, M., Kyozuka, J. and Shimamoto, K. (1993) Insect resistant rice generated by introduction of a modified d-endotoxin gene of Bacillus thuringiensis. *Bio/Technology*, 11, 1151-1155.

Ghorbani, R., Wilcockson, S., Koocheki, A. and Leifert, C. (2008) Soil management for sustainable crop disease control: a review. *Environmental Chemistry Letters*, 6, 149-162.

Goddu, R.R. and Delker, D.A. (1958) Determination of terminal epoxides by near-infrared spectrophotemetry. *Analytical Chemistry*, 30, 2013-2016.

Gong, W., Yan, X.-Y., Wang, J.-Y., Hu, T.-X. and Gong, Y.-B. (2009) Long-term manuring and fertilization effects on soil organic carbon pools under a wheat-maize cropping system in North China Plain. *Plant and Soil*, 314, 67-76.

Gordy, W. and Martin, P.C. (1937) The Infrared Absorption of HCl in Oxygenated Solvents. *Physical Review*, 52, 1075.

Govorun, V.M. and Archakov, A.I. (2002) Proteomic Technologies in Modern Biomedical Science. *Biochemistry (Moscow)*, 67, 1109-1123.

Greiff, W.R., Morgan, W.T. and Ponte, J.M. (2002) The role of variance in term weighting for probabilistic information retrieval, *Proceedings of the eleventh international conference on Information and knowledge management*, ACM.

Harrigan, G., Martino-Catt, S. and Glenn, K. (2007a) Metabolomics, metabolic diversity and genetic variation in crops. *Metabolomics*, 3, 259-272.

Harrigan, G.G., Stork, L.G., Riordan, S.G., Reynolds, T.L., Ridley, W.P., Masucci, J.D., MacIsaac, S., Halls, S.C., Orth, R., Smith, R.G., Wen, L., Brown, W.E., Welsch, M., Riley, R., McFarland, D., Pandravada, A. and Glenn, K.C. (2007b) Impact of Genetics and Environment on Nutritional and Metabolite Components of Maize Grain. *Journal of Agricultural and Food Chemistry*, 55, 6177-6185.

Harrigan, G.G., Stork, L.G., Riordan, S.G., Ridley, W.P., MacIsaac, S., Halls, S.C., Orth, R., Rau, D., Smith, R.G., Wen, L., Brown, W.E., Riley, R., Sun, D., Modiano, S., Pester, T., Lund, A. and Nelson, D. (2007c) Metabolite Analyses of Grain from Maize Hybrids Grown in the United States under Drought and Watered Conditions during the 2002 Field Season. *Journal of Agricultural and Food Chemistry*, 55, 6169-6176.

Hazebroek, J., Harp, T., Shi, J. and Wang, H. (2007) Metabolomic Analysis of Low Phytic Acid Maize Kernels in Nikolau, B.J. and Wurtele, E.S. (Eds), *Concepts in Plant Metabolomics*, Springer, Berlin. pp. 221-237.

Herencia, J.F., Ruiz-Porras, J.C., Melero, S., Garcia-Galavis, P.A., Morillo, E. and Maqueda, C. (2007) Comparison between Organic and Mineral Fertilization for Soil Fertility Levels, Crop Macronutrient Concentrations, and Yield. *Agronomy Journal*, 99, 973-983.

Herschel, W. (1800a) Experiments on the refrangibility of the rays of the sun. *Philosophical Transactions of the Royal Society London*, 90, 284-292.

Herschel, W. (1800b) Investigation of the powers of the prismatic colours to heat and illuminate objects; with Remarks, that prove the different refrangibility of radiant heat. To which is added, and inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. *Philosophical Transactions of the Royal Society London*, 90, 255-283.

Hirel, B., Andrieu, B., Valadier, M.-H., Renard, S., Quilleré, I., Chelle, M., Pommel, B., Fournier, C. and Drouet, J.-L. (2005) Physiology of maize II: Identification of physiological markers representative of the nitrogen status of maize (Zea mays) leaves during grain filling. *Physiologia Plantarum*, 124, 178-188.

Hruschka, W.R. (2001) *Near-infrared technology in the agricultural and food industries*. AACC Press, St Paul, MN, USA.

Ikegaya, K. (1990) Determination of chemical constituents in processed green tea by near infrared analysis. *Japan Agricultural Research Quarterly*, 24, 49-53.

IPS (2009) Agrarmeteorologisches Messnetz Bayern, Institut für Pflanzenschutz, Bayerische Landesanstalt für Landwirtschaft (LfL).

Jackson, J.E. (1991) *A user's guide to principal components*. Wiley-Interscience.

Jacob, J.P. (2007) Nutrient content of organically grown feedstuffs. *Journal of Applied Poultry Research*, 16, 642-651.

James, C. (2003) *Global Review of Commercialized Transgenic Crops: 2002 Feature: Bt Maize*. ISAAA, Ithaca, NY.

Jugenheimer, R.W. (1976) *Corn. Improvement, seed production, and uses*. Wiley and Sons. New York, NY, USA.

Kamal-Eldin, A., Appelqvist, L.Å., Yousif, G. and Iskander, G.M. (1992) Seed lipids of Sesamum indicum and related wild species in Sudan. The sterols. *Journal of the Science of Food and Agriculture*, 59, 327-334.

Kaye, W., Canon, C. and Devaney, R.G. (1951) Modification of a Beckman model DU spectrophotometer for automatic operation at 210-2700 µm. *J Opt Soc Am*, 41, 658-664.

Kirchhoff, E. (2008) *Series-Fachmann-Kraut: Food Composition and Nutrition Tables*, 7th edn, Stuttgart, Germany.

Kopka, J. (2006) Current challenges and developments in GC-MS based metabolite profiling technology. *Journal of Biotechnology*, 124, 312-22.

Kopka, J., Fernie, A., Weckwerth, W., Gibon, Y. and Stitt, M. (2004) Metabolite profiling in plant biology: platforms and destinations. *Genome Biology*, 5, 109.

Kopka, J., Schauer, N., Krüger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R. and Steinhauser, D. (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, 21, 1635-1638.

Kuiper, H.A., Kleter, G.A., Noteborn, H.P.J.M. and Kok, E.J. (2001) Assessment of the food safety issues related to genetically modified foods. *The Plant Journal*, 27, 503-528.

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

Kusano, M., Fukushima, A., Kobayashi, M., Hayashi, N., Jonsson, P., Moritz, T., Ebana, K. and Saito, K. (2007) Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *Journal of Chromatography B*, 855, 71-79.

Lampkin, N., Padel, S. and Foster, C. (2000) Organic farming. *Cap Regimes and the European Countryside: Prospects for Integrations Between Agricultural, Regional and Environmental Policies*, 221.

Langenkämper, G., Zörb, C., Seifert, M. and Betsche, T. (2006) Concentration of minerals and antioxidants in ecologically and conventionally grown wheat. *Getreidetechnologie*, 60, 295-300.

Lehesranta, S.J., Koistinen, K.M., Massat, N., Davies, H.V., Shepherd, L.V.T., McNicol, J.W., Cakmak, I., Cooper, J., Lück, L., Kärenlampi, S.O. and Leifert, C. (2007) Effects of agricultural production systems and their components on protein profiles of potato tubers. *Proteomics*, 7, 597-604.

Leigh, J.H. and Kinnear, T.C. (1980) On Interaction Classification. *Educational and Psychological Measurement*, 40, 841-843.

Loewus, F.A. and Murthy, P.P.N. (2000) myo-Inositol metabolism in plants. *Plant Science*, 150, 1-19.

Lozovaya, V., Ulanov, A., Lygin, A., Duncan, D. and Widholm, J. (2006) Biochemical features of maize tissues with different capacities to regenerate plants. *Planta*, 224, 1385-1399.

Lu, C., Hawkesford, M.J., Barraclough, P.B., Poulton, P.R., Wilson, I.D., Barker, G.L. and Edwards, K.J. (2005) Markedly Different Gene Expression in Wheat Grown with Organic or Inorganic Fertilizer. *Proceedings: Biological Sciences*, 272, 1901-1908.

Lucca, P., Hurrell, R. and Potrykus, I. (2001) Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. *TAG Theoretical and Applied Genetics*, 102, 392-397.

Ma, J., Morrow, D., Fernandes, J. and Walbot, V. (2006) Comparative profiling of the sense and antisense transcriptome of maize lines. *Genome Biology*, 7, R22.

Mäder, P., Fliessbach, A., Dubois, D., Gunst, L., Fried, P. and Niggli, U. (2002) Soil Fertility and Biodiversity in Organic Farming. *Science*, 296, 1694-1697.

Manetti, C., Bianchetti, C., Casciani, L., Castro, C., Di Cocco, M.E., Miccheli, A., Motto, M. and Conti, F. (2006) A metabonomic study of transgenic maize (Zea mays) seeds revealed variations in osmolytes and branched amino acids. *Journal of Experimental Botany*, 57, 2613-2625.

Martens, H. and Naes, T. (1992) *Multivariate calibration*. John Wiley & Sons Ltd.

Matissek, R. and Steiner, G. (2005) *Lebensmittelanalytik: Grundzüge, Methoden, Anwendungen*. Springer-Verlag, Berlin Heidelberg.

McClure, W.F. and Tsuchikawa, S. (2007) Instruments in Ozaki, Y., McClure, W.F. and Christy, A.A. (Eds), *Near-infrared spectroscopy in food science and technology*, John Wiley & Sons, Inc., New Jersey. pp. 75-107.

Meyna, S. (2005) Freie und triglycerid-gebundene Hydroxyfettsäuren in Gerste und Malz und ihre Bedeutung für die Geschmacksstabilität des Bieres, *Fakultät III (Prozesswissenschaften)*, Technische Universität Berlin, pp. 126.

Miller, L.T. (1982) Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *Journal of Clinical Microbiology*, 16, 584-586.

Miralbés, C. (2004) Quality control in the milling industry using near infrared transmittance spectroscopy. *Food Chemistry*, 88, 621-628.

Miyanishi, T., Tutimoto, S., Ogura, M. and Iio, T. (1991) Studies on the taste and flavor of sweet corn. I. Changes in chemical components in sweet corn (cv. Golden Earlipak) kernels during maturation. *Nippon Shokuhin Kogyo Gakkaishi*, 38, 758-64.

Moose, S.P., Dudley, J.W. and Rocheford, T.R. (2004) Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends in Plant Science*, 9, 358-364.

Mounet, F., Lemaire-Chamley, M., Maucourt, M., Cabasson, C., Giraudel, J.-L., Deborde, C., Lessire, R., Gallusci, P., Bertrand, A., Gaudillère, M., Rothan, C., Rolin, D. and Moing, A. (2007) Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics*, 3, 273-288.

Na Jom, K., Frank, T. and Engel, K.-H. (2010) A metabolite profiling approach to follow the sprouting process of mung beans (Vigna radiata). *Metabolomics*, 1-16.

Norris, K.H. (1963) A simple spectroradiometer for the 0.4- to 1.2-micron region. *ASAE Paper No. 63-329*, 1-5.

Oard, J.H., Linscombe, S.D., Braverman, M.P., Jodari, F., Blouin, D.C., Leech, M., Kohli, A., Vain, P., Cooley, J.C. and Christou, P. (1996) Development, field evaluation, and agronomic performance of transgenic herbicide resistant rice. *Molecular Breeding*, 2, 359-368.

OECD (1993) Safety evaluation of foods derived by modern biotechnology: concept and principles.

OECD (2002) Consensus Document on Compositional Considerations for New Varieties of Maize (Zea Mays): Key Food and Feed Nutrients, Anti-Nutrients and Secondary Plant Metabolites, Environment Directorate.

Osborne, B.G. (2000) Near-infrared spectroscopy in food analysis, *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Chichester.

Osborne, B.G. (2008) NIR analysis of cereal products. *Practical Spectroscopy*, 35, 399-413.

Perez-Marin, D., Garrido-Varo, A. and Guerrero, J.E. (2007) Non-linear regression methods in NIRS quantitative analysis. *Talanta*, 72, 28-42.

Perez, D.P., Sanchez, M.T., Cano, G. and Garrido, A. (2001) Authentication of Green Asparagus Varieties by Near-Infrared Reflectance Spectroscopy. *Journal of Food Science*, 66, 323-327.

Pierce, K.M., Hope, J.L., Hoggard, J.C. and Synovec, R.E. (2006) A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC x GC-TOFMS) separations of metabolites in plant samples. *Talanta*, 70, 797-804.

Pingali, P.L. (2001) Meeting world maize needs: Technological opportunities and priorities for the public sector. *CIMMYT*, 1999/2000 World Maize Facts and Trends.

Plyler, E.K. and Williams, D. (1936) The Infrared Absorption Spectra of HCl in Benzene. *Physical Review*, 49, 215.

Qin, Z.-Q., Cai, S.-S., Xie, H., Si, H., Zhan, J.-L. and Yan, T.-J. (2007) Prediction of total nicotine, total sugar, reducing sugar and total nitrogen in tobacco with AOTF-NIR. *Yancao Keji*, 2, 30-34.

Quilitzsch, R., Baranska, M., Schulz, H. and Hoberg, E. (2005) Fast determination of carrot quality by spectroscopy methods in the UV-VIS, NIR and IR range. *Journal of Applied Botany and Food Quality*, 79, 163 - 167.

Raghothama, K.G. (1999) Phosphate Acquisition. *Annual Review of Plant Physiology and Plant Molecular Biology*, 50, 665-693.

Reynolds, T.L., Nemeth, M.A., Glenn, K.C., Ridley, W.P. and Astwood, J.D. (2005) Natural variability of metabolites in maize grain: Differences due to genetic background. *Journal of Agricultural and Food Chemistry*, 53, 10061-10067.

Ridley, W.P., Sidhu, R.S., Pyla, P.D., Nemeth, M.A., Breeze, M.L. and Astwood, J.D. (2002) Comparison of the Nutritional Profile of Glyphosate-Tolerant Corn Event NK603 with That of Conventional Corn (Zea mays L.). *Journal of Agricultural and Food Chemistry*, 50, 7235-7243.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N. and Willmitzer, L. (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant Journal*, 23, 131-142.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A. and Jent, N. (2007) A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, 44, 683-700.

Rosipal, R. and Krämer, N. (2006) Overview and Recent Advances in Partial Least Squares in Saunders, C. (Ed), *Subspace, Latent Structure and Feature Selection*, Springer-Verlag, Berlin Heidelberg. pp. 34-51.

Sachs, L. and Hedderich, J. (2006) *Angewandte Statistik*. Springer.

Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. *Trends in Plant Science*, 13, 36-43.

Sandorfy, C., Buchet, R. and Lachenal, G. (2007) Principles of Molecular Vibrations for Near-Infrared Spectroscopy in Ozaki, Y., McClure, W.F. and Christy, A.A. (Eds), *Near-infrared spectroscopy in food science and technology*, John Wiley & Sons, Inc., New Jersey. pp. 11-46.

Schauer, N. and Fernie, A.R. (2006) Plant metabolomics: Towards biological function and mechanism. *Trends in Plant Science*, 11, 508-516.

Scherz, H. and Senser, F. (2000) *Food Composition and Nutrition Tables*, 6 edn.

SciFinder (2009) Query on the concept "organic farming", CAS.

Seebauer, J.R., Moose, S.P., Fabbri, B.J., Crossland, L.D. and Below, F.E. (2004) Amino acid metabolism in maize earshoots. Implications for assimilate preconditioning and nitrogen signaling. *Plant Physiology*, 136, 4326-4334.

Semel, Y., Schauer, N., Rössner, U., Zamir, D. and Fernie, A. (2007) Metabolite analysis for the comparison of irrigated and non-irrigated field grown tomato of varying genotype. *Metabolomics*, 3, 289-295.

Shenk, J.S., Workman, J.J., Jr. and Westerhaus, M.O. (2008) Application of NIR spectroscopy to agricultural products. *Pract. Spectrosc. FIELD Full Journal Title:Practical Spectroscopy*, 35, 347-386.

Shenk, J.S., Workman, J.J. and Westerhaus, M.O. (2001) Practical Spectroscopy in Burns, D.A. and Ciurczak, E.W. (Eds), *Handbook of Near-Infrared Analysis*, Marcel Dekker, New York. pp. 419-474.

Shu, X.-L., Frank, T., Shu, Q.-Y. and Engel, K.-H. (2008) Metabolite Profiling of Germinating Rice Seeds. *Journal of Agricultural and Food Chemistry*, 56, 11612-11620.

Sidhu, R.S., Hammond, B.G., Fuchs, R.L., Mutz, J.-N., Holden, L.R., George, B. and Olson, T. (2000) Glyphosate-Tolerant Corn: The Composition and Feeding Value of Grain from Glyphosate-Tolerant Corn Is Equivalent to That of Conventional Corn (Zea mays L.). *Journal of Agricultural and Food Chemistry*, 48, 2305-2312.

Skeie, S., Feten, G., Almoy, T., Ostlie, H. and Isaksson, T. (2006) The use of near infrared spectroscopy to predict selected free amino acids during cheese ripening. *International Dairy Journal*, 16, 236-242.

Steinhauser, D. and Kopka, J. (2007) Methods, applications and concepts of metabolite profiling: primary metabolism, *Plant Systems Biology*, Birkhäuser, Basel. pp. 171-194.

Tarpley, L., Duran, A.L., Kebrom, T.H. and Sumner, L.W. (2005) Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. *BMC Plant Biology*, 5.

Ter Braak, C.J.F. and Gremmen, N.J.M. (1987) Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moisture. *Plant Ecology*, 69, 79-87.

Tian, H.-q., Ying, Y.-b., Lu, H.-s., Fu, X.-p. and Yu, H.-y. (2007) Measurement of soluble solids content in watermelon by Vis/NIR diffuse transmittance technique. *Journal of Zhejiang University - Science B*, 8, 105-110.

Voelker, T. and Kinney, A.J. (2001) Variations in the Bionsynthesis of Seed-Storage Lipids. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52, 335-361.

Wang, W. and Paliwal, J. (2007) Near-infrared spectroscopy and imaging in food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, 1, 193-207.

Warman, P.R. and Havard, K.A. (1998) Yield, vitamin and mineral contents of organically and conventionally grown potatoes and sweet corn. *Agriculture, Ecosystems & Environment*, 68, 207-216.

Weber, E.J. (1969) Lipids of maturing grain of corn (Zea mays). I. Changes in lipid classes and fatty acid composition. *Journal of the American Oil Chemists' Society*, 46, 485-8.

Welle, R., Greten, W., Rietmann, B., Alley, S., Sinnaeve, G. and Dardenne, P. (2003) Near-Infrared Spectroscopy on Chopper to Measure Maize Forage Quality Parameters Online. *Crop Science*, 43, 1407-1413.

Wendland, M., Diepolder, M. and Capriel, P. (2007) *Leitfaden für die Düngung von Acker- und Grünland*. Bayerische Landesanstalt für Landwirtschaft (LfL).

Wheeler, O.H. (1959) Near infrared spectra of organic compounds. *Chem Rev*, 59, 629-666.

Whetsel, K.B. (1968) Near-infrared spetrophotometry in Edward, G.B.J. (Ed), *Applied Spectroscopy Reviews*, New York. pp. 1-67.

Willer, H. and Kilcher, L. (2009) *The World of Organic Agriculture - Statistics and Emerging Trends 2009*. IFOAM, Bonn; FiBL, Frick; ITC, Geneva.

Williams, D. (1936) The Infrared Absorption of Hydrogen Chloride in Nonionizing Solvents. *Physical Review*, 50, 719.

Williams, D. and Rogers, L.H. (1937) The Infrared Absorption Spectrum of Vitamin C1. *Journal of the American Chemical Society*, 59, 1422-1423.

Williams, P.C. and Sobering, D.C. (1996) How do we do it: a brief summary of the methods we use in developing near infrared calibrations. *Near infrared spectroscopy: The future waves*, 185–188.

Winter, C.K., Davis, S.F. and Tarver, T.G. (2006) Organic foods. *Food Technology*, 60, 44-48.

Woodcock, T., O'Donnell, C.P. and Downey, G. (2008) Review: Better quality food and beverages: the role of near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 16, 1-29.

Xu, Z. and Godber, J.S. (1999) Purification and Identification of Components of g-Oryzanol in Rice Bran Oil. *Journal of Agricultural and Food Chemistry*, 47, 2724-2728.

Xue, Z.-Y., Zhi, D.-Y., Xue, G.-P., Zhang, H., Zhao, Y.-X. and Xia, G.-M. (2004) Enhanced salt tolerance of transgenic wheat (Tritivum aestivum L.) expressing a vacuolar Na+/H+ antiporter gene with improved grain yields in saline soils in the field and a reduced level of leaf Na+. *Plant Science*, 167, 849-859.

Ye, X., Al-Babili, S., Kl, ouml, ti, A., Zhang, J., Lucca, P., Beyer, P. and Potrykus, I. (2000) Engineering the Provitamin A (-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science*, 287, 303-305.

Young, T.E. and Gallie, D.R. (2000) Programmed cell death during endosperm development. *Plant Molecular Biology*, 44, 283-301.

Zhang, H., Wang, B. and Xu, M. (2008) Effects of Inorganic Fertilizer Inputs on Grain Yields and Soil Properties in a Long-Term Wheat-Corn Cropping System in South China. *Communications in Soil Science and Plant Analysis*, 39, 1583-1599.

Zhou, G.X., Crocker, L., Xu, J., Tabora, J. and Ge, Z. (2006) In-line measurement of a drug substance via near infrared spectroscopy to ensure a robust crystallization process. *Journal of Pharmaceutical Sciences*, 95, 2337-2347.

Zörb, C., Betsche, T. and Langenkämper, G. (2009) Search for Diagnostic Proteins To Prove Authenticity of Organic Wheat Grains (Triticum aestivum L.). *Journal of Agricultural and Food Chemistry*, 57, 2932-2937.

Zörb, C., Langenkämper, G., Betsche, T., Niehaus, K. and Barsch, A. (2006) Metabolite Profiling of Wheat Grains (Triticum aestivum L.) from Organic and Conventional Agriculture. *Journal of Agricultural and Food Chemistry*, 54, 8301-8306.

8 PUBLICATIONS AND PRESENTATIONS

Röhlig, M.R., Eder, J. and Engel, K.-H. (2009) Metabolite profiling of maize grain: differentiation due to genetics and environment. *Metabolomics, 5*, 457–477.

Röhlig, M.R. and Engel, K.-H. (2010) Influence of the Input System (Conventional versus Organic Farming) on Metabolite Profiles of Maize (*Zea mays*) Kernels. *Journal of Agricultural and Food Chemistry, 58*, 3022–3030.

Barros, E., Lezar, S., Anttonen, M.J., van Dijk, J.P., Röhlig, M.R., Kok, E.J. and Engel, K.-H. (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechology Journal, 8*, 1–16.

Anttonen, M.J., Lehesranta, S., Auriola, S., Röhlig, M.R., Engel, K.-H. and Kärenlampi, S.O. (2010) Genetic and environmental influence on maize kernel proteome. *Journal of Proteome Research, 9*, 6160–6168.

Davies, H.V., Shepherd, L.V.T., Stewart, D., Frank, T., Röhlig, R. and Engel, K.-H. (2010) Metabolome variability in crop plant species – When, where, how much and so what? *Regulatory Toxicology and Pharmacology, 58*, 54–61.

Röhlig, M.R. and Engel, K.-H. (2006) Beiträge der Lebensmittelanalytik zur Bewertung von Lebensmittelqualität und -sicherheit. *Euroscience Open Forum (ESOF2006),* Ernährungstag München (Presentation).

Röhlig, M.R. and Engel, K.-H. (2009) Metabolite Profiling von Mais. *Lebensmittelchemische Gesellschaft (LChG), Regionalverband Bayern,* 60. Arbeitstagung 2009 (Presentation).

Presentations at meetings of the EU scientific project „SAFE FOODS—Promoting Food Safety through a New Integrated Risk Analysis Approach for Foods“ (EU FP6 project contract no. Food-CT-2004-50644).

„DIE DEFINITION
VON WAHNSINN IST,
IMMER WIEDER
DAS GLEICHE ZU
TUN UND ANDERE
ERGEBNISSE ZU
ERWARTEN.“

ALBERT EINSTEIN

LEBENSLAUF

Zur Person

Name	Richard Martin Röhlig
Geburtstag	17. Oktober 1978 in Leipzig
Staatsangehörigkeit	deutsch
Familienstand	verheiratet

Wissenschaftliche Tätigkeit

04.2005 – 10.2009	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Allgemeine Lebensmitteltechnologie (Prof. Dr. K.-H. Engel), Technische Universität München, Freising-Weihenstephan Anfertigen der vorliegenden Arbeit: „Metabolite Profiling of Maize Grain: Differentiation due to Genetics, Environment and Input System“
10.1999 – 03.2005	Studium „Technologie und Biotechnologie der Lebensmittel" an der Technischen Universität München. Diplomarbeit: „Analytische Charakterisierung von Raucharomen mittels Profiling (GC-FID)“

Beruf und Ausbildung

10.2009 –	Technologie und Applikation, Nestlé Deutschland AG
09.1998 – 09.1999	Zivildienst in der Blindenanstalt Nürnberg
09.1989 – 06.1998	Erlangung der allgemeinen Hochschulreife am Willstätter-Gymnasium Nürnberg
09.1985 – 09.1989	43. Polytechnische Oberschule „Walter Ulbricht“, Leipzig