universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Characterization of genetically modified maize by quantitative PCR and Next-Generation Sequencing"

verfasst von / submitted by

Alexandra Schamann BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 838
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Ernährungswissenschaften
Betreut von / Supervisor:	Doz. Dr. Alexander Haslberger

Danksagung

An dieser Stelle möchte ich mich herzlich bei allen Personen bedanken, die mich durch das Studium begleitet haben und mich bei der Erstellung der Masterarbeit unterstützt haben.
Zuerst möchte ich mich bei Dr. Christian Brandes bedanken, der die Masterarbeit ermöglicht hat und mich dabei bestmöglich unterstützt hat. Weiters gilt mein Dank Agnes Draxler, BSc. und Sina Ben Ali, MSc., die mir nicht nur die Methoden der Molekularbiologie praktisch gelernt haben, sondern auch die Wartezeiten im Labor mit interessanten und humorvollen Unterhaltungen gefüllt haben. Außerdem möchte ich mich bei Mag. Rupert Hochegger und der Agentur für Gesundheit und Ernährungssicherheit für die Genehmigung dieser Masterarbeit und bei der gesamten Abteilung „Mikrobiologie und Molekularbiologie“ für Hilfestellungen bedanken.
Weiters bedanke ich mich bei Doz. Dr. Alexander Haslberger für die Betreuung der Masterarbeit seitens der Universität Wien.
Ganz besonders möchte ich mich bei meinen Eltern Ursula und Martin bedanken, die mir das Studium ermöglicht haben. Vielen Dank dafür, dass ihr mich immer motiviert habt, geduldig diverse Präsentationen angehört habt, Arbeiten Korrektur gelesen habt und mich in allen Entscheidungen stets unterstützt habt.
Zuletzt gilt mein Dank meinem Freund Florian. Vielen Dank für deine Geduld, wenn ich neben dem Studium und der Arbeit nur wenig Zeit für dich gefunden habe und dafür, dass du immer für mich da warst, wenn ich dich gebraucht habe.

Table of contents

List of tables

Figure index

List of abbreviations

bp	base pair
°C	degree Celsius
cDNA	complementary deoxyribonucleic acid
CRISPR/Cas	clustered regularly interspaced short palindromic repeats/CRISPR-associated systems
Ct	cycle threshold
CTAB	hexadecyltrimethylammoniumbromide
CV	variation coefficient
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double – stranded DNA
EC	European Commission
EDTA	ethylenediaminetetraacetic acid
EFSA	European Food Safety Authority
e.g.	*exempli gratia,* for example
e.i.	*id est,* that is
EPSPS	5-enolpyruvyl-shikimate-3-phosphate synthase
ERA	environmental risk assessment
EU	European Union
FAO	Food and Agriculture Organization
fwd	forward
g	times gravity
GM	genetically modified
GMO	genetically modified organism
GO	Gene Ontology
ha	hectare
HRM	High-Resolution Melting
kbp	kilobase pair

M	molar
mg	milligram
ml	milliliter
mM	millimolar
mRNA	messenger RNA
mol.	molecular
µl	microliter
n.d.	no date
NGS	Next-Generation Sequencing
NHEJ	non-homologous end-joining
nm	nanometer
no.	*numero*, number
NTC	no templet control
OECD	Organization for Economic Co-operation and Development
PCR	Polymerase Chain Reaction
PVP	polyvinylpyrrolidon
rev	reverse
RNA	ribonucleic acid
RT-qPCR	quantitative Reverse Transcription-Polymerase Chain Reaction
SD	standard deviation
SDN	side-directed nucleases
SNP	single nucleotide polymorphism
T-DNA	transfer DNA
TALEN	transcription-activator-like effector nucleases

1 Introduction

Since the beginning of the commercialization of genetically modified (GM) plants in 1996, these had been grown on accumulated 2.1 billion hectares of cultivated area worldwide until 2016. Maize is the second most frequently produced GM crop, after soybean [ISAAA, 2016]. In the European Union (EU), 27 GM maize varieties are authorized as food and feed [European Commission, 2017], whereas only the event MON810 is authorized for cultivation in eleven countries and regions - Czech Republic, England, Estonia, Finland, Flanders, Ireland, Portugal, Romania, Slovakia, Spain, and Sweden [EU, 2016].

However, an extensive risk assessment is essential for the authorization of GM plants to identify potential risks of the modification [EFSA, 2011]. The insertion of a transgene might lead to intended effects (e.g. herbicide tolerance) as well as to unintended effects (e.g. variations in the transcriptome) [Sorochinskii et al., 2011]. The EU requests the comparison of, for example, specific components, metabolites, and toxins between a GM plant and a non-GM comparator following the principle of substantial equivalence as control for unintended effects [EFSA, 2011]. However, unpredicted unintended effects probably cannot be found in this way. As additional methods, profiling technologies can be applied to detect unintended effects [Heinemann et al., 2011].

Further, according to the European Directive 2001/18/EC, the insert of the GM plant must be stable [EU, 2001]. Genetic stability is usually controlled by Southern blot analysis. This low-resolution method can identify large rearrangements, but it is not suitable for the detection of small variations like Single Nucleotide Polymorphisms (SNPs). Better options for the discovery of small modifications are, for instance, sequencing-based methods like Next-Generation Sequencing (NGS) [Kohli et al., 2010].

In previous investigations, SNPs, which might lead to unintended effects, were identified in the NK603 and the MON810 event by High-Resolution Melting (HRM) analysis and Sanger sequencing [Ben Ali et al., 2014, Castan et al., 2017]. In this study, unintended effects were investigated in a transcriptome analysis by comparing the gene expression of endogenous genes of a GM maize variety containing the event NK603 with the nearly isogenic, non-GM maize variety. The quantitative Reverse-Transcription Polymerase Chain Reaction (RT-qPCR)

was used for this analysis. Further, the genetic stability of the 3'border region of the transgene of a maize variety containing the event MON810 was controlled by performing a real-time PCR in combination with an HRM analysis as a screening method for the identification of SNPs in a large number of maize grains. Divergent samples were subsequently verified by NGS.

2 Literature overview

2.1 Definition: genetically modified organism

Referring to the EU Directive 2001/18/EC, a genetically modified organism (GMO) is an "organism, with the exception of human beings, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination" [EU, 2001].

Referring to the Austrian law on genetic engineering (BGBl. Nr. 510/1994), GMOs are "organisms, in which the genetic material was changed in a way that does not occur under natural conditions by crossing or natural recombination or other conventional breeding techniques […]" [Republic Austria, 1994]. Techniques of genetic modifications are recombinant nucleic acid techniques by using vector systems, the direct introduction of genetic information prepared outside of the target organism (e.g. macro- and microinjection), and cell fusion leading to living cells with genetic material, which could not occur under natural conditions. In contrast, techniques like undirected mutagenesis, *in vitro* fertilization, natural processes like conjugation and transduction, and polyploidy induction are not defined as genetic modifications [Republic Austria, 1994].

2.2 Legal background

2.2.1 International Regulation: Cartagena Protocol on Biosafety to the Convention on Biological Diversity

The Cartagena Protocol is an international treaty supplementing the Convention on Biological Diversity of the United Nations and enforced on September 11, 2003. The aim of the protocol is to ensure the safe transport, handling, and usage of living modified organisms, which were produced by methods of modern biotechnology and which might have negative consequences on the biological diversity. As an international treaty, it focuses on the transboundary movements of living modified organisms. The protocol determines

an advanced informed agreement procedure. This ensures that countries, which import living modified organisms that could be released into the environment, obtain all necessary information to decide whether the import is accepted or not. Further, the treaty defines the requirement of a risk assessment of living modified organisms and the risk management. In all, 170 countries, including Austria, have already ratified the Cartagena Protocol [Secretariat of the Convention on Biological Diversity, 2000].

2.2.2 European Directive 2001/18/EC

Directive 2001/18/EC was enforced on March 12, 2001, after being passed in the European Parliament and the Council. Since it was an EU directive, it had to be implemented in the national law of each Member State. In Austria, it was implemented in the law on genetic engineering (BGBl. Nr. 510/1994). Directive 2001/18/EC represents the legal basis of the EU concerning the release of a GM organism in the environment for experimental purposes, as well as for placing commercial products on the market, which includes import. A goal of this Directive is the harmonization of the different laws of all EU Member States in the field of GMO. It follows the precautionary principle and should ensure the protection of human health and the environment as its main goal [EU, 2001].

Directive 2001/18/EC regulates GMO as a plant. Risk assessment and authorization are operated mainly by the respective authorities of the Member States. Before releasing a GMO, a notification has to be submitted to the national competent authority, including all required information listed in Annexes III and IV of Directive 2001/18/EC. This includes, among others, a detailed description of the genetic modification and the presentation of an environmental risk assessment (ERA), as well as of a monitoring plan. The national competent authority revises the notification and answers it by composing an assessment report, saying whether the GMO can be placed on the market, and under which conditions. The authorization is valid for a maximum of ten years, after which it has to be renewed [EU, 2001].

As mentioned above, the description of an extensive ERA is obligatory for the notification, for which the principles listed in Annex II of Directive 2001/18/EC must be followed. The aim of the ERA is to identify all potentially adverse direct, indirect, immediate, and delayed

effects of the GMO on the health of humans and the environment. This is done on a case-by-case basis. Adverse effects could, for instance, occur due to the spread of GMO in the environment or due to genetic instability [EU, 2001].

For the notification, a case-by-case monitoring plan following the principles of Annex VII of Directive 2001/18/EC is also obligatory. Its aims are the conformation of any supposition of potentially adverse effects of the GMO mentioned in the ERA and the occurrence of negative effects on the environment or human health, which were not considered in the ERA [EU, 2001].

2.2.3 European Directive 2015/412

The safeguard clause in Article 23 of Directive 2001/18/EC permits the Member States to temporarily prohibit the selling of a GM product in their territory if they have reasons (e.g. due to new information or additional scientific knowledge) to believe that this product can negatively influence the health of humans or the environment [EU, 2001]. Some Member States, including Austria, used this clause to restrict the cultivation of GMO. To amend this clause and also to give Member States the opportunity to prohibit the cultivation of a certain GM product in the long term, Directive (EU) 2015/412 was legislated. This Directive enables the Member States to demand the exclusion of its own territory from the geographical scope of the cultivation-permission from the applicant during the renewal of the authorization of a certain GMO [EU, 2015b].

2.2.4 European Regulation No. 1829/2003

As mentioned before, Directive 2001/18/EC regulates the GMO as a plant, whereas the European Regulation 1829/2003 on GM food and feed regulates GMO as food and feed. Since it is a regulation, it has to be directly enforced in all Member States. The scope of this Regulation includes GMOs that are used as food (e.g. GMO maize), food containing GMOs or consisting of GMOs (e.g. convenience products, including GMO-soy lecithin), and food or at least ingredients of food produced from GMOs (e.g. rape oil made from GMO rape). The

scope does not include food produced using GM auxiliary materials and products of GMO-fed animals [EU, 2003].

The Regulation 1829/2003 prescribes the authorization, supervision, and labeling of these products. The process of authorization is more centralized in the EU than that of Directive 2001/18/EC. To receive an authorization, applications are verified by the European Food Safety Authority (EFSA). The EFSA composes an assessment, based on which the European Commission accepts or rejects the application. The authorization also needs to be renewed after ten years. The Regulation 1829/2003 states that food containing GMO in a proportion higher than 0.9% must be correctly labeled [EU, 2003].

2.2.5 Austrian law on genetic engineering (BGBl. Nr. 510/1994)

Apart from the mentioned European laws, the Austrian law on genetic engineering (BGBl. Nr. 510/1994) regulates GMOs in Austria. This law, enforced on January 1, 1995, has been adapted several times over the last years, especially to include the European Directive 2001/18/EC. Different fields of applications of genetic engineering in medicine, as well as in agriculture, are regulated, mainly to protect human health. Some of these are the work with GMO in contained systems, the release of GMO in the environment, the placing of GMO on the market, gene therapy, and official controls [Republic Austria, 1994].

The Austrian legislation comprises many safety precautions for GMOs. Directive (EU) 2015/412 was included in the Austrian law on genetic engineering. It enables the exclusion of the Austrian territory from the geographical scope of GMO authorization under Directive (EU) 2001/18/EC. Consistent with this, Austria was excluded from the geographical scope of the cultivation of MON810 maize [EU, 2015b]. Further, temporary prohibitions and restrictions are possible if the state reasonably assumes, due to new information or new scientific knowledge, that an authorized product is a safety risk. However, only the products that were authorized under Directive (EU) 2001/18/EC, and not those under Regulation (EU) 1829/2003, can be restricted in this way [Republic Austria, 1994].

Moreover, every federal state of Austria has its own genetic engineering precaution law to regulate specific local precautionary measures. These laws comprise, for example, certain

requirements for the coexistence of different agricultural cultivation systems (e.g. safety zones between the cultivation of GM plants and conventional plants or barriers for the pollens of GM plants), and the guarantee that the regional goals for natural conservation are not impaired in the case of the cultivation of GM plants [Land Salzburg, 2004].

2.3 Maize

Maize (*Zea mays ssp. mays*) was domesticated about 9,000 years ago from teosinte (*Zea mays ssp. parviglumis*) in the southwest of Mexico [van Heerwaarden et al., 2011]. In 1493, Columbus brought Caribbean tropical maize to Europe. However, it is believed that temperatures were too low for this cultivar in Northern Europe and therefore, further varieties were introduced from Nord America later [Tenaillon and Charcosset, 2011]. Maize is an annual and monoclinous plant with male as well as female flowers in each plant. The male germ cells are in the tassel, which is located at the top of the maize plant, whereas the female germ cells are in the ears at the base of the leaves. Thus, cross-pollination, as well as self-pollination, is possible [Nannas and Dawe, 2015].

According to the Food and Agriculture Organization (FAO) of the United Nations, maize is that kind of grain which is produced in the highest amounts worldwide nowadays. Around 1,060 million tons of maize was produced worldwide on 188.0 million hectares (ha) in 2016, compared to 749 million tons of wheat, and 741 million tons of rice. The three main production countries are the Unites States of America (385 million tons on 35.1 million ha in 2014), China (232 million tons on 39.0 million ha) and Brazil (64 million tons on 15.0 million ha) [FAO, 2018]. In 2016, GM maize was grown on 60.6 million ha globally in 16 different countries, including the United States of America (30.1 million ha), Brazil (15.6 million ha), and Argentina (4.7 million ha) as the top three GM maize-producing countries. Focusing on the inserted traits of the maize plants, insecticide- and herbicide-resistant maize is cultivated on 78.5% of the globally cultivated area, herbicide-resistant varieties on 11.5%, and insecticide-resistant ones on 10% [ISAAA, 2016].

The produced maize is not only used for human consumption, but mostly for feeding livestock and for industrial processing, for example, for the production of bio-ethanol. In

developed countries, maize is mostly cultivated to feed animals, whereas in developing countries, a higher percentage is used for human consumption [Shiferaw et al., 2011].

The genome of maize had repeatedly been the objective of scientific works, and many essential genetic discoveries were made on the basis of this genome. For instance, Barbara McClintock first discovered the transposable elements of the genome when she studied the genome of maize with differently colored grains [McClintock, 1948]. In 1983, McClintock was awarded the Nobel Prize for finding the “jumping genes” [Ravindran, 2012]. Schnable et al. sequenced the whole genome of the maize inbred line B73 with the help of bacterial artificial chromosomes and fosmid clones. The genome of maize consists of ten chromosomes, which are structurally very different, and comprises about 2.3 gigabases [Schnable et al., 2009]. It includes around 40,000 protein-coding genes [Law et al., 2015].

2.4 MON810 event

The GM maize line MON810 was developed by Monsanto and is known by the trade name YieldGard™. MON810 maize was initially authorized in the EU in 1998, when the Commission Decision 98/294/EC came into force [EU, 1998]. Consistent with Directive (EU) 2015/412, 19 Member States demanded the exclusion of their state or at least a part of it from the geographical scope of the cultivation of MON810 maize. Thus, in the EU, the cultivation is only accepted in Belgium (apart from Wallonia), Czech Republic, England, Estonia, Finland, Ireland, Portugal, Romania, Slovakia, Spain, and Sweden at the moment [EU, 2016]. In the Commission Implementing Decision (EU) 2017/1207, the authorization of food and feed containing MON810 maize was renewed for ten years [EU, 2017]. In 2016, MON810 maize was cultivated in Spain, Portugal, Slovakia, and Czech Republic [ISAAA, 2016].

The insertion of the transgene was done by microprojectile bombardment of embryonic maize cells [Hernández et al., 2003]. MON810 maize is an insecticide-resistant variety. The insertion of a gene of the soil bacterium *Bacillus thuringiensis subsp. kurstaki* enables the GM plants to produce an insecticidal protein. *Bacillus thuringiensis* is a gram-positive, facultative-aerobic bacterium that produces spores. During the sporulation process, the

bacterium releases crystals containing δ-endotoxins (Cry-proteins). A large number of different δ-endotoxins exist. These are categorized into different classes (Cry 1, 2, 3,…), as well as subclasses (Cry1A, Cry1B,…). Each Cry-protein is toxic for only a limited number of insect species. In MON810 maize, Cry1A(b) is produced. It is toxic for lepidopterans, especially, the European corn borer (*Ostrinia nubilalis*) and the Mediterranean corn borer (*Sesamia nonagrioides*). When the crystals reach the intestine of the insects, they dissolve and the δ-endotoxins are split into smaller fragments, which can bind to the intestinal receptors. The toxic effect might be the formation of pores in the cell membrane of intestinal cells, which might lead to a disruption of the ion flow [Sanchis and Bourguet, 2008].

Besides the Cry1A(b) gene (3.64kb), the MON810 insert consists of the enhanced 35S promotor and the heat shock protein (hsp) 70 intron, which are upstream of the Cry1A(b) gene. The 35S promoter (0.61kb) is derived from the cauliflower mosaic virus and widely used in GM crops. The hsp70 intron (0.8kb) increases the gene expression [Rosati et al., 2008]. A truncation at the 3'end of the Cry1A(b) was identified together with the loss of the complete NOS terminator [Hernández et al., 2003].

2.5 NK603 event

The GM maize variety NK603 was developed by Monsanto. Its request for the NK603 containing feed to be placed on the market in the EU was accepted in the Commission Decision 2004/643/EC [EU, 2004]. In 2005, the Commission Decision 2005/448/EC authorized the sale of food and food ingredients containing NK603 maize [EU, 2005]. Both authorizations were only valid for the NK603 maize products to be placed on the market, but not for the cultivation of NK603 maize [EU, 2004, EU, 2005]. In 2015, these two authorizations were renewed in the Commission Implementing Decision (EU) 2015/684 for ten years [EU, 2015a].

NK603 maize is herbicide-tolerable, which enables farmers to use herbicides based on glyphosate. Glyphosate is a widely used broad-spectrum herbicide, which is present in many different herbicide formulations on the market. The mode of action is based on the

inhibition of the enzyme 5-enolpyruvyl-shikimate-3-phosphate synthase (EPSPS), which is part of the shikimate pathway. This pathway leads to the production of the three aromatic amino acids, phenylalanine, tyrosine, and tryptophan, in plants, bacteria, and fungi, whereas humans and animals cannot produce it. Since these three amino acids are essential, the stopping of their production is one of the lethal effects of glyphosate in plants [Duke and Powles, 2008].

The NK603 transgene consists of two cassettes, which were inserted by microparticle bombardment with the aid of PV-ZMGT32 plasmids. In total, the transgene comprises around 6.7kb. Both cassettes include a 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene obtained from the gram-negative soil bacterium *Agrobacterium sp.* strain CP4. In the first cassette, the gene is regulated by the rice (*Oryza sativa*) actin 1 gene promoter and intron, the chloroplast transit peptide obtained from the EPSPS gene of *Arabidopsis thaliana*, and stopped by the 3'terminator sequences of the nopaline synthase gene obtained from *Agrobacterium tumefaciens*. The second cassette has nearly the same composition as the first one, but it is regulated by an enhanced 35S promoter obtained from the cauliflower mosaic virus and the hsp70 intron instead of the rice actin gene promoter [CERA, 2017, EU, 2004].

Thus, the GM maize plants additionally express the CP4 EPSPS of *Agrobacterium tumefaciens*, which has very low affinity to glyphosate compared to the native EPSPS of maize plants. Therefore, the shikimate pathway is not blocked and the plants can still produce aromatic amino acids instead of being killed by glyphosate [Dill, 2005]. The CP4 EPSPS is a polypeptide comprising 455 amino acids. About 50% of the amino acid sequence of the CP4 EPSPS is equal to that of the native maize EPSPS [CERA, 2017].

Heck et al. analyzed the nucleotide sequence of the insert and detected two nucleotide variations in the EPSPS gene of the second cassette compared to the first one. One nucleotide variation is a silent mutation, which does not change the amino acid sequence, whereas the other one leads to the translation of proline instead of leucine at position 214. However, this mutation is not located in the active center of the gene [Heck et al., 2005].

2.6 Genetic stability

The optimal GM plant has only one copy of an unaltered transgene, which is inserted without disrupting the functional endogenous genes of the plant and without leading to further alterations of the plant DNA. However, this theory is hardly found in practice, in which a higher copy number of the transgene might be inserted [OECD, 2015], the insertion site might interrupt functional genes, and alterations of the plant DNA might occur. Often, the transformation of a plant already leads to alterations at the insertion location (insertion-site mutations), such as the insertion of filler DNA or vector sequences, and deletions or rearrangements of plant DNA sequences, or elsewhere in the plant genome (genome-wide mutations). Consequences of transformation-induced mutations might be a loss of function - e.g. when a transgene is inserted into a functional gene or a regulatory sequence - or alterations of the gene expression of endogenous genes [Wilson et al., 2006]. However, these transformation-induced modifications are not part of the genetic stability of transgenes which focuses on post-transformational modifications.

According to Directive 2001/18/EC, the genetic stability of the insert is an obligatory requirement for the admission of a GM plant [EU, 2001]. In the application for the authorization of a GM plant, the applicant must prove that the transgenic locus is stable. Applications for the market release of GM crops with a single event must show the genetic stability in five sequential generations or vegetative cycles. Data of the first and the last generations must be submitted. In applications for GM crops with stacked events, comparisons of the sequence of the inserts, as well as of the flanking regions between the stacked variety and its corresponding plants containing the single events, must be submitted [EU, 2013].

Methods for the control of the genetic stability include, among others, Southern blot analysis, PCR, real-time PCR, Fluorescence in-situ hybridization, and DNA sequencing, whereas the common method is Southern blot analysis. However, small rearrangements, for example, small insertions or deletions, are partly too subtle to be detected by low-resolution detection methods like Southern blot analysis [Kohli et al., 2010], though they might have certain effects on the plant [Ben Ali et al., 2014]. Better options for the detection

of small variations are DNA sequencing and PCR, which are rarely used in routine control [Kohli et al., 2010]. Over the last years, the question has arisen as to whether Next-Generation Sequencing can replace Southern blot analysis for the molecular characterization of GM plants. NGS is a high-throughput method for DNA sequencing, showing advantages while controlling the genetic stability, which is part of the molecular characterization. The identification of small sequence modifications like small insertions and deletions is possible using NGS in contrast to Southern blot analysis [Pauwels et al., 2015]. Further, NGS requires less starting material for the analysis (about 1/10) compared to Southern blot analysis [Guttikonda et al., 2016]. However, NGS delivers a huge amount of data, which has to be correctly analyzed and interpreted, which is still a challenge [Pauwels et al., 2015].

Genetic instabilities, such as changes in the DNA sequence or structural variations, might occur due to meiotic instabilities, epigenetic factors, or probably due to viral sequences, leading to differences between single plants and different generations [Neumann et al., 2011]. The stability of the inserted transgene can be influenced by a variety of factors, of which the position effects and the locus structure effects are the most important. Position effects might occur due to the genomic DNA next to the integrated transgene. For instance, nearby regulatory elements, such as enhancers, can modify the expression level of the transgene, or the methylation level of the transgene can increase due to the surrounding DNA [Kohli et al., 2010]. Further, when the transgene is inserted into a transposon, it might change its position in the genome [Pla et al., 2012]. Locus structure effects are influenced by the copy number of the inserted transgene and their intactness and arrangement. The insertion of a higher copy number can, for instance, lead to transgene silencing instead of an increased expression [Kohli et al., 2010] due to higher methylation levels. Further, a homologous recombination of the copies of the inserted transgene might lead to a reduction of the stability of the transgene [Pla et al., 2012].

While discussing genetic stability, natural mutations existing in every plant (influenced, for example, by the occurrence of wrongly incorporated base pairs during the replication of the DNA, of crossing over of homologous chromosomes during the meiosis, or of chromosome doubling before the cell division of the mitosis starts) should also be considered [OECD,

2015]. However, the question arises as to whether the mutation rate of the transgenes is increased compared to the natural mutation rate of endogenous genes [Pla et al., 2012]. Ogasawara et al. compared the mutation rate of the EPSPS transgene in Roundup Ready soybeans and an endogenous plant gene, coding a storage protein (B-conglycinin gene). One mutation was detected per 1144bp in the transgene compared to one mutation per 1079bp in the reference gene. Thus, the mutation rate showed hardly any difference. However, while comparing the number of amino acid substitutions, noticeable differences were detected. Only four amino acid substitutions were found in the transgene, whereas 25 were detected in the B-conglycinin gene. This indicates that most mutations in the transgene were silent mutations [Ogasawara et al., 2005].

Post-transformational modifications of transgenes and corresponding border regions are hardly analyzed. Especially, investigations of commercial GM plants are missing. Therefore, further analyses would be very important [Neumann et al., 2011]. The results of some studies investigating genetic stability are presented below.

As mentioned before, Hernández et al. detected a truncation of the MON810 transgene at the 3'end, leading to a loss of the complete NOS terminator and of the end of the Cry1A(b) gene [Hernández et al., 2003]. Rosati et al. investigated the possible effects of the truncation on the read-through transcription downstream of the truncation by analyzing the transcriptional activity of two MON810 maize varieties. Genetic instabilities were detected. The truncation led to the expression of only a partial Cry1A endotoxin. Due to the loss of the stop codon, the read-through changed, resulting in the production of new proteins without homology to already known proteins [Rosati et al., 2008].

La Paz et al. analyzed the overall structure of the MON810 transgene and the flanking regions in three different maize varieties by Southern blot analysis without finding any new rearrangements. Further, the genetic stability of the transgene and the flanking regions was investigated in 28 different maize varieties by performing DNA mismatch endonuclease assays focusing mainly on the detection of point mutations. Neither the transgene nor the 3'flanking region showed any polymorphism in all the investigated varieties. However, six SNPs were detected in the 5'flanking region, whereas each SNP appeared in a different variety respectively. The SNPs were located between 784bp and 531bp upstream of the

transgene. Therefore, the qualitative and quantitative control of GMO based on PCR methods was not affected by all these SNPs. The six detected SNPs result in a mutation rate of $1.6*10^{-5}$ substitutions per site and generation, which is comparable to the natural mutation rate for hypervariable regions in maize ($8*10^{-4}$) [La Paz et al., 2010].

Ben Ali et al. analyzed the genetic stability of the flanking regions of the MON810 event in a stacked maize variety (MON88017 x MON810) and of a single-event oilseed rape variety (GT73) using HRM analysis and Sanger sequencing. In the 3'flanking region of the MON810 event in maize, two out of 100 samples showed a substitution of cytosine by thymine in the same position. In both samples, the detected SNP was heterozygous and silent, which means that the amino acid sequence was not altered. No SNPs were found in the 5'flanking region of MON810 or in either flanking region of the rape variety [Ben Ali et al., 2014].

Castan et al. investigated the whole NK603 transgene, including the flanking regions of a stacked maize variety (NK603 x MON810), by performing HRM analysis and Sanger sequencing. Two single nucleotide insertions were found in the rice actin 1 promoter in the first cassette of the NK603 transgene. The inserts were detected in all samples, as well as in the certified reference material for NK603. The SNPs may be explained by a sequencing error of the NK603 patent or by the introduction of the SNPs during the breeding process for the production of commercial lines. Due to the location in the promoter region, the protein sequence might not be affected, whereas a modification of the gene expression might be possible [Castan et al., 2017].

Compared to these results showing genetic instabilities, the following investigations could prove genetic stability. Neumann et al. controlled the genetic stability in the 5' and 3' junction region in 567 maize seeds containing a single MON810 event by performing Scorpion primers analysis and subsequent Sanger sequencing. No variations were detected in this variety [Neumann et al., 2011]. Using the same methods as Neumann et al., Madi et al. analyzed the 3'end of a single-event soybean variety (RR 40-3-2) in 1,034 samples. The genetic stability was proven in the selected region in all the samples [Madi et al., 2013].

Guttikonda et al. compared Southern blot analysis and Sanger sequencing to NGS (whole genome sequencing and target capture sequencing) as paired-end sequencing to perform a molecular characterization of GM soybeans. Two different GM soybean varieties containing

a single event, as well as a stacked variety containing both events, which was produced by conventional breeding of the single-event varieties, were included. The study demonstrated that NGS was successful in identifying transgene copy number, T-DNA integrity, stability of the T-DNA inserts, and verification of the presence of plasmid backbone sequences. The results of NGS were verified by traditionally used methods (Southern blot analysis and Sanger sequencing). The single-event varieties contained only one copy of the transgene and the stacked variety contained each transgene only once. This shows that no changes of the copy number occurred due to breeding. A stable inheritance of the T-DNA was shown across generations, which was proven by Southern blot analysis. The authors concluded that NGS is suitable for performing a molecular characterization of GM crops. However, further improvements, for example, an increase of reading length, are still necessary [Guttikonda et al., 2016].

2.7 Unintended effects in GM plants

The insertion of a transgene into the plant genome might lead to intended (e.g. herbicide resistance or insecticide resistance) as well as unintended effects. In addition to the insertion effects, unintended effects might appear due to further mechanisms, such as pleiotropy, a mechanism that describes the effects of an individual gene on different characteristics, and somaclonal variations, which explain genetic variations in *in vitro* culture [Ladics et al., 2015]. Unintended effects occur as statistically significant variations in the phenotype, the metabolism, or the composition, besides the intended difference due to the transgene between a GM plant and the non-GM ancestral plant [Cellini et al., 2004]. However, unintended effects can also occur due to traditional breeding methods [Schnell et al., 2015].

Unintended effects can be categorized into predictable and unpredictable effects. Predictable unintended effects are consequences of, for example, the insertion of the transgene, which might have been assumed based on our current knowledge of genetics and interconnections in the metabolism of plants. In contrast, unpredictable unintended effects cannot be explained by our state of knowledge [Cellini et al., 2004]. Unintended effects might occur due to many different causes. The influence of changing environmental

conditions or stress factors on GM plants might lead to unintended effects [Sorochinskii et al., 2011]. A further cause can be the insertion effects due to the transformation of the plant, which can also lead to the insertion of, for example, filler DNA, as already mentioned before. Insertion effects may change the gene expression of endogenous plant genes, which can lead to unintended traits of the GM plant [Schnell et al., 2015].

As assumed, it is hardly possible to directly test for unintended effects [Ladics et al., 2015]. In the EU, GM plants are controlled for the manifestation of intended as well as unintended effects in the context of the risk assessment of GM plants and the derived food and feed in a weight for evidence approach. First, a molecular characterization is performed, which focuses on the flanking regions of the inserted transgene for the explicit detection of unintended effects [EFSA, 2011], to control, for example, the deterioration of the endogenous genes of the plant due to the insertion of the transgene and the production of unintended proteins by the GM plant [Ladics et al., 2015]. Next, the compositional, agronomic, and phenotypic characterization is performed in a comparative approach, in which GM plants are compared to non-GM comparators with a history of safe use after cultivation under the same conditions. Especially, the content of macro- and micronutrients, certain metabolites, and known toxins are compared between the GM plant and its comparator. The testing of equivalence determines if the variation between the GM plant and the comparator is higher than the natural variations between non-GM reference varieties [EFSA, 2011]. If significant differences are detected in the hazard identification step, further investigation follows to identify possible harmful effects [Devos et al., 2016]. This comparative approach, or rather, this concept of substantial equivalence, i.e. the comparison of new food to long-term consumed food, is not only in the EU but worldwide often used as the basis for the safety assessment of GM plants and has been described by the Organization for Economic Co-operation and Development [OECD, 1993].

Additionally, profiling technologies, including transcriptomics, proteomics, and metabolomics, are applied for the identification of unintended effects [Ladics et al., 2015]. Profiling comprises many different methods to analyze a plant at various biochemical levels. For instance, in transcriptomics, the expression of active genes is analyzed by measuring the mRNA. In proteomics, the totality of the proteins that are translated from the mRNA is

analyzed, and in metabolomics, the metabolites of a plant are investigated. Risk assessment of the EU follows mainly a targeted approach to identify unintended effects, whereas omic-technologies are performed by a rather untargeted approach, which enables the identification of unintended effects, for example, metabolites, which are not measured in the comparative approach [Kok et al., 2010]. The detection and subsequent characterization of unknown substances, for example, proteins or metabolites, the identification of eventually toxic responses of GM plants to environmental factors, and the supplemental information gained in the hazard identification step of GM plants are seen as some strengths of profiling methods [Heinemann et al., 2011].

However, profiling technologies used in the food safety assessment also seem to have some limitations, such as a low inter- and intra-laboratory reproducibility of results, as well as a low standardization of the performed methods. Furthermore, there is a lack of definitions that explain which differences detected by omic-technologies are safe and which might be unsafe [Chassy, 2010]. In the following paragraphs, the results of some publications focusing on the performance of profiling techniques with GM maize are summarized.

Zolla et al. investigated the proteomic profile of seeds of two subsequent generations of MON810 maize and their two wild-type comparators using 2D-electrophoresis and mass spectrometry. The two progeny varieties were grown under the same environmental conditions. The differences between the two wild-type generations, as well as between the two GM generations, were interpreted as environmental effects. While comparing the wild-type seeds to their own progenies, 100 spots on the gel were identified to have different expression levels and 78 were differently expressed between the GM seeds and their progenies. While comparing the proteomic profile of the two parental generations (GM maize vs. conventional maize), 27 protein spots were identified to be differently expressed, which is quite low compared to the number of spots mentioned before. This might indicate that the effects of the gene insertion were smaller than those of the environment. The comparison of the proteome profile of the two progenies showed 43 differently expressed spots. One newly expressed protein in GM progenies was zein, which is an allergenic protein. Furthermore, same storage proteins had a truncated form in the GM seeds compared to the wild-type seeds [Zolla et al., 2008].

Benevenuto et al. analyzed the proteomic and metabolomic profile of GM maize plants containing NK603 and of their conventional comparators under stress conditions, such as water deficiency and the application of Roundup herbicide. By comparing the two control groups - GM maize and conventional maize without stress factors - six differently expressed proteins were identified, which might be attributed to the genetic modification. These proteins were especially related to the carbohydrate and energy metabolism, indicating a higher energy demand of GM maize plants. Furthermore, differences in phytohormones and further components related to the defense mechanism of the plant were detected. Drought led to a stronger down-regulation of proteins related to photosynthesis in GM plants compared to the conventional plants under the same stress condition. Under herbicide application as a further stress factor besides drought, four proteins mainly related to increases in energy metabolism were differently expressed in GM maize plants compared to GM maize under drought but without herbicide application. This might indicate a higher energy requirement for the adaptation to an increasing number of stress factors. The authors concluded that environmental factors were the major source of change, but genetic modifications can also lead to variations in the proteomic and metabolomic profile [Benevenuto et al., 2017].

Coll et al. compared transcriptome profiles of leaves of GM maize varieties containing the event MON810 to their non-GM comparators. To eliminate the environment-influencing factors, all plants were grown *in vitro* under the same conditions. Gene expression was compared using microarray, and the 40 selected sequences were verified by performing real-time PCR. In all, 282 genes were differently expressed in the variety Aristis Bt (GM modified) vs. Aristis (conventional comparator) and 24 genes in the variety PR33P67 (GM modified) vs. PR33P66 (conventional comparator). The observed differences between GM maize and conventional maize were less significant than those between the varieties resulting from conventional breeding [Coll et al., 2008]. However, while growing the same varieties in the field, the genes which were differently expressed in the first investigation were similarly expressed [Coll et al., 2009].

La Paz et al. compared the gene expression between the immature embryos of a GM maize variety containing MON810 and the near-isogenic, non-GM maize variety, both grown

under controlled conditions using mRNA sequencing, microarray hybridization, and as a control real-time PCR. Around 140 genes, mainly related to carbohydrate metabolism and protein modification, were differently expressed, which might be due to the delayed maturation of the GM variety compared to the conventional variety. To investigate if the detected genes were also differently expressed in other varieties containing MON810, 30 genes were analyzed in two additional variety pairs. Of these, 22 genes were differently expressed in all three GM varieties compared to their comparators [La Paz et al., 2014].

Further investigations using profiling methods with GM maize [e.g. Agapito-Tenfen et al., 2013, Mesnage et al., 2016, Tan et al., 2017, Tan et al., 2016, Vidal et al., 2015] also found differences between the GM maize and its conventional comparator.

2.8 New biotechnological methods for plant modifications

In the mid-1990s, the traditional plant-breeding techniques were complemented by genetic modification to obtain new traits of plants. The majority of GM plants were produced by T-DNA integration using *Agrobacterium tumefaciens* or by particle bombardment. However, due to the uncontrolled integration of DNA into the plant genome, risks like a disruption of endogenous genes or an alteration of the plant composition may occur. Over the years, the number of new biotechnological methods for plant breeding has increased manifold. The aim of these methods is to achieve precise genetic modification so as to avoid the risks of transgenic plants [Hartung and Schiemann, 2014]. In the following paragraphs, a selection of new methods will be briefly introduced.

In cisgenesis and intragenesis, genes from the target plant or from cross-compatible plant species are transferred. In cisgenesis, the whole gene, including the regulatory sequences, is transferred into the plant genome in sense orientation, whereas in intragenesis, divergent regulatory elements and coding sequences are assembled and can be transferred in both sense and antisense orientations. Cisgenic plants can be obtained by traditional breeding methods, whereas intragenic plants cannot be bred in this way [Cardi, 2016].

Genome editing comprises a huge variety of methods, which enable the cutting, deletion, and replacement of a certain DNA sequence, as well as the insertion of a desired sequence

[Cardi, 2016]. One group of genome editing methods can be summarized as site-directed nucleases (SDN). SDNs can be used, for example, for specific mutations of the genome, for insertions or deletions, as well as for replacement of certain genes. SDNs are proteins, which can recognize and bind a certain DNA sequence [Hartung and Schiemann, 2014]. Then, double-strand breaks of the DNA are induced by nucleases [Rinaldo and Ayliffe, 2015], which are repaired by endogenous DNA-repairing mechanisms like the non-homologous end-joining (NHEJ) or the homologous recombination [Puchta and Fauser, 2014]. NHEJ occurs more often. It is an error-prone method, which can lead to small insertions or deletions at the precise break site of the DNA [Rinaldo and Ayliffe, 2015]. Double-strand breaks can be achieved by enzymes such as meganucleases, zinc-finger nucleases, transcription-activator-like effector nucleases (TALEN) and clustered regularly interspaced short palindromic repeats/CRISPR-associated systems (CRISPR/Cas) [Rinaldo and Ayliffe, 2015]. Zinc-finger nucleases contain two functional domains [Porteus and Carroll, 2005]. The zinc-finger DNA recognition domain consists of Cry_2-His_2 zinc-finger domains, of which each finger binds specifically a nucleotide triplet of the DNA, whereas the non-sequence-specific nuclease domain cuts the DNA [Rinaldo and Ayliffe, 2015]. Zinc-finger nucleases were, for instance, used by Ainley et al. to create a stacked maize variety with the integration of a transgene at a precise genomic locus [Ainley et al., 2013]. CRISPR/Cas9 is a further mechanism that enables precisely double-strand breaks [Doudna and Charpentier, 2014]. The CRISPR/Cas system is part of the defense mechanism in archaea and bacteria and protects them against penetrating viral or plasmid DNA [Horvath and Barrangou, 2010]. The bacteria or the archaea integrates small fragments of the invading viral DNA into the CRISPR region of the genome, transcribes it, and forms the RNA duplex (tracrRNA:crRNA). When this RNA duplex (single guide RNA) binds at complementary viral DNA, the CRISPR-associated protein Cas9, which is an endonuclease, performs precise double-strand breaks in the viral DNA at the binding position and, therefore, inactivates the virus. This principle can be applied to genome editing, as Cas9 can perform very precise breaks in a target DNA (e.g. plant DNA), which enables, for example, the deletion, insertion, or replacement of a certain DNA sequence or the modulation of the transcription of a specific gene [Doudna and Charpentier, 2014].

A further method for genome editing is RNA-dependent DNA methylation. Without altering the DNA sequence, the gene expression can be changed by epigenetic variations like promoter methylation or silencing of transcriptional genes. These epigenetic variations can be inherited to the next generations [Hartung and Schiemann, 2014]. Further examples for gene editing methods used for plant breeding are oligonucleotide-directed mutagenesis, reverse breeding, grafting on genetically modified rootstock, and agro-infiltration [Hartung and Schiemann, 2014].

3 Materials and Methods

3.1 Materials

3.1.1 Test material

The following maize grains were used as test material:

Genetic stability was investigated in GM hybrid maize grains of the variety MON810 DKB 350YG containing a single MON810 event and having the unique identifier MON-ØØ81Ø-6.

For the transcriptome analysis, GM maize grains of the hybrid variety AG9045 NK603 containing a single NK603 event and having the unique identifier MON-ØØ6Ø3 of the year 2012/13 were analyzed. Further, grains of the conventional, non-GM maize variety AG9045 of the year 2008/09 were investigated. All three maize varieties were obtained from the Brazilian seed company Sementes Agroceres.

3.1.2 Primers

3.1.2.1 Primers used for the verification of MON810

Primer name	Sequence	Reference
VW01 (fwd)	5‘-TCGAAGGACGAAGGACTCTAACG-3‘	[ISO, 2005]
VW03 (rev)	5‘-TCCATCTTTGGGACCACTGTCG-3‘	

Table 1: List of primer names and sequences for the verification of the event MON810

3.1.2.2 Primers used for the zygosity testing of MON810

Primer name	Sequence	Reference
CRYfwd	5‘-TCTTCACGTCCAGCAATCAG-3‘	[Rosati et al., 2008]
HECTExfwd	5’-TCAATCATCAAAGCATCATCG-3’	
HECTupRev3	5‘-TTTGGGAAGGAAAAGGTATC-3‘	

Table 2: List of primer names and sequences for the zygosity testing of the MON810 transgene

3.1.2.3 Primers used for the HRM analysis

Primer name	Sequence	Reference
3m810 fwd.	5'-CCAAGCACGAGACCGTCAA-3'	[Neumann et al., 2011]
3m810 rev.	5'-CTCGCAAGCAAATTCGGAA-3'	

Table 3: List of primer names and sequences for the HRM analysis

3.1.2.4 Primers used for the Amplicon PCR

Primer name	Sequence	Reference
Amplicon-3m810 fwd.	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCAAGCACGAGACCGTCAA-3'	[modified after Neumann et al., 2011]
Amplicon-3m810 rev.	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GCTCGCAAGCAAATTCGGAA-3'	

Table 4: List of primer names and sequences for the Amplicon PCR

3.1.2.5 Primers used for the gene expression testing

Gene number	Primer name	Direction	Zm no. [Cannon et al., 2011] / Entrez no. [NCBI, n.d.] of coding genes	Sequence
1	G1B	fwd.	Zm00001d033846	5`-ACCAGCGAACCAGAGATT-3'
		rev.		5`-CTTGCTTTCCAGCTCCTTATC-3'
2	Up6A	fwd.	Zm00001 d031127	5`-TACTCCATAGAGGAGGACAGGG-3'
		rev.		5`-CTGCACATCTTGAGCTCTCTCG-3'
3	G3B	fwd.	Zm00001d020025	5`-GTTCTAGATCCGAGGATGACCTA-3'
		rev.		5`-ATGCTGAGCTGCTTGTTCT-3'
4	U14B	fwd.	Zm00001d028814	5`-AGCGTCAGGCAGTTCAA-3'
		rev.		5`-CCTCGATGAGCGTGTTCT-3'
5	G5B	fwd.	100192063	5`-CTACCGTGCTCGATTCTTC-3'
		rev.		5`-GAGCGACAGGGAGAAATG-3'
6	8D3	fwd.	Zm00001d022464	5`-GTCGCTCAAGAACTTCCTC-3'
		rev.		5`-ACTCGAACTCCATTTCTTAGAC-3'
7	D6B	fwd.	Zm00001d036222	5`-GCGTCTCTGGTGTTCAT-3'
		rev.		5`-CAGAAGTCAAACTCCGTCTC-3'
8	G10B	fwd.	Zm00001d028816	5`-GTGATGGACTGGCACAC-3'
		rev.		5`-GACTTGCACTCGCACTT-3'

Table 5: List of primer names, sequences, and coding genes for the gene expression testing

Gene number	Primer name	Direction	Zm no. [Cannon et al., 2011] / Entrez no. [NCBI, n.d.] of coding genes	Sequence
9	1U1	fwd.	Zm00001d022084	5`-GGATGTGCGTGTTCCAG-3’
		rev.		5`-CGGAGAAGAAGCTGTTGAG-3’
10	R1G	fwd.	Zm00001d012221	5`-TGTCGCTAGCTGTCAGTGTC-3’
		rev.		5`-CCAATCTGGGTTCCAAATCGTA-3’
11	U13C	fwd.	Zm00001d025081	5`-ACTCAACGTGCCCTATCT-3’
		rev.		5`-CATGCTAATTCAGACCTCACT-3’
12	G13A	fwd.	Zm00001d048947	5`-GCGACGTACCACCTGTA-3’
		rev.		5`-CACGGTTCTTCACCCTTATG-3’
13	G12A	fwd.	Zm00001d048949	5`-GAACAACTGGGACCTCAAC-3’
		rev.		5`-TGGTCCACGATCCTCAC-3’
14	ZB	fwd.	Zm00001d035559	5`-ACTTCAGCAGACTCGCCTTC-3’
		rev.		5`-GATCCATCCGTCACCACTCC-3’
15	G2C	fwd.	Zm00001d002160	5`-TTCTCTGCACCTCAGCACAA -3’
		rev.		5`-GACGCTCATCGCCACCA-3’
16	11D2	fwd.	103652895	5`-CAGCAGATACCGTGCATAC-3’
		rev.		5`-GAAAGCCCTCACTTTCTTTAATC-3’
17	UP7C	fwd.	Zm00001d038717	5`-AACCCCTCAAGGGAAAGGCTA-3’
		rev.		5`-GCTCAAGCTCGACGACCATC-3’
18	D13C	fwd.	Zm00001d036293	5`-ACATGCTCATCAAGCTCATC-3’
		rev.		5`-TCGTGGTAGTTGAGGTAGTT-3’
19	G9B	fwd.	Zm00001d035737	5`-AAGGATGGATGCTTGGATTT-3’
		rev.		5`-GCTTTCCATCTGCTCTCATAG-3’
20	G4E	fwd.	Zm00001d047260	5`-ACGGGAGGGACATCTTCAT -3’
		rev.		5`-CACTAGACCAGAGAAGACGGAT-3’
21	D10B	fwd.	Zm00001d035597	5`-CCAAGGTCGCCTACGTC-3’
		rev.		5`-CCCGTGAGCTGGAAGTT-3’
22	UD	fwd.	Zm00001d037941	5`-CGCAGCTGTTTTGGATCGAG-3’
		rev.		5`-AGCCGGCAATTAACAGACCA-3’

Table 5: List of primer names, sequences, and coding genes for the gene expression testing (continuing)

3.1.2.6 Primers for the reference gene used for the gene expression testing

Primer name and coding protein	Direction	Sequence	Reference
Ubiquitin	fwd.	5'-AGACCCTGACTGGAAAAACC-3'	[Thiebaut et al., 2014]
	rev.	5'-CGACCCATGACTTACTGACC-3'	

Table 6: Primer sequence of ubiquitin as reference gene for the gene expression testing

3.1.3 Kits

Wizard® DNA Clean-Up System (Promega)

RNase-Free DNase Set (Qiagen)

RNeasy Mini Kit (Qiagen)

Qubit® dsDNA BR Assay Kit (ThermoFisher Scientific)

Qubit® RNA BR Assay Kit (ThermoFisher Scientific)

5X Green GoTaq® Reaction Buffer (Promega)

GoTaq® DNA Polymerase (Promega)

GoTaq® 1-Step RT-qPCR System (Promega)

Type-it® HRM™ PCR Kit (Qiagen)

HotStarTaq Master Mix Kit (Qiagen)

Nextera XT Index Primer (Illumina)

3.1.4 Equipment list

- Pipettes (1000µl, 200µl, 100µl, 20µl, 10µl, 2µl, Gilson Pipetman)
- Scales (sartorium, models R160P and BP2100)
- Thermomixer (Eppendorf, model comfort)
- Centrifuges (Eppendorf, models 5415D, 5424R, and 5810R)
- Vortexer (VELP Scientifica, model Wizard Advanced IR Vortex Mixer)
- Vacuum pump (Welch-Ilmvac, model 2522)
- ThermoStat (Eppendorf, model C)
- Nanophotometer™ (IMPLEN, version V1.6.1.)

- Fluorometer (ThermoFisher Scientific, model Qubit® 2.0)
- Gel electrophoresis with Electrophoresis Power Supply (Amersham Pharmacia Biotech, model EPS 601)
- Gel Doc™ XR (Bio Rad, Universal Hood II)
- Microwave (Hitachi, MR-5400)
- Thermocycler (Eppendorf, model Mastercycler® ep. gradient)
- Shaker (Eppendorf, model Mix Mate)
- Varioklav (H+P)
- Refrigerators (Liebherr, at +4°C and -20°C)
- Real-time PCR cycler (Qiagen, Rotor-Gene Q)

3.2 Methods

3.2.1 DNA extraction

For the DNA extraction, single grains were crushed using a garlic press and a mortar; 150mg of the maize powder of each grain were mixed with 820µl TNE buffer (5ml Tris [10mM, pH=8], 4.39g NaCl, 2 ml EDTA [2mM], and 50ml SDS [1%]), 300µl proteinase K (600 µg/mL), and 150µl Guanidin-HCl (5M). The samples were shaken in a thermomixer overnight at 60°C. The following morning, the samples were centrifuged for 5 minutes (16100 × g) at room temperature; 560µl of the upper aqueous phase was transferred into a new tube and 300µl chloroform was added. After 20 seconds of vortexing, the two phases were separated by centrifuging for 8 minutes, and 500µl of the upper phase was transferred to a new tube. To eliminate the RNA, 2µl of RNAse (8µg/ml) was added to each sample. Next, the samples were incubated in the thermomixer for 30 minutes under constant shaking at 60°C. The DNA was purified with the help of the Wizard® DNA Clean-Up Kit, which eliminates restriction enzymes, mononucleotides, DNA polymerases, exonucleases, and endonucleases. 1ml of the Wizard® DNA Clean-Up Resin was added to each sample and the resin/DNA mix was transferred to minicolumns withholding the DNA plus resin in filters. Using a vacuum, the solution was drawn through the minicolumns. Subsequently, the columns were purified twice with 1ml 80% isopropanol each. For drying the filters, the

vacuum was reapplied for 1 minute. Then, the minicolumns were put in a 1.5ml collection tube and 20µl of nuclease-free water was pipetted in every filter. The samples were centrifuged for 1 minute. Finally, the minicolumns were transferred from the collection tube in a new 1.5ml tube and 70°C warm 10mM Tris (pH=7,4) was pipetted into every filter of the minicolumns to eluate the DNA. After 10 minutes' incubation at room temperature, the samples were centrifuged for 1 minute to transfer the purified DNA into the tubes.

3.2.2 RNA extraction

For the RNA extraction, a protocol based on [Chang et al., 1993] and [Barros et al., 2010] in a slightly altered form was performed.

Maize grains were homogenized in fluid nitrogen using a garlic press and a mortar. 500mg of the homogenized grains was transferred into a 15ml tube and 5ml of 60°C warm extraction buffer (2% CTAB, 2% PVP [mol. weight: 40,000g/mol], 2M NaCl, 0.9mM DEPC, 0.5mM spermidine, 100mM Tris [pH=8]) and 100µl of mercaptoethanol were added per sample. After vortexing, 5ml of chloroform-isoamylalcohol (24:1; CIA) was added. The samples were vortexed and centrifuged for 15 minutes at 15°C (3184 × g). The upper aqueous phase was transferred into a new tube and mixed with 5ml of CIA. The samples were centrifuged for 15 minutes as before. This CIA extraction was repeated once. After the third centrifugation, the upper phase was again transferred into a new tube and 900µl of 10M lithium chloride with 1mM DEPC was added. The samples were vortexed and stored in the refrigerator at 4°C overnight.

The next morning, the samples were centrifuged for 30 minutes at 5°C (3184 × g) and the supernatant was eliminated. The formed pellet was solved with 500µl of 60°C warm SSTE buffer (1M NaCl, 0.5% SDS solution, 1mM EDTA, 1mM DEPC, 10mM Tris [pH=8]). Then, the buffer/pellet mixture was transferred into a 2ml tube and 500µl of CIA was added. After vortexing, the samples were centrifuged for 10 minutes at 21°C (16363 × g). The upper aqueous phase was transferred to a new tube and 96% ethanol was added in the threefold amount of the sample (approximately 1.5ml). Then, the samples were incubated on ice for 5 minutes and centrifuged at 4°C for 35 minutes (16363 × g). After eliminating the

supernatant, 250µl of 75% ethanol was added and the samples were vortexed shortly. The samples were centrifuged at 4°C for 10 minutes (16363 × g) and the supernatant was removed. For drying the pellet, the tubes were incubated with open lids for 10 minutes at room temperature, and 150µl of 10mM Tris (pH>7) was added. Next, the samples were incubated at 65°C for 10 minutes in the thermostat. To dissolve the RNA, the samples were pipetted up and down. The extracted RNA was subsequently purified or stored immediately at -80°C.

3.2.2.1 DNA digestion and RNA purification

DNA digestion was done using the RNase-Free DNase Set and RNA purification using the RNeasy Mini Kit [Qiagen, 2012]. For the DNA digestion, 70µl of the extracted RNA, 10µl of RDD buffer, 2.5µl DNAse stock solution, and 17.5µl RNase-free water were pipetted into a 1.5ml tube. To degrade the DNA by DNAse, the samples were incubated in the thermostat at 23°C for 10 minutes. For the RNA purification, 350µl of RLT buffer was added and the samples were vortexed. Then, 250µl of 96% ethanol was added and the samples were mixed by pipetting them up and down. Each sample was transferred to an RNeasy spin column withholding the RNA in the silica membrane and centrifuged for 15 seconds at room temperature (11363 × g). 500µl of RPE washing buffer was pipetted into each spin column before the samples were centrifuged again for 15 seconds, as before. After a second washing step with 500µl of RPE washing buffer, the samples were centrifuged for 2 minutes (11363 × g). Then, the spin columns were placed in a new 1.5ml tube. For the elution of the RNA, 50µl of RNase-free water was pipetted into the spin columns and the samples were centrifuged for 1 minute (11363 × g) to transfer the purified RNA into the tubes.

3.2.3 Photometer

The purity and concentration of 3µl of the extracted DNA and RNA were measured photometrically with a nanophotometer. To determine the purity of the DNA, the absorbance ratio of 260nm/280nm wavelength was measured. This ratio indicates protein contamination, since the optimal absorbance of DNA is at 260nm, while that of proteins is at 280nm wavelength. The ideal A_{260}/A_{280} ratio is 1.8, showing pure DNA.

To check the purity of the RNA, the absorbance ratios of 260nm/280nm and of 260nm/230nm wavelength, as well as the absorbance at 320nm wavelength, were measured. The ideal A_{260}/A_{280} ratio for RNA is 2.0. The A_{260}/A_{230} ratio should be higher as the A_{260}/A_{280} ratio, and the ideal A_{320} value is near 0, as a higher value indicates protein contamination [Matlock, n.d.].

3.2.4 Fluorometer

The DNA concentration was also measured with a fluorometer using the Qubit® dsDNA BR Assay Kit. One component of this kit is the dsDNA BR reagent containing a fluorescent dye, which has a very low fluorescence, when it is unattached. As soon as it binds to a DNA molecule (or an RNA molecule, when RNA is measured), the fluorescence becomes very intensive and can be measured. For the calibration, two standards with known concentrations are used. Standard 1 has a dsDNA concentration of 0ng/µl and standard 2 of 100ng/µl. After measuring the fluorescence of the sample, the fluorometer calculates the concentration of the dsDNA.

Before performing the measurements, a working solution was prepared. For each sample and each standard, 1µl dsDNA BR reagent was diluted with 199µl of dsDNA BR buffer to achieve a 1:200 dilution. For the preparation of the standards, 190µl of the working solution was pipetted into a Qubit® assay tube and 10µl of dsDNA BR Standard 1 and, respectively 2, was added. For the samples, 2µl of the DNA sample was merged into 198µl of the working solution. After vortexing, the standards and samples were incubated in the dark for 10 minutes. Then, the fluorescence was measured and the concentration was calculated [Thermo Fisher Scientific, 2015a]. The RNA concentration was measured similarly using the Qubit® RNA BR Assay Kit, including the RNA BR buffer and the RNA BR reagent [Thermo Fisher Scientific, 2015b].

3.2.5 Gel electrophoresis

3.2.5.1 1% agarose gel

The degradation and contamination of the DNA were controlled using gel electrophoresis on a 1% agarose gel. DNA fragments were separated by length; thus, the size of the fragments could be estimated, and the degradation and contamination could be controlled. To prepare the gel, 1.5g agarose was dissolved in 150ml TAE buffer in the microwave. Before the gel hardened, 4µl of ethidium bromide was added for the detection of the DNA. 8µl of the samples, together with 2µl of 5x DNA loading buffer, was loaded into the swells of the hardened gel. 6µl of a DNA ladder (200 bp to 10 kbp) was also loaded on to the gel. The gel electrophoresis was run for 23 minutes with 140 Volt. Then, the DNA was visualized with the help of the UV light of a Gel Doc.

3.2.5.2 2.5% agarose gel

After the performance of a PCR, the PCR products were controlled using a gel electrophoresis on a 2.5% agarose gel. After the PCR described in 3.2.6.1, 9µl of each sample and 6µl of DNA ladder (100bp-2kbp) were loaded on to the gel and compared against each other after running the gel with the same instrument settings as in 3.2.5.1.

After the PCR described in 3.2.6.2, 6µl of each sample and 6µl of DNA ladder (100bp-2kbp) were loaded on to the gel. The gel electrophoresis was run for 175 minutes with 90 Volt.

3.2.5.3 2% agarose gel

The PCR products of the Amplicon PCR described in 3.2.9.1, as well as the Index PCR described in 3.2.9.3, were controlled by performing a gel electrophoresis on a 2% agarose gel. 3g of agarose were dissolved in 150ml of 0,5X TBE Buffer, and 10µl of GelRed were used to detect the DNA. After the gel hardened, 5µl of each PCR product, together with 1µl of 5x DNA loading buffer and 8µl of DNA ladder (100bp-15kbp), were loaded on to the gel. The gel electrophoresis was run for 45 minutes with 125 Volt.

3.2.5.4 5% formamide, 1% agarose gel

The purity of the RNA was checked using gel electrophoresis on a 1% agarose gel. The gel was prepared as described in 3.2.5.1. Additionally, 5% formamide was added before the gel hardened. About 1000ng RNA of each sample, together with 5µl of 5x loading buffer, was pipetted into a swell of the gel. As a marker, 6µl of DNA ladder (100bp to 2kbp) was used. The gel electrophoresis was run for 90 minutes with 80 Volt.

3.2.6 PCR

3.2.6.1 PCR to check the presence of the MON810 transgene

To control the presence of the DNA sequence of the MON810 transgene, a PCR was conducted using the primers VW01 (forward primer) and VW03 (reverse primer) (see 3.1.2.1), which bind to the 35S promoter of the transgene.

The reaction mixture of a reaction volume of 20µl is presented in the table below.

Component	Volume per sample	Final concentration
5X Green GoTaq® Reaction Buffer	4µl	1x
each primer (100µM)	0,5µl	2,5µM
dNTP (each 10mM)	0,5µl	each 250µM
GoTaq® DNA Polymerase	0,3µl	1,5u
ddH2O	13,2 µl	-
DNA sample	1µl	variable

Table 7: Reaction mixture for the PCR to check the presence of MON810

The following temperature program was chosen for the PCR:

1. Initialization step: 94°C for 10 minutes
2. Denaturation step: 94°C for 30 seconds
3. Annealing step: 58°C for 30 seconds
4. Extension step: 72°C for 30 seconds
5. Final extension: 72°C for 1 minute
6. Final hold: 16°C

Steps 2 to 4 were repeated 40 times.

The PCR products were loaded on to a 2.5% agarose gel as described in 3.2.5.2.

3.2.6.2 PCR to determine the zygosity of the MON810 transgene

The degree of zygosity of the MON810 transgene was tested on the principles of the PCR-based zygosity-testing method of [Liu and Chen, 2009]. The method was adapted using the three MON810 specific primers HECTExfwd, CRYfwd, and HECTupRev3 (see 3.1.2.2) published by [Rosati et al., 2008]. HECTExfwd binds to the 5'flanking region of the plant genome and HECTupRev3 to the 3'flanking region. This primer pair is used to detect wild-type maize without a MON810 transgene, as well as hemizygous samples. CRYfwd binds to the 3'end of the transgene. The CRYfwd/HECTupRev3 primer pair spans from the transgene to the endogenous plant genome and detects, therefore, homozygous and hemizygous samples.

The reaction mixture of a reaction volume of 20µl is presented in the table below.

Component	Volume	Final concentration
5X Green GoTaq® Reaction Buffer	4 µl	1x (1.5 mM $MgCl_2$)
GoTaq® G2 DNA Polymerase	0,3 µl	1,5u
dNTPs (each 10mM)	0,5 µl	250µM
HECTExfwd primer (10µM)	0,5µl	250nM
Cryfwd primer (10µM)	0,5µl	250nM
HECTupRev3 primer (10µM)	1µl	500nM
nuclease-free water	11,2µl	-
DNA sample	2 µl	variable

Table 8: Reaction mixture of the PCR to determine the zygosity of the MON810 transgene

The following temperature program was chosen for the PCR:

1.) Initialization step: 94°C for 10 minutes
2.) Denaturation step: 94°C for 30 seconds
3.) Annealing step: 54°C for 1 minute
4.) Extension step: 72°C for 1 minute
5.) Final extension: 72°C for 8 minutes
6.) Final hold: 16°C

Steps 2 to 4 were repeated 30 times.

The PCR products were loaded on to a 2.5% agarose gel, as described in 3.2.5.2.

3.2.7 Real-time PCR

3.2.7.1 Real-time PCR efficiency testing

Before an RT-qPCR was performed, the PCR efficiency was determined. A dilution series with five different concentrations (mostly 25ng/µl, 5ng/µl, 1ng/µl, 0.2ng/µl, and 0.04ng/µl) of a sample was generated (standard 1-5). As a reference gene, the maize-specific housekeeping gene ubiquitin was chosen. Each standard was tested with ubiquitin and with a specific primer in duplicate. The efficiency was tested with all 22 specific primers listed in 3.1.2.5.

For the performance of the RT-qPCR for the efficiency testing, the GoTaq® 1-Step RT-qPCR System was used with the following reaction mixture and a reaction volume of 15µl:

Component	Volume for efficiency testing	Volume for analyzing the gene expression	Final concentration
GoTaq® qPCR Master Mix, 2X	7,5µl	10µl	1x
each primer (2µM)	1,5 µl	2µl	0,2µM
GoScript™ RT Mix 50X	0,3µl	0,4µl	1x
ddH2O	1,2µl	1,6µl	-
RNA sample	3µl	4µl	variable

Table 9: Reaction mixture for the efficiency testing and the analysis of the gene expression by RT-qPCR

The following temperature program was chosen for the performance of the RT-qPCR:

1.) Hold: 37°C for 15 minutes
2.) Initialization step: 95°C for 10 minutes
3.) Denaturation step: 95°C for 10 seconds
4.) Annealing step: 60°C for 30 seconds
5.) Extension step: 72°C for 30 seconds
6.) Melt step: 60-99°C, rising 1°C per 1.5 minutes

Steps 2 to 4 were repeated 45 times.

3.2.7.2 Analysis of the gene expression by RT-qPCR

For the comparison of the gene expression of endogenous genes of GM maize grains and conventional maize grains, a RT-qPCR was performed using the primers listed in 3.1.2.5. As a reference, the housekeeping gene ubiquitin was included in every run. Since housekeeping genes are genes, which are necessary for maintaining basic cellular functions, they are expected to have the same expression in all cells of a specific organism independently of influencing factors like the type of tissue, or the cell cycle state, and are therefore used as reference genes [Eisenberg and Levanon, 2013]. Each investigated gene was tested in an individual RT-qPCR run. The difference in the gene expression of a specific gene, as well as of ubiquitin, was analyzed by comparing three sample pools (each consisting of ten different maize grains) of GM maize grains (2.5µg/ml) with three sample pools of conventional maize grains (2.5µg/ml). Each sample pool was tested in triplicate. To obtain the standard straight line and the slope, a dilution series with five different concentrations of a sample was included after the efficiency was proven in the efficiency testing (see 3.2.7.1). Each standard of the dilution series was tested in duplicate with the specific primer as well as with ubiquitin.

The reaction mixture (see Table 9, volume for analyzing the gene expression) and the temperature program for the RT-qPCR were the same as in 3.2.7.1, but a reaction volume of 20µl was used.

3.2.8 High-Resolution Melting analysis

A quantitative PCR, in combination with an High-Resolution Melting (HRM) analysis, was performed using the Type-it HRM PCR Kit and the primer pair 3m810 (forward and reverse) (see 3.1.2.3). One component of the HRM PCR Mastermix is the fluorescent dye EvaGreen®, which can be integrated into dsDNA. Thus, based on the intensity of the fluorescence, the denaturation of the DNA can be measured when the temperature is constantly increased in the HRM step.

The reaction mixture of a reaction volume of 16µl is presented in the table below.

Component	Volume per sample	Final concentration
HRM PCR Mastermix (2x)	8 µl	1x
each primer (10µM)	1,12 µl	0,7 µM
ddH_2O	4,16 µl	-
DNA sample (40ng/µl)	1,6 µl	4 ng/µl

Table 10: Reaction mixture of the qPCR and the HRM analysis

The following temperature program was chosen for the qPCR and the HRM analysis:

1.) Initialization step: 95°C for 5 minutes
2.) Denaturation step: 95°C for 10 seconds
3.) Annealing step: 55°C for 30 seconds
4.) Extension step: 72°C for 20 seconds
5.) HRM: 80–92°C

Steps 2 to 4 were repeated 40 times. The increase in the temperature in the HRM step was 0.2°C per 4 seconds.

3.2.9 Amplicon sequencing

3.2.9.1 Amplicon PCR

In the Amplicon PCR, the 3´border region of the MON810 transgene was amplified using the primers Amplicon-3m810 (forward and reverse) (see 3.1.2.4), which have overhanging adapters.

The reaction mixture of a reaction volume of 25µl is presented in the table below.

Component	Volume per sample	Final concentration
each Amplicon PCR primer (1µM)	5µl	0,2µM
HotStarTaq Master Mix (2x)	12,5µl	1x
DNA sample (5ng/µl)	2,5µl	0,5ng/µl

Table 11: Reaction mixture of the Amplicon PCR

The following temperature program was chosen:

1.) Initialization step: 95°C for 15 minutes
2.) Denaturation step: 95°C for 30 seconds
3.) Annealing step: 62°C for 30 seconds
4.) Extension step: 72°C for 30 seconds
5.) Final extension: 72°C for 10 minutes
6.) Final hold: 4°C

Steps 2 to 4 were repeated 40 times.

The Amplicon PCR products were loaded on to a 2% agarose gel, as described in 3.2.5.3.

3.2.9.2 PCR clean-up

The amplicons obtained in the Amplicon PCR (see 3.2.9.1) were purified to eliminate free primers. After centrifuging the PCR tubes for 1 minute at room temperature (1000 × g), 20µl of each PCR product were transferred to a new 96-well MIDI plate. 90µl of the vortexed AMPure XP beads (magnetic beads) was added to each sample. Then, the MIDI plate was shaken for 2 minutes on a shaker and incubated for 5 minutes at room temperature. Next, it was placed on a magnetic stand, on which the magnetic beads, together with the DNA, were magnetically attached to the bottom. While incubating the samples for 2 minutes, the supernatant was cleared. Then, the complete supernatant of each sample was removed and discarded. 200µl of 80% ethanol was added to each sample to wash the magnetic beads. After 30 seconds of incubation, the ethanol was removed. This washing step was repeated once. Next, the magnetic beads were dried by incubating the MIDI plate for 10 minutes. Then, the MIDI plate was removed from the magnetic stand and 52.5 µl of 10mM Tris (pH=8) was added to each sample. The plate was shaken for 2 minutes and incubated for 2 minutes. Then, the plate was placed again on the magnetic stand and incubated for 2 minutes. After

the attachment of the magnetic beads to the bottom of the MIDI plate, 50µl of the supernatant (DNA dissolved in Tris) of each sample was transferred to a new 96-well PCR plate.

3.2.9.3 Index PCR

After the PCR clean-up described in 3.2.9.2, the Index PCR was performed to attach dual indices to the overhanging adapters of the amplicons. An index was attached to each end of every sample, whereas the index sequence of eight nucleotides was different at both ends. Each sample was marked with an individual index combination, which enabled each sample to be clearly identified after the Amplicon sequencing. The Nextera XT Index 1 and 2 primers of the Nextera XT Index Kit were used [Illumina, 2013]. Index 1 primers were adjacent to further nucleotides of the P7 sequence and Index 2 primers to the P5 sequence. Since 12 different Index 1 primers and 8 different Index 2 primers were used, 96 samples could be merged in the preparation of the NGS run and were analyzed together.

The reaction mixture of a reaction volume of 50µl is presented in the table below.

Component	Volume	Final concentration
HotStarTaq Master Mix (2x)	25µl	1x
nuclease-free water	10µl	-
each primer	5µl	*
sample	5µl	variable

Table 12: Reaction mixture of the Index PCR

*primer concentration was not provided by Illumina

After pipetting the reaction mixture, the 96-well plate was centrifuged for 1 minute at room temperature (1000 × g).

For the Index PCR, the following temperature program was chosen:

1.) Initialization step: 95°C for 15 minutes
2.) Denaturation step: 95°C for 30 seconds
3.) Annealing step: 55°C for 30 seconds
4.) Extension step: 72°C for 30 seconds
5.) Final extension: 72°C for 10 minutes
6.) Final hold: 4°C

Steps 2 to 4 were repeated eight times.

After the performance of the Index PCR, the 96-well plate was centrifuged for 1 minute (280 × g) and 50µl of each Index PCR product was transferred to a new 96-well plate. Then, the PCR clean-up was repeated as described in 3.2.9.2. In the last step, only 25µl of the supernatant was transferred to a new 96-well plate.

3.2.9.4 Dilution, pooling, and denaturation

The DNA concentration of the samples was measured after the second PCR clean-up using the fluorometer, as described in 3.2.4. Then, the samples were diluted with 10mM Tris (pH=8) to achieve a concentration of 4nM.

5µl of all the diluted samples were pooled in a new 1.5ml tube and vortexed. 5µl of this pool was mixed with 5µl of 0.2M NaOH to denature the DNA. After centrifuging for 2 minutes (280 × g), the pool was incubated at room temperature for 5 minutes. Next, 990µl of pre-chilled Hybridization Buffer was added and 2µl of this dilution was spiked to an NGS run.

4 Results

4.1 Approach and aim of the study

4.1.1 Investigation of the 3'border region of the MON810 transgene

The first part of this master's project focused on the analysis of a border region spanning the end of a transgene to the endogenous plant genome. Border regions are used for the official qualitative and quantitative control of GMOs. However, modifications in these regions might lead to wrong results if PCR-based methods are performed [Ben Ali et al., 2014]. Therefore, the investigation of the genetic stability in these sections is essential.

In earlier investigations of the working group in which this master's project was performed, genetic instabilities were found in the 3'border region of the MON810 transgene in a stacked maize variety (MON88017 x MON810) [Ben Ali et al., 2014], whereas the border regions of maize containing a single MON810 event showed genetic stability [Neumann et al., 2011]. It was hypothesized that instabilities occur more frequently in GM crops containing stacked events as single events. This study was performed to test this hypothesis by controlling the genetic stability of the 3'border region of the MON810 transgene in a single-event maize variety. The methods used for this investigation were HRM analysis and Amplicon sequencing.

The target region of the analysis is located at the 3'insert-to-plant junction of the MON810 transgene. It comprises the 3'end of the Cry1A(b) gene and the adjacent nucleotides of questionable purpose merging into the maize genome. Therefore, the 180bp long target region is event-specific [Neumann et al., 2011].

The work flow of the investigation was as follows: First, the genomic DNA of individual grains of MON810 maize was extracted and the quality of the samples was checked. Second, the target sequence of 202 samples was amplified in a qPCR and prescreened for the detection of mutations in the following HRM analysis. The deviation of the melting behavior of the samples to the reference sample was calculated and declared as confidence values. Samples with the highest deviation and, therefore, the lowest confidence values were most

suspected to have mutations. Thus, the 20 samples that showed the lowest mean of the confidence values of the two HRM analyses were selected for the sequencing using Amplicon sequencing. Additionally, the four reference samples were sequenced. Finally, the resulting reads of the target sequence were compared to the reference sequence to identify mutations.

4.1.2 Comparison of the gene expression of a NK603 maize variety to the nearly isogenic, conventional maize variety

The introduction of a transgene into the endogenous plant genome leads to intended effects, for example, herbicide or insecticide resistance. Additionally, unintended effects, for instance, differences in gene expression, might occur [Sorochinskii et al., 2011]. Regrettably, unintended effects of GM plants have not been extensively analyzed yet. In the second part of the master's project, the hypothesis that the insertion of the NK603 transgene unintentionally changes the gene expression of endogenous maize genes was tested. The gene expression of a NK603-containing maize variety was compared with that of the nearly isogenic non-GM maize variety.

In previous investigations of the working group, the maize variety AG8025 NK603 was compared to the non-transgenic comparator variety AG8025 by RNA sequencing. In the following differential gene expression analysis using the programs DESeq2, EdgeR, and CLC, 286 differently expressed genes were identified by all the three programs. In all, 236 genes were up-regulated in GM maize grains, whereas 50 genes were down-regulated [Ben Ali et al., unpublished]. To verify these results, the gene expression of 22 genes was additionally analyzed by RT-qPCR. 18 of these 22 genes were confirmed in this way, showing a statistically significant result [Draxler, unpublished].

To investigate whether the differently expressed genes are unintended effects caused by the NK603 transgene, a further NK603 maize variety was compared with the non-transgenic comparator variety in this master's project.

For this investigation, the RNA of maize grains containing NK603 as a single event, as well as of conventional maize grains, was extracted and sample pools were merged. By

performing RT-qPCR analyses, the gene expression of the same 22 genes as in the earlier investigated NK603 maize variety was compared between the GM maize and the conventional maize. The relative expression ratio and the log2fold change of each gene were calculated by relative quantification. Next, the statistically significant expressed genes were summarized as GO categories.

Finally, the resulting log2fold changes of the 22 genes were compared with those of the first investigated NK603 maize variety to determine whether the discovered unintended effects occurred due to the transgene NK603.

4.2 Verification of the genetic stability

4.2.1 Characteristics of the extracted DNA

4.2.1.1 DNA concentration and quality

Altogether, the genomic DNA of 264 grains of the maize variety MON810 DKB 350YG was extracted, as described in 3.2.1. While measuring the DNA concentration using the photometer, the range of all the samples was from 15µg/ml to 184µg/ml. The mean was 80.5µg/ml (SD: 28.3 µg/ml, CV: 35.1%) and the median was 77.3µg/ml.

The A_{260}/A_{280} ratio indicates the purity of the DNA and was measured and calculated by the photometer. The mean of this ratio of all the samples was 1.796 and the median was 1.8. This signals that the DNA was mostly very pure.

Since a fluorometer measures only the concentration of the target molecules (dsDNA) of the sample and less contaminating molecules as a photometer, the results are more accurate compared to those of a photometer. To control the photometrically measured DNA concentrations, 72 samples were additionally measured by a fluorometer. The DNA concentrations ranged from 4.5µg/ml to 153.0µg/ml with a mean of 54.6µg/ml (SD: 26.2µg/ml, CV: 48.0%) and a median of 53.7µg/ml. On an average, the concentrations were lower than those measured photometrically.

4.2.1.2 Verification of the DNA on a 1% agarose gel

The purity and degradation of the DNA extracts were controlled by performing a gel electrophoresis with a 1% agarose gel (see 3.2.5.1). The bands of the DNA extracts were approximately at the same position as the 10kbp band of the DNA ladder, which is the supreme band of the ladder. All the 264 samples showed a clear band in this position. As already seen by checking the A_{260}/A_{280} ratio, the DNA was very pure, and the samples had, at the most, just a weak degradation.

While comparing the intensity of the DNA bands of the different samples, the approximate concentration of the DNA could be estimated. Most samples that had a high concentration measured photometrically also showed a more intensive band on the gel, whereas the samples with a low concentration showed lighter bands.

In Figure 1, the agarose gel with the DNA bands of the samples 41–48 is shown. In the first position, the DNA ladder (200bp-10kbp) is visible. Sample 43 shows the lightest band of these eight samples and has the lowest concentration measured with the photometer (42.5µg/ml), whereas the samples 46 and 48 show the most intensive bands and have the highest concentrations (46: 82.0µg/ml and 48: 101µg/ml).

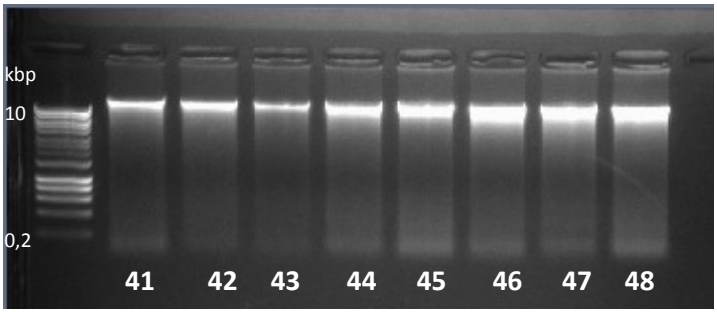

Figure 1: Checking the purity of the DNA of the samples 41–48 on a 1% agarose gel

4.2.2 Verification of the presence of the MON810 transgene

A qualitative PCR was conducted to control the presence of the MON810 transgene in the maize grains of DKB 350YG, as described in 3.2.6.1. This PCR was performed as verification with 36 of the 264 samples with the primer pair VW01/VW03. Then, the PCR products were

checked on a 2.5% agarose gel using gel electrophoresis. The amplified DNA fragment has a length of 170bp [Margarit et al., 2006] and should, therefore, be located between the bands at 100bp and 250bp of the DNA ladder. All the 36 samples showed an intensive DNA band in this position, whereas the no-template control (NTC) next to the DNA ladder showed no DNA band. In Figure 2, the agarose gel with the PCR products of the samples 1–8 can be seen.

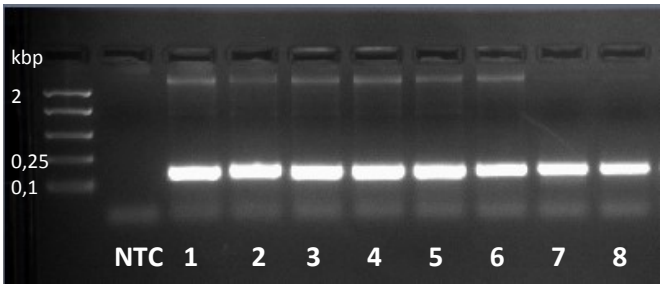

Figure 2: Checking the presence of MON810 in the samples 1–8 on a 2.5% agarose gel

4.2.3 Zygosity of the MON810 transgene

The zygosity describes the similarity of the alleles for a certain trait in the genome of an organism. Genes can be heterozygous, homozygous, or hemizygous. Heterozygosity means that a certain gene is present in the genome with different alleles, whereas homozygosity means the presence of the same allele of a gene twice in the genome. A hemizygous gene is present as only one copy in organisms with diploid cells, which might be the result of a loss of chromosomes or the absence of a second related gene (e.g. the X chromosome in men) [Sauermost, 2001].

The MON810 transgene in hybrid maize grains can be homozygous as well as hemizygous. A hemizygous transgene, which is more common, can be inherited from a transgenic female parent or a transgenic male parent [Zhang et al., 2008]. When the female as well as the male parent plant contain the transgene, the progenies inherit the transgene homozygously.

The degree of zygosity of the MON810 transgene was tested by performing a qualitative PCR with the three primers HECTExfwd, CRYfwd, and HECTupRev3 [Rosati et al., 2008] (see 3.2.6.2). Subsequently, the PCR products were loaded on to a 2.5% agarose gel (see 3.2.5.2).

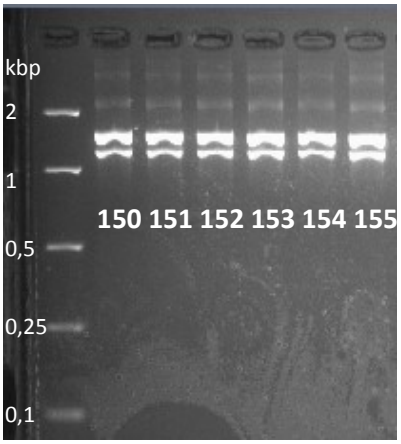

Figure 3: Verification of hemizygosity of MON810 in the samples 150-155 on a 2.5% agarose gel

In wild-type maize grains without the MON810 transgene, a fragment of 1502bp (HECTExfwd to HECTupRev3) is amplified, whereas in grains with a homozygous MON810 transgene, a fragment of 1246bp (CRYfwd to HECTupRev3) is detected [Rosati et al., 2008]. In grains with a hemizygous MON810 transgene, both the mentioned fragments are amplified and, therefore, two DNA bands are visible on the gel.

20 samples of MON810 DKB 350YG were tested. All the samples showed two DNA bands, indicating that the samples were hemizygous. Figure 3 presents the agarose gel of the zygosity testing of the samples 150–155, showing the two DNA bands between the 2kbp and the 1kbp bands of the DNA ladder.

4.2.4 HRM analysis

4.2.4.1 Investigation of the 3'border region of MON810

An HRM analysis aims to detect variations of single bases of the DNA as well as insertions or eliminations of short DNA sequences [Druml and Cichna-Markl, 2014]. Thus, the HRM analysis was already successfully used to detect natural as well as induced mutations in many different plant species. SNPs can be identified when they are homozygous as well as heterozygous [Simko, 2016]. While performing an HRM analysis, the length of the analyzed amplicon should not exceed 300bp [Druml and Cichna-Markl, 2014], as the method has a higher sensitivity in small fragments. Vossen et al. recommend for fragment screening the use of fragments of 150–200bp and, for SNP typing, the use of fragment sizes of 80–100bp [Vossen et al., 2009].

For the performance of an HRM analysis, the target DNA sequence is first amplified by qPCR in the presence of the fluorescent dye EvaGreen. After finishing the last PCR cycle, the produced amplicons are melted by slowly raising the temperature and the DNA strands separate. Since the fluorescent dye only binds double-stranded DNA, the gradually denaturing amplicons liberate the dye and the measured fluorescence signal decreases [Simko, 2016].

HRM analysis was performed to detect variations in the 3'border region of the MON810 transgene to the endogenous plant genome. High quality and purity of the DNA samples are essential for working with HRM analysis [Derveaux et al., 2010]. The 202 samples of the 264 DNA extracts, which showed the highest purity on the 1% agarose gel and on the A_{260}/A_{280} ratio (approximately 1.8) and which had neither a too low (<40µg/ml) nor a very high DNA concentration, were selected. The samples were diluted with nuclease-free water to obtain a DNA concentration of 40µg/ml. Then, the 180bp long target sequence spanning from the 3'end of the Cry1A(b) gene to the endogenous maize genome was amplified in the qPCR and denatured in the HRM analysis. The qPCR and the HRM analysis of each sample were performed twice on different days. The investigations were conducted in four runs in the first screening and in three runs in the second screening with a larger number of samples per run. The data was analyzed with the software Rotor-Gene Q Series Version 2.3.1 from Qiagen. The results were presented as amplification curves, melting curves, and confidence values by the software.

4.2.4.1.1 Analysis of the amplification curves

The results of the qPCR were controlled by analyzing the amplification curves, which should reach the plateau. This was achieved by performing 40 qPCR cycles. Figure 4 shows, as an example, the amplification curves of the samples 112–181 in the second screening. Since the concentrations of the samples are approximately the same, the Cycle threshold (Ct) values, which are usually in the area of the start of the exponential increase of the curves, and the course of the amplification curves are similar. The mean of the Ct values of theses samples is 21.44 Ct values (SD: 0.21 Ct values, CV: 0.98%). The Ct values range from 21.08

(sample 123) to 22.24 (sample 176). The means of the Ct values of all seven runs (four in the first screening and three in the second) range from 19.71 to 21.44 Ct values.

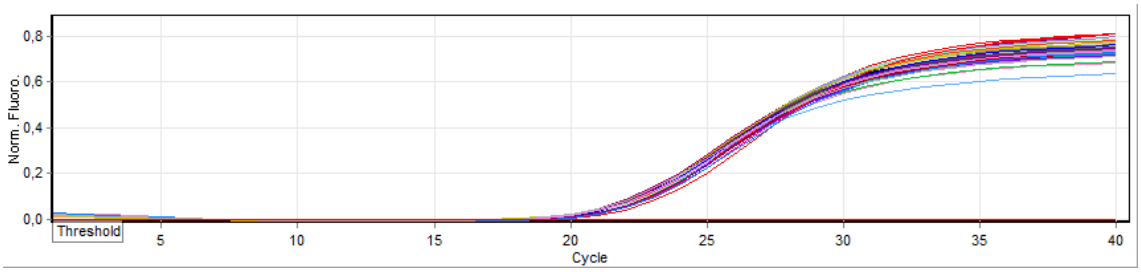

Figure 4: Amplification curves of the qPCR of the samples 112-181 in the second screening

Noticeably, divergent Ct values showed sample 32 in both screenings (30.41 in the first screening and 30.36 in the second), though it was newly diluted for the second screening based on the DNA concentration measured by the fluorometer instead of the photometer to surely obtain the DNA concentration of 40µg/ml.

4.2.4.1.2 Analysis of the melting curves and the difference plots

To detect nucleotide variations, the melting curves and the confidence values were considered. While plotting the negative derivative of the fluorescence signal measured over the temperature against the temperature, the melting temperature and the melting domains can be analyzed. The melting temperature is defined as the temperature at which half of the DNA strands are separated and varies depending on the length, the sequence, and the GC (guanosine/cytosine) content of the amplicons [Reed et al., 2007]. The higher the GC content, the higher is the stability and, therefore, the melting temperature of this DNA sequence. Even small variations of the nucleotides shift the melting temperature. Thus, a sample with a divergent melting curve is most likely to have a genetic variation [Druml and Cichna-Markl, 2014].

201 samples showed a similar melting curve and melting temperature of about 87°C; thus, theses samples did not seem to have a variation. Again, the only sample showing a divergent melting curve was number 32. In both screenings, the melting curve was slightly shifted and smaller than that of all the other samples. This is demonstrated in Figure 5, which shows

the melting curves of the samples 1–55 in the first screening. The small, black curve is the melting curve of sample 32.

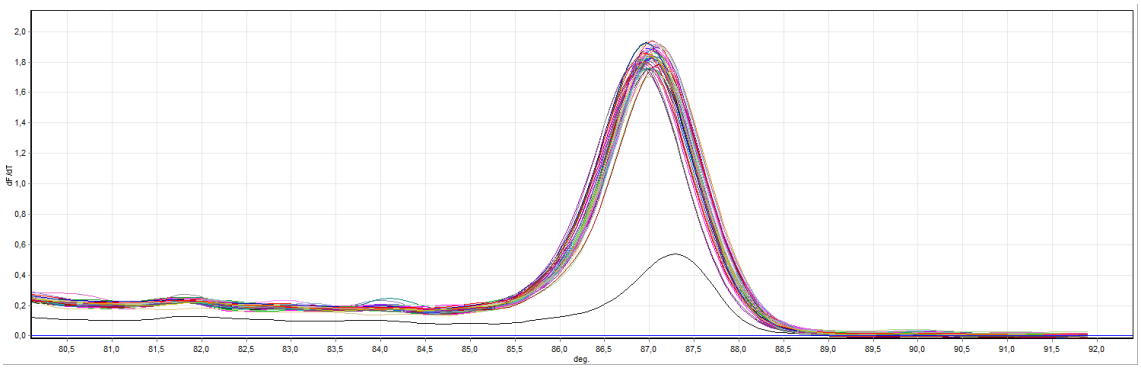

Figure 5: Melting curves resulting from the HRM analysis of the samples 1–55

Small differences of the melting curves of the different samples can be detected in a difference plot by subtracting the normalized melting curve of a sample—defined as the reference sample—from the normalized melting curves of the remaining samples and plotting it against the temperature [Wittwer et al., 2003]. As an example, Figure 6 shows the difference plot of the samples 1–55 in the first screening. Sample 32, which is the black curve, has the largest deviation of the reference sample number 44.

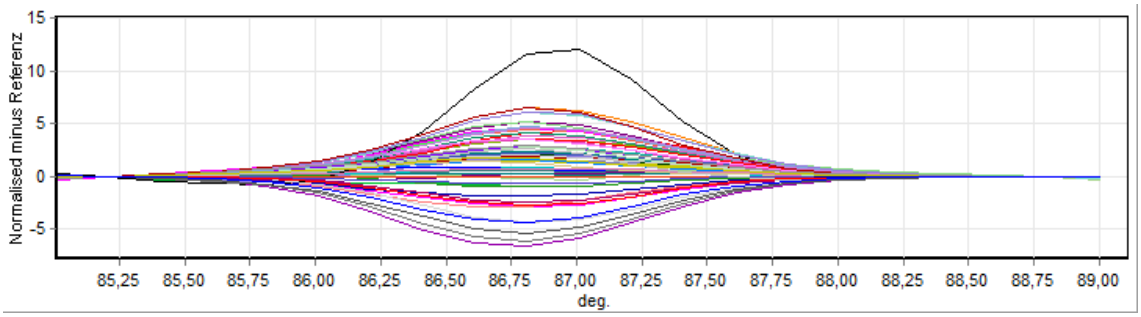

Figure 6: Normalized difference plot obtained by HRM analysis of the samples 1–55 with sample 44 as a reference

While interpreting the melting curves and the difference plots of all the samples, sample 32 seemed to be the only suspect sample for having variations.

4.2.4.1.3 Analysis of the confidence values

The identification of nucleotide variations was mainly based on the analysis of the calculated confidence values. For comparing, a pool reference containing 20 different DNA samples in the same proportion was included in every HRM analysis run. However, when the results were interpreted, one sample of each run that showed an average amplification curve and an average melting curve was chosen as a reference, since the pool reference could not be sequenced in the following Amplicon sequencing. Samples number 44, 92, 169, and 193 were selected as reference samples.

In every run, the confidence value of the reference sample was adjusted to 100%, whereas the sample with the highest deviation in the melting behavior compared to the reference had the lowest confidence value. A low confidence value indicated the possibility of a variation in the DNA sequence.

As an example, Table 13 shows the confidence values of the second screening of samples 112–181. Sample number 169 was chosen as the reference and, thus, had a confidence value of 100%. The confidence values of these samples, except for the reference, have a mean of 82.79% (SD: 13.71%, CV: 16.56%) and range from 52.36% (sample 141) to 99.92% (sample 175).

No.	Colour	Sample	Confidence %
1		112	75,14
2		113	71,05
3		114	77,96
4		115	81,17
5		116	75,79
6		117	82,24
7		118	69,84
8		119	74,32
9		120	69,18
10		121	79,63
11		122	74,71
12		123	81,08
13		124	57,15
14		125	76,56
15		126	87,07
16		127	72,33
17		128	95,51
18		129	79,43
19		130	91,70
20		131	80,91
21		132	75,04
22		133	71,21
23		134	82,76
24		135	63,78
25		136	54,12
26		137	70,90
27		138	65,44
28		139	56,87
29		140	63,78
30		141	52,36
31		142	76,95
32		143	53,35
33		144	54,48
34		145	69,01
35		146	78,53
36		147	73,60
37		148	87,36
38		149	91,35
39		150	85,53
40		151	88,50
41		152	77,48
42		153	98,66
43		154	97,22
44		155	99,57
45		156	98,39
46		157	97,36
47		158	94,69
48		159	87,44
49		160	77,59
50		161	90,09
51		162	97,04
52		163	97,55
53		164	95,35
54		165	88,63
55		166	96,63
56		167	93,94
57		168	93,33
58		169	100,00
59		170	99,68
60		171	99,84
61		172	96,40
62		173	99,18
63		174	97,06
64		175	99,92
65		176	89,29
66		177	94,96
67		178	95,81
68		179	98,60
69		180	96,68
70		181	96,23
71		Pool2	99,81
72		ntc	38,14

Table 13: Confidence values of the samples 112–181 in the second screening obtained by qPCR and HRM analysis

Table 14 and Table 15 summarize the results of all HRM analyses. The mean of the confidence values, together with the standard derivation and the variation coefficient, the median of the confidence values and the sample with the lowest confidence value of each run of the first screening are listed in Table 14 and of the second screening, in Table 15.

Run no.	Included samples	Mean confidence value [%]	SD [%]	CV [%]	Median confidence value [%]	Lowest confidence value [%] (sample)	Reference sample
1st run	1-55	93,92	7,16	7,62	96,47	58,16 (32)	44
2nd run	56-105, 112-123	94,04	8,15	8,67	96,8	55,80 (105)	92
3rd run	106-111, 124-181	93,96	6,07	6,46	96,62	75,48 (139)	169
4th run	182-202	90,46	20,56	22,73	97,14	5,91 (183)	193

Table 14: Results of the HRM analyses of the first screening

Run no.	Included samples	Mean confidence value [%]	SD [%]	CV [%]	Median confidence value [%]	Lowest confidence value [%] (sample)	Reference sample
1st run	45-111	92,05	13,60	14,78	96,28	15,90 (48)	92
2nd run	112-181	82,79	13,71	16,56	82,76	52,36 (141)	169
3rd run	1-44, 182-202	96,84	3,61	3,72	97,71	85,28 (41)	193

Table 15: Results of the HRM analyses of the second screening

Two samples had a low confidence value in only one screening: sample 183 in the first (confidence value of 5.91%) and sample 48 in the second screening (confidence value of 15.90%). Both samples were screened a third time, resulting in confidence values of 98.82% for sample 183 and 99.66% for sample 48. The low confidence values of both samples in only one screening were most likely due to an insufficient amplification. Therefore, it could be assumed that none of the 202 samples were suspected to have nucleotide variations according to the confidence values.

The mean of the confidence values of the two—and for samples 48 and 183, three—HRM analyses was calculated. The 20 samples with the lowest mean were selected for the following Amplicon sequencing. In Table 16, these 20 samples are listed in the ascending order of the mean of the confidence values.

Sample number	Mean of 2 (3*) confidence values [%]
105	56,24
139	66,18
141	66,33
183*	68,13
143	69,01
107	70,03
48*	71,04
144	71,34
136	72,61
145	74,73

Sample number	Mean of 2 (3*) confidence values [%]
32	75,71
124	75,83
138	76,12
140	76,39
113	76,59
116	76,70
108	78,23
118	78,29
95	79,06
142	80,50

Table 16: Presentation of the confidence values of the 20 samples selected for Amplicon sequencing (without reference samples)

4.2.5 Amplicon sequencing

Amplicon sequencing is a recently developed method of NGS. Instead of sequencing the whole genome, for which NGS is often applied, Amplicon sequencing can be used to analyze a certain DNA sequence in numerous samples [Bybee et al., 2011]. Thus, in many high-throughput DNA sequencing studies, Amplicon sequencing is the method of choice [Murray et al., 2015]. NGS technologies can be used to detect small modifications in the DNA sequence, such as small deletions or inserts as well as nucleotide substitutions [Pauwels et al., 2015].

Amplicon sequencing was done to detect variations in the 180bp long target sequence in the 3'border region of the MON810 transgene. As HRM analysis was used as a prescreening, only the 20 samples, which had the lowest mean of the confidence values, were sequenced

(see Table 16). Additionally, the four samples, which served as a reference in the HRM analyses (samples 44, 92, 169, and 193), were also sequenced.

In general, sequencing technologies include the following three steps: preparation of the samples, sequencing and imaging, and data analysis [Metzker, 2010]. For the preparation of the samples, a protocol of Illumina [Illumina, 2013] in a slightly altered form was followed. The preparation of Amplicon sequencing includes two PCRs [Bybee et al., 2011]. In the Amplicon PCR, the target sequence was amplified, and adopters were attached (see 3.2.9.1). After the performance of a PCR clean-up, dual indices were attached to the adaptors as an individual barcode for each sample in the Index PCR (see 3.2.9.3). The PCR clean-up was repeated, and the samples were pooled. The denatured and diluted pool was spiked to an NGS run.

4.2.5.1 Quality control of the samples by using gel electrophoresis

A high quality of each sample is essential for pooling up to 96 samples in one NGS run. Therefore, the PCR products resulting from the Amplicon PCR and the Index PCR were controlled on a 2% agarose gel (see 3.2.5.3). The analyzed sequence has a length of 180bp [Neumann et al., 2011]. In the Amplicon PCR, adapters of 33 and 34 nucleotides were attached by the forward and reverse primer, respectively [Illumina, 2013]. Thus, the amplicons obtained in the Amplicon PCR had a length of 247bp. All the samples showed an intensive band on the agarose gel between the 200bp and the 300bp band of the DNA ladder. Figure 7 presents the DNA bands of the amplicons of the samples 141–145, 107, 108, 169, and 183 after the Amplicon PCR on the agarose gel.

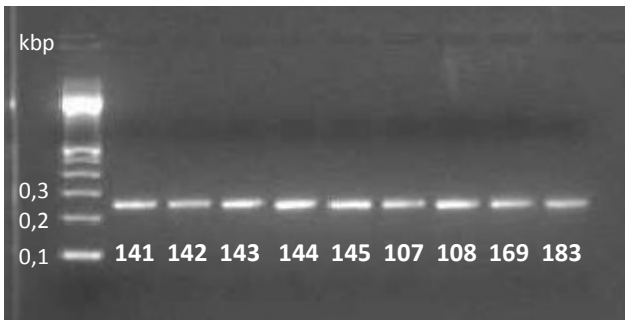

Figure 7: Verification of the amplicons after the Amplicon PCR on a 2% agarose gel

In the Index PCR, two indices of eight nucleotides, as well as the P5 and P7 sequences of 24 and 29 nucleotides, were attached [Illumina, 2013], resulting in an amplicon length of

316bp. Since all the samples showed a clear band on a 2% agarose gel at approximately the position of the 300bp band of the DNA ladder, the samples could be pooled and sequenced.

4.2.5.2 Sequencing results of the 3'border region of the MON810 transgene

The data of the sequencing was analyzed using the software CLC Genomics Workbench 7.5. The number of reads per amplicon, which could be mapped to the reference sequence, ranged from 8,900 to 16,700. The reads of the amplicons were paired and trimmed to eliminate the terminal nucleotides, such as the indices. Then, the reads were mapped to the reference sequence published by [Neumann et al., 2011]. A threshold value for the detection of nucleotide variations of 1% was defined. This means that the software only identifies a variation when more than 1% of the reads of a sample harbor a mutation in a certain position.

Hardly any single nucleotide polymorphism was detected in the analyzed DNA sequence at the 3'border region of the MON810 transgene. 23 samples had no SNP at all. Only one sample, number 32, showed a nucleotide variation in position 91 of the target sequence in a very low frequency. In 1.46% of the reads, adenine was replaced by guanine. This led to the transformation of the codon CAC to CGC. Consequently, the amino acid histidine was replaced by arginine.

The reference sequence of the 3'border region of MON810 [modified after Neumann et al., 2011]:

"CCAAGCACGAGACCGTCAACGTGCCCGGTACTGGTTCCCTCTGGCCGCTGAGCGCCCCCAGCCC
GATCGGCAAGTGTGCCCACCACAGCCACCACTTCTCCTTGGACATCGATGTGGGCTGCACCGACCT
GAACGAGGACTTTCGGTAGCCTTCTTTCATTTCCGAATTTGCTTGCGAGC"

The sequence of the 180bp target region at the 3'junction of the MON810 transgene is presented above. The sequence of the forward primer used for the HRM analysis is marked in blue and the sequence of the reverse primer in red. The position of the SNP detected in sample number 32 is marked in yellow. In 1.46% of the reads, this adenine was replaced by guanine.

4.3 Transcriptome analysis

In the second part of the master's thesis, the transcriptome of GM maize grains of the variety AG9045 containing NK603 and of the nearly isogenic, non-GM maize grains of the variety AG9045 was compared. The reference laboratory for GMO of the Austrian Agency for Health and Food Safety verified the presence of NK603 in the GM maize grains and the absence of NK603 in the conventional variety by performing a qPCR with event-specific primers.

4.3.1 External comparison of the GM maize grains to the conventional maize grains

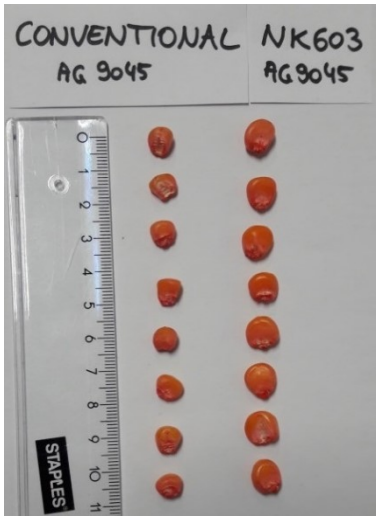

Figure 8: Comparison of the conventional maize grains to the GM grains

The GM maize grains (AG9045 NK603) were visually compared to the conventional maize grains (AG9045). The GM grains were bigger than the conventional grains and had a typical maize grain form, whereas the conventional grains were smaller and had a rather spherical form. Figure 8 presents the GM as well as the conventional maize grains and shows these visual differences.

Further, 35 single maize grains of each variety were weighted. The mean of the weight of one grain of the conventional maize was 0.23g (SD: 0.021g, CV: 9.17%), ranging from 0.20g to 0.28g. The mean of the weight of one grain of the GM maize was 0.34g (SD: 0.020g, CV: 5.81%), ranging from 0.30g to 0.38g. Thus, the visually detected difference in the size can also be proven by the weight of the grains.

4.3.2 Characteristics of the extracted RNA

4.3.2.1 RNA concentration and quality

The RNA of 68 grains of the conventional maize variety AG9045 and of 46 grains of the GM maize variety AG9045 containing NK603 was extracted as described in 3.2.2. Always, three to ten grains were extracted together to form one sample. Thus, 20 RNA samples of the conventional grains and 14 RNA samples of the GM grains were obtained.

Next, the RNA concentrations were measured photometrically. The RNA concentrations of the conventional grains ranged from 149µg/ml to 480µg/ml with a mean of 304.6µg/ml (SD: 84.55µg/ml, CV: 27.76%), and a median of 283.5µg/ml. The RNA concentrations of the GM grains had a mean of 560.36µg/ml (SD: 82.55µg/ml, CV: 14.73%), a median of 576.5µg/ml, and ranged from 421µg/ml to 700µg/ml.

For the determination of the purity of the RNA, the A_{260}/A_{280} ratio, the A_{260}/A_{230} ratio, and the A_{320} were measured by photometer. The following measurement results were considered as ideal for RNA extracts: the A_{260}/A_{280} ratio between 1.9 and 2.2, the A_{260}/A_{230} ratio >2.2, and the A_{320} value near 0. The 20 RNA samples of the conventional maize grains had the following measurement results: a mean of the A_{260}/A_{280} ratio of 2.13, a mean of the A_{260}/A_{230} ratio of 2.29, and a mean of the A_{320} value of 0.003. All samples fulfilled the specified quality criteria. The following values of the 14 RNA samples of the GM maize grains were measured: a mean of the A_{260}/A_{280} ratio of 2.09, a mean of the A_{260}/A_{230} ratio of 2.26, and a mean of the A_{320} value of 0.004. Only two samples had a slightly too low A_{260}/A_{230} ratio; apart from that, all the quality criteria were fulfilled.

After the RNA purification, the RNA concentrations were also measured by the fluorometer. The RNA concentrations of the conventional maize samples had a mean of 362.69µg/ml (SD: 107.86µg/ml, CV: 29.74%). The concentrations ranged from 85.2µg/ml to 580µg/ml and had a median of 357µg/ml. The RNA concentrations of the GM maize samples had a mean of 550.71µg/ml (SD: 164.87µg/ml, CV: 29.94%), ranging from 282µg/ml to 858µg/ml with a median of 536µg/ml.

Since the RNA samples should be free of DNA contamination [Derveaux et al., 2010], the DNA concentration of the samples was measured by the fluorometer after the DNA digestion was performed. None of the samples had a higher concentration than 4.47µg/ml. This indicates that the DNA digestion worked well.

4.3.2.2 Verification of the RNA on a formamide agarose gel

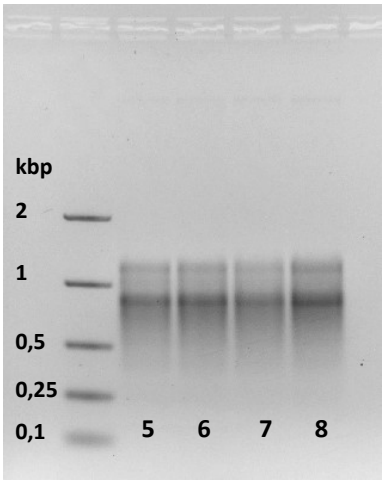

Figure 9: Verification of RNA bands of the GM samples 5-8 on a 5% formamide 1% agarose gel

To control the degradation of the RNA extracts, the samples were applied on a 5% formamide 1% agarose gel (see 3.2.5.4) before the DNA digestion and the RNA purification were performed. Two bands were seen on the gel: the one with the longer fragments showing the 28S rRNA and the other one, the 18S rRNA. Ideally, the band of the 28S rRNA is twice as intense as the band of 18S rRNA [Sambrook and Russell, 2001]. The GM samples showed two RNA bands on the gel, but in all the samples, the band of 18S rRNA was more intense compared to the band of 28S rRNA. On the agarose gel of the conventional samples, the two bands were not clearly visible because of smear of partially degraded RNA. Figure 9 shows the two RNA bands of the GM samples number 5–8 on the agarose gel. The 18S rRNA band is more intense as the 28S rRNA band and the degraded RNA smear is visible.

4.3.3 Real-time PCR efficiency

While comparing different samples, a constant amplification efficiency is necessary to obtain reliable results [Pfaffl, 2004]. Thus, the efficiency of a dilution series was always tested before the proper RT-qPCR was performed. By plotting the Ct values of the five standards of the dilution series against the known concentration of the standards, the standard straight line and their slope were determined by the software Rotor Gene Q Series Version 2.3.1. The efficiency (E) was calculated by substituting the slope (M) into the following formula:

$$E = 10^{(-1/M)}$$ [Rasmussen, 2001]

$$E\ (\%) = (10^{(-1/M)} - 1) * 100$$

The ideal slope is close to -3.322, resulting in an efficiency of 2, or rather, 100%. This indicates an exact doubling of the amplicons in every qPCR cycle. A slope of -3.1 to -3.6, or rather, an efficiency of 90% to 110% was accepted in the efficiency test. Then, the gene expression was tested with the same dilution series.

4.3.4 Comparison of the gene expression of GM maize to conventional maize

To measure the gene expression, RT-qPCR was used. To perform a RT-qPCR, a very small amount of mRNA as starting material is enough. The extracted mRNA was first translated into the corresponding single-strand complementary DNA (cDNA) with the aid of the retroviral enzyme reverse transcriptase. Next, the cDNA was amplified exponentially [Freeman et al., 1999]. The amplification was detected in "real time" using the fluorescent dye BRYT Green®, which binds to double-stranded DNA. By measuring the fluorescence, the concentration of the DNA was calculated.

The gene expression was compared between the GM maize variety AG9045 containing NK603 and the conventional, nearly isogenic maize variety AG9045. It is necessary to take the nearly isogenic maize variety as a comparator and not a variety, which differs in more than just the absence of the transgene to the GM variety, to eliminate the possibility that the observed differences in the gene expression occur due to genes other than the transgene.

Since high RNA quality and purity are necessary for obtaining reliable results from a RT-qPCR [Derveaux et al., 2010], the samples, which, based on the quality criteria described in 4.3.2.1, had the highest quality after the extraction and the purification of the RNA, were chosen for the analysis. The samples were pooled resulting in sample pools consisting of 10 different maize gains each and diluted to obtain a concentration of 2.5µg/ml. Next, an efficiency testing was performed with each pool. Three pools of each variety showing the

best efficiencies were selected for the test of the gene expression of 22 endogenous maize genes by RT-qPCR. In each RT-qPCR run, each pool was tested in triplicate with one of the 22 target primers listed in 3.1.2.5 and as a reference with the housekeeping gene ubiquitin.

Additionally, a dilution series of a sample was included in each RT-qPCR to control the efficiency. This sample was either the RNA of the GM or the conventional maize, showing an optimal efficiency when tested. The five standards of the dilution series were tested in duplicate with a target primer as well as with ubiquitin.

After the performance of the RT-qPCR, the results were analyzed using the software Rotor Gene Q Series. First, the efficiency of the straight line of the dilution series was calculated for the target gene as well as for ubiquitin. As described in 4.3.3, an efficiency of 90–110% of at least three standards was an obligatory prerequisite to guarantee reliable results of the RT-qPCR. Second, the melting curves of the amplicons of each pool were compared to control that in each pool (GM and conventional), the same fragment was amplified. Melting curves at different temperatures can indicate the amplification of different fragments. Only the pools amplifying the same fragment were compared with each other.

The difference of the gene expression of the 22 target genes was determined by relative quantification. It was expressed relatively to the expression level of the maize-specific housekeeping gene ubiquitin. The calculation was based primarily on the comparison of the Ct values of the conventional pools to those of the GM pools of the target gene as well as of ubiquitin. In general, the mean of the Ct values of the three pools (each in triplicate) was calculated. Pools having Ct values with a higher deviation than two standard deviations of the mean of all pools of this variety were excluded. The difference between the mean of the Ct values of the conventional pools and that of the Ct values of the GM pools was calculated and compared between the target gene and ubiquitin. A correction for the efficiencies of the dilution series was included for a better estimation of the relative expression ratio (R) [Pfaffl, 2004]. The log2fold change was obtained by computing the logarithm for the base 2 of the relative expression ratio.

$$E = 10^{(-1/M)}$$ [Rasmussen, 2001]

$$\Delta Ct(t) = \bar{x}_{(Ct(t\ CON-pools))} - \bar{x}_{(Ct(t\ GM-pools))}$$
$$\Delta Ct(ubi) = \bar{x}_{(Ct(ubi\ CON-pools))} - \bar{x}_{(Ct(ubi\ GM-pools))}$$

$$R = \frac{E(t)^{\Delta Ct(t)}}{E(Ubi)^{\Delta Ct(ubi)}}$$ [Pfaffl, 2001]

$$log2fold\ change = log_2(R)$$

M…slope

E…efficiency

R…relative expression ratio

t…target gene

ubi…ubiquitin as reference

CON-pools…conventional pools

GM-pools…genetically modified pools

Relative expression ratios between 0 and 0.9999, and consequently, negative log2fold values indicate a down-regulation of the target gene in the GM samples compared to the conventional samples. Relative expression ratios > 1 and, therefore, positive log2fold values indicate an up-regulation of the target gene in the GM samples. Generally, log2fold values ≥ 1 and ≤ -1 were taken as significant results, as this threshold was taken by other authors too [e.g. La Paz et al., 2014].

As an example, the gene expression of gene number 22 is presented in more detail. The means of the Ct values of the conventional pools were 29.77 for the target gene and 20.70 for ubiquitin, whereas the means of the GM pools were 27.36 for the target gene and 20.01 for ubiquitin. The efficiency of the target gene was 2.1012 (110%) and of ubiquitin, 2.0088 (101%). While substituting these values into the formula quoted before, a relative expression ratio of 3.6932 and a log2fold change of 1.8849 were calculated. Thus, the gene expression was strongly up-regulated in the GM maize grains compared to the conventional grains.

Figure 10 demonstrates the difference in the gene expression of the target gene number 22. It presents the amplification curves of gene 22 (orange: GM pools, violet: conventional pools) and ubiquitin as the reference (gray: GM pools, green: conventional pools). The amplification curves of the GM pools of the target gene (orange) increased clearly earlier compared to those of the conventional pools (violet), leading to lower Ct values of the GM pools compared to the conventional pools. This shows an increase in the gene expression of gene 22 in the GM grains.

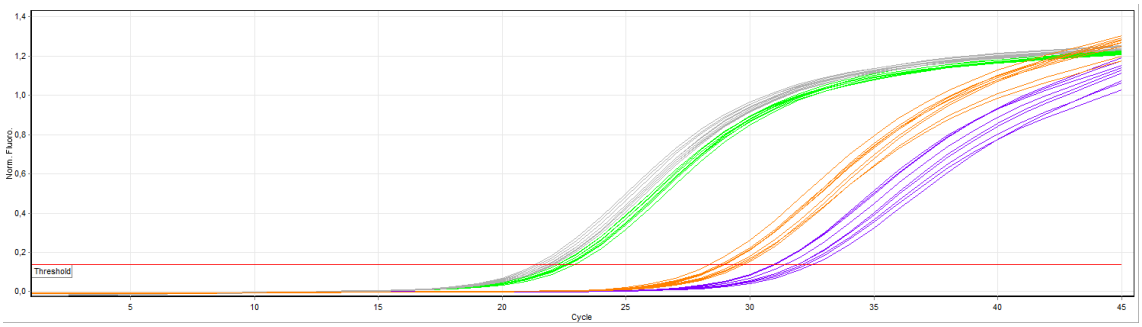

Figure 10: Amplification curves of ubiquitin (gray: GM pools, green: conventional pools) and gene 22 (orange: GM pools, violet: conventional pools) obtained by RT-qPCR

Table 17 presents the efficiencies of the straight line of ubiquitin and the target genes, as well as the log2fold changes of the gene expression testing of all 22 target genes.

Gene no.	Primer name	Zm number [Cannon et al., 2011]/ Entrez no. [NCBI, n.d.]	Efficiency ubiquitin [%]	Efficiency target gene [%]	Log2fold change
13	G12A	Zm00001d048949	102	104	-2,6722
12	G13A	Zm00001d048947	94	100	-1,9789
7	D6B	Zm00001d036222	107	109	-1,2786
9	1U1	Zm00001d022084	98	109	-1,2411
18	D13C	Zm00001d036293	101	99	-1,2765
16	11D2	103652895	98	97	-0,6751
15	G2C	Zm00001d002160	105	108	-0,4037
20	G4E	Zm00001d047260	98	98	-0,3411
8	G10B	Zm00001d028816	106	100	-0,2295
17	Up7C	Zm00001d038717	105	107	-0,2241
4	U14B	Zm00001d028814	100	94	-0,1804
3	G3B	Zm00001d020025	105	96	-0,0559
6	8D3	Zm00001d022464	103	110	-0,0024
19	G9B	Zm00001d035737	101	104	0,0671
21	D10B	Zm00001d035597	100	103	0,2936
2	Up6A	Zm00001d031127	100	94	0,4839
1	G1B	Zm00001d033846	100	106	0,5463
5	G5B	100192063	97	105	0,8213
14	ZB	Zm00001d035559	96	103	1,7219
22	U-D	Zm00001d037941	101	110	1,8849
10	R1G	Zm00001d012221	101	105	2,7616
11	U13C	Zm00001d025081	97	106	2,9627

Table 17: Efficiencies and log2fold changes of the gene expression testing of 22 target genes listed in ascending order of the log2fold changes

4.3.5 Interpretation of the results of the transcriptome analysis

4.3.5.1 Differently expressed genes

While defining log2fold changes ≥ 1 as statistically significant results, nine of the 22 target maize genes were differently expressed between the GM and the conventional maize grains. Four of these nine genes were up-regulated, whereas five genes were down-regulated in the GM maize grains. Table 18 shows the chromosomes on which these nine genes are located and the proteins they code.

Gene no.	Primer pair	GRMZM no.	Zm no. [Cannon et al., 2011]	Coding protein [The UniProt Consortium, 2017]	Chromosome no. [The UniProt Consortium, 2017]
13	G12A	GRMZM2G117971	Zm00001d048949	Hevein-like preproprotein	4
12	G13A	GRMZM2G117989	Zm00001d048947	Pathogenesis-related protein3	4
7	D6B	GRMZM2G005499	Zm00001d036222	Putative sugar phosphate/phosphat e translocator	6
9	1U 1	GRMZM2G045155	Zm00001d022084	B12D protein	7
18	D13C	GRMZM2G127948	Zm00001d036293	Caffeoyl-CoA O-methyltransferase 1	6
11	U13C	AC204711,3_FG00 3	Zm00001d025081	Rhodanese-like domain-containing protein 19 mitochondrial	10
10	R1G	GRMZM2G316362	Zm00001d012221	Acyl-desaturase	8
22	U-D	GRMZM2G085964	Zm00001d037941	AP2-EREBP transcript factor	6
14	ZB	GRMZM2G112238	Zm00001d035559	Dirigent protein	6

Table 18: Coding proteins and the location on the chromosomes of the nine differently expressed genes

To obtain a classification of the identified genes, they were collated to the appropriate Gene Ontology (GO) termini using the tool GORetriever of the AgBase website (www.agbase.msstate.edu) [McCarthy et al., 2006]. GO is a worldwide effort to unify the different expressions for gene products in divergent databases. GO termini are collated to genes and proteins. Gene products are categorized into the three domains of cellular component, molecular function, and biological process [Ashburner et al., 2000].

Overall, 21 different GO termini were collated to the nine genes. Of these, five were collated to the domain cellular component, eight to molecular function, and eight to biological process. The GO termini of each gene are listed in the appendix in Table 22. Next, the GO termini were summarized as a Slim View using the tool GOSlimViewer of the same website [McCarthy et al., 2006]. Table 19 shows the resulting Slim View, including the summarized GO categories, and the corresponding annotations collated to the three domains mentioned before.

GO categories [McCarthy et al., 2006]	Annotation
Cellular component	
GO:0005575	Cellular component
GO:0005576	Extracellular region
GO:0005634	Nucleus
GO:0016020	Membrane
Molecular function	
GO:0003677	DNA binding
GO:0003700	Transcription factor activity, sequence-specific DNA binding
GO:0003824	Catalytic activity
GO:0005488	Binding
GO:0016740	Transferase activity
GO:0030246	Carbohydrate binding
Biological process	
GO:0006139	Nucleobase-containing compound metabolic process
GO:0006629	Lipid metabolic process
GO:0006950	Response to stress
GO:0008150	Biological process
GO:0008152	Metabolic process
GO:0009058	Biosynthetic process
GO:0009605	Response to external stimulus
GO:0009607	Response to biotic stimulus
GO:0009987	Cellular process

Table 19: Slim View of the differently expressed genes including GO categories and their annotation

4.3.5.2 Non-differently expressed genes

13 of the 22 tested genes did not show a statistically significant difference in the gene expression between the GM and the conventional maize grains. Nevertheless, the coding proteins, when known, of the 13 genes and their location on the chromosomes are presented in Table 20.

Gene no.	Primer pair	GRMZM no.	Zm no. [Cannon et al., 2011] / Entrez no. [NCBI, n.d.]	Coding protein [The UniProt Consortium, 2017]	Chromosome no. [The UniProt Consortium, 2017]
16	11D2	GRMZM2G024527	103652895	Putative ubiquitin family protein	4
15	G2C	GRMZM2G106445	Zm00001d002160	Wound induced protein	2
20	G4E	GRMZM2G425482	Zm00001d047260	Uncharacterized protein	9
8	G10B	GRMZM2G112488	Zm00001d028816	Pathogenesis-related protein 10	1
17	Up7C	GRMZM2G386674	Zm00001d038717	Seed specific protein Bn15D17A	6
4	U14B	GRMZM2G112538	Zm00001d028814	Pathogenesis-related protein 10	1
3	G3B	GRMZM2G092137	Zm00001d020025	Uncharacterized protein	7
6	8D3	GRMZM2G472852	Zm00001d022464	Ultraviolet-B-repressible protein	7
19	G9B	GRMZM2G018786	Zm00001d035737	D-glycerate 3-kinase chloroplastic	6
21	D10B	GRMZM2G034724	Zm00001d035597	Legumin-like protein	6
2	Up6A	AC203989.4_FG001	Zm00001d031127	Cupin, RmlC-type	1
1	G1B	GRMZM2G026780	Zm00001d033846	Uncharacterized protein	1
5	G5B	GRMZM2G300424	100192063	Uncharacterized protein	1

Table 20: Coding proteins and the location on the chromosomes of the 13 genes without different expression levels

5 Discussion

5.1 Genetic stability of the 3'border region of the MON810 transgene

The aim of the first part of this master's project was the verification of the genetic stability of the 3'border region of the MON810 transgene in a single-event maize variety by performing HRM analysis and NGS. Altogether, the DNA of 202 maize grains was analyzed. A substitution of adenine by guanine was detected in one sample, which led to the replacement of histidine by arginine in the amino acid sequence. However, this SNP was detected in a low frequency of 1.46%. In the remaining 201 samples, no variations were identified.

As mentioned before, Ben Ali et al. found SNPs in two samples in the 3'flanking region of the MON810 transgene of a stacked maize variety (MON88017 x MON810) [Ben Ali et al., 2014]. In contrast, while analyzing the border regions of GM plants containing a single event (maize [Neumann et al., 2011], soybean [Madi et al., 2013], and oilseed rape [Ben Ali et al., 2014]), no SNPs were detected. Consequently, it was hypothesized that SNPs occur more frequently in GM crops containing stacked events as single events. [Ben Ali et al., unpublished]. To test this hypothesis, this study was performed with a single-event maize variety.

PCR-based methods used for the official GMO control are performed with primers binding in the border regions of transgenes, which are event-specific. Mutations in this region might lead to wrong results in the control. Therefore, the detection of variations in the border regions is of vital importance [Ben Ali et al., 2014]. As instabilities and mutations were detected in the 3'flanking region of the MON810 transgene, but not in the 5'flanking region [Ben Ali et al., 2014, Rosati et al., 2008], the 3'border region of this transgene was analyzed in this project.

The screening for mutations was performed by HRM analysis, which shows numerous advantages. HRM analysis is quite inexpensive, simple to perform, and highly sensitive

[Vossen et al., 2009]. Further, a large number of samples (with the used qPCR cycler up to 70 samples) can be analyzed simultaneously. HRM analysis was already successfully used to identify natural and induced mutations in a wide variety of plants. As mentioned before, even variations of single bases can be detected due to differences in the melting curves [Simko, 2016]. However, two amplicons could have the same melting curve without having a completely identical sequence. This might occur, for instance, due to an erasing effect, when two mutations compensate for each other (e.g. T → G and G → T) [Druml and Cichna-Markl, 2014], leading to a false negative result, even though this will rarely happen. Furthermore, the identification of the number of variations and the exact localization are not possible using only HRM analysis. Therefore, the combination with a sequencing method is necessary [Simko, 2016]. Even better would be the sequencing of all samples by NGS. However, this was impossible in this study due to the high costs and the huge expenditure of time.

The last study, focusing on the genetic stability of GM crops of the working group in which this research was conducted, was performed by applying HRM analysis followed by Sanger sequencing of divergent samples [Castan et al., 2017]. As an innovative approach in this study, Sanger sequencing was replaced by NGS. A major advantage of using NGS compared to Sanger sequencing was the enormous number of produced reads. For a better imagination of the difference, Treangen et al. mentioned the comparison that about six billion shorter reads can be produced in one NGS run, whereas around 30 million reads were generated in the whole original human genome project, which was performed by Sanger sequencing [Treangen and Salzberg, 2012]. The large number of reads per sample enables a better quantification of the frequency of the detected SNPs. SNPs appearing in a very low frequency cannot always be detected by Sanger sequencing. The target sequence of sample 32 was additionally analyzed by Sanger sequencing. However, no SNP was detected by this method, whereas a SNP with a low frequency (1.46%) was found by performing NGS. Beck et al. compared the use of NGS and Sanger sequencing in a clinical study. In all, 5,660 variations were detected by NGS. Of these, 19 could not be found by Sanger sequencing first. Using newly designed primers, 17 of these variations were detected in a second Sanger sequencing run. However, two variations, both of which were in non-coding regions, could not be confirmed by Sanger sequencing [Beck et al., 2016].

A further advantage of NGS compared to Sanger sequencing is the simultaneous evaluation of millions of base pairs [Beck et al., 2016], which reduces the time for the analysis by a large extent. For instance, the whole human genome can be sequenced in only a few days using NGS [Treangen and Salzberg, 2012]. NGS generates shorter reads than Sanger sequencing, which can make the assembly and the analysis of NGS data more complex [Varshney et al., 2009]. This might be a problem, for example, for the detection of variations in repetitive DNA sequences [Treangen and Salzberg, 2012], but this is not relevant for the short amplicon, which was analyzed in this study.

Sample 32 showed a SNP with a very low frequency of 1.46% by NGS. Some explanations for this SNP, without the guarantee of completeness, could be the following: First, SNPs that are present in low frequencies could be somatic mutations [Xu et al., 2017]. Somatic mutations occur due to wrong incorporations of bases during the replication of the DNA or due to exogenous or endogenous mutagens (e.g. radiation) in dividing cells usually in the form of point mutations [Greenman et al., 2007]. Second, the maize grain may contain a further partial copy of the transgene, including the analyzed sequence at the 3'border region of the transgene. This copy could harbor a mutation at position 91, leading to the identified SNP. Third, an artifact due to the qPCR or NGS cannot be excluded. Errors can accumulate during the whole NGS protocol, from the library preparation to the sequencing and the read alignment [Xu et al., 2017].

As mentioned before, this project was performed to test the hypothesis that mutations occur more frequently in GM crops containing stacked events as single events. Additionally to the maize variety DKB 350YG containing MON810 as a single event, which is presented in this master's thesis, the 3'border region of two further maize varieties, one containing MON810 as a single event and one as a stacked event having also the transgene NK603 (variety 631 RR/Bt), were analyzed. Of each variety, 24 samples were sequenced by NGS. In 14 samples of the variety 631 RR/Bt, SNPs were detected. Of these, the majority was localized in position 71 of the analyzed target sequence, which was the same as in this investigation. Five of the 14 samples had a higher frequency of 10% (up to 53%), which were all in position 71. The further maize variety containing MON810 as a single event showed two SNPs with a very low frequency. Consequently, it seems to be correct that GM maize

varieties containing stacked events harbor more SNPs than varieties with single events. However, for the confirmation of this hypothesis, a higher number of different GM crops has to be analyzed. [Ben Ali et al., unpublished].

Since the cultivation of GM crops with stacked events is increasing [ISAAA, 2016], the investigation of the genetic stability of further stacked event varieties is necessary. The presented method is approved for the identification of small variations to be part of the molecular characterization of GM plants in addition to controls by Southern blot analysis or whole genome sequencing.

5.2 Comparison of the gene expression of a GM maize variety to the conventional comparator

The aim of the second part of this master's thesis was the identification of unintended effects in the transcriptome of GM maize due to the transgene NK603. It was hypothesized that the gene expression of endogenous genes can change unintentionally due to the insertion of a transgene. To test this hypothesis, the gene expression of 22 maize genes was compared between a GM maize variety containing the transgene NK603 (AG9045 NK603) and the nearly isogenic, non-GM maize variety (AG9045). This was done by performing RT-qPCR. Nine of these genes had a statistically significant difference in the gene expression. These 22 genes were selected because they showed a different expression in the maize variety pair investigated earlier (AG8025 NK603 vs. AG8025) [Draxler, unpublished].

RT-qPCR is a good method for analyzing gene expression because it is accurate, sensitive, simple to perform, relatively cheap, and fast [Derveaux et al., 2010, Pfaffl, 2001]. The experience of this study showed that it is a fast method to test single genes. However, to test the gene expression of many different genes, it is rather time-consuming. Since it is relatively easy to perform a RT-qPCR, even inexperienced researchers can rapidly obtain results, which appear to be of a high quality but are probably not. To obtain reliable results, numerous important points have to be considered, of which a selection is presented in the following paragraphs [Derveaux et al., 2010, Nestorov et al., 2013].

First, while planning the experimental design, it is possible to choose between the sample maximization method, which means to test one gene with as many samples as possible in one run, and the gene maximization method, which tests different genes in the same run, but spreads different samples across runs. To minimize run-to-run variations between samples, the first method is recommended while performing studies with relative quantification [Derveaux et al., 2010]. Thus, the sample maximization method was chosen in this study and each gene was tested in a separate run. To maximize the sample number, three sample pools, each consisting of ten different maize grains, were used for the analysis. Hence, the DNA of 30 different maize grains of each variety was compared.

Second, as already mentioned in 4.3.2, high RNA quality and low DNA contamination are important [Derveaux et al., 2010], which was considered in this study.

Third, the selection and verification of the used primers are essential [Derveaux et al., 2010]. As recommended, all primers used in the study were controlled by performing a RT-qPCR with a subsequent verification of the amplicon length by gel electrophoresis. Further, the standards of a dilution series were tested before the gene expression was analyzed to verify the qPCR efficiency.

Fourth, the method chosen for the quantification of differences in the gene expression is relevant [Derveaux et al., 2010]. Absolute and relative quantification are possible. For the absolute quantification, an external quantification curve is included. Absolute quantification might be necessary e.g. for the control of the percentage of GMO in food. Relative quantification compares the gene expression of the target gene to a reference gene. This approach was selected in this study because it is adequate for the comparison of physiological changes of the gene expression [Pfaffl, 2001]. Normalization with a reference gene is the most common approach. An advantage is that no differences between the reference gene and the target gene occur due to the preparation of the samples [Nestorov et al., 2013].

The selection of adequate reference genes is essential. Reference genes have to be constitutively expressed and the expression level should be comparable to that of the target genes and must be constant [Nestorov et al., 2013]. In an optimal way, reference genes are selected using software programs like NormFinder [Andersen et al., 2004], which shows,

based on a statistical model, which housekeeping gene is the optimal reference gene in a specific experiment [Nestorov et al., 2013]. In this study, ubiquitin was chosen as a reference gene because it was shown to be suitable in the previous, comparable experiment. Further, the use of the same reference gene improves the comparison of both experiments. It is recommended that at least two different reference genes are included [Chapman and Waldenström, 2015], for example, because smaller differences can be identified and a higher accuracy is obtained, compared to the use of only one reference gene [Derveaux et al., 2010]. However, Chapman et al. concluded after a literature research that only 13% of the investigated studies included more than one reference gene [Chapman and Waldenström, 2015]. In this study, only one reference gene was included, as it is not possible to include more reference genes in one run due to a lack of available tube positions. However, to obtain more accurate results, a repeat of the testing of the 22 genes with a further reference gene is worth considering.

For the comparison of the gene expression, the log2fold change was calculated. A value ≥1 was defined to be a statistically significant difference. This significance level was chosen because it was also used by other authors [e.g. La Paz et al., 2014]. However, it is not based on a mathematical model, which ensures that genes with a log2fold change ≥ 1 are really differently expressed or, rather, that genes with a log2fold change, for example, of 0.8 are not differently expressed. The derivation of a mathematical model of the significance level for such transcriptome analyses would be necessary. Thus, in this study, the expression of genes was tested, which were already investigated in a further maize variety by two different methods (RNA sequencing [Ben Ali et al., unpublished] and qPCR [Draxler, unpublished]).

After the finding of some genes with a difference in the gene expression, the mechanisms still have to be clarified. The functions of the coding proteins have to be investigated better to determine whether the production of allergens or toxins may have increased in the GM grains.

A limitation of the study is that the two compared maize varieties were grown in two different years. The GM variety AG9045 NK603 was grown in 2012/2013 and the conventional comparator AG9045 in 2008/2009. This is a suboptimal basis for the

comparison of the transcriptome between these varieties, since environmental factors might have a strong influence on gene expression, as Benevenuto et al. concluded (described in 2.7) [Benevenuto et al., 2017]. It is supposed that more differently expressed genes would have been identified if the maize varieties had been grown in the same year and, therefore, under comparable climatic influences. A better test material for this comparable approach would have been maize grains of plants grown under the same conditions next to each other in a controlled environment. However, the results show that differences in the gene expression exist between the GM and the conventional maize grains, which are obviously independent of external influences like the environment, the climate, and the year. Thus, the difference of the year in which the grains were grown is not necessarily a limitation. The reason for the test material of different years is the difficulty to obtain GM seeds for investigational purposes because companies producing and selling GM seeds refused to provide test material. We were grateful for obtaining GM maize varieties and their conventional comparators from the Brazilian market.

Before this master's project was performed, the gene expression of endogenous genes of a further GM maize variety containing NK603 (AG8025 NK603) was compared to the non-GM comparator (AG8025) by RNA sequencing. In all, 286 genes were identified by three different programs (DESeq2, EdgeR, and CLC) used for the differential gene expression analysis to have a significant difference in the gene expression [Ben Ali et al., unpublished]. Of these, 22 genes were selected for the verification by RT-qPCR. A statistically significant difference in the gene expression was proven in 18 genes [Draxler, unpublished]. It was hypothesized that the observed differences in the gene expression occurred due to the insertion of the NK603 transgene. However, to control whether the detected differences are specific for a certain NK603 variety or if it is possible to generally conclude that the NK603 transgene unintentionally influences the gene expression of many endogenous genes in different maize varieties, as it was observed for the MON810 transgene [La Paz et al., 2014], it was necessary to control further NK603 varieties, which was started in this master's thesis.

Since the same 22 genes were tested in both variety pairs (AG9045 NK603 vs. AG9045 and AG8025 NK603 vs. AG8025) by RT-qPCR, the results can be compared to each other. Overall,

four of the 22 genes (number 10, 11, 14, and 22) were identified to have a statistically significant difference in both varieties in the same direction of the gene expression. All four genes were up-regulated in the GM maize varieties compared to the conventional varieties. The log2fold changes of these four genes of both varieties are compared in Table 21. The log2fold change of AG8025 presents the mean of two RT-qPCR runs, whereas the log2fold change of AG9045 was determined in only one run.

Gene no.	Zm no. [Cannon et al., 2011]	Log2fold change AG9045	Log2fold change AG8025 [Draxler, unpublished]	Coding protein [The UniProt Consortium, 2017]
11	Zm00001d025081	2,9627	2,318	Rhodanese-like domain-containing protein 19 mitochondrial
10	Zm00001d012221	2,7616	2,588	Acyl-desaturase
22	Zm00001d037941	1,8849	1,438	AP2-EREBP transcript factor
14	Zm00001d035559	1,7219	1,847	Dirigent protein

Table 21: Genes showing comparable differences in the gene expression between GM and conventional maize grains in two maize variety pairs

While investigating the GO classes to which these genes belong, the genes can be collated to the following functions: Gene number 10 is relevant for the fatty acid metabolic process, metal ion binding, and oxidation reduction process, gene number 22 for DNA binding and transcription, and gene number 14 for carbohydrate binding, isomerase activity, and apoplast. For gene number 11, no GO classes could be found [McCarthy et al., 2006].

Further, in two genes (number 7 and 18) that had a statistically significant difference in the gene expression in the variety pair AG9045 vs. AG9045 NK603, no statistically significant difference was measured in the variety pair AG8025 vs. AG8025 NK603 by RT-qPCR. Three genes (numbers 9, 12, and 13) showed a significant difference in the gene expression in both varieties, but in different directions. In the variety pair AG9045, the genes were down-regulated in the GM grains, whereas in the variety pair AG8025, the genes were up-regulated. Table 24 in the appendix shows the log2fold changes of all 22 genes of both

variety pairs measured by RT-qPCR, and additionally, those of the variety pair AG8025 NK603 vs. AG8025 measured by RNA sequencing.

Probably, more genes would have shown a comparable difference in the gene expression in both variety pairs if all four varieties had grown next to each other under controlled conditions. Further, it would have been optimal if all 286 genes, which showed a different gene expression in the RNA sequencing of the variety pair AG8025 NK603 vs. AG8025, had been tested also in the variety pair AG9045 NK603 vs. AG9045. However, this was impossible due to the high costs and expenditure of time.

La Paz et al. performed a comparable study (described in 2.7) with maize containing the MON810 event. Of the 30 genes that were differently expressed in the first investigated variety pair, 22 were also differently expressed in two further variety pairs with the MON810 event. Compared to the four genes out of 22 that were identified to have a comparable difference in the expression in two variety pairs in this study, 22 genes out of 30 is high. An explanation for the different results of these two studies might be that the plants of La Paz et al. were grown under the same conditions, but obviously not next to each other, because they were obtained from different seed companies. On the contrary, the maize grains chosen for this study were not from plants grown under the same conditions. Moreover, the environmental influence on the test material of this study may have been higher because maize grains were chosen as test material, which usually show more environmentally influenced differences as maize embryos, which were analyzed by [La Paz et al., 2014].

It can be concluded from the results of the experiments that the NK603 transgene seems to influence many maize genes unintentionally, as it was observed from the MON810 transgene [La Paz et al., 2014]. The difference in the gene expression of the variety pair AG9045 NK603 vs. AG9045 does not agree with the results of the NK603 variety investigated earlier (AG8025 NK603 vs. AG8025) in all the analyzed genes. This suggests that the genetic background strongly influences which genes are unintentionally changed by a genetic modification. Further influencing factors are obviously environmental effects, such as the year of the harvest of the test material. Moreover, the direction of the effects (up-regulation or down-regulation of the gene expression) seems to depend on the NK603 variety. For the

clarification of these aspects, it would be important to perform a more accurate analysis of the whole transcriptome by NGS (RNA sequencing) followed by RT-qPCR for verifying eventually identified genes. Furthermore, it would be necessary to analyze the individual genes that have a difference in the gene expression and that are important for the plant health more detailed by molecular and biochemical methods with the consideration of genetic effects as well as environmental influences. This might have an impact on parameters that would be important for the authorization of GMO and should be tested routinely.

6 Conclusion

Over the last years, the cultivation of GM plants has substantially increased worldwide. At the same time, the methods for the analysis of GMO advanced and became more precise. Due to the increasing number of GM varieties, especially with stacked events, the performance of research focusing on risk assessment is very important to ensure a safe food market.

In this study, GM maize was investigated by different approaches to determine the genetic stability and to detect unintended effects in the transcriptome. For the approval of a new GM event in the EU, the applicant must prove the genetic stability of the insert according to the European Directive 2001/18/EC. However, few studies have analyzed post-transformational modifications. In the first part of the study, the genetic stability was investigated in maize grains of a variety containing a single MON810 event. The analysis focused on the border region of the transgene because mutations in this location might lead to wrong results in the official GMO control by PCR-based methods with primers binding in this position. A 180bp long sequence, spanning from the 3'end of the transgene to the endogenous maize genome, was amplified and screened for the presence of mutations by performing a real-time PCR coupled with an HRM analysis. Samples with a divergent melting behavior compared to the reference were suspected of harboring mutations and, therefore, also sequenced by NGS. Of the 202 samples, 201 showed genetic stability, whereas in one sample, a SNP with a low frequency (1.46%) was detected. The reason for the occurrence of SNPs with such a low frequency could not be clearly determined. Further, the degree of zygosity of the MON810 transgene was analyzed. In all the controlled samples, the transgene was hemizygous.

The presented combination of HRM analysis and NGS is well suited for the detection of small variations in the DNA. The verification of the genetic stability by low-resolution methods like Southern blot analysis is enough for the approval of a new GM plant in the EU. Southern blot analysis is an appropriate method for the analysis of, for example, the copy number or large variations of the transgene, but not for small modifications. Thus, a combination of

both approaches—the presented one and Southern blot analysis—would be recommendable for an extensive verification of the genetic stability of the insert.

As a further part of this study, the occurrence of unintended effects due to the genetic modification was investigated. Unintended effects should usually be detected in the molecular characterization of the GM plant or in its compositional, agronomic, and phenotypic characterization, which follows a comparative approach by searching for differences between the GM plant and the non-GM comparator. Additionally, unintended effects can be found in non-targeted approaches, for example, by omic-technologies. In this project, the gene expression of 22 selected endogenous genes was compared between a GM maize variety containing a single NK603 transgene and the nearly isogenic, non-GM maize variety by RT-qPCR. Altogether, nine genes had a statistically significant difference in the gene expression. Of these, four genes were up-regulated and five were down-regulated in the GM grains.

Obviously, transgenes cannot be inserted into the plant genome without leading to unintended changes, for example, in the gene expression. It is necessary to clarify whether these changes can be detrimental to humans, animals, or the environment. A detrimental consequence for humans and animals could, for instance, be an up-regulated expression of genes coding toxins or allergens and, for the environment, an increase in unexpected resistances.

There are still many open questions about the safety of GM plants on which research could focus. The transcriptome of further GM plants, especially with stacked events, should be analyzed by NGS. This could probably enable the characterization of phenotypic and polygenic traits of GM plants, which might influence the plant health. Further, the identification of similarities in the different GM plants would probably be possible. Interesting questions for further research would be if GM plants in other stages of vegetation (e.g. leaves) show comparable differences in the gene expression as in the grains. Furthermore, differently produced proteins should be analyzed in more detail. It might be also possible to investigate the consequences of specific modifications by inducing selective variations via CRISPR/Cas. Moreover, the investigation of the genetic stability of GM plants could be expanded to analyze mutations in the whole genome.

Useful data was generated to help appropriate authorities to make responsible decisions concerning the risk assessment of GM plants. In the constant enhancement of the use of biotechnology to produce food and feed, the guarantee of the health of humans, animals, and the environment should have the highest priority.

7 Abstract

7.1 English version

Since the beginning of the commercialization of genetically modified (GM) plants more than 20 years ago, the worldwide cultivation of GM plants as well as the number of events that are authorized to be placed on the European market as food or feed or as food or feed ingredients has substantially increased. This requires extensive controls and a precautionary safety assessment to ensure a safe food market.

In this study, GM maize was analyzed at different levels, focusing on the genome and the transcriptome. The first part of the study deals with genetic stability because the European Directive 2001/18/EC requires the genetic stability of the insert for the approval of a GM plant. The 3'border region of the transgene to the endogenous genome of maize containing the event MON810 was analyzed. High-Resolution Melting analysis was used to screen mutations in 202 single maize grains. Samples showing a divergent melting behavior compared to the reference were subsequently verified by Next Generation Sequencing. In one sample, a single nucleotide polymorphism (SNP) with a low frequency (1.46%) was detected, whereas the remaining samples showed genetic stability. The combination of HRM analysis and NGS is well suited for the identification of small variations of the DNA, such as SNPs, deletion, and rearrangements.

Since the insertion of a transgene might lead to intended effects (e.g. herbicide tolerance) as well as to unintended effects, the second part of the study focuses on the detection of unintended effects by performing a transcriptome analysis. Unintended effects are usually analyzed following the concept of substantial equivalence by comparing GM plants to their non-GM comparators. The gene expression of 22 endogenous maize genes was compared between a maize variety containing the event NK603 and the nearly isogenic, non-GM maize variety by quantitative Reverse Transcription-Polymerase Chain Reaction. Nine differently expressed genes were identified. Of these, four were up-regulated and five down-regulated in the GM maize grains compared to the non-GM maize grains. Since this GM maize variety

is already on the market, a further analysis of the differently produced proteins would be necessary.

7.2 German version

Seit dem Beginn der Kommerzialisierung von gentechnisch veränderten (GV) Pflanzen vor mehr als 20 Jahren hat sich sowohl der weltweite Anbau von GV Pflanzen als auch die Anzahl an Events, welche in Europa für das Inverkehrbringen als Lebensmittel oder Futtermittel oder als Lebensmittel- bzw. Futtermittelinhaltsstoff zugelassen sind, erheblich gesteigert. Das erfordert umfangreiche Kontrollen und eine vorsorgliche Sicherheitsbeurteilung, um einen sicheren Lebensmittelmarkt zu gewährleisten.

In dieser Studie wurde GV Mais auf unterschiedlichen Ebenen untersucht, wobei der Fokus auf dem Genom und dem Transkriptom lag. Der erste Teil der Studie beschäftigt sich mit genetischer Stabilität, da die Europäische Richtlinie 2001/18/EG für die Zulassung einer GV Pflanze die genetische Stabilität des eingeführten Genabschnitts voraussetzt. Die 3‘ Grenzregion des Transgens zum endogenen Genom von Mais mit dem Event MON810 wurde analysiert. Mit Hilfe einer hochauflösenden Schmelzkurvenanalyse wurden 202 einzelne Maiskörner auf Mutationen gescreent. Proben, die ein abweichendes Schmelzverhalten im Vergleich zu einer Referenz zeigten, wurden anschließend mit einer Next Generation Sequenzierung überprüft. Bei einer Probe wurde ein Einzelnukleotid-Polymorphismus (SNP) mit einer sehr niedrigen Häufigkeit (1.46%) gefunden, während die restlichen Proben genetische Stabilität zeigten. Die Kombination aus hochauflösender Schmelzkurvenanalyse und NGS ist gut geeignet, um kleine Veränderungen an der DNA, wie SNPs, Deletionen und Umordnungen, zu identifizieren.

Da das Einführen des Transgens sowohl zu erwünschten Effekten (z.B. Herbizidtoleranz) als auch zu unerwarteten Effekten führen kann, beschäftigt sich der zweite Teil der Studie mit der Detektion von unerwarteten Effekten anhand einer Transkriptionsanalyse. Unerwartete Effekte werden üblicherweise dem Konzept der Substanziellen Äquivalenz folgend über den Vergleich von GV Pflanzen zu ihren nicht GV Vergleichspflanzen ermittelt. Die Genexpression von 22 endogenen Maisgenen wurde mit Hilfe der quantitativen Reverse

Transkriptase-Polymerase Kettenreaktion zwischen einer Maissorte mit dem Event NK603 und der beinahe isogenen, nicht GV Maissorte verglichen. Neun unterschiedlich exprimierte Gene wurden identifiziert. Von diesen waren in den GV Maiskörnern vier Gene höher und fünf Gene niedriger exprimiert als in den nicht GV Maiskörnern. Da diese GV Maissorte bereits auf dem Markt ist, wäre eine weitere Analyse der unterschiedlich produzierten Proteine notwendig.

8 References

Agapito-Tenfen SZ, Guerra MP, Wikmark O-G, Nodari RO. Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. Proteome Science, 2013; 11: 46.

Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R, Corbin DR, Miles RR, Arnold NL, Strange TL, et al. Trait stacking via targeted genome editing. Plant Biotechnology Journal, 2013; 11 (9): 1126–1134.

Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Research, 2004; 64: 5245–5250.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics, 2000; 25 (1): 25–29.

Barros E, Lezar S, Anttonen MJ, van Dijk JP, Röhlig RM, Kok EJ, Engel K. Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnology Journal, 2010; 8 (4): 436–451.

Beck TF, Mullikin JC, Biesecker LG. Systematic evaluation of Sanger validation of next-generation sequencing variants. Clinical Chemistry, 2016; 62 (4): 647–654.

Ben Ali S-E, Madi ZE, Hochegger R, Quist D, Prewein B, Haslberger AG, Brandes C. Mutation scanning in a single and a stacked genetically modified (GM) event by real-time PCR and high resolution melting (HRM) analysis. International Journal of Molecular Sciences, 2014; 15 (11): 19898–19923.

Benevenuto RF, Agapito-Tenfen SZ, Vilperte V, Wikmark O-G, van Rensburg PJ, Nodari RO. Molecular responses of genetically modified maize to abiotic stresses as determined through proteomic and metabolomic analyses. PloS one, 2017; 12 (2): e0173069.

Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, Crandall KA. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biology and Evolution, 2011; 3: 1312–1323.

Cannon EKS, Birkett SM, Braun BL, Kodavali S, Jennewein DM, Yilmaz A, Antonescu V, Antonescu C, Harper LC, Gardiner JM, et al. POPcorn: An online resource providing access to distributed and diverse maize project data. International Journal of Plant Genomics, 2011; 2011, article ID 923035: 10 pages.

Cardi T. Cisgenesis and genome editing: combining concepts and efforts for a smarter use of genetic resources in crop breeding. Plant Breeding, 2016; 135 (2): 139–147.

Castan M, Ben Ali S-E, Hochegger R, Ruppitsch W, Haslberger AG, Brandes C. Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing. European Food Research and Technology, 2017; 243 (3): 353–365.

Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Gatehouse AM, Kärenlampi S, Kok EJ, Leguay J-J, et al. Unintended effects and their detection in genetically modified crops. Food and Chemical Toxicology, 2004; 42: 1089–1125.

CERA. GM Crop Database. NK603 (MON-ØØ6Ø3-6), 2017. Internet: http://cera-gmc.org/GmCropDatabaseEvent/NK603 (accessed 2 November 2017).

Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter, 1993; 11 (2): 113–116.

Chapman JR, Waldenström J. With reference to reference genes: A systematic review of endogenous controls in gene expression studies. PloS one, 2015; 10 (11): e0141853.

Chassy BM. Can – omics inform a food safety assessment? Regulatory Toxicology and Pharmacology, 2010; 58: S62-S70.

Coll A, Nadal A, Collado R, Capellades G, Messeguer J, Melé E, Palaudelmàs M, Pla M. Gene expression profiles of MON810 and comparable non-GM maize varieties cultured in the

field are more similar than are those of conventional lines. Transgenic Research, 2009; 18: 801–808.

Coll A, Nadal A, Palaudelmàs M, Messeguer J, Melé E, Puigdomènech P, Pla M. Lack of repeatable differential expression patterns between MON810 and comparable commercial varieties of maize. Plant Molecular Biology, 2008; 68: 105–117.

Derveaux S, Vandesompele J, Hellemans J. How to do successful gene expression analysis using real-time PCR. Methods, 2010; 50 (4): 227–230.

Devos Y, Álvarez-Alfageme F, Gennaro A, Mestdagh S. Assessment of unanticipated unintended effects of genetically modified plants on non-target organisms: a controversy worthy of pursuit? Journal of Applied Entomology, 2016; 140 (1-2): 1–10.

Dill GM. Glyphosate-resistant crops: history, status and future. Pest Management Science, 2005; 61 (3): 219–224.

Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science, 2014; 346 (6213): 1258096-1-9.

Druml B, Cichna-Markl M. High resolution melting (HRM) analysis of DNA – Its role and potential in food analysis. Food Chemistry, 2014; 158: 245–254.

Duke SO, Powles SB. Glyphosate: a once-in-a-century herbicide. Pest Management Science, 2008; 64 (4): 319–325.

EFSA. Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal, 2011; 9 (5): 2150.

Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends in Genetics, 2013; 29: 569–574.

EU. 98/294/EC: Commission Decision of 22 April 1998 concerning the placing on the market of genetically modified maize (Zea mays L. line MON 810), pursuant to Council Directive 90/220/EEC. Official Journal of the European Union, 1998; L131: 32–33.

EU. Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and

repealing Council Directive 90/220/EEC. Official Journal of the European Union, 2001; L106: 1–39.

EU. Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. Official Journal of the European Union, 2003; L268: 1–23.

EU. 2004/643/EC: Commission Decision of 19 July 2004 concerning the placing on the market, in accordance with Directive 2001/18/EC of the European Parliament and of the Council, of a maize product (Zea mays L. line NK603) genetically modified for glyphosate tolerance. Official Journal of the European Union, 2004; L295: 35–37.

EU. 2005/448/EC: Commission Decision of 3 March 2005 authorising the placing on the market of foods and food ingredients derived from genetically modified maize line NK 603 as novel foods or novel food ingredients under Regulation (EC) No 258/97 of the European Parliament and of the Council. Official Journal of the European Union, 2005; L158: 20–22.

EU. Commission Implementing Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006. Official Journal of the European Union, 2013; L157: 1–48.

EU. Commission Implementing Decision (EU) 2015/684 of 24 April 2015 authorising the placing on the market of genetically modified maize NK603 (MON-ØØ6Ø3-6) and renewing the existing maize NK603 (MON-ØØ6Ø3-6) products, pursuant to Regulation (EC) No 1829/2003 of the European Parliament and of the Council. Official Journal of the European Union, 2015a; L112: 6–10.

EU. Directive (EU) 2015/412 of the European Parliament and of the Council of 11 March 2015 amending Directive 2001/18/EC as regards the possibility for the Member States to restrict or prohibit the cultivation of genetically modified organisms (GMOs) in their territory. Official Journal of the European Union, 2015b; L68: 1–8.

EU. Commission Implementing Decision (EU) 2016/321 of 3 March 2016 adjusting the geographical scope of the authorisation for cultivation of genetically modified maize (Zea mays L.) MON 810 (MON-ØØ81Ø-6). Official Journal of the European Union, 2016; L60: 90–92.

EU. Commission Implementing Decision (EU) 2017/1207 of 4 July 2017 renewing the authorisation for the placing on the market of genetically modified maize MON 810 (MON-ØØ81Ø-6) products pursuant to Regulation (EC) No 1829/2003 of the European Parliament and of the Council. Official Journal of the European Union, 2017; L173: 18–22.

European Commission. EU Register of authorised GMOs. Internet: http://ec.europa.eu/food/dyna/gm_register/index_en.cfm (accessed 2 November 2017).

FAO. FAOSTAT. Crops: Maize, 2018. Internet: http://www.fao.org/faostat/en/#data/QC (accessed 2 February 2018).

Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: pitfalls and potential. Biotechniques, 1999; 26 (1): 112–125.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, et al. Patterns of somatic mutation in human cancer genomes. Nature, 2007; 446 (7132): 153–158.

Guttikonda SK, Marri P, Mammadov J, Ye L, Soe K, Richey K, Cruse J, Zhuang M, Gao Z, Evans C, et al. Molecular characterization of transgenic events using next generation sequencing approach. PloS one, 2016; 11 (2): e0149515.

Hartung F, Schiemann J. Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. The Plant Journal, 2014; 78 (5): 742–752.

Heck GR, Armstrong CL, Astwood JD, Behr CF, Bookout JT, Brown SM, Cavato TA, DeBoer DL, Deng MY, George C, et al. Development and characterization of a CP4 EPSPS-based, glyphosate-tolerant corn event. Crop Science, 2005; 44: 329–339.

Heinemann JA, Kurenbach B, Quist D. Molecular profiling - a tool for addressing emerging gaps in the comparative risk assessment of GMOs. Environment International, 2011; 37 (7): 1285–1293.

Hernández M, Pla M, Esteve T, Prat S, Puigdomènech P, Ferrando A. A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard® based on the 3'-transgene integration sequence. Transgenic Research, 2003; 12 (2): 179–189.

Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science, 2010; 327 (5962): 167–170.

Illumina. 16S metagenomic sequencing library preparation. Preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system, 2013. Internet: https://support. illumina.com/content/dam/illumina-support/documents/documentation/chemistry_ documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf (accessed 2 November 2017).

ISAAA. Global status of commercialized biotech/GM crops: 2016. Ithaca (NY): ISAAA Brief No. 52, 2016. Internet: http://africenter.isaaa.org/wp-content/uploads/2017/06/ ISAAA-Briefs-No-52.pdf (accessed 2 November 2017).

ISO. ISO 21569:2005: Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — Qualitative nucleic acid based methods. Switzerland, 2005.

Kohli A, Miro B, Twyman RM. Transgene integration, expression and stability in plants: strategies for improvements. In: Kole C, Michler CH, Abbott AG, Hall TC, ed. Biotech Plants. Berlin, Heidelberg: Springer, 2010: 201–238.

Kok EJ, van Dijk JP, Prins TW, Kleter GA. Safety assessment of GMO-derived foods. Collection of Biosafety Reviews, 2010; 5: 82–111.

La Paz JL, Pla M, Centeno E, Vicient CM, Puigdomènech P. The use of massive sequencing to detect differences between immature embryos of MON810 and a comparable non-GM maize variety. PloS one, 2014; 9 (6): e100895.

La Paz JL, Pla M, Papazova N, Puigdomènech P, Vicient CM. Stability of the MON 810 transgene in maize. Plant Molecular Biology, 2010; 74 (6): 563–571.

Ladics GS, Bartholomaeus A, Bregitzer P, Doerrer NG, Gray A, Holzhauser T, Jordan M, Keese P, Kok E, Macdonald P, et al. Genetic basis and detection of unintended effects in genetically modified crop plants. Transgenic Research, 2015; 24 (4): 587–603.

Land Salzburg. Gesetz über Maßnahmen der Gentechnik-Vorsorge (Gentechnik-Vorsorgegesetz). Landesgesetzblatt, 2004; 18. Stück: 151–154.

Law M, Childs KL, Campbell MS, Stein JC, Olson AJ, Holt C, Panchy N, Lei J, Jiao D, Andorf CM, et al. Automated update, revision, and quality control of the maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. Plant Physiology, 2015; 167 (1): 25–39.

Liu N, Chen H. An accurate and rapid PCR-based zygosity testing method for genetically modified maize. Molecular Plant Breeding, 2009; 7 (3): 619–623.

Madi ZE, Brandes C, Neumann G, Quist D, Ruppitsch W, Hochegger R. Evaluation of Adh1 alleles and transgenic soybean seeds using Scorpion PCR and HRM analysis. European Food Research and Technology, 2013; 237 (2): 125–135.

Margarit E, Reggiardo MI, Vallejos RH, Permingeat HR. Detection of BT transgenic maize in foodstuffs. Food Research International, 2006; 39 (2): 250–255.

Matlock B. Assessment of nucleic acid purity - Technical note 52646 of Thermo Fisher Scientific. Internet: https://tools.thermofisher.com/content/sfs/brochures/TN52646-E-0215M-NucleicAcid.pdf (accessed 2 November 2017).

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, et al. AgBase: a functional genomics resource for agriculture. BMC Genomics, 2006; 7: 229.

McClintock B. Mutable loci in maize. In: Carnegie Institution of Washington, ed. Annual Report of the Director of the Department of Genetics, Carnegie Institution of Washington Year Book No. 47. New York: Cold Spring Harbor, 1948: 155–169.

Mesnage R, Agapito-Tenfen SZ, Vilperte V, Renney G, Ward M, Séralini G-E, Nodari RO, Antoniou MN. An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process. Scientific Reports, 2016; 6: 37855.

Metzker ML. Sequencing technologies - the next generation. Nature Reviews Genetics, 2010; 11 (1): 31–46.

Murray DC, Coghlan ML, Bunce M. From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. PloS one, 2015; 10 (4): e0124671.

Nannas NJ, Dawe RK. Genetic and genomic toolbox of Zea mays. Genetics, 2015; 199 (3): 655–669.

NCBI. Gene. Internet: https://www.ncbi.nlm.nih.gov/gene (accessed 2 November 2017).

Nestorov J, Matić G, Elaković I, Tanić N. Gene expression studies: How to obtain accurate and reliable data by quantitative real-time RT PCR. Journal of Medical Biochemistry, 2013; 32 (4): 325–338.

Neumann G, Brandes C, Joachimsthaler A, Hochegger R. Assessment of the genetic stability of GMOs with a detailed examination of MON810 using Scorpion probes. European Food Research and Technology, 2011; 233 (1): 19–30.

OECD. Safety evaluation of foods derived by modern biotechnology: Concepts and principles, 1993. Internet: http://dbtbiosafety.nic.in/guideline/OACD/Concepts_and_Principles_1993.pdf (accessed 2 November 2017).

OECD. Molecular characterisation of plants derived from modern biotechnology. In: OECD, ed. Safety Assessment of Foods and Feeds Derived from Transgenic Crops, Volume 2. Paris: OECD Publishing, 2015: 23–36.

Ogasawara T, Chikagawa Y, Arakawa F, Nozaki A, Itoh Y, Sasaki K, Umetsu H, Watanabe T, Akiyama H, Maitani T, et al. Frequency of mutations of the transgene, which might result in the loss of the glyphosate-tolerant phenotype, was lowered in Roundup Ready® soybeans. Journal of Health Science, 2005; 51 (2): 197–201.

Pauwels K, de Keersmaecker S, de Schrijver A, Du Jardin P, Roosens NH, Herman P. Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: Added value or not? Trends in Food Science & Technology, 2015; 45 (2): 319–326.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Research, 2001; 29 (9): 2002-2007.

Pfaffl MW. Quantification strategies in real-time PCR. In: Bustin SA, ed. AZ of quantitative PCR. La Jolla, USA: International University Line, 2004: 87–112.

Pla M, La Paz JL, Vicient C. Insert stability and transgenic plant risk. In: Heldman DR, Hoover DG, Wheeler MB, ed. Encyclopedia of Biotechnology in Agriculture and Food: CRC Press, 2012: 1–4.

Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. Nature Biotechnology, 2005; 23 (8): 967–973.

Puchta H, Fauser F. Synthetic nucleases for genome engineering in plants: prospects for a bright future. The Plant Journal, 2014; 78 (5): 727–741.

Qiagen. RNeasy® Mini Handbook, 2012. Internet: http://www.bea.ki.se/documents/EN-RNeasy%20handbook.pdf (accessed 2 November 2017).

Rasmussen R. Quantification on the LightCycler. In: Meurer S, Wittwer C, Nakagawara K, ed. Rapid cycle real-time PCR, methods and applications. Berlin, Heidelberg: Springer, 2001: 21–34.

Ravindran S. Barbara McClintock and the discovery of jumping genes. Proceedings of the National Academy of Sciences, 2012; 109 (50): 20198–20199.

Reed GH, Kent JO, Wittwer CT. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics, 2007; 8 (6): 597–608.

Republic Austria. Bundesgesetz, mit dem Arbeiten mit gentechnisch veränderten Organismen, das Freisetzen und Inverkehrbringen von gentechnisch veränderten Organismen und die Anwendung von Genanalyse und Gentherapie am Menschen

geregelt werden (Gentechnikgesetz). Bundesgesetzblatt für die Republik Österreich, 1994; Nr. 510/1994: 4111–4149.

Rinaldo AR, Ayliffe M. Gene targeting and editing in crop plants: a new era of precision opportunities. Molecular Breeding, 2015; 35: 40.

Rosati A, Bogani P, Santarlasci A, Buiatti M. Characterisation of 3' transgene insertion site and derived mRNAs in MON810 YieldGard maize. Plant Molecular Biology, 2008; 67 (3): 271–281.

Sambrook J, Russell D. Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor (NY): Cold spring harbor laboratory press, 2001.

Sanchis V, Bourguet D. Bacillus thuringiensis: applications in agriculture and insect resistance management. A review. Agronomy for Sustainable Development, 2008; 28 (1): 11–20.

Sauermost R. Spektrum der Biologie. Heidelberg: Spektrum Akademischer Verlag, 2001.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. The B73 maize genome: complexity, diversity, and dynamics. Science, 2009; 326 (5956): 1112–1115.

Schnell J, Steele M, Bean J, Neuspiel M, Girard C, Dormann N, Pearson C, Savoie A, Bourbonnière L, Macdonald P. A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. Transgenic Research, 2015; 24 (1): 1–17.

Secretariat of the Convention on Biological Diversity. Cartagena Protocol on Biosafety to the Convention on Biological Diversity. Montreal: Secretariat of the Convention on Biological Diversity, 2000. Internet: https://www.cbd.int/doc/legal/cartagena-protocol -en.pdf (accessed 2 November 2017).

Shiferaw B, Prasanna BM, Hellin J, Bänziger M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Security, 2011; 3: 307–327.

Simko I. High-resolution DNA melting analysis in plant research. Trends in Plant Science, 2016; 21 (6): 528–537.

Sorochinskii BV, Burlaka OM, Naumenko VD, Sekan AS. Unintended effects of genetic modifications and methods of their analysis in plants. Cytology and Genetics, 2011; 45 (5): 324–332.

Tan Y, Tong Z, Yang Q, Sun Y, Jin X, Peng C, Guo A, Wang X. Proteomic analysis of phytase transgenic and non-transgenic maize seeds. Scientific Reports, 2017; 7: 9246.

Tan Y, Yi X, Wang L, Peng C, Sun Y, Wang D, Zhang J, Guo A, Wang X. Comparative proteomics of leaves from phytase-transgenic maize and its non-transgenic isogenic variety. Frontiers in Plant Science, 2016; 7: 1211.

Tenaillon MI, Charcosset A. A European perspective on maize history. Comptes Rendus Biologies, 2011; 334 (3): 221–228.

The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Research, 2017; 45: D158-D169.

Thermo Fisher Scientific. Qubit® dsDNA BR Assay Kits, 2015a. Internet: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/Qubit_dsDNA_BR_Assay_UG.pdf (accessed 2 November 2017).

Thermo Fisher Scientific. Qubit® RNA BR Assay Kits, 2015b. Internet: https://tools.thermofisher.com/content/sfs/manuals/Qubit_RNA_BR_Assay_UG.pdf (accessed 2 November 2017).

Thiebaut F, Rojas CA, Grativol C, Motta MR, Vieira T, Regulski M, Martienssen RA, Farinelli L, Hemerly AS, Ferreira PCG. Genome-wide identification of microRNA and siRNA responsive to endophytic beneficial diazotrophic bacteria in maize. BMC Genomics, 2014; 15: 766.

Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews. Genetics, 2012; 13 (2): 36–45.

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Gonzalez JdJS, Ross-Ibarra J. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proceedings of the National Academy of Sciences, 2011; 108 (3): 1088–1092.

Varshney RK, Nayak SN, May GD, Jackson SA. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology, 2009; 27 (9): 522–530.

Vidal N, Barbosa H, Jacob S, Arruda M. Comparative study of transgenic and non-transgenic maize (Zea mays) flours commercialized in Brazil, focussing on proteomic analyses. Food Chemistry, 2015; 180: 288–294.

Vossen RH, Aten E, Roos A, den Dunnen JT. High-resolution melting analysis (HRMA): More than just sequence variant screening. Human Mutation, 2009; 30 (6): 860–866.

Wilson AK, Latham JR, Steinbrecher RA. Transformation-induced mutations in transgenic plants: analysis and biosafety implications. Biotechnology and Genetic Engineering Reviews, 2006; 23: 209–234.

Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. High-resolution genotyping by amplicon melting analysis using LCGreen. Clinical Chemistry, 2003; 49 (6): 853–860.

Xu C, Ranjbar MRN, Wu Z, DiCarlo J, Wang Y. Detecting very low allele fraction variants using targeted DNA sequencing and a novel molecular barcode-aware variant caller. BMC Genomics, 2017; 18 (1): 5.

Zhang D, Corlet A, Fouilloux S. Impact of genetic structures on haploid genome-based quantification of genetically modified DNA: theoretical considerations, experimental data in MON 810 maize kernels (Zea mays L.) and some practical applications. Transgenic Research, 2008; 17 (3): 393–402.

Zolla L, Rinalducci S, Antonioli P, Righetti PG. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. Journal of Proteome Research, 2008; 7 (5): 1850–1861.

9 Appendix

9.1 GO termini of the nine differently expressed genes

Presentation of the coding proteins and the GO termini (including their classification of the three domains cellular component [C], molecular function [F], and biological process [P]) of the nine genes showing a statistically significant difference in the gene expression between the GM maize grains of the variety AG9045 NK603 and the conventional maize grains of the variety AG9045 [McCarthy et al., 2006].

Gene no.	Zm no.	GO:ID	GO Term Name	Domain	DB object name
7	Zm00001d036222	GO:0016020	Membrane	C	Putative sugar phosphate/phosphate translocator
		GO:0016021	Integral component of membrane	C	
9	Zm00001d022084	GO:0016020	Membrane	C	B12D protein
		GO:0016021	Integral component of membrane	C	
10	Zm00001d01222	GO:0006631	Fatty acid metabolic process	P	Acyl-desaturase
		GO:0045300	Acyl-[acyl-carrier-protein] desaturase activity	F	
		GO:0046872	Metal ion binding	F	
		GO:0055114	Oxidation-reduction process	P	
11	Zm00001d025081	No GO:ID found			
12	Zm00001d048947	GO:0042742	Defense response to bacterium	P	Pathogenesis-related protein3
		GO:0050832	Defense response to fungus	P	

Table 22: GO termini of the nine differently expressed genes

Gene no.	Zm no.	GO:ID	GO Term Name	Domain	DB object name
13	Zm00001d048949	GO:0042742	Defense response to bacterium	P	Hevein-like preproprotein
		GO:0050832	Defense response to fungus	P	
14	Zm00001d035559	GO:0005576	Extracellular region	C	Dirigent protein
		GO:0016853	Isomerase activity	F	
		GO:0030246	Carbohydrate binding	F	
		GO:0048046	Apoplast	C	
18	Zm00001d036293	GO:0008168	Methyltransferase activity	F	Caffeoyl-CoA 3-O-methyltransferase
		GO:0008171	O-methyltransferase activity	F	
		GO:0016740	Transferase activity	F	
		GO:0032259	Methylation	P	
22	Zm00001d037941	GO:0003677	DNA binding	F	AP2-EREBP transcription factor
		GO:0003700	Transcription factor activity, sequence-specific DNA binding	F	
		GO:0005634	Nucleus	C	
		GO:0006351	Transcription, DNA-templated	P	
		GO:0006355	Regulation of transcription, DNA-templated	P	

Table 22: GO termini of the nine differently expressed genes (continuing)

9.2 Slim view of the genes having a different gene expression in two NK603 maize varieties

Presentation of the Slim view, including the GO categories and their annotation of the four genes that have a statistically significant difference in the gene expression between GM and conventional maize grains in two variety pairs (AG9045 NK603 vs. AG9045 and AG8025 NK603 vs. AG8025).

GO categories	Annotation
Cellular component	
GO:0005576	Extracellular region
GO:0005634	Nucleus
Molecular function	
GO:0003677	DNA binding
GO:0003700	Transcription factor activity, sequence-specific DNA binding
GO:0003824	Catalytic activity
GO:0005488	Binding
GO:0030246	Carbohydrate binding
Biological process	
GO:0006139	Nucleobase-containing compound metabolic process
GO:0006629	Lipid metabolic process
GO:0008150	Biological process
GO:0008152	Metabolic process
GO:0009058	Biosynthetic process
GO:0009987	Cellular process

Table 23: Slim view of genes having a different gene expression in two maize varieties

9.3 Comparison of the log2fold changes of two maize variety pairs

Presentation of the difference of the gene expression as log2fold changes between GM maize grains and conventional maize grains in the two maize variety pairs AG9045 NK603 vs. AG9045 (measured by qPCR) and AG8025 NK603 vs. AG8025 (measured by qPCR and RNA sequencing), starting with the four statistically significant expressed genes.

Gene no.	Zm no. [Cannon et al., 2011]	Log2fold change (qPCR) AG9045	Log2fold change (qPCR) AG8025 [Draxler, unpublished]	Log2fold change (RNA seq.) AG8025 [Ben Ali et al., unpublished]	Coding protein [The UniProt Consortium, 2017]
10	Zm00001d012221	2,7616	2,588	4,906	Acyl-desaturase
11	Zm00001d025081	2,9627	2,318	2,202	Rhodanese-like domain-containing protein 19 mitochondrial
14	Zm00001d035559	1,7219	1,847	3,691	Dirigent protein
22	Zm00001d037941	1,8849	1,438	1,765	AP2-EREBP transcript factor
1	Zm00001d033846	0,5463	1,766	3,648	Uncharacterized protein
2	Zm00001d031127	0,4839	3,392	2,731	Cupin, RmlC-type
3	Zm00001d020025	-0,0559	1,917	2,118	Uncharacterized protein
4	Zm00001d028814	-0,1804	2,497	2,017	Pathogenesis-related protein 10
5	100192063	0,8213	-1,619	-2,351	Uncharacterized protein
6	Zm00001d022464	-0,0024	-1,115	-2,479	Ultraviolet-B-repressible protein
7	Zm00001d036222	-1,2786	-0,268	-1,147	Putative sugar phosphate/phosphate translocator
8	Zm00001d028816	-0,2295	2,566	1,545	Pathogenesis-related protein 10
9	Zm00001d022084	-1,2411	2,317	12,343	B12D protein
12	Zm00001d048947	-1,9651	1,804	1,195	Pathogenesis-related protein3

Table 24: Comparison of the log2fold changes of two maize variety pairs

Gene no.	Zm no. [Cannon et al., 2011]	Log2fold change (qPCR) AG9045	Log2fold change (qPCR) AG8025 [Draxler, unpublished]	Log2fold change (RNA seq.) AG8025 [Ben Ali et al., unpublished]	Coding protein [The UniProt Consortium, 2017]
13	Zm00001d048949	-2,6722	2,553	1,484	Hevein-like preproprotein
15	Zm00001d002160	-0,4037	0,914	3,536	Wound induced protein
16	103652895	-0,6751	-1,369	-1,532	Putative ubiquitin family protein
17	Zm00001d038717	-0,2241	2,229	2,686	Seed specific protein Bn15D17A
18	Zm00001d036293	-1,2765	-0,814	-1,806	Caffeoyl-CoA O-methyltransferase 1
19	Zm00001d035737	0,0671	-0,139	-1,374	D-glycerate 3-kinase chloroplastic
20	Zm00001d047260	-0,3411	1,737	2,03	Uncharacterized protein
21	Zm00001d035597	0,2936	-0,138	-1,384	Legumin-like protein

Table 24: Comparison of the log2fold changes of two maize variety pairs (continuing)