universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

"Influence of genetic engineering on genetics, epigenetics, and transcriptomics of *Zea mays*"

verfasst von / submitted by

Sina-Elisabeth Ben Ali, Bakk.rer.nat. MSc

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

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 610 474
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Ernährungswissenschaften
Betreut von / Supervisor:	Doz. a.o. Prof. Dr. Alexander Haslberger

This work was cooperation between the University of Vienna and the Austrian Agency for Health and Food Safety (AGES). It was carried out from October 2013 to April 2018 at the Department of Molecular Biology and Microbiology, Institute for Food Safety, AGES, under the supervision of Doz. a.o. Prof. Dr. Alexander Haslberger, Department of Nutritional Sciences, University of Vienna, and Dr. Christian Brandes, Department of Molecular Biology and Microbiology, Institute for Food Safety, AGES.

Danksagungen

An erster Stelle möchte ich mich ganz herzlich bei
Dr. Christian Brandes für die Ermöglichung und professionelle sowie persönliche Unterstützung und Betreuung meiner Dissertation bedanken.

Außerdem möchte mich recht herzlich bei **Doz. a.o. Prof. Dr. Alexander Haslberger** für die Betreuung und Unterstützung meiner Arbeit bedanken.

Für die Genehmigung und finanzielle Unterstützung meines Vorhabens bedanke ich mich bei der **AGES**, insbesondere bei **Mag. Rupert Hochegger**.

Ich danke außerdem allen
Kolleginnen und Kollegen der Gruppe Molekularbiologie, die mir stets unterstützend zur Seite standen und für eine angenehme Arbeitsatmosphäre gesorgt haben.

Ein spezieller Dank geht außerdem an **Magali Castan, Marko Raskovic, Agnes Draxler und Alexandra Schamann**. In diesem Zusammenhang möchte ich mich auch beim **FFG** für die Förderung mehrerer Studentinnen-Praktika bedanken.

Vielen Dank auch an den **Hochschuljubiläumsfonds der Stadt Wien**, der mein Projekt finanziell unterstützt hat.

Meinen Eltern, Brigitte und Uli, möchte ich für die mentale Stärke danken, die sie mir mit auf den Weg gegeben haben.

Außerdem danke ich meinem Mann **Mohamed** für die Unterstützung meines Vorhabens sowie meinen Kindern **Yassin und Leyla** für die Durchhaltekraft.

“Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less."
Marie Curie

Table of contents

1. INTRODUCTION

1.1. Genetic engineering and commercial use of genetically modified plants

Genetic modification, also described as genetic engineering, is the manipulation of genetic material using recombinant DNA techniques. These techniques raised between 1972 and 1974, when Paul Berg, Herbert Boyer, Annie Chang, and Stanley Cohen of Stanford University and University of California San Francisco created the first recombinant DNA molecule (1973) [1]. In later experiments, Boyer and Cohen introduced recombinant DNA into a bacterial cell (*Escherichia coli*) [2], thereby creating the first genetically modified organism (GMO). These experiments formed the basis of recombinant DNA technology, which introduced a new era in the field of molecular biology. Genetic engineering enabled breeding beyond species boundaries. In 1974, the first transgenic animal (mouse) was created by Rudolf Jaenisch [3]. Michael W. Bevan, Richard B. Flavell, and Mary-Dell Chilton developed the first genetically modified (GM) plant in 1983 [4]. A chimaeric antibiotic resistance gene was inserted into a tobacco plant through infection with *Agrobacterium tumefaciens*. In 1985, Rob Horsch and colleagues of Monsanto reported gene insertion into tobacco, petunia, and tomato using *Agrobacterium*-mediated gene transfer [5].

Meanwhile, gene transfer into eukaryotic cells has been established for the production of GMOs. Since its market introduction in 1996, the use of commercial GM crops has drastically expanded. A high percentage of economically important crops as soybean, maize, cotton, and canola are transgenic [6]. Transgenic organisms contain recombined DNA from unrelated species. Other (rare) GMO categories are intragenic (introduced DNA is from the same genome), famigenic (introduced DNA is from the same family), linegenic (introduced DNA is from a species of the same lineage), and xenogenic organisms (introduced DNA is synthetic or an unnatural combination of protein domains from various species) [7].

The most commonly cultivated GM plants confer herbicide tolerance, insect tolerance, or a combination of both. In the year 2016, 185.1 million hectares of GM crops were grown in 26 countries worldwide [6]. In the European Union (EU), only the event MON810 is authorized for cultivation in eleven countries and regions – Spain, Portugal, Czech Republic, Slovakia, Sweden, Finland, Estonia, Ireland, Romania, Flanders, and England [8]. In 2016, only four of these countries cultivated MON810 maize (Spain, Portugal, Slovakia, and Czech Republic) [6].

More recently, the commercialization of GM crops containing single transgenic traits have given way to the introduction of stacked GM crops containing multiple transgenes [9,10]. In this work, the term single event is used for GM plants containing a single transgene, and the term stacked event is used for plants produced by conventional breeding (cross-breeding) of single events. Next to conventional breeding, re-transformation, co-transformation, or multigene cassette transformation can be used for gene stacking, but these are not addressed in the current work [9]. Approximately 75.4 million hectares of stacked events were grown in 2016, which means +29% compared to 2015 [6]. Moreover, there has been an increase in the number of approved stacked events and the number of transgenes stacked in a single plant (e.g., SmartStax™ maize containing eight transgenic events) [9].

For classical plant transformation (the “genetic alteration of a cell resulting from the direct uptake and incorporation of exogenous genetic material from its surroundings and taken up through the cell membrane(s)” [11]), *Agrobacterium*-mediated and biolistic methods are the two most widely used approaches [12,13]. *Agrobacterium tumefaciens* is a soil bacterium, which causes crown gall disease in infected plants. This disease leads to uncontrolled tumor growth induced through a circular DNA plasmid, which is called tumor-inducing (Ti) plasmid. When a plant cell is infected, a part of the Ti plasmid, which is called T-DNA (single-stranded), is inserted into the plant genome [14,15]. With *Agrobacterium*-mediated transformation, the desired gene is inserted into the Ti plasmid of *Agrobacteria*. Further, by mixing the bacteria with plant cells, the Ti plasmid moves into the plant cell nucleus and inserts the manipulated T-DNA into the plant genome. Of course, the pathogenic property is removed prior to transformation [16]. Biolistic gene transfer, also called particle bombardment, is a method, in which DNA (constituting a gene cassette) is precipitated on microparticles (gold or tungsten) and directly transferred into the nucleus of target cells by a gene gun [12]. This method was originally developed for plant cells by Sanford and Klein in the late 1980s, but has evolved over time, and is now used for a wide range of cells and tissues. Advantages of particle bombardment are that multiple transgenes can be introduced at once and that the size of the transgenic constructs is not a large limiting factor [17]. Transformation rates are relatively similar (5-40%) between *Agrobacterium*-mediated and biolistic methods [18] . After successful transformation using either particle bombardment or *Agrobacterium*-mediated transformation, the resulting cell is genetically modified and can be regenerated into a plant. Figure 1 shows the principle of the two major transformation methods used in plant genetic engineering.

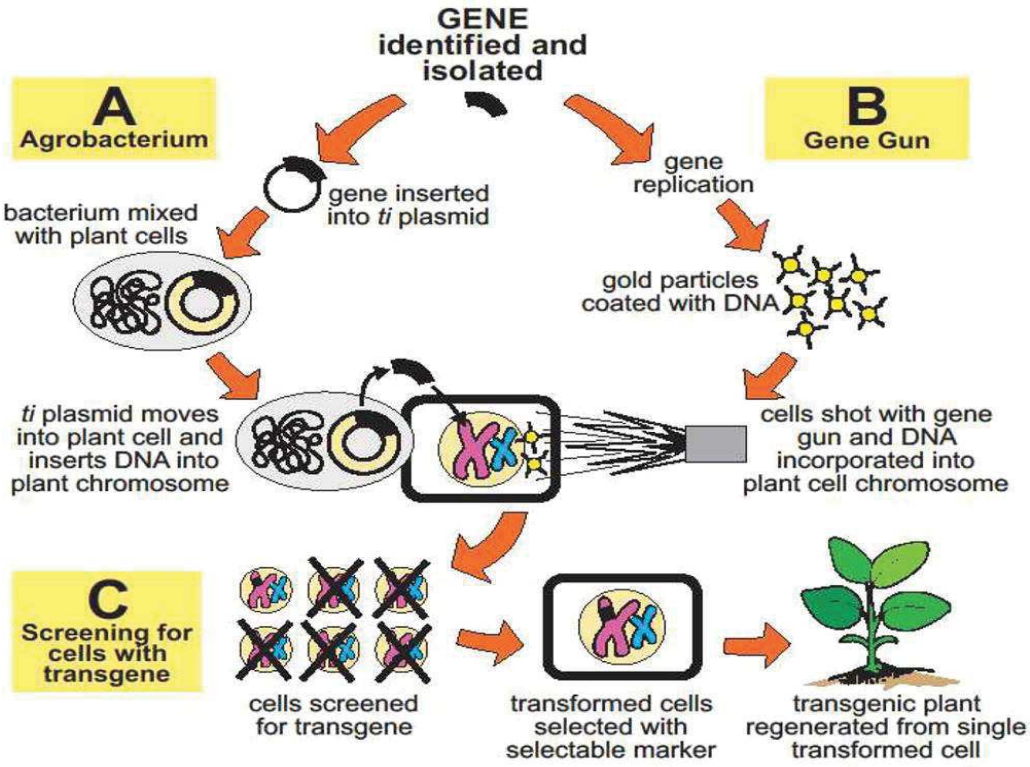


Figure 1. Gene transformation via *Agrobacterium* or gene gun [16].

1.1.1. Transgene integration

Transgene integration in plants transformed by either *Agrobacterium*-mediated transformation or particle bombardment occurs mainly through illegitimate recombination, particularly through non-homologous end-joining (NHEJ). Homologous recombination is very unlikely in transgene integration. However, there may be alternative pathways, since transgene integration in plants is not clearly elucidated until now [19,13,14]. There is much evidence that double-strand break (DSB) repair enzymes are involved in the transgene integration process and that DSBs create the integration target sites [20,19]. Further, Salomon and Puchta [20] demonstrated that T-DNA integration is increased when DSBs are intentionally induced, indicating that DSBs are a limiting factor for transgene integration.

NHEJ is the main pathway for the repair of DSBs in higher eukaryotes [13]. It mediates the (re-)ligation of broken DNA ends and does not require a homologous template [21]. Recombination through NHEJ often leads to mutations at the repair site similar to those observed at DNA integration sites in plants having undergone genetic engineering [13]. Other observed commonalities between NHEJ repair sites and transgene integration sites, which indicate that NHEJ is involved in transgene integration, are the presence of short repeats

(1-8 bp) and plant organelle DNA, for example mitochondrial or chloroplast DNA [13,19]. However, Park et al. [14] reported that T-DNA integration can also occur without the activity of the most important NHEJ proteins, strongly suggesting that an alternative pathway for transgene integration exists. In animal cells, an alternative NHEJ pathway is already explored, and in plants a similar route, differing slightly by some enzymes, may exist [14].

Most studies report that transgenes with stable transgene expression tend to integrate into gene-rich regions [19,22-24], whereas transgene integration in general (without selection for stable transgene expression) is more random. In other words, genetic modification offers no control over the transgene insertion locus [14,25,26]. Characterization of transgene integration sites is of major interest, because DNA insertions are often accompanied with additional molecular changes, such as deletion of endogenous plant DNA, insertion of additional DNA or DNA rearrangements (see chapter 1.1.2.) [26,13]. These changes are also called unintended effects and can have potential adverse effects (see chapter 1.2.).

1.1.2. Insertional effects in genetically modified plants

Genetic changes resulting from gene insertion via genetic modification are designated insertional effects. As already mentioned above, unintended insertion of additional DNA is a common phenomenon of genetic engineering. Additional DNA can originate from the natural plant genome, from plant organelles, from the plasmid used for transformation (e.g. vector backbone sequences), or from the intended DNA insert [13,27]. Deletions near the border regions from the transgene to the natural plant genome are also common insertional effects [28]. Interestingly, frequency and form of insertional effects are associated with the transformation method used for genetic modification [29].

Particle bombardment is well known for multi-copy events in tandem or inverted repeat structure leading to complex integration patterns [30,29]. Large scale rearrangements of the full-length transgene with varying copy numbers are not rare either [13,31]. However, rearrangements may also be caused by regenerating transformed plants through tissue culture [13,32,30]. *Agrobacterium*-mediated plant transformation is often accompanied by insertion-site mutations and incomplete DNA integration [28,33]. However, the *Agrobacterium*-mediated method results more often in transformations with one or a few transgene integrations per genome. This may lower the probability for transgene silencing since an inverse correlation between transgene copy number and expression level is reported. In

addition, rearrangements are less common and transgene expression tends to be more stable when *Agrobacteria* are used for plant transformation [34-36].

One example for insertional effects is the commercialized maize MON810 transgene, which underwent a truncation event that led to the complete loss of the nopaline synthase (NOS) terminator at the 3´end of the *cry1ab* gene [37]. The MON810 transgene also interrupts exon 8 of the putative maize HECT E3 ligase gene 8 and has an additional insertion with 100% homology to the *Zea mays* mitochondrial genome within the first 44 bp of its 3´ flanking region [38]. The maize NK603 transgene shows DNA rearrangements and includes chloroplast DNA [39]. Maize MON863 contains mitochondrial DNA at both flanking regions. In addition, open reading frames were identified at the border regions of the transgene [40]. Interestingly, all three events (MON810, NK603, and MON863) were produced by particle bombardment [41].

Insertional effects can lead to unintended properties in the GM plant, for instance the integration of a transgene into a gene-rich region can affect or disrupt endogenous genes. DNA insertion can lead to altered gene expression, modified protein levels or alternative splice variants of endogenous plant genes [13]. Moreover, read-through transcripts due to deleted stop codons at the flanking regions, which can lead to fusion proteins with unknown properties, are possible [38]. However, not all insertional effects result in unintended properties, but it is generally assumed that transgene integration within or nearby genes or regulatory elements increases the probability of unintended properties [13].

1.2. Unintended effects in genetically modified plants

Modification of plants by genetic engineering can result in various unintended effects (next to the intended effect of the novel trait), such as altered gene expression, differing protein profiles, metabolomics changes, and compositional changes [13]. The European Food Safety Authority (EFSA) defines unintended effects as follows: “Unintended effects of the genetic modification are considered to be consistent (non-transient) differences between the GM plant and its appropriate comparator, which go beyond the primary intended effect(s) of introducing the transgene(s).” Biological relevance (which includes natural variation) and potential hazard have to be assessed in case of significant differences between a GM plant and its isogenic comparator [42].

There are already several studies investigating unintended effects in insect-resistant MON810 maize on different OMIC levels. La Paz et al. [43] performed RNA sequencing and microarray with different MON810 maize varieties including near-isogenic comparators, which revealed 140 differentially expressed genes mainly involved in carbohydrate metabolism, protein metabolism, and chromatin organization. Gullì et al. [44] compared drought stress response and gene expression between a MON810 maize variety and a near-isogenic variety. Despite the provoked drought stress, *cry1ab* expression level was stable. Many stress-responsive genes were up-regulated under drought-stress conditions in both, the GM and the non-GM variety, but more genes were up-regulated in the GM variety, which indicates a more efficient stress response.

A proteome analysis was conducted by Coll et al. [45], which revealed virtually identical proteomic patterns between MON810 and comparators, with a very low number of differing spots. Agapito-Tenfen et al. [46] conducted a comparative analysis on MON810 maize and comparator grown under different agroecosystems conditions in Brazil using two-dimensional gel electrophoresis combined with mass spectrometry. 32 differentially expressed proteins were identified and their molecular functions were mainly assigned to carbohydrate and energy metabolism, genetic information processing, and stress response.

NK603 maize, which is tolerant to the broad spectrum herbicide Roundup, was also studied for unintended effects. Agapito-Tenfen et al. [47] analyzed stacked maize varieties (NK603 x MON810). The research group used proteomics for the identification of unintended effects and detected 22 proteins to be differentially expressed in stacked and single events versus non-GM near-isogenic maize and a landrace variety. Two major pathways were affected, energy/carbohydrate metabolism and detoxification metabolism. Benevenuto et al. [48] performed proteomic and metabolomic analyses on Roundup Ready maize plants under abiotic stress conditions. 20 differentially modulated proteins were identified between GM and non-GM hybrids under different water deficiency conditions and herbicide sprays. Again, especially energetic/carbohydrate metabolic processes were affected.

Barros et al. [49] compared two Roundup Ready maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics, and metabolomics. The results showed that the environment had a major effect in protein and gene expression and metabolite production as well. However, until now there is no transcriptomics study on Roundup Ready maize and comparator which used high-throughput RNA sequencing.

Potential sources of unintended changes are insertional effects (see chapter 1.1.2.), pleiotropic effects of the transgene (which means that the transgene influences multiple phenotypic traits), and random mutations resulting from the transformation or tissue culture process (somaclonal variations). The influence of environmental or stress factors on GM plants might also lead to unintended effects [50,13]. In addition, there has been the concern that transgene products of stacked events may interact with each other (e.g. via protein-protein binding) [10].

However, unintended effects are not limited to genetic engineering. Conventional breeding, as well as the new plant breeding techniques (NPBT) described in chapter 1.6., can lead to unintended effects [30,13]. Any technique, leading to changes of DNA sequences and gene expression in an organism, can induce unintended effects in this organism [51]. Moreover, “unintended” does not necessarily mean harmful [52]. Unintended changes can have potential adverse effects and it is important to identify and evaluate these effects.

In the EU, GM plants need to be assessed for biologically relevant parameters in an Environmental Risk Assessment (ERA) prior to authorization [53]. EU biosafety regulations (EU Directive 2001/18/EC, Regulation (EC) 503/2013, Regulation (EC) 1829/2003) require an ERA of GM plants to prevent potential adverse effects on human health or the environment. EFSA has issued separate guidance documents for the risk assessment of GM plants, in which molecular characterisation is described as starting point for the detection of unintended effects, and the comparative approach as general principle for the identification of intended and unintended differences between GM plants and its conventional counterparts. This comparative approach includes the investigation of targeted compositional, phenotypic, and agronomic as well as environmental characteristics [42,54]. The comparative approach, which is based on the concept of substantial equivalence between GM plants and conventional counterparts, is worldwide used as basis for GM crop safety assessments and has also been described by the Organization for Economic Co-operation and Development [55], and the Food and Agriculture Organization [56].

Currently used targeted approaches have the disadvantage that important, unintended differences can be overlooked, as only selected features or chemical/nutritional compounds are investigated. Especially unexpected effects, such as pleiotropic effects, resulting from the genetic modification, are not taken into account to a sufficient extent by the

current approach [57]. Most applicants apply methods that target specific sequences or expressed products in risk assessment of GM plants. New profiling techniques are rarely used in risk assessment because their application in GMO context is still under development. However, these techniques are usually performed as untargeted approach, which enables the identification of unintended effects, which are hard to identify using targeted approaches [58,59]. Recently, the use of new profiling methods technologies in the comparative risk assessment of GM plants was discussed in an EFSA colloquium and most experts agreed that these new technologies can add value to risk assessment [60].

1.2.1. Detection of unintended effects using new profiling methods

New profiling methods (also known as OMIC technologies) allow the universal detection of genes, mRNA transcripts, proteins, or metabolites in biological samples. Different biological levels can be analyzed (genomics, transcriptomics, proteomics, and metabolomics). These techniques include DNA/RNA microarrays, microRNA fingerprinting, RNA sequencing, proteome/metabolome/toxicological profiling using mass spectrometry, differential image gel electrophoresis for proteome analysis, and nuclear magnetic resonance spectroscopy for the identification of metabolites, etc. [61,62].

The mentioned profiling methods allow a more holistic analysis of differences between GM plants and their conventional counterparts compared to the currently used targeted approaches (agronomic comparisons, compositional analyses, animal nutrition, and toxicology evaluations) [62]. In addition, new profiling techniques can serve as supplementary tools in risk assessments of GM plants and offer valuable information [59]. Further, EFSA started mapping the use of new profiling tools in the risk assessment related to food and feed safety [63]. However, one challenge of new profiling methods is the bioinformatics evaluation of the produced data. These methods can generate an enormous amount of data and appropriate software tools as well as specific knowledge is required for statistical data evaluation [59]. Furthermore, it has to be noted that obtained OMIC profiles vary, if the examined organism is exposed to different environmental conditions (e.g. geography or season) [62].

The dominant technique for transcriptome analysis is RNA sequencing which has overtaken previously used microarrays. RNA sequencing is the direct sequencing of transcripts (cDNA) by high-throughput sequencing technologies. This technique has been widely used for differential gene expression studies and has many advantages compared to conven-

tional microarrays [64]. Transcription is the first step in gene regulation and it is essential for the understanding of gene regulatory networks. Proteomics is the study of all proteins present in a cell or organism at a specific time point [65]. For proteome analysis, two-dimensional gel electrophoresis is commonly used [59,47]. Using this method, proteins are separated in a two-dimensional gel. Speed and direction of the proteins depend on their electric charge and separation is based on the iso-electric point and the molecular size of the proteins. Metabolomics is the study of the totality of the metabolites present in an organism at a defined time point [65]. For metabolome analysis, the ideal method depends on the compounds analyzed. Several methods are described, such as gas chromatography coupled to mass spectrometry, liquid chromatography coupled to ultraviolet diode-array detection, and liquid chromatography coupled to nuclear magnetic resonance [59].

1.3. Genetic stability of transgenic inserts

An essential safety criterion for the admission of GM plants in the EU is the verification of the genetic stability of the transgenic locus for five generations or vegetative cycles [29,54]. However, for stacked events produced by conventional breeding of single events, which were positively assessed, it is sufficient to prove that the integrity of the single events has been conserved [54]. Although Southern blot analysis does not enable the identification of minor DNA changes [66], genetic stability and copy number of transgenes present in stacked events are usually demonstrated by Southern blot analysis. Thus, small DNA changes, such as point mutations, small deletions, and rearrangements that might arise during breeding, might be overlooked. However, even minor changes can have an important impact, since they can lead to unintended effects affecting the properties, ingredients or morphology of the plant [67,33]. In addition, minor DNA changes can lead to incorrect GMO detection and quantification results (mainly based on PCR methods) in the regulatory control of food and feed [68]. Small DNA changes can be precisely identified by DNA sequencing methods [66], for example Sanger sequencing or next generation sequencing.

In recent years, it could be shown that the plant transformation process is often accompanied by unintended DNA modifications [69,70,38,32,71]. Possible consequences of plant transformation are described in chapter 1.1.2. in detail. Moreover, maintenance, breeding, and cultivation of transgenic seeds may also lead to unintended DNA modifications [72,29]. Investigations of the post-transformational stability of commercial GM plants show different results: Morisset et al. [73], Zaulet et al. [74], and Ogasawara et al. [75] reported

genetic instabilities of commercial GM plants, whereas la Paz et al. [76], Madi et al. [77], and Neumann et al. [29] detected no instabilities in their investigations of transgenic loci. Interestingly, until now no study investigated the genetic stability of herbicide-resistant NK603 maize. Further, all mentioned investigations were performed with single events (and none was performed using stacked events). Furthermore, it should be noted in this context that the identification of low-frequency mutations is highly dependent on the employed method. If large numbers of biological replicates are pooled for the experiments, mutated samples may be masked, due to their low frequency. Thus, Neumann et al. [29] recommends analyzing a large number of individual samples separately, which requires much more time, but on the other hand enables the identification of even low frequency variants (provided an appropriate method is employed).

Most applicants claim that transgenic inserts and its flanking regions are inherited intactly during breeding, but they do not provide appropriate evidence [66]. In contrast, Syngenta demonstrated that nucleotide changes can occur during breeding: Inbred Bt11 field maize was crossed with elite sweet maize. Subsequently, the Bt11 insert (obtained by DNA sequencing) of the inbred field maize was compared to the Bt11 sequence of the progenies (from the above mentioned cross). The flanking regions of the Bt11 insert were also compared. A total of eight nucleotide changes were identified. Two nucleotide changes were located in the flanking regions and the other six were detected in the Bt11 insert (four of them in the intervening sequences and two in the NOS terminator). Because there was no mutation in the promoter or coding region, the Biosafety Advisory Council concluded that the expressed protein remains unchanged [78].

These findings clearly demonstrate that transgenic inserts as well as its flanking regions can mutate during the breeding process. However, spontaneous mutations are natural and endogenous plant genes are also subject of this natural process [66]. Hence, it cannot be excluded that post-transformational mutations occur in the promoter or coding region, which might have a significant influence on the expressed protein, and thus on its safety. De Schrijver et al. [66] argued that DNA sequencing data are irrelevant in case of stacked events, since major protein changes, potentially affecting the product safety, would be detected in later stages of the risk assessment. However, toxicological tests are usually performed with *Cry* proteins purified from laboratory strains of bacteria engineered to express *Cry* protein instead of using *Cry* proteins extracted from the assessed GM plant [79]. Therefore, DNA sequencing data of stacked events are very important.

The increasing number of stacked events has raised the question whether the stacking of transgenes may increase genetic instability [10]. Many transgenes share similar promoter sequences: More than 80% of transgenic plants contain the Cauliflower Mosaic Virus 35S promoter [80]. Sequence homologies can lead to instabilities via homologous recombination [10]. However, repetitive DNA sequences occur also naturally in all plants and are one of the driving factors of genome diversity (next to mobile elements) [81]. In addition, plant genomes have generally the ability to change their genetic structure (at the DNA and chromatin level). If transgenes are integrated into plant genomes, the general genetic reorganization does not stop. Thus, integrated transgenes are involved into the reorganization of plant genomes [10]. The question is, if transgenes, particularly stacked transgenes, increase genetic reorganization, which would mean an increased mutation rate compared to the natural mutation rate.

Spontaneous mutations are a natural phenomenon in all living organisms [13]. The average single-base substitution rate for maize is estimated to be 5.39×10^{-8} per site per year and for genic maize regions, 4.79×10^{-8} per site per year [82]. Spontaneous mutations can occur during DNA replication, homologous or intrachromosomal recombination, mitotic and meiotic cell division, but also as a consequence of genome duplication [29,13]. Further, it has to be considered that transposable elements ("jumping genes") have the ability to alter genomes. Position changes of transposable elements can create or reverse mutations. Transposable elements can also increase gene copy number or even lead to the creation of new genes through exon shuffling [83,84]. *Zea mays* has a relatively high mutation rate due to its big genome size and its high proportion of mobile elements [13]. Thus, it is important to consider the natural mutation rate when analyzing the genetic stability of transgenic loci.

1.4. Methylation analyses of genetically modified plants

Genetic changes, such as insertion of transgenes, can lead to epigenetic alterations in plants [85]. Epigenetic variations are heritable variations, which do not involve a change of the DNA sequence [86]. Plant epigenetics is an increasing research field, which can reveal new possibilities for plant improvement. In order to understand phenotypic variations, it is necessary to elucidate the association between genetic and epigenetic changes at particular loci. The best studied epigenetic mark is DNA methylation, a chromatin modification, which can be the result of genetic or epigenetic influences. Other chromatin modifications are histone tail modifications and small RNAs [87].

The effect of transgene integration on epigenetic modification of the host genome is not fully elucidated. However, it is known that epigenetics plays an important role for successful genetic modification and stable transgene expression. There are several reports of undesired epigenetic silencing in GM plants. Epigenetic silencing leads to inactivation of genes when multiple copies of the same or homologous sequence are introduced into a genome. Interaction of homologous sequences can lead to highly methylated DNA in promoter or coding sequences, which induces transcriptional or post-transcriptional gene silencing. Nevertheless, it has to be mentioned that gene silencing is a natural host defense mechanism against viruses, which existed long before genetic engineering. Hence, transgenes are similar cellular invaders, which can induce gene silencing processes. Despite this knowledge, it is hard to exclude epigenetic silencing in GMOs [85].

Plant DNA is highly methylated, mostly 5-methylcytosine, but also N6-methyladenine. Cytosine can be methylated in symmetric CG and CHG (where H is A or T), and asymmetric CHH (where H is A, C or T) context. DNA methylation of cytosine in symmetrical CG motifs is the most common methylation pattern in plants. Gene expression and DNA replication are strongly influenced by DNA methylation. Thus, DNA methylation also influences plant growth and development [88]. Further, mutational hotspots can arise through the promutagenic activities of cytosine modifications such as methylation, deamination and halogenation [89]. Moreover, C → T transitions frequently occur at 5-methylcytosine sites because deamination of 5-methylcytosine leads to thymine, which is not recognized by the enzyme uracil-DNA glycosylase and, therefore, not repaired [15]. Thus, cytosine methylation might also play a role in genetic stability of transgenes.

Despite the knowledge that epigenetic markers affect gene expression and DNA replication, risk assessment for GM plants (chapter 1.2.) does not require investigations of these markers [90]. Further, there are only few studies that have investigated DNA methylation in commercial GM plants. La Paz et al. [76] investigated DNA methylation patterns of the 35S-promoter and the *cry1ab* coding region of the MON810 transgene in different transgenic maize varieties. No statistically significant differences were detected in the promoter and the coding region. The average total methylation was 0.6% (promoter region) and 1.2% (coding region), which represents a low methylation status. Vilperte et al. [90] analyzed DNA methylation levels in five different transgenic maize varieties (four single events and one stacked variety) and identified significant differences in cytosine methylation levels in the Figwort mosaic virus (FMV) promoter and *cry2ab2* coding region between the

four single events. In the promoter region, significant differences were detected between the stacked event variety and the single event varieties. However, the average total methylation was also low for the analyzed regions.

1.5. Zea mays

Zea mays belongs to the grass family (*Poaceae*) and was domesticated from its wild ancestor teosinte (*Zea mays ssp. parviglumis*) more than 9,000 years ago. Mexico is its country of origin [91]. The maize genome is very diverse and almost as large as the human genome (2.3 Gb). It is the largest plant genome ever sequenced. The genome sequence is regularly updated. Maize contains ten chromosomes (2n=20) and includes many long terminal repeat transposons. *Zea mays* underwent genome duplication in the past, which led to several chromosomal breakages and fusions, when the genome returned into a diploid state [92]. Further, maize is a well-established model plant. It can easily be grown outdoor, but also in greenhouse or growth chambers. However, cultivation is restricted to mild temperatures. Generally, self- and crosspollination are possible. Male and female reproductive organs are separated, male germ cells are produced in the tassel at the top of the plant and female germ cells are located in the ears in the middle section of the plant. Two male gametes from a single pollen grain fertilize two female gametes, the egg and the central cell, to form the embryo (2n) and endosperm (3n), respectively [17]. Every single seed of an ear is an independent event [93].

Many economically important crops such as maize, soybeans, cotton, and canola, are also available as genetically modified. Maize is the second-most common GM crop after soybean. The majority of commercial GM maize contains stacked transgenes (insect resistance and herbicide resistance) [6]. Moreover, several new plant breeding methods for genetic modification are applied on maize and other crops (see chapter 1.6.). Hence, the analysis of safety issues regarding GM maize is of major importance to guarantee safety for humans, animals, and the environment.

1.6. New plant breeding techniques (NPBT)

Progress in plant breeding has not stopped and plant breeders try to invent techniques, which are more precise than the classical genetic engineering technique described above. NPBT are used to enhance, silence, insert or remove traits of interest; thereby the DNA of plant cells becomes modified. In contrast to classical (or first generation) genetic engineering, the majority of NPBT do not override the species barrier. NPBT can reduce time ef-

forts and costs necessary for crop variety development. Biotechnology companies and plant breeders describe their techniques as faster version of conventional breeding techniques and hope for lower regulation of NPBT products than for first generation GMOs, since NPBT are closer to natural processes. However, legal classification of NPBT is uncertain in the EU. At the moment it is not clear, if NPBT products become regulated as GMOs. Moreover, in the EU, the definition of GMOs goes back to the year 1990 and needs to be updated [94,51]. Biosafety considerations conducted for plants produced by NPBT have indicated that the general approach, developed for the risk assessment of GM crops in principle, would also be appropriate to address the currently identified risk issues for NPBT crops [53]. Another problem with products of most NPBT is that genetic changes cannot be distinguished from natural occurring mutations. Until now, there is no appropriate detection method for several NPBT products available [95].

Many new biotechnology-based plant breeding techniques with great potential have been developed more recently, among them

- Genome-editing techniques (Oligonucleotide Directed Mutagenesis, Site-Directed Nucleases),
- Cisgenesis and Intragenesis,
- RNA-dependent DNA methylation,
- Reverse Breeding,
- Grafting techniques (grafting on GM rootstock), and
- Agro-infiltration techniques (Agro-infiltration “sensu stricto”, Agro-infection, Floral dip) [94,51].

Of these, genome-editing techniques, particularly CRISPR-Cas9 (Clustered Regulatory Interspaced Short Palindromic Repeats), are thought to be the most powerful techniques. CRISPR-Cas9 is the most simple, efficient, and versatile technique of all genome-editing methods. It enables DNA editing at precise locations [51,96,97]. CRISPR-Cas9 is a bacterial system and belongs to the site-directed nuclease (SDN) techniques (a subgroup of the genome-editing techniques), which cuts DNA at target sites, and thereby produces DSBs, which enable the insertion of mutations. There are three different SDN applications, SDN1 enables the insertion of random mutations, SDN2 enables the insertion of non-random mutations, and SDN3 enables the insertion of large segments (genes). All three SDN applications work at precise locations and rely on natural DSB repair mechanisms. SDN1 relies on the NHEJ pathway, whereas SDN2 and SDN3 rely on the homology directed re-

pair (HDR) pathway. NHEJ is error-prone, thus leading to random mutations. Naturally, HDR uses the sister chromatid as template for the repair of DSB, thereby avoiding the introduction of any mutations. This mechanism is tricked, when employing SDN2 or SDN3 by intentionally introducing an exogenous homologous nucleotide (with the desired change) into the cell, which is used as template instead of the sister chromatid [98,51].

A common application is the delivery of DNA encoding CRISPR-Cas9 into plant cells by classical genetic engineering (by gene gun or *Agrobacterium*-mediated transformation). The expressed CRISPR-Cas9 system leads to cleavage of target sites and causes mutations [99,100]. However, using this method, CRISPR-Cas9 constructs become stably integrated into plant genomes, leading to persistent nuclease activity, which may cause unintended effects in the plant organism [101,102,51,100]. Further, this application overrides the species barrier, since the CRISPR-Cas9 system is of bacterial origin. Furthermore, it is obvious to define resulting organisms as GMOs because classical genetic engineering is used as prior step. One way around this problem is to select negative segregants (without the CRISPR-Cas9 DNA) which carry the desired change on the DNA level [100] How negative segregants become classified legally is not yet clear (in the EU).

2. AIM OF THE DOCTORAL THESIS

The cultivation areas of GM plants intended for human and animal consumption are increasing globally. Insect-resistance and herbicide-resistance are the most dominant GM traits worldwide. Hence, the investigation of safety issues regarding GM plants is of major importance to guarantee safety for humans, animals, and the environment. In addition, updated scientific information is very important for an appropriate risk assessment of GM plants. Since the first release of GM plants, several new molecular biology methods were developed, which have the ability to provide new scientific information. In the era of next generation sequencing, it was obvious to use this powerful tool, in addition to other appropriate molecular biology methods, to provide a deeper inside into the genetics, epigenetics and transcriptomics of frequently marketed GM maize varieties.

The following aspects were studied in detail:

i. The post-transformational genetic stability of single and stacked GM maize varieties,
ii. the analysis of DNA methylation patterns in different GM maize varieties, and
iii. the occurrence of possible unintended effects of GM maize.

Regarding the genetic stability, it was examined if stacked events tend to be more instable compared to single events. In addition, the gap of non-existing published data on the post-transformational stability of herbicide-resistant NK603 maize was filled. For these investigations, high resolution melting (HRM) analysis, Sanger sequencing, and amplicon sequencing (targeted next generation sequencing) were employed.

Concerning the methylation analyses, it was investigated if DNA methylation is connected to genetic instabilities in the insect-resistant MON810 transgene. Bisulfite sequencing was used for this purpose. Despite the strong impact on gene expression, epigenetic analysis is not part of the GM plant risk assessment. Further, only few data on DNA methylation patterns in commercial GM plants are available.

Possible unintended effects were analyzed on the transcriptomic level in herbicide-resistant NK603 maize varieties. In order to detect significant differences between NK603 maize and near-isogenic non-GM maize, high-throughput RNA sequencing and reverse transcriptase (RT-) quantitative (q) PCR were employed.

3. RESULTS REPORTED IN MANUSCIPTS

Paper 1 (Chapter 3.1.)

Mutation Scanning in a Single and a Stacked Genetically Modified (GM) Event by Real-Time PCR and High Resolution Melting (HRM) Analysis

Sina-Elisabeth Ben Ali, Zita Erika Madi, Rupert Hochegger, David Quist, Bernhard Prewein, Alexander G. Haslberger and Christian Brandes

International Journal of Molecular Sciences Volume 15, Issue 11 (2014) 19898-19923; doi:10.3390/ijms151119898

Paper 2 (Chapter 3.2.)

Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing

Magali Castan, Sina-Elisabeth Ben Ali, Rupert Hochegger, Werner Ruppitsch, Alexander G. Haslberger and Christian Brandes

European Food Research and Technology Volume 143, Issue 3 (2017) 353-365; https://doi.org/10.1007/s00217-016-2749-2

Paper 3 (Chapter 3.3.)

Genetic and epigenetic characterization of the cry1ab coding region and its 3´ flanking genomic region in MON810 maize using next-generation sequencing

Sina-Elisabeth Ben Ali, Alexandra Schamann, Stefanie Dobrovolny, Alexander Indra, Sarah Z. Agapito-Tenfen, Rupert Hochegger, Alexander G. Haslberger and Christian Brandes

European Food Research and Technology (2018); https://doi.org/10.1007/s00217-018-3062-z

Paper 4 (Chapter 3.4.)

Analysis of transcriptomic differences between NK603 maize and near-isogenic maize using RNA sequencing and RT-qPCR

Sina-Elisabeth Ben Ali, Agnes Draxler, Alexandra Schamann, Sarah Z. Agapito-Tenfen, Rupert Hochegger, Alexander G. Haslberger and Christian Brandes (in preparation)

Int. J. Mol. Sci. **2014**, *15*, 19898-19923; doi:10.3390/ijms151119898

OPEN ACCESS
International Journal of
Molecular Sciences
ISSN 1422-0067
www.mdpi.com/journal/ijms

Article

Mutation Scanning in a Single and a Stacked Genetically Modified (GM) Event by Real-Time PCR and High Resolution Melting (HRM) Analysis

Sina-Elisabeth Ben Ali [1,2,†], Zita Erika Madi [1,†], Rupert Hochegger [1], David Quist [3], Bernhard Prewein [1], Alexander G. Haslberger [2] and Christian Brandes [1,*]

[1] Austrian Agency for Health and Food Safety, Spargelfeldstrasse 191, 1220 Vienna, Austria; E-Mails: sina-elisabeth.ben-ali@ages.at (S.-E.B.A.); zmadi@gmx.net (Z.E.M.); rupert.hochegger@ages.at (R.H.); bernhard.prewein@ages.at (B.P.)

[2] Department of Nutritional Sciences, University of Vienna, Althanstraße 14, 1090 Vienna, Austria; E-Mail: alexander.haslberger@univie.ac.at

[3] Centre for Biosafety–GenØk, PB 6418 Science Park, 9294 Tromsoe, Norway; E-Mail: david.quist@uit.no

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: christian.brandes@ages.at; Tel.: +43-5-0555-32202; Fax: +43-5-0555-32219.

External Editor: Virginia Garcia-Cañas

Received: 13 August 2014; in revised form: 11 October 2014 / Accepted: 21 October 2014 / Published: 31 October 2014

Abstract: Genetic mutations must be avoided during the production and use of seeds. In the European Union (EU), Directive 2001/18/EC requires any DNA construct introduced via transformation to be stable. Establishing genetic stability is critical for the approval of genetically modified organisms (GMOs). In this study, genetic stability of two GMOs was examined using high resolution melting (HRM) analysis and real-time polymerase chain reaction (PCR) employing Scorpion primers for amplification. The genetic variability of the transgenic insert and that of the flanking regions in a single oilseed rape variety (GT73) and a stacked maize (MON88017 × MON810) was studied. The GT73 and the 5' region of MON810 showed no instabilities in the examined regions. However; two out of 100 analyzed samples carried a heterozygous point mutation in the 3' region of MON810 in the stacked variety. These results were verified by direct sequencing of the amplified PCR products

as well as by sequencing of cloned PCR fragments. The occurrence of the mutation suggests that the 5' region is more suitable than the 3' region for the quantification of MON810. The identification of the single nucleotide polymorphism (SNP) in a stacked event is in contrast to the results of earlier studies of the same MON810 region in a single event where no DNA polymorphism was found.

Keywords: HRM analysis; Scorpion primer; MON810; MON88017 × MON810; GT73; genetic stability; stacked event; GMO

1. Introduction

Since its introduction in 1996, the use of commercial genetically modified (GM) crops has drastically expanded. In the year 2013, 175.2 million hectares of GM crops were grown. Of these, less than 0.3 million hectares were cultivated in the EU (Spain, Portugal, Czech Republic, Romania, and Slovakia) to one singular maize variety (MON810) [1]. More recently, the commercialization of single events (for example MON810 maize or Roundup Ready Soy), have given way to the introduction of stacked events containing multiple transgenes [2]. Approximately 47 million hectares of stacked traits were grown in 2013, which is 3.3 million hectares more than that reported in 2012 [1]. There has been an increase in the number of approved stacked events and the number of transgenes stacked in a single plant (e.g., SmartStax™ maize, containing eight events) [2].

Authorization and compliance measures of GM food and feed require that GM events be genetically characterized and quantified. Additionally, transgene stability is important aspect to ensure food safety [3,4]. Reports on the genetic stability of genetically modified organisms (GMOs) have presented variable results. Initially, it was presumed that the production of transgenic plants involved the insertion of a single transgene and the flanking regions without further genomic disruption. However, instances of genomic alterations, including complex nucleotide rearrangements (e.g., deletions, duplications, insertions of organelle and filler DNA, and translocations) have occurred as a result of the transformation process [5–13]. The stability of the transgene, and gene expression as well as performance of the plant can be influenced by various factors, including host biology, genome organization, by epigenetic factors, or by nucleotide changes within the introduced DNA construct [4].

In addition to the requirement of genetic stability during GMO development, transgenic plants should demonstrate genetic stability subsequently during cultivation and propagation. Recent studies have identified the occurrence of DNA rearrangements after the transformation process and during the post-release monitoring. Choffnes *et al.* [14] studied transgene integration patterns of several hundred plants in subsequent generations. It was demonstrated that the original transgene integration patterns of regenerated primary transgenic plants (T_0) were not stable when passed to progenies, possibly due to a high level of homologous recombination. Another study performed cytogenetic analyses of GM maize and revealed a fragile phenotype of 45S rDNA as a consequence of genetic modification [15]. McCabe *et al.* [11] detected differences in transgene expression between T_0 plants and plants grown from selfed T_0 seeds (T_1 generation) of transgenic lettuce. And it was suggested that these differences were due to inactivation or suppression of transgene expression. Ulian *et al.* [16]

reported instabilities of transfer DNA (T-DNA) insertions in T_0 plant genomes of petunia. Differences in plant protein expression that originated from different locations were also discovered [17,18].

In recent years, new methods have been adapted for profiling the epigenome, proteome, and transcriptome of GMOs [19]. La Paz, *et al.* [20] analyzed cytosine methylation of the MON810 transgene in different MON810 varieties. The methylation level of the transgene was very low and a comparison between the different varieties revealed no significant differences in symmetric DNA methylation. In contrast, significant differences were observed in the asymmetric sites that play a minor role in epigenetics. Most of the reported comparative transcriptome analyses of MON810 maize were performed using microarrays [21–23]. La Paz *et al.* [24] analyzed MON810 and isogenic maize varieties using high-throughput RNA sequencing and found 140 differentially expressed genes. The authors suggested that the differences were due to a slightly delayed maturation process of MON810 compared to the conventional varieties. A proteomic approach was performed by Agapito-Tenfen *et al.* [25]. Using two-dimensional gel electrophoresis combined with mass spectrometry, 32 differentially expressed proteins were identified in Brazilian MON810 maize when compared to isogenic control maize. Rang *et al.* [26] detected different RNA variants transcribed from the transgene in Roundup Ready soybean which may code for 5-Enol-pyruvylshikimate-3-phosphate synthase (EPSPS) fusion proteins.

According to the European Food Safety Authority (EFSA) guidance document and the EU directive 2001/18/EC, the introduced DNA construct must be stable and no changes in the DNA construct should occur during the cultivation and propagation of the plants [27,28]. Even minor changes in the construct are critical because they may lead to unintended changes in the plant properties, content, and/or morphology [29]. Additionally, genetic heterogeneity in the samples may render the analysis for GMO detection unreliable or equivocal [30]. The EFSA guidance document details the requirements for the authorization of GMOs. Accordingly, the GMO events must be analyzed for DNA rearrangements and instabilities of the insert [27,28]. Genetic stability is commonly verified by Southern Blot analysis, however, Southern Blots have drawbacks, including low sensitivity (e.g., only major DNA changes can be detected). In many cases minor changes may occur, which can have an impact on the plant. A variety of methods have been used for analyzing minor nucleotide changes. Genomic samples can be sequenced directly or GMOs can be analyzed by methods such as Sensitive Capillary Electrophoresis (CSCE) or long-range polymerase chain reaction (PCR). Further, real-time PCR employing Scorpion primers for amplification or high resolution melting (HRM) analysis can be used to distinguish between different alleles or to filter out mutated samples [31,32].

According to EFSA, applicants must demonstrate the genetic stability of the transgenic locus over five generations or vegetative cycles before a product can be authorized. For stacked events, the applicant must establish the integrity of the inserts [33]. Further, EFSA determined that the risk assessment of stacked events consisting of events approved by the EU should focus on the genetic stability of such plants, the expression level of the transgenes, and unintended interactions between the stacked events [34]. The genetic stability of the single insert is verified by Southern Blot and segregation analyses before authorization [4,35,36]. The EU directive 2001/18/EC requires that the GMOs be inspected every 10 years [28]. Aguilera *et al.* recommended the validation of the genetic stability of inserts for the whole lifespan of a product [37].

In this study, the genetic stability of a stacked maize (MON88017 × MON810) and that of a single oilseed rape event (GT73) were investigated. MON88017 × MON810 is a maize variety produced

by conventional crossing of a double-stacked maize (MON88017) that expresses a *CP4 EPSPS* gene from Agrobacterium sp. CP4, conferring resistance to glyphosate, and the coleopteran-active delta-endotoxin Cry3Bb1 gene from Bacillus thuringiensis (Bt) (*cry3Bb1*) to provide protection against coleopteran insects, particularly the maize rootworm, and the lepidopteran-active delta-endotoxin Cry1Ab (*cry1Ab)*-containing MON810 for protection against lepidopterans such as *Ostrinia nubilalis* and *Sesamia* spp. GT73 is a glyphosate-tolerant oilseed rape [38]. We used real-time PCR with Scorpion primers for amplification (Scorpion PCR) and HRM analysis to examine whether DNA alterations occurred amongst individual seeds (of MON88017 × MON810 maize) or plants (of GT73 oilseed rape). Earlier studies have adapted these methods to analyze the genetic stability and organization of transgenic insertion events. In Neumann *et al.* [32] it could be shown that single nucleotide polymorphism (SNPs) in plasmids can be demonstrated by Scorpion analysis. As a further work for optimizing the method, different alleles of the alcohol dehydrogenase 1 (*ADH1*) gene in maize were investigated by Madi *et al.* [31]. In single seeds, the homozygous as well as heterozygous state of the *ADH1* gene was found. Alleles can be discriminated using Scorpion PCR as well as by HRM analysis. They can be separated by cloning the PCR-amplified DNA. Testing of these clones allows a better differentiation between the alleles compared to testing only the genomic DNA.

In the present work, real-time PCR in combination with subsequent HRM analysis is used as a highly sensitive mutation scanning of selected DNA regions as a standardized means for determining the level of the genetic stability. Samples that potentially carried mutations were cloned into plasmids and sequenced by the Sanger method to verify the DNA alterations.

2. Results and Discussion

2.1. Basics of Screening Genetically Modified Organisms (GMOs) for Genetic Stability Using Scorpion Polymerase Chain Reaction (PCR) and High Resolution Melting (HRM) Analysis

We analyzed border regions between the endogenous plant genome and the exogenous transgene, known as "event-specific" regions. These regions are being used for the official GMO control. Unidentified mutations may render the test results unreliable. Therefore, we examined the border regions for establishing the genetic stability of GT73 in oilseed rape and MON810 in MON88017 × MON810 maize.

In contrast to the earlier investigations on MON810 [20,32], our study examined MON810 in a stacked event. Combinations of transgenic traits increase the complexity of risk assessment because more undesired scenarios may occur. At first the independent single events present in the stacked event should be safe and stable. In addition, interactions between the events as well as synergistic effects need to be analyzed [33]. Recent studies have demonstrated the suitability of real-time PCR with Scorpion primers and HRM analysis for investigating the nucleotide polymorphisms in genetically modified plants. These tests can be performed readily and are cost effective. The characterization of the Scorpion primers used for the analysis of GMOs has been described by Neumann *et al.* [32] and Thelwell *et al.* [39]. Madi *et al.* [31] have shown that Scorpion PCR combined with HRM analysis is a highly sensitive method for the detection of DNA alterations, including SNPs. Several studies have demonstrated the functionality of HRM analysis [40,41].

The potential genomic alterations were analyzed via a two- to three-step screening procedure, followed by sequencing. In the first step, a large number of individual seeds (MON88017 × MON810 maize samples) or leaves (GT73 oilseed rape samples) were tested using either real-time PCR with Scorpion primers or HRM analysis. The results of this step helped identify the samples that might harbor a genetic divergence. High C_t values obtained from Scorpion PCR analysis and low confidence values obtained from HRM analysis suggest a modification (SNP or other small sequence alterations) in the target region of the sample (see supplementary Figures S1–S4). Values from the first screening tests were used to select samples for a second round of screening with a lower number of genomic samples. Madi *et al.* have shown that PCR-cloning can be used to separate alleles from genomic samples containing heterozygous mutations [31]. Therefore, in the subsequent screening step, the same primers were used to clone the PCR fragments of the selected samples into plasmids. These plasmids were subjected to HRM analysis alone or HRM analysis and Scorpion PCR. The results from the last screening were used to filter out plasmids that harbor mutations. However, the HRM or Scorpion analysis of genomic samples or plasmids could yield false positive results due to variance among samples or because of variations in DNA quality. With the screening procedure alone, the number of true positive samples could not be determined. To avoid such artifacts and accurately identify samples that harbor alterations, the selected samples from the screening were subjected to sequencing as a secondary validation method.

2.2. Screening of 5′ Flanking Region of the GT73 (5′-GT73)

The screening of the 5′ junction of the oilseed rape variety was performed using a three-step procedure employing real-time PCR with Scorpion primers and HRM analysis. The analyzed target sequence of the 5′-GT73 region is shown in Figure 1. The primer sequences are marked in color and are additionally described in Table 1.

Figure 1. Illustration of the target sequence of the Scorpion primers and the High Resolution Melting (HRM) analysis employed for examining the 5′ flanking region of the GT73 (5′-GT73) transgene. The forward primer is shown in red, and the blue letters indicate the reverse primer. The probe sequence (green) lies exactly along the transition from the transgene to the oilseed rape genome. The analyzed sequence contained 156 nucleotides.

5′-CCAATCTGGAATGCTGCTAAATCCTGAGCTCAAGCTTGATGGGGATCAGATTGTCGTT
TCCCGCCTTCAGTTTAAACT**ATCAGTGTTCGACTTTTT**ATGTAACAACCCGCCCCGGATC
CAACCCCGAATCC**CCGTATATTAATAGTTAAGGGGTCT-3′**

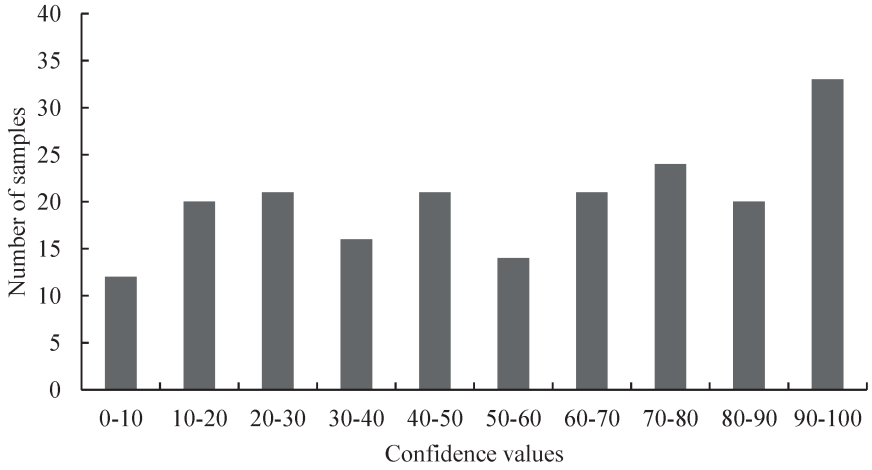
In total, 202 individual GT73 oilseed rape samples were tested by real-time PCR employing Scorpion primers and by HRM analysis on three different days. The C_t values ranged from 38.01 to 51.84. The average C_t value and the standard deviation were 42.78 ± 3.80. A pool of 20 genomic GT73 oilseed rape samples (5′-GT73 pool) served as reference to calculate the confidence values. Further, the confidence values ranged from 0.67 to 99.67, with an average value and a standard deviation of 55.44 ± 29.44. The distribution of the confidence values of the first screening was calculated according to Madi *et al.* [31]. Figure 2 shows the graphical distribution of the HRM measurements.

Table 1. Sequences of primers used for real-time PCR screening and for PCR cloning.

Primer	Sequence
3m810fw	5'-CCAAGCACGAGACCGTCAA-3'
3m810rev	5'-GCTCGCAAGCAAATTCGGAA-3'
VW01 [41]	5'-TCGAAGGACGAAGGACTCTAACG-3'
VW03 [41]	5'-TCCATCTTTGGGACCACTGTCG-3'
5GT73fw	5'-CCAATCTGGAATGCTGCTAAA-3'
5GT73rev	5'-AGACCCCTTAACTATTAATATACGG-3'
5GT73Scorprev	5'-[F]AAAAAGTCGAACACTGAT[Q][HEG]AGACCCCTTAACTATTAATATACGG-3'
3GT73fw	5'-CCATATTGACCATCATACTCATTGCT-3'
3GT73rev	5'-GCTTATACGAAGGCAAGAAAAGGA-3'
M13fw lo2	5'-AACGACGGCCAGTGAATTGTAATACG-3'
M13rev lo2	5'-CAGGAAACAGCTATGACCATGATTACG-3'

[F] Fluorophore, [Q] quencher, [HEG] stopper, the F/Q of 5GT73Scorprev are Cy5/BHQ1.

Figure 2. The distribution of the HRM measurements of 5'-GT73 oilseed rape tested by HRM analyses. For the HRM analysis, the confidence values (range: 0.67–99.67) were divided into 10 classes. The values of these classes are on the *x*-axis and the number of samples for the classes on the *y*-axis.



A total of 74 samples from the first round of screening showing high Scorpion C_t values or low confidence values were selected for further screening. In this screening step, confidence values of the HRM measurements and the C_t values of the Scorpion analysis were calculated. The C_t values varied between 37.69 and 46.44, with an average value and a standard deviation of 42.00 ± 1.97. For calculation of the confidence values the 5'-GT73 pool was used again. The confidence values ranged from 2.86 to 99.67. The average confidence value and the standard deviation were 62.29 ± 30.12. Based on the results of the first screening, eight samples with high C_t values and five samples with low confidence values as observed in the HRM analysis were selected for the second screening. Additionally, two control samples with average C_t and confidence values were determined. The data of the selected samples are presented in Table 2 (additional information of Table 2 is provided as supplementary file (Table S1). In the supplementary table, all C_t and confidence values of selected samples are presented).

Table 2. (**a**) After the first screening, eight samples (13, 14, 18, 20, 30, 40, 152 and 170) were chosen for further analysis based on the C_t values. Sample 13 showed the highest C_t value; (**b**) Six samples (137, 152, 153, 156, 157 and 333) were selected because of low confidence values. Sample 152 was chosen because of high C_t and very low confidence values; (**c**) The two control samples, 118 and 175, with average C_t and confidence values are also shown.

(a) Sample	C_t Value	(b) Sample	Confidence Value	(c) Sample	C_t Value	Confidence Value
13	51.84	137	11.80	118	40.54	35.13
14	47.64	152	2.86	175	41.40	37.49
18	48.30	153	9.17			
20	45.77	156	9.11			
30	45.93	157	17.55			
40	46.67	333	10.15			
152	46.57					
170	45.11					

Based on the results of Madi *et al.*, we assumed that nucleotide changes could also occur as heterozygous alleles [13]. Therefore, prior to the second screening, the selected samples were cloned into plasmids using PCR-cloning (see Experimental Section) to characterize the possible heterozygous alleles. From each selected sample (Table 2), two to three plasmids were generated and the DNA preparations were diluted to a concentration of 10^{-4} ng/μL. These were tested in the second round of screening by real-time PCR employing Scorpion primers and were subsequently subjected to HRM analysis. A plasmid pool consisting of 20 plasmids served as reference. The results of the second screening are shown in Table 3.

Table 3. Illustration of the Scorpion C_t and HRM confidence values obtained from the second round of screening using plasmids.

Sample	C_t Value	Confidence Value	Sample	C_t Value	Confidence Value
13-2	27.40	99.51	137-2	29.12	94.98
13-3	29.33	98.45	137-3	30.67	99.02
14-1	28.50	98.49	152-1	38.42	90.38
14-3	28.26	98.11	152-2	29.33	99.04
18-1	28.02	94.65	152-3	28.36	86.10
18-2	28.64	92.18	153-1	29.30	99.18
20-1	32.16	99.51	153-2	28.07	93.83
20-2	34.58	95.70	153-3	29.19	96.05
20-4	32.45	10.07	156-2	29.92	88.81
30-2	31.12	70.30	156-3	27.35	97.89
30-3	29.89	94.64	157-2	33.44	81.05
30-5	29.58	97.83	157-3	27.74	98.92
40-1	32.92	4.60	170-2	31.27	95.93
40-2	32.41	99.42	170-3	31.36	25.01
40-3	32.42	93.27	175-2	31.44	16.65
118-1	29.21	98.37	175-3	28.48	0.00
118-2	24.60	97.01	333-1	28.52	93.82
118-3	27.89	99.58	333-2	38.83	78.62
137-1	29.09	97.96			

Because plasmids with low confidence values (<50) could possibly contain SNPs, samples 20-4, 40-1, 170-3, 175-2, and 175-3 were selected for sequencing. Sequencing results showed no genetic instabilities. To exclude the possibility that the low number of plasmids generated per genomic sample prevented the identification of heterozygous DNA samples, we generated 30 more plasmids from the genomic samples 40, 175 and 333 (10 plasmids per sample). These plasmids were tested by Scorpion PCR and HRM analyses. The above-mentioned plasmid pool served as reference again. Results of this screening are depicted in Table 4.

As shown in Table 4, the C_t values ranged from 20.69 to 23.44 and the confidence values ranged from 69.21 to 99.79. With both, C_t values and confidence values, the differences between samples were very low. Nevertheless, a minimum of one plasmid per group was sequenced. However, differences in the nucleotide sequence were not detected. Thus, real-time PCR and HRM analyses revealed no genetic instabilities in the analyzed 5' flanking region of the GT73 oilseed rape.

Table 4. Scorpion C_t and HRM confidence values obtained from the screening of additional plasmids (containing 5'-GT73 region). For each genomic sample, 10 cloned fragments were obtained (differentiated by letters a to k).

Sample	C_t Value	Confidence Value	Sample	C_t Value	Confidence Value
40-a	23.44	99.64	175-f	21.25	97.37
40-b	22.08	97.46	175-g	22.31	84.30
40-c	22.25	95.40	175-h	21.16	90.70
40-d	22.78	98.56	175-i	21.08	98.68
40-e	22.32	96.64	175-k	23.05	98.79
40-f	21.99	97.56	333-a	22.25	99.79
40-g	22.40	99.00	333-b	22.15	97.95
40-h	22.10	98.95	333-c	22.30	96.00
40-i	22.14	95.38	333-d	21.77	91.31
40-k	22.19	97.35	333-e	21.72	71.89
175-a	21.09	99.26	333-f	22.27	85.75
175-b	21.02	96.66	333-g	21.88	85.34
175-c	23.38	98.21	333-h	21.60	85.15
175-d	22.05	93.81	333-i	21.70	99.57
175-e	21.30	97.86	333-k	20.69	69.21

2.3. Screening of 3' Flanking Region of the GT73 (3'-GT73)

Although we attempted to use Scorpion PCR followed by HRM analysis to screen the 3' flanking region of the GT73 transgene for mutations, we were unable to develop an adequate Scorpion primer with the desired elements. Specifically, we were unable to distinguish a wild-type plasmid from a plasmid with a 2-bp deletion. Therefore, the first screening was performed using HRM analysis alone. A 108-bp nucleotide sequence at the 3'-GT73 flanking region was analyzed using real-time PCR and HRM analyses. Figure 3 shows the 3'-GT73 target sequence. The primer sequences are presented in Table 1.

Figure 3. Illustration of the target sequence used in the HRM analysis of the 3' junction of GT73 transgene. The transgene sequence is at the 3' region, and the transition to the genome lies in the underlined region. The forward primer is shown in red and the blue letters indicate the reverse primer. The sequence contains 108 nucleotides.

5′-CCATATTGACCATCATACTCATTGCTGATCCATGTAGATTTCCCGGACATGAAGATC
ATCCTCCTTCCTTTCCTTGCCTTTCCT**TCCTTTTCTTGCCTTCGTATAAGC-3′**

All in all, 510 individual samples were screened by HRM analysis. The average confidence value and the standard deviation were 55.16 ± 33.60. A pool of 20 genomic GT73 oilseed rape samples (3'-GT73 pool) served as reference to calculate the confidence values. The distribution of the confidence values was also calculated. The confidence values of the screening were divided into 10 groups and the cluster frequencies were determined.

As illustrated in Figure 4, the clusters with the lowest and the highest confidence values had the highest frequencies. The distribution of the 3'-GT73 confidence values was congruent to a bimodal sample pool containing wild-type (confidence values > 90) and mutant (confidence values < 10) samples. For further investigations, 47 samples with low confidence values were selected and screened two times using HRM analysis. The genomic 3'-GT73 pool served as reference. Based on the average values of this screening, 10 samples with the lowest confidence values were selected (Table 5).

Figure 4. The distribution of the HRM measurements of 3'-GT73 oilseed rape. For the HRM analysis, the confidence values (range: 0.00–99.97) were divided into 10 classes. The values of these classes were drawn on the *x*-axis and the number of samples for the classes on the *y*-axis.

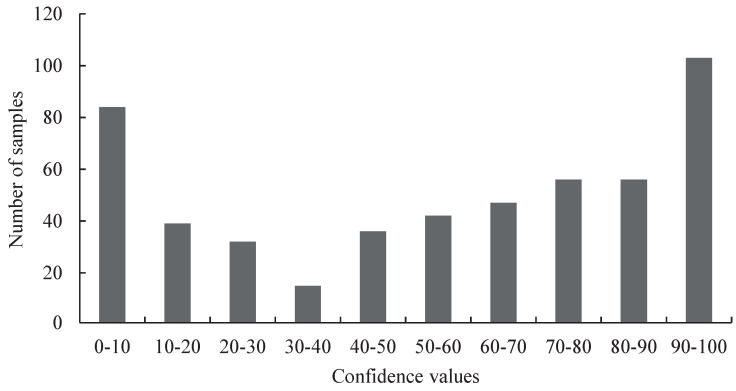


Table 5. The samples 50, 51, 63, 70, 72, 99, 145, 379, 388 and 494 were chosen based on the confidence values of the second screening. Sample 63 had the lowest confidence value.

Sample	Confidence Value
50	1.31
51	1.69
63	0.94
70	4.14
72	1.00
99	3.16

Table 5. *Cont.*

Sample	Confidence Value
145	10.44
379	10.05
388	25.31
494	55.22

These genomic samples were cloned into plasmids by PCR-cloning (Zero Blunt TOPO PCR Cloning kit). For the PCR, a high-fidelity proofreading enzyme (Phusion Hot Start II DNA Polymerase) was used and the same primer pair employed for the first 3'-GT73 screening was utilized. One or two plasmids were generated per selected genomic sample. Plasmid preparations were diluted to a concentration of 10^{-4} ng/μL and analyzed in a second screening step using the 3'-GT73 primer pair. A pool of 10 plasmids served as reference. The results of the second screening are shown in Table 6.

Table 6. In the second screening, 17 plasmids were analyzed. The confidence values ranged from 86.96 to 99.92.

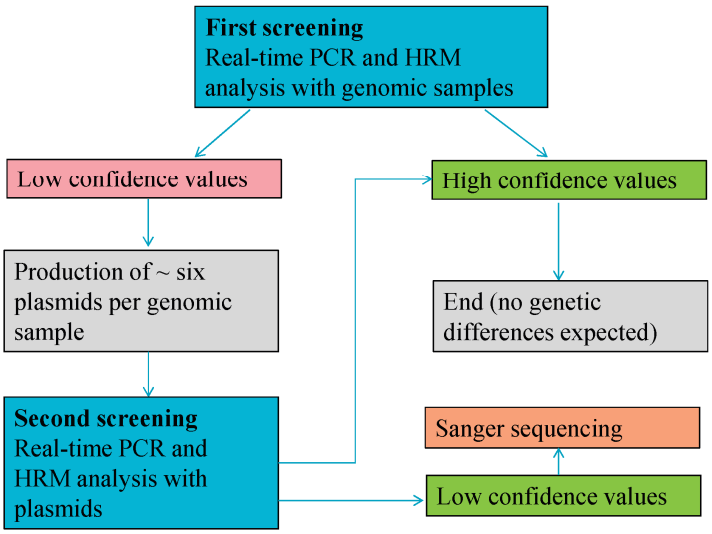
Sample	Confidence Value
50-3	92.04
50-4	99.92
51-1	95.27
51-2	97.84
63-1	86.96
70-1	96.85
72-1	89.00
72-2	99.54
99-2	98.34
99-3	98.83
145-1	94.89
145-2	98.18
379-3	96.40
379-4	99.78
388-1	99.67
494-1	99.81
494-2	99.78

The confidence values obtained for the plasmids in the second screening are different from those of the first screening with genomic samples. Whereas the confidence values of the selected samples in the first HRM screening were low (eight samples below 11 and two samples below 60; Table 5), the HRM analysis of the second screening yielded very high confidence values (>85). Based on these results, we concluded that none of the samples had a mutation in the targeted region. To verify this, two plasmids were sequenced. Sequencing confirmed that mutations were absent. Thus, the analysis of the 3' flanking region of the GT73 transgene in oilseed rape leaves revealed no genetic instabilities.

2.4. Screening of 5' Junction of MON810 (5'-MON810)

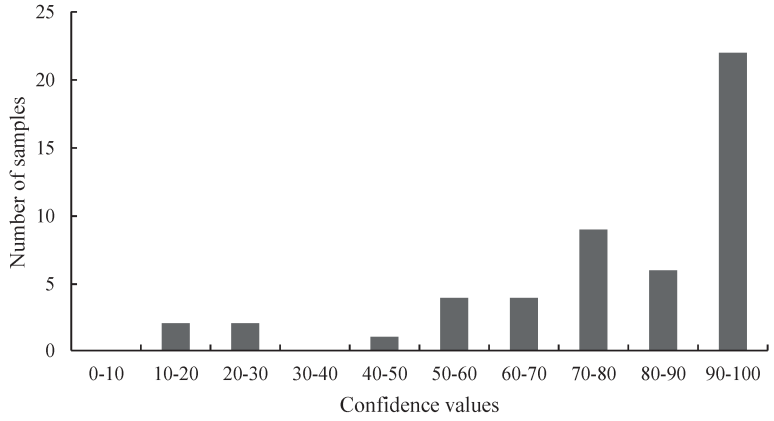
For the investigation of 5' junction of MON810 (5'-MON810), a stacked GM maize variety (MON88017 × MON810) containing three genetic elements was used. The analyzed target region has been described by Neumann *et al.* [32] and the primers are shown in Table 1. It was possible to test this region using Scorpion PCR [32] and HRM analysis. However, because Scorpion analysis only reveals DNA alterations in the probe and primer regions and HRM analysis identifies changes in the whole target sequence, only HRM analysis was performed. For the screening of the 5'-MON810 region, a two-step procedure was employed. In the first step, individual genomic samples of seeds were tested using real-time PCR and HRM analyses. In the second step, the genomic samples that showed the possible presence of mutations were cloned into plasmids and examined by HRM analysis. Finally, these plasmids were sequenced to identify possible mutations. Figure 5 gives an overview of the screening schema.

Figure 5. The two-step screening schema for the analysis of MON810. Only samples with low confidence values in the first screening were selected for the second screening. Samples with high confidence values were associated with wild-type and were not analyzed. Approximately six or more plasmids were generated per selected genomic sample. After the final screening step, selected plasmids were sequenced to identify DNA alterations, and the type and the exact location of the mutations were determined.



In the first step, 50 maize samples (MON88017 × MON810) were tested twice (separately on different days). A pool of 20 genomic maize samples (MON88017 × MON810) served as reference to calculate the confidence values. Figure 6 shows the graphical distribution of the HRM measurements. If the samples harbored mutations, a distribution similar to that in Figure 4 would be expected and very high (wild-type) and very low (mutants) confidence values would be very frequent. However, this was not observed. Cluster 10 showed the highest frequency, whereas cluster one had a very low frequency. The confidence values ranged from 13.16 to 99.83. The average confidence value and the standard deviation were 77.88 ± 23.36.

Figure 6. The distribution of the HRM measurements of the 5' junction of MON810 (5'-MON810) region tested with real-time PCR. For the HRM analysis the confidence values were divided into 10 classes. The values of these classes were plotted on the *x*-axis and the numbers of the samples for the classes on the *y*-axis.



Samples with confidence values below 10 were not present. Four samples, which had confidence values below 25, were chosen for the second screening. Additionally, sample 383 was selected because a pre-screening revealed that this sample contained an SNP in the coding region of the MON810 transgene (see Section 2.5.). The selected samples and the confidence values of the first 5'-MON810 screening are shown in Table 7.

Table 7. Four samples, 272, 275, 322 and 354 were chosen based on the confidence values of the first screening. Sample 383 was selected because it showed a DNA alteration at the 3'-MON810 junction.

Sample	Confidence Value
272	22.41
275	23.18
322	15.95
354	13.16
383	75.80

The 5'-MON810 target region of the chosen maize samples was cloned into plasmids by PCR-cloning (see Experimental Section). Results from our analysis of GT73 suggested that a high number of clones are necessary to distinguish the authentic mutations from artifacts. Therefore, six plasmids with cloned PCR fragments were obtained for each sample. Thus, a total 30 plasmids were obtained and tested in the second 5'-MON810 screening round using HRM analysis employing the same primer pair as that used for the first screening. The confidence values ranged from 17.88 to 99.94. The average value and the standard deviation were 95.02 ± 15.15. Table 8 shows the confidence values of the second screening.

Table 8. Plasmid 272-5 served as reference (confidence value of 100). Plasmids 275-8 and 272-4 showed the lowest confidence values.

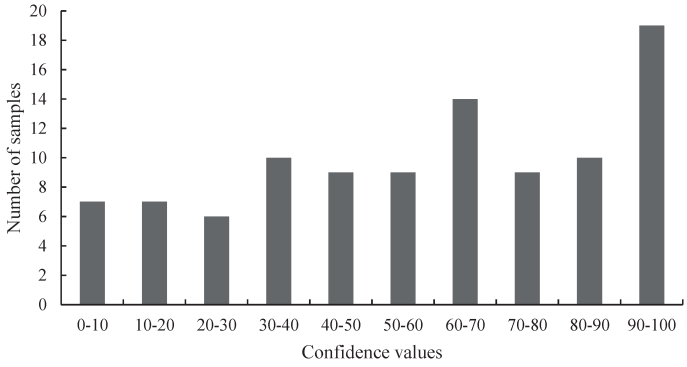
Plasmid	Confidence Value	Plasmid	Confidence Value
272-2	99.66	322-4	90.11
272-3	99.94	322-5	96.53
272-4	79.48	322-6	99.14
272-5	100.00	354-1	99.75
272-6	99.62	354-2	99.17
272-8	99.23	354-4	91.91
275-1	96.43	354-5	99.56
275-2	95.71	354-7	99.55
275-3	98.19	354-8	99.62
275-5	98.91	383-2	97.87
275-6	99.68	383-3	96.63
275-8	17.88	383-4	98.40
322-1	99.88	383-5	99.63
322-2	98.81	383-6	99.67
322-3	99.71	383-7	99.92

As shown in Table 8, only plasmid 275-8 had a low confidence value (17.88). Plasmid 272-4 showed the second lowest value (79.48). All other plasmids showed a confidence value of >90 and therefore were associated with the wild-type. To verify if the plasmids 275-8 and 272-4 harbored mutations, these samples were subjected to Sanger sequencing. The sequencing revealed an SNP, a C→T substitution, at position 147 in plasmid 275-8, but not in 272-4. Because no other plasmid generated from the sample 275 showed this DNA alteration, this SNP was assessed as due to polymerase error. Polymerase errors can occur even when high-fidelity enzymes are used. The theoretical probability of having a polymerase error in the present cloning sequence was 0.003 (calculated according to Foissac) [42]. Thus, three out of 1000 fragments may contain an error. The calculated error frequency is markedly lower than that expected based on the results of our experiments. A somatic mutation can also be the reason for the altered DNA sequence of sample 275-8. However, false positive results demonstrate the importance of the cloning step and the subsequent detailed characterization of these plasmids after the first screening.

2.5. Screening of 3' Junction of MON810 (3'-MON810)

We analyzed the 3' junction of MON810 (3'-MON810) region of the GMO variety MON88017 × MON810. The 3'-MON810 target sequence has been reported previously by Neumann *et al.* [32]. The primers are shown in Table 1. The two-step procedure shown in Figure 5 was used for screening. The distribution of the HRM measurements is shown in Figure 7. Based on this distribution, no or very few mutations would be expected. This is because a bimodal distribution was not evident. A pool of 20 genomic maize samples (MON88017 × MON810) served as reference to calculate the confidence values. The confidence values ranged from 0.54 to 99.03. The average confidence value and the standard deviation were 57.91 ± 29.59.

Figure 7. The distribution of the HRM measurements of the 3' junction of MON810 (3'-MON810) region tested with real-time PCR. For the HRM analysis the confidence values were divided into 10 classes. The values of these classes were plotted against the number of samples.



For the second screening, eight samples from clusters one and two were chosen. Table 9 shows the selected samples and the confidence values obtained from the HRM analysis performed for the first screening.

Table 9. The eight samples shown were selected for the second screening because they resulted in low confidence values in the first screening.

Sample	Confidence Value
263	10.95
264	1.23
345	0.54
352	7.91
354	8.45
355	3.01
374	10.88
383	4.81

The 3'-MON810 target regions of the chosen maize samples were cloned into plasmids (see Experimental Section). For each sample, six plasmids were obtained and are shown in Table 9. Thus, in the second 3'-MON810 screening, 48 plasmids were subjected to HRM analysis using the same primer pair as that employed for the first screening. The confidence values ranged from 2.94 to 99.14. The average value and the standard deviation were 70.46 ± 33.69. Table 10 shows the confidence values obtained from the second screening.

Table 10. A total of 48 plasmids (containing the 3'-MON810 target region) were analyzed in the second screening. The plasmid 383-2 served as reference (confidence value of 100).

Plasmid	Confidence Value	Plasmid	Confidence Value
263-2	67.69	354-1	44.77
263-3	72.02	354-2	15.97
263-4	66.57	354-3	90.48
263-5	62.97	354-4	64.39

Table 10. *Cont.*

Plasmid	Confidence Value	Plasmid	Confidence Value
263-7	82.78	354-5	62.92
263-8	88.24	354-6	81.43
264-1	88.51	355-1	91.75
264-2	7.99	355-2	90.81
264-3	4.61	355-3	89.48
264-4	4.95	355-4	95.72
264-6	6.01	355-5	94.10
264-7	7.97	355-7	94.76
345-2	64.71	374-2	99.14
345-3	75.71	374-3	99.01
345-4	78.59	374-4	98.39
345-5	86.28	374-6	98.75
345-6	91.02	374-7	98.61
345-7	84.85	374-8	96.73
352-1	81.06	383-1	3.14
352-2	96.06	383-2	100.00
352-3	95.95	383-5	97.56
352-4	97.95	383-7	95.53
352-5	93.88	383-8	6.40
352-6	92.63	383-9	2.9

Several plasmids were chosen for sequencing. These are depicted in Table 11. Altered nucleotide sequences were detected in 14 plasmids. These changes were verified by forward and reverse sequencing. The plasmids with altered nucleotide sequences belonged to three different genomic samples, 264, 354 and 383. In the case of sample 264, six out of seven plasmids showed a C→T substitution at position 71 of the analyzed 3'-MON810 target sequence, whereas four of seven plasmids of sample 383 showed a C→T substitution at position 71. A heterozygous SNP appeared to be present in sample 383. For sample 264 it was not clear whether the SNP was homozygous or heterozygous. Thus, the PCR product of the 3'-MON810 target region in sample 264 was subjected to direct sequencing. As shown in Figure 8, the result indicated that the SNP was heterozygous. In the sequencing chromatogram, an overlay of two peaks was present at position 71. A thymine peak was found to overlay a cytosine peak. Direct sequencing of the 3'-MON810 target region of the sample 383 also revealed an overlay of a cytosine and a thymine peak (data not shown). The results for sample 354 favor the notion that this was a false positive. This is because there were many different types of SNPs. In addition to the wild-type three different substitutions (G→T, C→T und C→A) at three different positions were found in the plasmids produced from the same genomic sample 354. Direct sequencing of a PCR product of the 3'-MON810 region of sample 354 showed wild-type. Although a plant cultivated from sample 354 would have provided useful information, this approach was not feasible because the whole maize kernel was homogenized for DNA extraction.

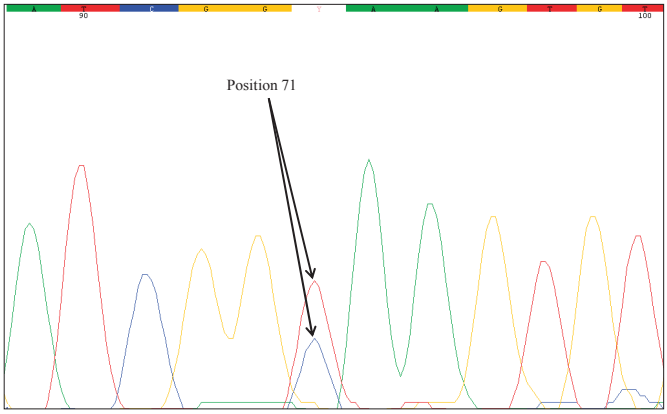
It is unknown whether the partially triploid state of the grain may have influenced the results. Endosperm, a triploid tissue, accounts for 80%–90% of the maize kernel's weight. Endosperm contains two maternal and one paternal genome copies. For interpreting the results, it is important to consider

the degree of zygosity of the analyzed maize variety. It makes a difference if the seeds are hemi- or homozygous for the transgenic locus and in the case of hemizygosity, it is important to know if the transgene donor is male or female [43]. The degree of zygosity in the analyzed maize variety (MON88017 × MON810) remains unknown.

Table 11. Overview of the sequencing results for 3'-MON810 plasmids. Plasmids sequenced because of the screening results are marked with *. All other plasmids were sequenced subsequently because of the prior sequencing results. The confidence values are shown together with sequencing results for a better overview.

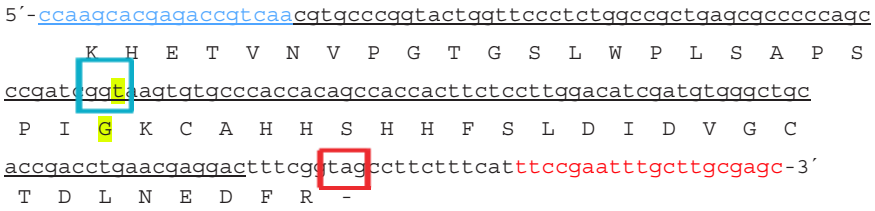
Plasmid	Confidence Value	SNP	Position
263-2	67.69	no	
263-3	72.02	no	
263-4	66.57	no	
263-5	62.97	no	
263-7	82.78	no	
263-8	88.24	no	
264-1	88.51	no	
264-2 *	7.99	C→T	71
264-3 *	4.61	C→T	71
264-4 *	4.95	C→T	71
264-6 *	6.01	C→T	71
264-7 *	7.97	C→T	71
264-9	not tested	C→T	71
345-2	64.71	no	
345-3	75.71	no	
345-4	78.59	no	
345-6	91.02	no	
352-1	81.06	no	
352-2	96.06	no	
354-1	44.77	G→T	74
354-2 *	15.97	C→T	86
354-3	90.48	no	
354-4	64.39	C→A	84
354-5	62.92	C→A	84
354-6	81.43	no	
355-3	89.48	no	
355-7	94.76	no	
374-2	99.14	no	
374-8	96.73	no	
383-1 *	3.14	C→T	71
383-2	100.00	no	
383-5	97.56	no	
383-7	95.53	no	
383-8 *	6.40	C→T	71
383-9 *	2.94	C→T	71
383-11	not tested	C→T	71

Figure 8. The sequencing chromatogram of the 3'-MON810 target region of sample 264 clearly showed a double peak at position 71. Double peaks occur when different alleles are present.



All SNPs detected in a study by Ogasawara *et al.* were silent mutations [44]. Silent mutations do not result in an altered amino acid sequence. The SNPs at position 71 identified in samples 264 and 383 were also silent mutations (the codons GGC and GGT both code for the amino acid glycine). This is illustrated in Figure 9. The detected SNP lies within the coding region of the *cry1Ab* toxin.

Figure 9. The analyzed 3'-MON810 target region is shown. The underlined DNA sequence belongs to the transgenic construct. The forward primer is shown in blue and the reverse primer in red. The SNP location is highlighted in yellow and the affected codon is framed blue. Cytosine is substituted by a thymine at position 71. The amino acid sequence of the *cry1Ab* gene is shown below the nucleotide sequence. The coded amino acid glycine (marked in yellow) is not affected by the SNP. The stop codon (framed red) lies within the genomic DNA because a truncation event at the 3' end of the *cryIAb* gene led to the loss of the nopaline synthase (NOS) terminator (for more details see [17]).



The results showed that a nucleotide change at the 3'-MON810 region was present in two out of 100 analyzed samples. However, the true rate of mutation at the 3' end of the MON810 transgene in MON88017 × MON810 remains unknown. For an estimation of this rate, a higher number of samples need to be analyzed.

3. Experimental Section

Unless stated otherwise, the chemicals, PCR reagents, and PCR primers used were of analytical grade and were purchased from Biozym (Hessisch Oldendorf, Germany) (PCR reagents) and Sigma–Aldrich (Vienna, Austria) (PCR primers and chemicals). The Zero Blunt TOPO PCR Cloning kit (Invitrogen, Vienna, Austria), EasyPrep Pro Plasmid Miniprep kit (Biozym), JM109 Competent Cells (>107 cfu/µg,

Promega, Mannheim, Germany), Wizard DNA Clean-up System (Promega), Phusion Hot Start II High-Fidelity DNA Polymerase kit (Biozym), and Type-it HRM PCR Kit (Qiagen, Hilden, Germany) were used according to the respective manufacturer's recommendations. The stacked event MON88017-3 × MON-00810-6 (hybrid 4421VT3) maize was obtained from Croplan Genetics® (Williston, ND, USA). And the GT73 (MON-00073-7) oilseed rape seeds were obtained by taking plants randomly from a field in Saskatchewan, Canada (season 2011). GT73 oilseed rape plants were cultivated in a closed system (in a separate room under controlled conditions: day/night 20 ± 2 °C, 12/12 h light/dark, r. h. 60% ± 5%). Leaves were harvested after four weeks and immediately frozen in liquid nitrogen.

3.1. Extraction of Genomic DNA

Maize seeds were homogenized with a household garlic crusher. Plant leaves of GT73 were crushed using a mortar and a pestle after cooling with liquid nitrogen. A total 100 mg of the homogenized leaves (GT73) or 150 mg of the homogenized seeds (MON88017 × MON810) were incubated with a mixture of 820 µL of TNE buffer (10 mM Tris (pH 8), 150 mM NaCl, 2 mM EDTA (pH 8)), 1% SDS, 150 µL of 5 M Guanidine-HCl, and 30 µL of proteinase K (600 µg/mL) overnight at 60 °C under constant shaking. Following this, the mixture was centrifuged for 5 min at 16,100× *g*. To 600 µL of the supernatant, 300 µL of chloroform (99%) was added and the mixture was vortexed for 20 s. The mixture was then centrifuged at 16,100× *g* for 8 min to separate the phases and 500 µL of the aqueous phase was transferred into a new tube. To this, 4 µL of RNAse (32 µg/mL) was added, and the mixture was incubated at 60 °C for 30 min under constant shaking. The extracted DNA was purified using the Wizard DNA Clean-up System (Promega). After a pre-elution with 20 µL of ddH_2O, the DNA was eluted with 10 mM Tris buffer (pH 7.4, 70 °C) for 10 min. The eluted DNA was immediately stored at −20 °C.

3.2. Instruments Used for HRM and Scorpion Analysis

Scorpion PCR experiments can be performed on any standard real-time PCR instrument. For HRM analysis, it is necessary to use a specific cycler optimized for precise melting analysis that uses dedicated heating algorithms and specific software [45,46]. For the Scorpion PCR as well as for the HRM experiments a Rotor-Gene Q (Qiagen) instrument was used. For the calculation of the scores (confidence values) of the HRM analysis, a specific software package (Rotor-Gene 2.0.2.4, Qiagen) of the Rotor-Gene Q cycler was used.

3.3. Screening by Real-Time PCR Using Scorpion Primers and HRM Analysis

A total 510 individual oilseed rape samples (GT73) and 100 individual maize samples (MON88017 × MON810) were analyzed on different days by three different operators. The mean of the values was calculated and was used for further investigation. Pools of genomic samples (for primary screenings) or plasmids (secondary screenings) served as reference to calculate the confidence values. The reference served as a control for the performance of all samples (including a positive (standard)

control as well as a non-template control), allowed for the verification of the validity of all parameters of the run, and provided a reference value that enabled a comparison of different runs.

Unimolecular, singular Scorpion primers consisting of forward/reverse PCR-primer sequences as well as additional elements were used for this work. The additional elements included a PCR stopper to prevent the PCR read-through of the probe element, and a specific probe sequence with a fluorescence-based detection module consisting of a fluorophore and a quencher. The PCR yields an amplicon containing a sequence that is complementary to the Scorpion probe. The specific probe sequence then binds to its complementary sequence within the amplicon, the hairpin loop opens up, and a signal is produced [32,39].

With HRM analysis, the melting point (*T*m) of the amplicon produced is determined after the PCR reaction. This is achieved by gradually increasing the temperature from the annealing temperature up to 95 °C. Simultaneously, the fluorescence of a DNA intercalating dye is measured. Because high-precision temperature control and fluorescence measurement are required for successfully implementing this method, such measurements can only be made using an HRM-certified equipment [40].

3.4. Qualitative PCR for Cloning

The PCR was performed with genomic DNA as template. A total 20 µL of the reaction mixture contained 4 µL of 5× Phusion HF-buffer (proprietary information, 1.5 mM $MgCl_2$, error rate was 4.4×10^{-7}), 500 nM of each primer, 250 µM of each deoxynucleoside triphosphate (dNTP), 0.4 U Phusion Hot Start II DNA Polymerase, and 50 ng of DNA. The temperature profile is shown in Table 12.

Table 12. The PCR temperature profile of the two-step protocol used for analysis. Touchdown PCR was used to eliminate non-specific amplification. Therefore, the annealing temperature started at 66 °C and was lowered by 0.5 °C per cycle for the first 18 cycles. The last 22 cycles had an annealing temperature of 60 °C.

Cycle Step	Temp.	Time	Number of Cycles
Initial denaturation	98 °C	2 min	
Denaturation	98 °C	10 s	
Annealing	66–57.5 °C Touchdown (−0.5 °C each cycle)	30 s	18
Extension	72 °C	30 s	
Denaturation	98 °C	10 s	
Annealing	60 °C	30 s	22
Extension	72 °C	30 s	
Final extension	72 °C	7 min	

A 1 µL aliquot of the PCR product was ligated and cloned into pCR-Blunt II-TOPO using the Zero Blunt TOPO PCR Cloning kit (Invitrogen) and transformed into JM109 Competent Cells. Colonies were grown in 5 mL of LB-medium and the recombinant DNA was isolated with the help of EasyPrep Pro Plasmid Miniprep kit (Biozym). Positive colonies were analyzed for the correct length of inserts using agarose gel electrophoresis on 2.5% gels. The validity of the insert was then verified by sequencing using M13 primers.

3.5. Real-Time PCR and HRM Conditions

Real-time PCR was performed on a Rotor-Gene Q (Qiagen) instrument. The total volume of the reaction mixture was 16 μL. The Type-it HRM PCR Kit (Qiagen), 700 nM primers (Scorpion primer 125 nM), and 1.6 μL of undiluted DNA (70–100 ng/μL) template were used for each reaction. The thermal cycling profile used is shown in Table 13.

Table 13. The thermal cycling profile of the real-time PCR (prior to HRM analysis) experiment.

Cycle Step	Temperature	Time	Number of Cycles
Initial Denaturation	95 °C	5 min	
Denaturation	95 °C	10 s	
Annealing	55 °C	30 s	55
Extension	72 °C	20 s	

The experiments were performed on different days and the mean of the values was calculated and was used for further analysis. HRM was performed with the temperature increasing at a rate of 0.2 °C per 4 s. The initial and final temperatures were 70 and 95 °C, respectively. The melting curves and the temperature-shifted curves were normalized to enable the comparison of samples. Modified curves and HRM scores were obtained using the Rotor-Gene Q series software (version 2.0.2.4, Qiagen). The normalized and temperature-shifted melting curves corresponded to the final curve after normalization. When an amplicon harbored a sequence variation, the normalized and temperature-shifted melting curves had a different shape from those of the wild-type amplicons (see Supplementary Figures S3 and S4).

3.6. Sequencing

All sequence analyses were performed using the Sanger sequencing method. The BigDye Fast cycle sequencing protocol, described by Platt *et al.* [47], was used for this purpose. The M13 sequences necessary for sequencing were present in the plasmids generated, as the cloning vector contained these sequences. To avoid sequencing errors, both forward and reverse sequencing were included in the method. This insured the reliability of the sequencing results.

3.7. Primers

Primers used for this work are shown Table 1.

4. Conclusions

The objective of the present work was to identify DNA alterations that may cause genetic instability in GMOs, particularly in stacked events. The number of mutated samples we found in our study was two (out of 100 stacked maize samples tested). A much higher number of samples was analyzed (n = 567 MON810 maize samples and n = 1034 RRS 40-3-2 samples of soy) in earlier studies that found no DNA alterations in soybean and maize [31,32]. The SNP frequency we observed in our study markedly exceeds the natural mutation rate. The average single-base substitution rate for maize is estimated to be 5.39×10^{-8} per site per year and for genic maize regions, 4.79×10^{-8} per site per

year [48]. One important point to mention is that the detected SNPs were most likely not identified by low-sensitivity methods such as Southern Blot analysis.

Increased rate of mutation at the border regions may have a negative influence on GMO quantification by real-time PCR. Madi *et al.* [31] concluded that analysis of DNA instabilities in regions commonly used for the quantification of GMOs is particular important. Therefore, in canola and maize, the border regions from the genome into the construct at the 5' and 3' regions of the relevant transgene were tested in our study. Mutations were detected only in the 3'-MON810 region. Therefore, the 5'-MON810 region appears to be more suitable for GMO quantification as no instabilities were identified in this region. Based on the definition that the first base (at the 5' end) of the analyzed 3'-MON810 target region is at position one, we found that both of the mutated samples harbored an SNP at position 71. However, by investigating a higher number of stacked maize samples (MON88017 × MON810), it cannot be excluded that SNPs would be identified at (other) positions which may change the transgenic protein. The inserted 3' flanking region of MON810 is a truncated version of MON810. The 3' end of the *cry1Ab* gene and the nopaline synthase (NOS) terminator were lost during the plant transformation [49]. It is also possible that the occurrence of the SNP detected in this study at the same region was random. One additional aim of this study was to evaluate the impact of the identified SNP on the amino acid sequence. Similar to that reported by Ogasawara *et al.* [44], the identified mutation was silent, meaning the amino acid sequence remained unchanged. Therefore, it is conceivable that the phenotype of the sample was not altered.

One unresolved question is the degree of zygosity of the transgene in the samples examined. Direct sequencing of the PCR products generated from the mutated samples indicated heterozygosity at nucleotide position 71. Two different nucleotides were identified at this position, a cytosine as wild-type and a thymine as mutant-type. The sequencing chromatograms of the PCR products clearly showed an overlay of these bases. However, sequencing errors could lead to similar chromatograms. Some positions may show overlapping nucleotides and it can be difficult to recognize heterozygous mutations just by direct sequencing of the PCR products. Therefore, we used the cloning approach described by Madi *et al.* [31]. Plasmids containing DNA of one allele can be obtained from genomic samples with heterozygous alleles. With clones comprised of one allele, the outcome of sequencing is unambiguous and sequencing error can be ruled out. PCR products of the affected samples were sequenced twice (forward and reverse). The plasmids of the MON88017 × MON810 samples 264 and 383 were also sequenced twice.

In Neumann *et al.* [32] and Madi *et al.* [31] the question was asked if the scores or the distribution of the values of the primary tests, (*i.e.*, the C_t values of the Scorpion PCR and the confidence values of the HRM analyses) can be used to detect mutations. Our results show that the distribution of the values does not necessarily indicate the presence of mutations (Figures 2, 4, 6 and 7). For samples harboring mutations, two groups can be expected: samples with very high confidence values (wild-type) and those with very low confidence values (mutated samples). However, only Figure 4 (belongs to the distribution of 3'-GT73 samples) showed a distribution consistent with this description and no mutations were identified in these samples. In contrast, the distribution of the values of the 3'-MON810 screening did not show the expected characteristics. However, mutations were found in the 3'-MON810 region.

Based on the mutations identified by this work, we conclude that cloning is required for the unambiguous characterization of SNPs in GMOs. In addition to the two mutations, we also encountered false positive results revealed through subsequent investigation. The different plasmids of the affected sample (354 of MON88017 × MON810 maize) showed inconsistent SNPs (Table 11). Although PCR error is a possible explanation, the precise nature of the factors that contributed to the false positive result is not known. Additionally, plasmid 275-8 of the 5'-MON810 region showed an SNP, which was assessed as due to error in the cloning sequence. A related aspect is the number of plasmids required. Our results show that minimum six to 10 plasmids are necessary for the characterization of heterozygous samples.

Whereas earlier studies investigated single MON810 maize events [20,32], the present work focused on stacked maize event containing MON810. Stacked events having the same promoter are more susceptible to unintended effects of the transgenic event expression [50]. Many stacked events carry transgenes that use the same promoter. The 35S promoter, a viral element, occurs twice in MON88017 × MON810 maize. Kohli *et al.* discussed that the 35S promoter contains a recombination hot spot that has an influence on GMOs [51]. Since stacked events contain multiple viral promoters the susceptibility to instabilities may be increased. Due to an increasing number of stacked events being commercialized, there is a proportionally-relevant need for the analysis of their genetic stability in the context of the specific stacked event in question on a case-by-case basis. Due to the present results one might assume that stacked events tend to be more instable than single events. To confirm this assumption, it would be necessary to demonstrate further instabilities of MON810 in other stacked events. Corresponding experiments are planned. In these, it will be important to analyze whether stacking of the transgenes, and the means by which the stacked varieties were created, may be a contributing factor. It also will be necessary to understand the potential characterization (regulatory) and biosafety issues that may be relevant.

For future studies, it may also be interesting to analyze the zygosity degree of the mutated samples in detail.

Supplementary Materials

Supplementary materials can be found at http://www.mdpi.com/1422-0067/15/11/19898/s1.

Acknowledgments

We would like to express our gratitude to all three anonymous reviewers for their helpful and detailed comments. The staff costs for this work were partially supported by the Austrian Research Promotion Agency (FFG) through FEMtech internships for the students.

Author Contributions

Christian Brandes, Rupert Hochegger and Alexander G. Haslberger conceived and designed the experiments. Christian Brandes supervised the project. Sina-Elisabeth Ben Ali and Zita Erika Madi performed the experiments. Bernhard Prewein performed sequencing and provided technical advice. Christian Brandes, Zita Erika Madi and Sina-Elisabeth Ben Ali analyzed the data. Christian Brandes,

Sina-Elisabeth Ben Ali and David Quist wrote the paper. David Quist provided material for the experiments.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. James, C. *Global Status of Commercialized Biotech/GM Crops*; ISAAA Briefs: Cornell University, Ithaca, NY, USA, 2013; p. 46.
2. Taverniers, I.; Papazova, N.; Bertheau, Y.; de Loose, M.; Holst-Jensen, A. Gene stacking in transgenic plants: Towards compliance between definitions, terminology and detection within the EU regulatory framework. *Environ. Biosaf. Res.* **2008**, *7*, 197–218.
3. Hernandez, M.; Rodriguez-Lazaro, D.; Ferrando, A. Current methodology for detection, identification and quantification of genetically modified organisms. *Curr. Anal. Chem.* **2005**, *1*, 203–221.
4. Pla, M. Insert stability and transgenic plant risk. In *Encyclopedia of Biotechnology in Agriculture and Food*; CRC Press Inc.: London, UK, 2012.
5. Forsbach, A.; Schubert, D.; Lechtenberg, B.; Gils, M.; Schmidt, R. A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. *Plant Mol. Biol.* **2003**, *52*, 161–176.
6. Latham, J.R.; Wilson, A.K.; Steinbrecher, R.A. The Mutational consequences of plant transformation. *J. Biomed. Biotechnol.* **2006**, *2006*, 1–7.
7. Wilson, A.K.; Latham, J.R.; Steinbrecher, R.A. Transformation-induced mutations in transgenic plants: Analysis and biosafety implications. *Biotechnol. Genet. Eng. Rev.* **2006**, *23*, 209–237.
8. Smith, N.; Kilpatrick, J.B.; Whitelam, G.C. Superfluous transgene integration in plants. *Crit. Rev. Plant. Sci.* **2001**, *20*, 215–249.
9. Nacry, P.; Camilleri, C.; Courtial, B.; Caboche, M.; Bouchez, D. Major chromosomal rearrangements induced by T-DNA transformation in Arabidopsis. *Genetics* **1998**, *149*, 641–650.
10. Windels, P.; de Buck, S.; van Bockstaele, E.; de Loose, M.; Depicker, A. T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. *Plant Physiol.* **2003**, *133*, 2061–2068.
11. McCabe, M.S.; Mohapatra, U.B.; Debnath, S.C.; Power, J.B.; Davey, M.R. Integration, expression and inheritance of two linked T-DNA marker genes in transgenic lettuce. *Mol. Breed.* **1999**, *5*, 329–344.
12. Belgian Biosafety Server. Available online: http://www.biosafety.be/gmcropff/EN/TP/MGC.html (accessed on 14 July 2014).
13. Tizaoui, K.; Kchouk, M.E. Genetic approaches for studying transgene inheritance and genetic recombination in three successive generations of transformed tobacco. *Genet. Mol. Biol.* **2012**, *35*, 640–649.
14. Choffnes, D.S.; Philip, R.; Vodkin, L.O. A transgenic locus in soybean exhibits a high level of recombination. *In Vitro Cell. Dev. Biol.* **2001**, *37*, 756–762.

15. Waminal, N.E.; Ryu, K.H.; Choi, S.-H.; Kim, H.H. Randomly detected genetically modified (GM) maize (*Zea mays* L.) near a transport route revealed a fragile 45S rDNA phenotype. *PLoS One* **2013**, *8*, e74060.
16. Ulian, E.C.; Magill, J.M.; Magill, C.W.; Smith, R.H. DNA methylation and expression of NPT II in transgenic petunias and progeny. *Theor. Appl. Genet.* **1996**, *92*, 976–981.
17. Rosati, A.; Bogani, P.; Santarlasci, A.; Buiatti, M. Characterisation of 3' transgene insertion site and derived mRNAs in MON810 YieldGard® maize. *Plant Mol. Biol.* **2008**, *67*, 271–281.
18. Nguyen, H.T.; Jehle, J.A. Quantitative analysis of the seasonal and tissue-specific expression of Cry1Ab in transgenic maize Mon810. *J. Plant Dis. Prot.* **2007**, *114*, 82–87.
19. Heinemann, J.A.; Kurenbach, B.; Quist, D. Molecular profiling—A tool for addressing emerging gaps in the comparative risk assessment of GMOs. *Environ. Int.* **2011**, *37*, 1285–1293.
20. La Paz, J.L.; Pla, M.; Papazova, N.; Puigdomènech, P.; Vicient, C.M. Stability of the MON810 transgene in maize. *Plant Mol. Biol.* **2010**, *74*, 563–571.
21. Coll, A.; Nadal, A.; Collado, R.; Capellades, G.; Kubista, M.; Messeguer, J.; Pla, M. Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. *Plant Mol. Biol.* **2010**, *73*, 349–362.
22. 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 Mol. Biol.* **2008**, *68*, 105–117.
23. Barros, E.; Lezar, S.; Anttonen, M.J.; van Dijk, J.P.; Röhlig, R.M.; Kok, E.J.; Engel, K.H. Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnol. J.* **2010**, *8*, 436–451.
24. La Paz, J.L.; Pla, M.; Centeno, E.; Vicient, C.M.; 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*, 1–13.
25. Agapito-Tenfen, S.Z.; Guerra, M.P.; Wikmark, O.G.; Nodari, R.O. Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. *Proteome Sci.* **2013**, *11*, 1–15.
26. Rang, A.; Linke, B.; Jansen, B. Detection of RNA variants transcribed from the transgene in Roundup Ready soybean. *Eur. Food Res. Technol.* **2005**, *220*, 438–443.
27. EFSA GMO Panel. Guidance document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. *EFSA J.* **2006**, *99*, 1–100.
28. European Parliament. 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-Commission Declaration. Available online: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32001L0018 (accessed on 12 March 2001).
29. Moch, K. Epigenetische Effekte bei transgenen Pflanzen: Auswirkungen auf die Risikobewertung. *BfN-Skripten* **2006**, *187*, 1–80.

30. Broothaerts, W.; Corbisier, P.; Schimmel, H.; Trapmann, S.; Vincent, S.; Emons, H. A single nucleotide polymorphism (SNP839) in the *adh1* reference gene affects the quantitation of genetically modified maize (*Zea mays* L.). *J. Agric. Food Chem.* **2008**, *56*, 8825–8831.
31. Madi, Z.E.; Brandes, C.; Neumann, G.; Quist, D.; Ruppitsch, W.; Hochegger, R. Evaluation of Adh1 alleles and transgenic soybean seeds using Scorpion PCR and HRM analysis. *Eur. Food Res. Technol.* **2013**, *237*, 125–135.
32. Neumann, G.; Brandes, C.; Joachimsthaler, A.; Hochegger, R. Assessment of the genetic stability of GMOs with a detailed examination of MON810 using Scorpion probes. *Eur. Food Res. Technol.* **2011**, *233*, 19–30.
33. EFSA GMO Panel. Scientific Opinion on Guidance for risk assessment of food and feed from genetically modified plants. *EFSA J.* **2011**, *9*, 2150.
34. EFSA GMO Panel. Guidance Document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants containing stacked transformation events. *EFSA J.* **2007**, *512*, 1–5.
35. EFSA GMO Panel. Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the notification (Reference C/GB/02/M3/3) for the placing on the market of glyphosate-tolerant and insect-resistant genetically modified maize NK603 × MON810, for import and processing, under Part C of Directive 2001/18/EC from Monsanto. *EFSA J.* **2005**, *308*, 1–22.
36. EFSA GMO Panel. Scientific Opinion on application EFSA-GMO-NL-2011–97 for the placing on the market of insect-resistant and herbicide-tolerant genetically modified cotton T304–40 for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Bayer CropScience AG. *EFSA J.* **2013**, *11*, 3251.
37. Aguilera, M.; Querci, M.; Balla, B.; Prospero, A.; Ermolli, M.; van den Eede, G. A qualitative approach for the assessment of the genetic stability of the MON810 trait in commercial seed maize varieties. *Food Anal. Methods* **2008**, *1*, 252–258.
38. ISAAA. Available online: http://www.isaaa.org/gmapprovaldatabase (accessed on 1 July 2014).
39. Thelwell, N.; Millington, S.; Solinas, A.; Booth, J.; Brown, T. Mode of action and application of Scorpion primers to mutataon detection. *Nucleic Acids Res.* **2000**, *28*, 3752–3761.
40. Druml, B.; Cichna-Markl, M. High resolution melting (HRM) analysis of DNA—Its role and potential in food analysis. *Food Chem.* **2014**, *158*, 245–254.
41. Montgomery, J.; Wittwer, C.T.; Palais, R.; Zhou, L. Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nat. Protoc.* **2007**, *2*, 59–66.
42. Foissac, X. Taq Polymerase Errors in PCR: Frequency and Management. COST School on bioinformatical analyses of phytoplasma sequences: Vilnius, Lithuania, 2011; Available online: http://costphytoplasma.ipwgnet.org/PDF%20files/Bioinformatics%20school/Foissac%202%20Taq%20errors-r.pdf (accessed on 1 July 2014).
43. 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 Res.* **2008**, *17*, 393–402.

44. 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. *J. Health Sci.* **2005**, *51*, 197–201.
45. Bush, S.M.; Krysan, P.J. iTILLING: A personalized approach to the identification of induced mutations in Arabidopsis. *Plant Physiol.* **2010**, *154*, 25–35.
46. Type-it HRM PCR Handbook. Available online: http://www.qiagen.com/resources/resourcedetail?id=a800ede3–2322–4341-a5c1-a9cbb6f630a3&lang=en (accessed on 14 July 2014).
47. Platt, A.R.; Woodhall, R.W.; George, A.L., Jr. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques* **2007**, *43*, 58–62.
48. Jiao, Y.; Zhao, H.; Ren, L.; Song, W.; Zeng, B.; Guo, J.; Wang, B.; Liu, Z.; Chen, J.; Li, W.; *et al*. Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **2012**, *44*, 812–815.
49. Hernandez, M.; Pla, M.; Esteve, T.; Prat, S.; Puigdomenech, 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 Res.* **2003**, *12*, 179–189.
50. Mumm, R.H. A look at product development with genetically modified crops: Examples from maize. *J. Agric. Food Chem.* **2013**, *61*, 8254–8259.
51. Kohli, A.; Griffiths, S.; Palacios, N.; Twyman, R.M.; Vain, P.; Laurie, D.A.; Christou, P. Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. *Plant. J.* **1999**, *17*, 591–601.

Eur Food Res Technol
DOI 10.1007/s00217-016-2749-2

ORIGINAL PAPER

Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing

Magali Castan[1,2] · Sina-Elisabeth Ben Ali[1,2] · Rupert Hochegger[1] · Werner Ruppitsch[1] · Alexander G. Haslberger[2] · Christian Brandes[1]

Received: 20 February 2016 / Revised: 28 June 2016 / Accepted: 16 July 2016

Abstract The examination of transgenic loci is an integral part of biosafety legislation in the European Union (EU). The EU directive 2001/18/EC requires any genetically modified (GM) event to be stable. Mutations in the nucleotide sequence of GM events must be avoided in the production and use of seeds. In the present work, an F1 and an F2 generation of the corn event NK603 were studied in stacked varieties (NK603 × MON810). The central aspect of this work was to utilize high-resolution melting analysis, real-time PCR, and Sanger sequencing for the examination of genetic stability of the entire construct of NK603 as well as in the regions flanking NK603. To perform such screening, it was necessary to develop specific PCR primers for the NK603 insert. Twenty-five specific primer pairs and PCR reactions were used to screen a total number of 340 samples. In addition to the screening, the NK603 zygosity was determined by a PCR-based testing method. Differences to the published patent sequence occurring in all samples were detected in two locations of the transgenic DNA sequence. These differences were also found in certified NK603 reference material.

Keywords Genetic stability · GMO · NK603 · High-resolution melting · Stacked event · Roundup Ready

Electronic supplementary material The online version of this article (doi:10.1007/s00217-016-2749-2) contains supplementary material, which is available to authorized users.

✉ Christian Brandes
christian.brandes@ages.at

Magali Castan
magali.castan@ages.at

Sina-Elisabeth Ben Ali
sina-elisabeth.ben-ali@ages.at

Rupert Hochegger
rupert.hochegger@ages.at

Werner Ruppitsch
werner.ruppitsch@ages.at

Alexander G. Haslberger
alexander.haslberger@univie.ac.at

[1] Austrian Agency for Health and Food Safety, Spargelfeldstrasse 191, 1220 Vienna, Austria

[2] Department of Nutritional Sciences, University of Vienna, Althanstraße 14, 1090 Vienna, Austria

Introduction

The transgenic event NK603 (Unique Identifier: MON-ØØ6Ø3) was produced by particle acceleration in maize embryos [1], which resulted in the random integration of recombinant DNA into the plant genome [2]. NK603 contains two (*Agrobacterium sp.* strain) CP4 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) expression cassettes, which confer plant tolerance to the herbicide glyphosate [1]. Glyphosate and glufosinate-tolerant varieties are the most common herbicide-tolerant (HT) crops in the world. HT varieties, which were first introduced in 1996 [3], now comprise the majority of cultivated genetically modified (GM) crops worldwide [4]. NK603 maize varieties are summarized under the term Roundup Ready (RR) plant cultivars, because their genetic modifications confer tolerance to the commercial herbicide Roundup [3].

Gene stacking or pyramiding is a breeding approach which is used by crop breeders to introduce multiple (trans) genes into an organism [5]. In the EU, stacked GM events are considered as new GMOs and require regulatory approval [6, 7]. Evaluation and authorization of stacked maize event NK603 × MON810 (Unique Identifier:

Published online: 06 August 2016

MON-ØØ6Ø3-6 × MON-ØØ81Ø-6), which we used for the analysis of the genetic stability of event NK603 in the present paper, took place under Regulation (EC) 1829/2003 on genetically modified food and feed. In 2005, the European Food Safety Authority (EFSA) expert panel on genetically modified organisms (GMOs) established the criteria for the scientific evaluation and assessment of the safety of NK603 × MON810 [8]. The safe use of such crops requires a thoroughly managed risk assessment [9] that includes molecular characterization, such as the detection of multiple-loci insertions, copy number of the transgene, rearrangements of the recombinant DNA, and plasmid backbone sequences [2, 10–12]. A further aspect of the molecular characterization is to demonstrate the genetic stability of GMO varieties in relation to the post-transformational characteristics of the insert. This means that the DNA sequence of the insert itself remains unchanged during cultivation and propagation of a GMO (after the transformation process). According to Annex III B of the Directive 2001/18/EC on the deliberate release of genetically modified organisms into the environment and the repealing Council Directive 90/220/EEC, the applicants must provide information on the genetic stability of the transgenic insert in their notifications. The Commission Implementation Regulation (EU) 503/2013 demands a separate assessment of the stability of the inserts in stacked GM events.

The insert of the NK603 transgene includes molecular rearrangements and a fragment of chloroplast DNA which, according to the EFSA GMO panel, pose no safety risk [13]. The rearrangements are likely to be caused by the transformation process and are also found in other widely used GMOs, such as MON810 maize and RR soybeans [14, 15]. Changes in the nucleotide sequence, which may cause genetic instability, are described in several publications [16–19]. The absence of post-transformational DNA changes in transgenic plants is commonly verified by Southern blot analyses and PCR methods [8, 13, 20, 21]. Segregation data, including data from Southern blot analyses, after several generations indicated that the NK603 insert is genetically stable [13, 20]. The same results were obtained by Southern blot analysis of the stacked event NK603 × MON810 [8]. Although the application of Southern blot analysis is common for the verification of the genetic stability in GMO notifications, it is not the method of choice for identifying small DNA alterations. A major disadvantage of Southern blotting is that it is only sensitive to alterations of a high number of nucleotides. However, the detection of single or several nucleotide polymorphisms is very important for the analysis of the phenotype of a plant. It is possible that even minor alterations of a sequence can have an impact on important characteristics. As confirmed by real-time PCR, three studies [18, 22, 23] showed that SNPs can cause significant errors in quantitative measurements of GM crops and therefore that SNPs can be a source of bias. Thus, it should be considered that Southern blot analysis alone is not sufficient to verify the genetic stability. Ben Ali et al. [24], Madi et al. [25], Neumann et al. [26], La Paz et al. [27], Papazova et al. [28], and Hayes et al. [2] have provided more accurate methods for investigating the genetic stability. Madi et al. [25] and Ben Ali et al. [24] were able to show that the combination of real-time PCR and HRM analysis [29, 30] is both a highly sensitive method and an excellent screening tool for mutation scanning in GM crops. Furthermore, additional Sanger sequencing for the verification of the real-time PCR results and the HRM analysis was used [24–26]. Hayes et al. [2] developed a Southern-by-Sequencing screening approach, which includes next-generation sequencing (NGS), for the molecular characterization of GM crops.

In recent studies, we have examined the genetic stability in single events of several transgenic maize [26], soybean, and canola [25] plants using real-time PCR and HRM screening, finding no differences in the nucleotide sequences of the analyzed GM varieties. However, we were able to detect sequence variations at the 3′ end of MON810 in a stacked variety of MON88017 × MON810 [24]. This difference between a single event and a stacked event may be explained by the fact that in stacked events, there are several DNA sequences in duplicate (promoter areas) that can influence event expression [31, 32]. According to [32], the 35S promoter which is used for NK603 can lead to genome instabilities. However, this hypothesis is highly controversial [33].

To verify this effect, the present study aimed to fill this gap and to provide information about SNPs and the genetic stability of a further stacked event containing NK603 × MON810 by using HRM and sequencing analyses. In our previous publications [24–26], we examined only the 3′ and 5′ junctions between the transgenic insert and the plant genome. The intention of the current study was to screen the entire construct of a GMO for genetic stability. In this approach, it was necessary to sequence the NK603 insert because only parts of the NK603 insert have been previously published. Using 25 different PCR reactions, we sequenced the NK603 event including the border regions and screened it using HRM analysis. We provide information on a large portion of the sequence of NK603 as well as primer pairs for the analysis and PCR amplification of NK603. This information is of great interest to other research groups: They can use this to improve the quantification and analysis of NK603. It is also possible to test the variability of the NK603 insert in other maize varieties to verify genetic stability. Overall, our knowledge about the genetic stability of GMOs is broadly improved with the publication of this work.

Materials and methods

Unless stated otherwise, the chemicals, PCR reagents, and PCR primers used were of analytical grade and were purchased from Biozym (Vienna, Austria) (PCR reagents and chemicals) and Sigma-Aldrich (Vienna, Austria) (PCR primers and chemicals). The Wizard DNA Clean-up System (Promega, Vienna, Austria), Phusion Hot Start II High-Fidelity DNA Polymerase Kit, and Type-it HRM PCR Kit (Qiagen, Hilden, Germany) were used according to the respective manufacturer's recommendations. Two corn varieties were analyzed: Progenies of hybrid DKC 26-79 and hybrid 631RR2/Bt. DKC 26-79 and 631RR2/Bt are stacked corn events (single-cross hybrids) with the Unique Identifier MON-ØØ6Ø3-6 × MON-ØØ81Ø-6. DKC 26-79 hybrid seeds (Monsanto) were obtained from the Canadian market, and in 2007, crops were grown on loamy soil in Nova Scotia, Canada. We bought the harvested corn and did not obtain details of how the crops were cultivated in detail. Only conventional agricultural practice including glyphosate treatment was specified. Harvested corn was shipped on treated pallets and served as investigation material. Hybrid 631RR2/Bt (Croplan Genetics®) was obtained from the US market (Indiana, 2005), and the hybrid seeds were used for our analyses. Genomic DNA was extracted from 140 individual corn grains of DKC 26-79 progenies (F2 generation) and from 200 individual corn grains of 631RR2/Bt (F1 generation; hybrid seeds). Furthermore, the certified NK603 reference material (GMO Standard ERM-BF415) was analyzed.

Extraction of genomic DNA

The corn grains were homogenized individually with a household garlic crusher. One hundred and fifty milligrams of the homogenized grains was incubated with a mixture of 820 μL of TNE buffer [10 mM Tris (pH 8), 150 mM NaCl, 2 mM EDTA (pH 8)], 1 % SDS, 150 μL of 5 M guanidine-HCl, and 30 μL of proteinase K (600 μg/mL) overnight at 60 °C under constant shaking. Next, the mixture was centrifuged for 5 min at 16,100 × g. Three hundred microliters of chloroform (99 %) was added to 600 μL of the supernatant, and the mixture was vortexed for 20 s. The mixture was then centrifuged at 16,100×*g* for 8 min to separate the phases, and 500 μL of the aqueous phase was transferred into a new tube. Four microliters of RNAse (32 μg/mL) was added to the aqueous phase, and the mixture was incubated at 60 °C for 30 min under constant shaking. The extracted DNA was purified using the Wizard DNA Clean-up System (Promega, Vienna, Austria). After a pre-elution with 20 μL of ddH2O, the DNA was eluted with 10 mM Tris buffer (pH 7.4, 70 °C) for 10 min. The eluted DNA was immediately stored at −20 °C.

DNA quality check

Agarose gel electrophoresis as well as spectrophotometry was used to check the DNA quality. DNA samples and Quantitas DNA Marker (Biozym) were loaded on a 1 % w/v agarose gel containing ethidium bromide and were then run for 30 min (140 V, 400 mA, 100 W). Afterward, the gel was viewed under UV light using a Bio-Rad Chemi XRS Gel Documentation system. Samples with clear and sharp bands in the upper gel region met the quality requirements. These samples were measured by a NanoPhotometer (Implen). The 260/280 ratio of absorbance was determined to assess the purity of the DNA samples. Samples with a ratio between 1.8 and 1.9 were employed for further investigations.

DNA quantitation

DNA concentrations were measured by Qubit® 2.0 Fluorometer (Life Technologies, Vienna, Austria) using a Qubit® dsDNA BR Assay Kit (Life Technologies, Vienna, Austria). The instrument uses fluorescent dyes that only bind to specific target molecules to determine the concentration of nucleic acids. Therefore, quantitation by the Qubit® 2.0 Fluorometer method is more sensitive than the widespread UV absorbance method [34].

PCR-based zygosity testing

NK603 zygosity was determined using a PCR-based testing method developed by Nan and Huabang [35]. Three primers (5′GP, 3′GP and 5′TP) were used to identify the degree of NK603 zygosity in the corn samples. The primer pair 5′GP and 3′GP were used to detect the wild type (NK603 transgene not present). 5′GP and 5′TP were used to identify the 5′ border region of NK603 containing genomic and transgenic DNA (in a hemizygous or homozygous state). Corn samples that were homozygous for NK603 were positive for the NK603 5′GP and 5′TP fragment but negative for wild type, whereas samples without NK603 were only positive for the wild type 5′GP and 3′GP fragment. Samples that were hemizygous for NK603 were positive for the wild type 5′GP and 3′GP fragment as well as for the NK603 5′GP and 5′TP fragment.

The PCR was performed using genomic DNA as a template. A total 20 μL of the reaction mixture contained 4 μL of 5X Colorless GoTaq® Reaction Buffer (1.5 mM MgCl2), 250 nM of each primer, 250 μM of each deoxynucleoside triphosphate (dNTP), 1.5 U GoTaq® G2 DNA Polymerase, and 50 ng of DNA. A Mastercycler® (Eppendorf) was used for this analysis. The temperature profile is shown in Table 1.

Table 1 Thermal cycling profile of the PCR zygosity experiment

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	94	10 min	
Denaturation	94	30 s	40
Annealing	58	30 s	
Extension	72	30 s	
Final extension	72	1 min	

Table 2 Thermal cycling profile of the real-time PCR (prior to HRM analysis) experiment

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	5 min	
Denaturation	95	10 s	55
Annealing	55	30 s	
Extension	72	20–30 s	

Overall, 30 samples of 631 RR2/Bt (F1) and 50 samples of DKC 26-79 progenies (F2) were investigated to identify the NK603 insert zygosity. A single event of MON810 was used as negative control (showed no bands). And during the investigation of DKC 26-79 progeny samples, 631 RR2/Bt was employed as positive control (two bands, one for NK603 and one for wild type).

Screening by real-time PCR using HRM analysis

HRM analysis is a cost-effective, time-saving and powerful method for SNP scanning and genotyping even in small laboratories [29, 30, 36, 37]. In the HRM analysis, the high-resolution melting point (Tm) of the amplicon produced was determined after the PCR reaction. This was achieved by gradually increasing the temperature from the annealing temperature up to 95 °C. Simultaneously, the fluorescence of a DNA intercalating saturation dye was measured. The Tm depended on the length and the guanine-cytosine (GC) content of the DNA template. Because high-precision temperature control and fluorescence measurement were required for the successful implementation of this method, these measurements could only be made using HRM-certified equipment [30]. For HRM analysis, it is necessary to use a specific cycler optimized for precise melting analysis that uses dedicated heating algorithms and specific software [38, 39]. A Rotor-Gene Q (Qiagen, Hilden, Germany) instrument was used for the HRM experiments. HRM was performed with the temperature increasing at a rate of 0.2 °C per 4 s. The initial and final temperatures were 70 and 95 °C, respectively. The melting curves and the temperature-shifted curves were normalized to enable sample comparisons. Modified curves and HRM scores (confidence values) were obtained using the Rotor-Gene Q series software (version 2.0.2.4, Qiagen). The normalized and temperature-shifted melting curves corresponded to the final curve after normalization. One sample (per run) was selected as reference genotype (confidence value of 100 %) and served as a reference for the difference plot. The software compared then melt curves of the other samples to the melt curve of the reference and generated confidence intervals relative to the selected genotype [40]. Amplicons with sequence variations have different shapes and therefore lower confidence values. This enables the identification of samples harboring sequence variations.

The total volume of the reaction mixture was 16 µL. The Type-it HRM PCR Kit (Qiagen, Hilden, Germany), 700 nM primers, and 1.6 µL of undiluted DNA (40 ng/µL) template were used for each reaction. The thermal cycling profile used is shown in Table 2.

A maximum of 70 DNA samples were analyzed per run (one replicate per run). For sections one and 23, every run was repeated on a different day. The mean of the values was calculated and was used for further investigation. A total 140 individual corn samples of DKC 26-79 (F2) and 200 individual samples of 631RR2/Bt (F1) were screened separately for sections one and 23 (see Table 6). For sections two to 25 (without six and 23, see Table 6), 20 individual corn samples of DKC 26-79 (F2) were screened (one replicate per run). The number of selected samples (for sequencing) varied depending on the confidence values. Our experience with HRM analysis has shown that it is useful to evaluate each run separately and sets a confidence threshold depending on the number of samples with low confidence values. The number of samples selected (for sequencing) per run as well as every individual confidence threshold is stated in the supplementary material (Tables S1–S26).

Sequencing

Sequencing was performed for the verification of the HRM analysis. As mentioned above, only the samples with the lowest confidence values (which indicate mutations) were sequenced to reveal possible mutations in the DNA sequence. All sequence analyses were performed using the Sanger sequencing method. The BigDye Fast cycle sequencing protocol [41] was used for this purpose. To avoid sequencing errors, both forward and reverse sequencing were included in this method to insure the reliability of the sequencing results. Prior to sequencing, all samples were prepared by an enzymatic PCR product cleanup, a cycle sequencing reaction as well as a formamide and filter step for purification. The PCR product cleanup mix was prepared as follows.

For purification, 2 µL of the mix (Table 3) were added to 5 µL of each sample in a 96-well microplate. The reaction mixture was incubated for 15 min at 37 °C followed by 80 °C for 15 min. For the cycle sequencing reaction, two master mixes (MM) were prepared as shown in Table 4.

For each sample, 8 µL of the MM forward primer and 8 µL of the MM reverse primer, each together with 2 µL of the purified sample, were added into two different 96-well microplate. Subsequently, the sequencing PCR was performed with an initial 60-s incubation step at 96 °C; followed by 25 cycles of 96 °C for 10 s; 50 °C for 5 s; and 60 °C for 75 s using a Mastercycler® (Eppendorf). The products were purified by the Performa® V3 96-Well Short Plate Kit (Edge Bio). Then, 5 µL formamide was added to each sample. Finally, the samples were transferred into the Sequencer 3500 Dx Genetic Analyzer (Applied Biosystems).

Table 3 Reaction mixture for PCR product cleanup (preparatory step of sequencing)

Reagents	Volume per sample (µL)
ddH_2O	1.55
Buffer for exonuclease I	0.1
Buffer for Antarctic phosphatase	0.1
Alkaline phosphatase	0.2
Exonuclease I (*E.Coli*)	0.05

Table 4 Reaction mixture for sequencing PCR

Reagents	MM forward primer per sample (µL)	MM reverse primer per sample (µL)
ddH_2O	4.5	4.5
BigDye® buffer	2	2
Primer forward (5 µM)	1	–
Primer reverse (5 µM)	–	1
BigDye®(contains dNTPs and ddNTPs)	0.5	0.5

Primers

The primers used in this work are shown in Tables 5 (primers for qualitative PCR) and 6 (primers for real-time PCR and qualitative PCR). Primers were designed using the specified references of Tables 5 and 6. The free software Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was employed for primer design.

Results

We performed a step-by-step analysis of the variability of the NK603 transgene and neighboring regions using HRM analysis and Sanger sequencing. The HRM analysis enables the identification of samples with DNA sequence variations. Different melt curves (compared to a reference sample) obtained from the HRM analysis suggest genetic divergences in the target region. Twenty-four newly designed primer pairs and one additional primer pair, which are used for the official detection of NK603, were employed for this purpose (see Table 6). The transgene (including the flanking regions) was divided into 25 sections which were investigated separately using the above-mentioned primer pairs (see Table 6). Eight of the sections appear twice in the transgene because CTP2, CP4 EPSPS, and T-NOS are repeated in NK603 (see Fig. 1).

Twenty single-grain DNA samples of corn hybrid DKC 26-79 progenies were examined for screening sections two to 25 (except six and 23; see Fig. 1 and Table 6). One hundred and forty single-grain DNA samples of DKC 26-79 progenies and 200 single-grain DNA samples of corn 631RR2/Bt hybrids were analyzed for the remaining two sections one and 23 (the border regions between the exogenous transgene and the endogenous maize genome; also known as "event-specific" regions). These regions are used for the official GMO control, and unidentified mutations in these regions may lead to unreliable test results [18, 22, 23]. Therefore, higher numbers of samples were studied for these sections. Samples with the lowest confidence values along with the reference sample (with the highest confidence value) were

Table 5 Primers for qualitative PCR

Primer	Sequence	Usage	References
5`GP	5′-GTCAAAGGATGCGGAACTGTT-3′	Zygosity testing	[35]
3`GP	5′-AAAGAACAAGTTGGATGCCGC-3′	Zygosity testing	[35]
5`TP	5′-GAGTAAGCTTGTTAACGCGG-3′	Zygosity testing	[35]
5′ flank f	5-TGGCATTTCCAACCCTAGAGAC-3`	Characterization of the 5′ flanking region	[1]
NK603 6 f	5′-CCTCCGCTTCCAAAGAAACGC-3′	Characterization of screening "section six"	
NK603 6 r	5′-CCTCTTCCTGATCTTGTTTAG-3′	Characterization of screening "section six"	ACCESSION EU155408

Table 6 Primers for real-time PCR with subsequent HRM analysis and for qualitative PCR

Primer	Sequence	Screening section	Target region	Source
NK603 1 f	5'-AGAGCCTCACGTTTCCAGGG-3'	1	5´ genomic flank → NK603 insert	[1]
NK603 1 r	5'-GCCGCCCTAGGGATATCAAG-3'			
NK603 2 f	5'-TATGCTTGAGAAGAGAGTCGGG-3'	2	P-ract1/I-ract1	
NK603 2 r	5'-TTGGGCCACCTTTTATTACCG-3'			
NK603 3 f	5'-TCGGTAATAAAAGGTGGCCCAA-3'	3		
NK603 3+4 r	5'-AGCACTTTGGGCTTTAGGAACT-3'			ACCESSION EU155408
NK603 4 f	5'-TAAAATAGCTTTCCCCCGTTGC-3'	4		
NK603 3+4 r	5'-AGCACTTTGGGCTTTAGGAACT-3'			
NK603 5 f	5'-CGTTGCAGCGCATGGGTATT-3'	5		
NK603 5 r	5'-GCGTTTCTTTGGAAGCGGAG-3'			
NK603 7 f	5'-GAATGGGGCTCTCGGATGTAGA-3'	7	P-ract1/I-ract1 → CTP2	ACCESSION EU155408
NK603 7 r	5'-TTCTGCACACCATTGCAGATTC-3'			ACCESSION JN400386
NK603 8 f	5'-TGACAAATGCAGCCTCGTGC-3'	8	P-ract1/I-ract1 → CP4 EPSPS	ACCESSION EU155408
NK603 8 r	5'-TGGGAGATCGACTTGTCGCC-3'			ACCESSION AY125353
NK603 9 f	5'-CGTCGTCGTGGGGATTGAAG-3'	9	CTP2 → CP4 EPSPS	ACCESSION JN400386
NK603 9 r	5'-GGCATTGCCGAAATCGAGCG-3'			ACCESSION AY125353
NK603 10 f	5'-AGGCGACACCTGGATCATCG-3'	10	CP4 EPSPS	
NK603 10 r	5'-CTCGATGACCGTCGTGATGC-3'			
NK603 11 f	5'-TCTACGATTTCGACAGCACCT-3'	11		
NK603 11 r	5'-CAGGCGGATGGTGCGCACGC-3'			
NK603 12 f	5'-CTCCGCACAGGTGAAGTCC-3'	12		
NK603 12 r	5'-GCGCGGGTTGATGACTTCG-3'			
NK603 13 f	5'-CGTCGAGACGGATGCGGACGGCG-3'	13		
NK603 13 r	5'-CGGTCGCCCCTTCCGCGAAGGCG-3'			
NK603 14 f	5'-ATATCCGATTCTCGCTGTCGCC-3'	14		
NK603 14 r	5'-GAGAGTTCGATCTTCGCGCC-3'			
NK603 15 f	5'-TCGCCACCCATCTCGATCAC-3'	15	CP4 EPSPS → T-NOS	
NK603 15 r	5'-TAATCATCGCAAGACCGGCA-3'			
NK603 16 f	5'-GTTGCCGGTCTTGCGATGAT-3'	16	T-NOS → P-e35S	
NK603 16 r	5'-GTCTCAATCGGACCATCACATC-3'			ACCESSION

Table 6 continued

Primer	Sequence	Screening section	Target region	Source
NK603 17 f	5'-AAGTGGATTGATGTGATGGTCCG-3'	17	P-e35S → I-*Zm*hsp70	KJ608140
NK603 17 r	5'-AGGCAGAGGGCGGAGTGAGCGCG-3'			[42]
NK603 18 f	5'-ACGCGCTCACTCCGCCCTCTGCC-3'	18	I-*Zm*hsp70	
NK603 18 r	5'-AATAAGCTCTGCAGACGAACAA-3'			
NK603 19 f	5'-TAATTTGTTCGTCTGCAGAGCTT-3'	19		
NK603 19 r	5'-AGAAGGCATCGAGCAAGATACG-3'			
NK603 20 f	5'-GAGTTTCCTTTTTGTTGCTCTC-3'	20	I-*Zm*hsp70 → CP4 EPSPS	
NK603 20 r	5'-GCTGCTTGCACCGTGAAG-3'			ACCESSION AY125353
NK603 21 f	5'-ACGAGCTTCCCGGAGTTCA-3'	21	CP4 EPSPS → T-NOS/partial P-Ract1	
NK603 21 r	5'-AAGCTTGGTACCGAATTCCCCG-3'			[1]
NK603 22 f	5'-AAATTATCGCGCGCGGTGTC-3'	22	T-NOS → 3′ genomic flank	
NK603 22 r	5'-CACTAGAGTGGAAGTGTGTCGC-3'			
NK603-F	5'-ATGAATGACCTCGAGTAAGCTTGTTAA-3'	23	Partial P-Ract1 → 3′ genomic flank	[1,43]
NK603-R	5'-AAGAGATAACAGGATCCACTCAAACACT-3'			
NK603 24 f	5'-ACACACTTCCACTCTAGTGTTTGAGTGG-3'	24	3′ genomic flank	[1]
NK603 24 r	5'-AAGTGGTGTACGGTTAAGTTGTATACG-3'			
NK603 25 f	5'-TTAGCAATGGCTCGTAATGCGGC-3'	25		
NK603 25 r	5'-AACCCCATCTTCGGCGTCGCTCCG-3'			

Screening sections one and 23 represent the 3′ and 5′ junctions between the transgenic insert and the plant genome. Sections two to 22 (without section six) belong to the NK603 insert. Screening sections 24 and 25 are beyond the transgene. These sections are located in the 3′ flanking region of the transgene

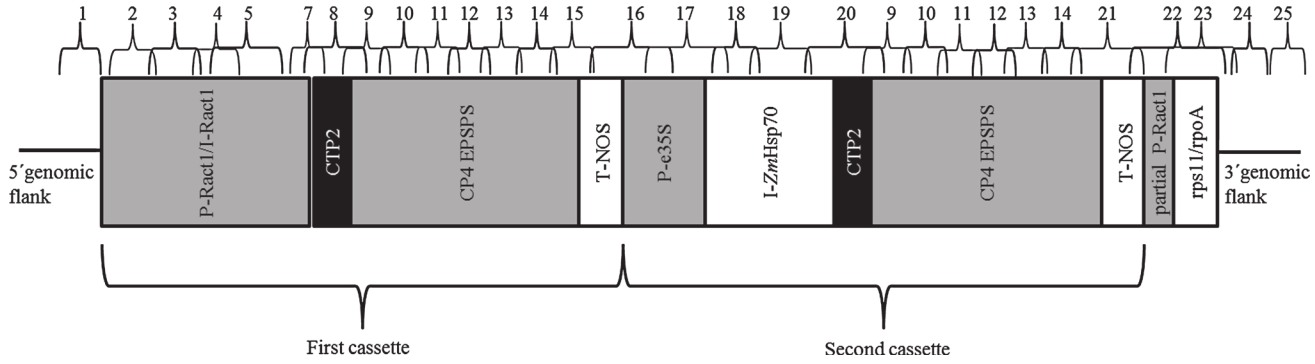


Fig. 1 Illustration of the transgenic locus and the sequencing sections. The main elements of NK603 are shown. NK603 contains two cassettes with similar structures. In the second cassette, the rice actin promoter and intron were replaced with the cauliflower mosaic virus promoter and hsp70 intron. The sequencing sections, which were used for the HRM screenings, are numbered (*1–5* and *7–25*). Their locations are depicted by clips. Figure modified after [44]

sequenced (null to eleven samples per screening run). For the non-border sections (except section six), one or two samples with the lowest confidence values in the HRM analysis were subjected to Sanger sequencing along with the reference sample. For section six, different primer pairs were tested, but none were suitable for HRM analysis. Therefore, twenty samples were subjected to Sanger sequencing (employing primers NK603 6 f and r, see Table 5) to obtain the sequence

information, as sequencing confirms sequence alterations and enables their characterization.

NK603 zygosity

As claimed by Monsanto, segregation of NK603 × MON810 progeny is consistent with Mendelian genetics [8]. Therefore, we did not test for independent segregation. Instead, NK603 zygosity was determined using a PCR-based testing method developed by Nan and Huabang [35]. Overall, 30 samples of 631 RR2/Bt (F1) and 50 samples of DKC 26-79 progenies (F2) were investigated to identify the NK603 insert zygosity. A single event of MON810 was used as negative control (showed no bands). And during the investigation of DKC 26-79 progeny samples, 631 RR2/Bt was employed as positive control (two bands, one for NK603 and one for wild type). In general, we postulate that the F1 seeds from the 631RR2/Bt hybrids are hemizygous for NK603. From the 30 seeds of 631RR2/Bt hybrids that we analyzed for NK603 zygosity, all samples were positive for both, wild type and NK603. This result is consistent with a hemizygosity state of 100 % of the 631RR2/Bt seeds. The results of the 50 samples of DKC 26-79 progenies were different. If it is assumed that the hemizygous DKC 26-79 individuals are allowed to self-pollinate, they are expected to segregate in a 3:1 ratio for herbicide-resistant and herbicide-sensitive progeny [45]. Assuming that the F1 parental generation of DKC 26-79 is hemizygous for the transgene, one would expect that ~25 % of the progenies (F2 generation) are homozygous for NK603, ~50 % are hemizygous for NK603, and ~25 % are null (no NK603 transgene). However, when 50 seeds of DKC 26-79 progenies were tested, no samples lacking the NK603 transgene were identified. Two samples were hemizygous, and 48 DNA samples were homozygous for NK603.

Analysis of the NK603 border regions

The aim of the HRM analysis screening was to identify possible mutations within the NK603 transgene. Montgomery et al. [29] stated that HRM analysis can be used to screen for mutations in amplicons of 1000 bp. However, for highly accurate mutation screening, the amplicon should be less than 400 bp. Though not always possible, most of the samples in the present study met the limit of 400 bp. One hundred and forty high-quality DNA samples from DKC 26-79 progenies and 200 DNA samples from 631RR2/Bt hybrids (from single corn grains) were used for the screenings of the border regions by analyzing sections one and 23 (screening sections are depicted in Table 6 and Fig. 1). DNA samples were diluted to 40 ng/µL (based on the DNA concentrations obtained by Qubit® 2.0 Fluorometer) and analyzed separately using real-time PCR and HRM analysis for screening sections one and 23. The screened fragments were 294 bp (screening section one) and 108 bp (screening section 23) in size. Both fragments included a transition from the plant genome to the transgene construct. Section one contained the 5′ NK603 transition and section 23 contained the 3′ NK603 transition.

Table 7 Selected samples of the first screening section for DKC 26-79 (F2) samples

DKC 26-79 (F2) samples	Mean Ct value	Mean confidence value (%)
105	19.3	100.0 (reference)
96	19.6	100.0 (reference)
166	29.7	31.9
170	18.0	31.0
20	30.5	30
110	19.5	17.4
100	19.2	6.0
75	25.9	3.8
22	28.1	3.4
34	20.9	3.0
59	20.6	2.9
161	32.2	2.3
98	25.1	1.8
183	21.0	1.5
83	27.8	1.5
72	20.8	1.3
112	19.3	0.3
38	20.8	0.0
97	27.5	0.0

The samples are ordered by decreasing confidence value

The Ct values of the real-time PCR of the first screening section for the DKC 26-79 (F2) samples ranged from 17.7 to 32.2 with an average Ct value of 20.4 and a standard deviation of 2.7 (CV = 13.2 %). The samples with the lowest confidence values as well as the reference samples were selected for Sanger sequencing. Table S1 (supplementary material) shows the Ct and confidence values of all DKC 26-79 (F2) samples in the first screening section. The selected samples are depicted in Table 7.

In the real-time PCR of the first screening section of the 631RR2/Bt (F1) samples, the Ct values ranged from 17.4 to 20.1, with an average Ct value of 18.2 and a standard deviation of 0.5 (CV = 2.8 %). The samples with the lowest confidence values as well as the reference samples were selected for Sanger sequencing. Table S2 (supplementary material) shows the Ct and confidence values of all 631RR2/Bt (F1) samples used in the first screening section. The selected samples are depicted in Table 8.

In the real-time PCR of screening section 23 for the DKC 26-79 (F2) samples, the Ct values ranged from 20.8 to 32.0 with an average Ct value of 22.6 and a standard deviation of 2.2 (CV = 9.7 %). The samples with the

Table 8 Selected samples of the first screening section for 631RR2/Bt (F1) samples

631RR2/Bt (F1) samples	Mean Ct value	Mean confidence value (%)
67	18.2	100.0 (reference)
107	17.8	100.0 (reference)
214	18.2	100.0 (reference)
26	18.1	75.6
205	18.4	72.9
42	18.2	72.5
56	17.9	71.9
115	17.7	42.9
117	17.6	38.7
82	19.1	35.7

The samples are ordered by decreasing confidence value

Table 9 Selected samples of screening section 23 for DKC 26-79 (F2) samples

DKC 26-79 (F2) samples	Mean Ct value	Mean confidence value (%)
96	21.1	100.0 (reference)
93	21.3	100.0 (reference)
112	22.4	22.4
110	22.0	14.6
182	22.5	11.6
100	21.0	9.6
180	21.1	9.2
72	22.9	2.5
183	22.5	0.3
38	22.4	0.0

The samples are ordered by decreasing confidence value

lowest confidence values as well as the reference samples were selected for Sanger sequencing. Table S3 (supplementary material) shows the Ct and confidence values of all DKC 26-79 (F2) samples in screening section 23. The selected samples are depicted in Table 9.

In the real-time PCR of screening section 23 for 631RR2/Bt (F1) samples, the Ct values ranged from 18.1 to 21.4, with an average Ct value of 19.5 and a standard deviation of 0.8 (CV = 4.1 %). The samples with the lowest confidence values as well as the reference samples were selected for Sanger sequencing. Table S4 (supplementary material) shows the Ct and confidence values of all 631RR2/Bt (F1) samples in screening section 23. The selected samples are depicted in Table 10.

The result of sequencing all selected samples of the NK603 border regions was that none of the samples were different from the DNA sequence of the NK603 patent [1].

Table 10 Selected samples of screening section 23 for 631RR2/Bt (F1) samples. The samples are ordered by decreasing confidence value

631RR2/Bt (F1) samples	Mean Ct value	Mean confidence value (%)
214	18.7	100.0 (reference)
108	18.9	100.0 (reference)
61	20.4	100.0 (reference)
16	20.5	66.2
25	20.2	61.2
26	20.4	49.3
82	20.0	46.3

Analysis of NK603 non-border sections 2–25 (except 23)

Different fragments of the NK603 transgene (see Table 6 and Fig. 1) with lengths ranging from 108 to 499 bp were scanned using HRM analysis after real-time PCR (see "Materials and Methods" section). Twenty DKC 26-79 (F2) samples were analyzed for sections two to 25 (without sections six and 23). Table 11 gives an overview of the samples chosen for sequencing. The selection was based on the lowest confidence value of each screening region. In addition, the reference sample of each screening region was sequenced.

HRM analysis of section six was not possible because no single peaks were present in the melting curves. However, as already mentioned, the primer pair of section six (see Table 5) was used to produce PCR products from 20 samples (DKC 26-79 progenies), which were then sequenced without selection. As depicted in Table S27 of the supplementary material, the sequencing information of section six revealed no mutations. However, sequencing of the samples in section two revealed SNPs in SEQ ID 7 of the NK603 maize patent [1]. Two nucleotide insertions were identified: a thymine between position 425 and 426 of SEQ ID 7 and a guanine between position 455 and 456 of SEQ ID 7. These positions are located in the rice actin 1 promoter [1]. As the rice actin 1 promoter occurs twice in NK603 (at least partially, see Fig. 1), it is possible that the primer pair of section two binds at both sides. For this reason, a primer pair specific to the 5′ flanking region of NK603 was employed. The forward primer (5′ flank f, see Table 5) binds at the genomic site, whereas NK603 3+4 r (which was used as a reverse primer) binds at P-ract1/I-ract1. The PCR products of eight samples (DKC 26-79 progenies) were produced (using 5`flank f and NK603 3+4 r) and sequenced to clarify whether the identified insertions belonged to the 5′ flanking region. The result of this sequencing confirmed the two nucleotide insertions described above and, thus, is markedly different to SEQ ID 7 of the NK603

Table 11 Selected samples of screening sections two to 25 (without sections six and 23) for DKC 26-79 (F2) samples

Screening section	Selected sample(s)	Confidence value(s) (%)
2	103	83.6
3	103 and 139	49.1 and 61.7
4	98	88.8
5	139	89.4
7	103	77.5
8	98	74.6
9	100	73.5
10	142	55.0
11	141	81.4
12	139	79.8
13	56	62.1
14	141	86.5
15	100	69.4
16	130 and 63	3.7 and 58.5
17	99 and 100	50.8 and 27.7
18	137	94.7
19	98	69.8
20	141	49.1
21	139	79.6
22	100	94.5
24	61	86.5
25	103	84.8

maize patent [1]. In addition, certified reference material of NK603 maize (eight samples) as well as eight samples of 631RR2/Bt maize was sequenced employing 5′ flank f and NK603 3+4 r. All samples, even the reference material showed the above-mentioned nucleotide insertions. The other 22 screening regions showed no mutations. All sequencing results can be found in Table S27 of the supplementary material.

Discussion and conclusion

The aim of this study was to sequence the NK603 construct as part of the stacked event NK603 × MON810. Thus, it was possible to examine the total NK603 construct for genetic stability and not just the 5′ and 3′ border regions between the GMO insert and the genomic sequence, which was undertaken in our previous publications [24–26]. By comparing the sequence of NK603 to the published patent sequence [1], two nucleotide insertions could be consistently detected in all sequenced samples of NK603 × MON810 (varieties 631RR2/Bt and DKC 26-79 progeny) as well as in the certified reference material of NK603. It is possible that these nucleotide insertions demonstrate a sequencing error in SEQ ID 7 of the published NK603 patent. This is supported by the fact that the nucleotide insertions were identified in all sequenced NK603 samples, including the reference material for NK603. Another explanation for the different sequence found in SEQ ID 7 of the published NK603 patent may be that the insertions were introduced during the breeding process for commercial lines (after the verification of the inserted sequence) [18]. The average single-base substitution rate for maize is estimated to be 1.63×10^{-8} per site per year for the whole genome as well as for genic maize regions [46]. To calculate a mutation rate based on the identified insertions, it would be necessary to know if the data are due to a sequencing error of the company or were introduced during crop breeding. As the nucleotide insertions are located in a promoter region, it can be assumed that there is no impact on the protein sequence of NK603. However, an influence on the promoter activity and thus on the expression rate of the transgene cannot be excluded. Since we have not found samples without the two described nucleotide insertions, it was not possible to compare gene expression between samples with and without nucleotide insertions.

Another important finding of the current study is the NK603 zygosity frequency imbalance of DKC 26-79 progenies (F2 seeds). In contrary to Mendel's law of independent assortment, no 3:1 ratio between dominant and recessive phenotypes could be detected. No sample without NK603 transgene could be identified. Similar results are reported in US patent 6,057,496 [47]. The result is explained by a preferential selection for transgene-positive gametes in HT plants after treatment with the appropriate herbicide. In the patent, it is described that a selective inhibition of phytotoxin-sensitive plant ovules, embryos, and pollen occurs in HT plants sprayed with herbicide. However, in the stated patent, only the herbicide-sensitive and herbicide-resistant progenies differed. The observed ratio between homozygous and hemizygous progenies is not mentioned. In our investigation, this ratio is imbalanced (2:48 between hemizygous and homozygous samples). One possible explanation for the high frequency of homozygotes would also be a selection pressure. If the appropriate herbicide is applied to hemizygous transgenic plants (F1 generation) during germ-cell development, it works as a gametocide and eliminates cells that lack resistance [47]. As observed in our case, the resulting progenies (F2 generation) are largely homozygous. Compared to the results of Lauer et al. [47], our level of homozygous progenies is much higher. After spraying (glyphosate treatment; 3.2 l/Ha @ V10 and V14 growth stage) and selfing plants hemizygous for the glyphosate resistance transgene, Lauer et al. identified 67.9 % of progeny were homozygous for the transgene. Our analysis revealed 96.0 % homozygous progeny. Probably, more pesticide was used in our case. Unfortunately, we have no information regarding the amount of

glyphosate treatment. However, knowledge of the degree of zygosity is crucial for adequate interpretation of genetic instabilities.

Another important aspect of the analysis of the genetic stability of transgenic loci is the GMO safety assessment in the EU, which requires the demonstration that each single GM event is genetically stable. In stacked GMOs, the structure and characteristics of the stacked events should be compared to the structures of single GM events [6]. The Commission Implementation Regulation (EU) 503/2013 also demands the separate assessment of the stability of the inserts in stacked GM events. However, Kok et al. [48] are opposed to an extra safety assessment of stacked GM events except in scientifically justified cases. They do not exclude instabilities in stacked events per se. It is also known that conventional breeding of plant varieties can lead to genetic changes. In our opinion, with a markedly different number of mutations in individual GM events compared to stacked GM events, experiments that are necessary for the GMO safety assessment of stacked events should be performed primarily using the DNA from stacked events. However, in the study presented no NK603 single variety was analyzed (except the NK603 reference material).

The present study is important for the discussion and interpretation of our previously published results, where we showed that genetic changes occur in commercial stacked GM events [24]. A heterozygous SNP was detected in the coding region of the Cry1Ab toxin of MON810 in stacked GM event MON88017 × MON810. An important difference to the current results is that the SNP of MON810 occurred only in a small proportion of samples, while the now-detected insertions occur in all analyzed samples. In another study, we examined the genetic stability of MON810 as single event [26], and we could not detect any differences even though the number of the samples was higher than in the examination of MON88017 × MON810. The examination of single events in soybeans and canola using real-time PCR and HRM screening also failed to show any SNPs [25]. This difference between the results of single events and those of stacked events may be explained by the high number of similar promoters in stacked events. However, it may also be a MON810-specific phenomenon. Conventional crossing of the stacked event can be eliminated as a cause because the mentioned SNP in the MON810 construct has been observed in another stacked variety (Ben Ali, unpublished). It does not seem to be a general phenomenon that the probability of genetic instabilities is higher in stacked events than that in single events. It would be necessary to analyze other stacked events containing SNPs and compare them to single events. Overall, there are very few examinations that have detected SNPs in commercialized GM events [18, 22–24]. Moreover, no studies have examined the frequency of genetic variations in large number of samples. In this study, the number of the screenings performed was rather low. Taken together, the findings of the current study confirm the importance of genetic stability investigations to exclude bias in GMO quantification and to further clarify whether stacked events need an extra safety assessment prior to authorization.

In future, it will be interesting to investigate the genetic stability of the whole genome of GM crops versus commercial crops without genetic modifications. Such an approach would be possible using next-generation sequencing (NGS) technology. Currently, this technology is still too expensive to use in analyzing genetic stability at a large scale, but we are convinced that this will change in the future.

Acknowledgments The material costs for this work were partially supported by the Hochschuljubiläumsstiftung (HJST) of the city of Vienna.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

Conflict of interest None.

References

1. Behr CF, Heck GR, Hironaka CH, You J (2012) Corn event PV-ZMGT32(nk603) and compositions and methods for detection thereof. US 8273959:B2
2. Zastrow-Hayes GM, Lin H, Sigmund AL, Hoffman JL, Alarcon CM, Hayes KR, Richmond TA, Jeddeloh JA, May GD, Beatty MK (2015) Southern-by-sequencing: a robust screening approach for molecular characterization of genetically modified crops. Plant Genome. doi:10.3835/plantgenome2014.08.0037
3. Cuhra M (2015) Review of GMO safety assessment studies: glyphosate residues in Roundup Ready crops is an ignored issue. Environ Sci Europe. doi:10.1186/s12302-015-0052-7
4. Bonny S (2011) Herbicide-tolerant transgenic soybean over 15 years of cultivation: pesticide use, weed resistance, and some economic issues. The case of the USA. Sustainability 3(12):1302–1322. doi:10.3390/su3091302
5. Taverniers I, Papazova N, Bertheau Y, De Loose M, Holst-Jensen A (2008) Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. Environ Biosafety Res 7(4):197–218. doi:10.1051/ebr:2008018
6. EFSA (2011) Scientific Opinion on Guidance for risk assessment of food and feed from genetically modified plants. EFSA J 9:2150
7. De Schrijver A, Devos Y, Van den Bulcke M, Cadot P, De Loose M, Reheul D, Sneyers M (2007) Risk assessment of GM stacked events obtained from crosses between GM events. Trends Food Sci Technol 18(2):101–109. doi:10.1016/j.tifs.2006.09.002
8. EFSA (2005) Opinion of the scientific panel on genetically modified organisms on a request from the commission related to the notification (Reference C/GB/02/M3/3) for the placing on the market of glyphosate-tolerant and insect-resistant genetically modified maize NK603 × MON810, for import and processing, under Part C of Directive 2001/18/EC from Monsanto. EFSA J 308:1–22

9. Agapito-Tenfen S, Vilperte V, Benevenuto R, Rover C, Traavik T, Nodari R (2014) Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biol 14(1):346
10. Latham JR, Wilson AK, Steinbrecher RA (2006) The mutational consequences of plant transformation. J Biomed Biotechnol 2:25376. doi:10.1155/JBB/2006/25376
11. Wilson AK, Latham JR, Steinbrecher RA (2006) Transformation-induced mutations in transgenic plants: analysis and biosafety implications. Biotechnol Genet Eng Rev 23:209–237
12. Smith N, Kilpatrick JB, Whitelam GC (2001) Superfluous transgene integration in plants. Crit Rev Plant Sci 20(3):215–249. doi:10.1080/20013591099218
13. EFSA (2003) Opinion of the scientific panel on genetically modified organisms on a request from the commission related to the safety of foods and food ingredients derived from herbicide-tolerant genetically modified maize NK603, for which a request for placing on the market was submitted under Article 4 of the Novel Food Regulation (EC) No 258/97 by Monsanto (QUESTION NO EFSA-Q-2003-002). EFSA J 9:1–14
14. Hernández M, Pla M, Esteve T, Prat S, Puigdomènech P, Ferrando A (2003) A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard® Based on the 3′-transgene integration sequence. Transgenic Res 12(2):179–189. doi:10.1023/A:1022979624333
15. Windels P, Taverniers I, Depicker A, Van Bockstaele E, De Loose M (2001) Characterisation of the Roundup Ready soybean insert. Eur Food Res Technol 213(2):107–112. doi:10.1007/s002170100336
16. Tizaoui K, Kchouk ME (2012) Genetic approaches for studying transgene inheritance and genetic recombination in three successive generations of transformed tobacco. Genet Mol Biol 35(3):640–649. doi:10.1590/s1415-47572012000400015
17. Choffnes DS, Philip R, Vodkin LO (2001) A transgenic locus in soybean exhibits a high level of recombination. Vitro Cell DevBiol Plant 37(6):756–762. doi:10.1007/s11627-001-0125-8
18. Morisset D, Demsar T, Gruden K, Vojvoda J, Stebih D, Zel J (2009) Detection of genetically modified organisms–closing the gaps. Nat Biotechnol 27(8):700–701. doi:10.1038/nbt0809-700
19. Zaulet M, Rusu L, Kevorkian S, Luca K, Mihacea S, Badea EM, Costache M (2009) Detection and quantification of GMO and sequencing of the DNA amplified products Rom. Biotechnol Lett 14(5):4733–4746
20. EFSA (2009) Applications (references EFSA-GMO-NL-2005-22, EFSA-GMO-RX-NK603) for the placing on the market of the genetically modified glyphosate tolerant maize NK603 for cultivation, food and feed uses, import and processing and for renewal of the authorisation of maize NK603 as existing products, both under Regulation (EC) No 1829/2003 from Monsanto. EFSA J 1137:1–50
21. EFSA (2009) Applications (EFSA-GMO-RX-MON810) for renewal of authorisation for the continued marketing of (1) existing food and food ingredients produced from genetically modified insect resistant maize MON810; (2) feed consisting of and/or containing maize MON810, including the use of seed for cultivation; and of (3) food and feed additives, and feed materials produced from maize MON810, all under Regulation (EC) No 1829/2003 from Monsanto. EFSA J 1149:1–85
22. Broothaerts W, Corbisier P, Schimmel H, Trapmann S, Vincent S, Emons H (2008) A single nucleotide polymorphism (SNP839) in the adh1 reference gene affects the quantitation of genetically modified maize (*Zea mays* L.). J Agric Food Chem 56(19):8825–8831. doi:10.1021/jf801636d
23. Gruden K, Allnutt TR, Ayadi M, Baeumler S, Bahrdt C, Berben G, Berdal KG, Bertheau Y, Bøydler Andersen C, Brodmann P, Buh Gašparič M, Burns MJ, Burrel AM, Cankar K, Esteve T, Holst-Jensen A, Kristoffersen AB, La Paz J, Lee D, Løvseth A, Macarthur R, Morisset D, Pla M, Rud RB, Skjœret C, Tengs T, Valdivia H, Wulff D, Zhang D, Žel J (2012) Reliability and cost of GMO detection. In: Genetically modified and non-genetically modified food supply chains: co-existence and traceability. Wiley, pp 307–332. doi:10.1002/9781118373781.ch18
24. Ben Ali SE, Madi ZE, Hochegger R, Quist D, Prewein B, Haslberger AG, Brandes C (2014) Mutation scanning in a single and a stacked genetically modified (GM) event by real-time PCR and high resolution melting (HRM) analysis. Int J Mol Sci 15(11):19898–19923. doi:10.3390/ijms151119898
25. Madi ZE, Brandes C, Neumann G, Quist D, Ruppitsch W, Hochegger R (2013) Evaluation of Adh1 alleles and transgenic soybean seeds using Scorpion PCR and HRM analysis. Eur Food Res Technol 237(2):125–135. doi:10.1007/s00217-013-1969-y
26. Neumann G, Brandes C, Joachimsthaler A, Hochegger R (2011) Assessment of the genetic stability of GMOs with a detailed examination of MON810 using Scorpion probes. Eur Food Res Technol 233(1):19–30. doi:10.1007/s00217-011-1487-8
27. La Paz JL, Pla M, Papazova N, Puigdomenech P, Vicient CM (2010) Stability of the MON 810 transgene in maize. Plant Mol Biol 74(6):563–571. doi:10.1007/s11103-010-9696-2
28. Papazova N, Ghedira R, Van Glabeke S, Bartegi A, Windels P, Taverniers I, Roldan-Ruiz I, Van Bockstaele E, Milcamps A, Van Den Eede G, Depicker A, De Loose M (2008) Stability of the T-DNA flanking regions in transgenic *Arabidopsis thaliana* plants under influence of abiotic stress and cultivation practices. Plant Cell Rep 27(4):749–757. doi:10.1007/s00299-007-0495-4
29. Montgomery J, Wittwer CT, Palais R, Zhou L (2007) Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. Nat Protoc 2(1):59–66. doi:10.1038/nprot.2007.10
30. Druml B, Cichna-Markl M (2014) High resolution melting (HRM) analysis of DNA—its role and potential in food analysis. Food Chem 158:245–254. doi:10.1016/j.foodchem.2014.02.111
31. Mumm RH (2013) A look at product development with genetically modified crops: examples from maize. J Agric Food Chem 61(35):8254–8259. doi:10.1021/jf400685y
32. Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P (1999) Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. Plant J Cell Mol Biol 17(6):591–601
33. Hull R, Covey SN, Dale P (2000) Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate. Microbial Ecol Health Dis 12(1):1–5. doi:10.1080/089106000435527
34. ThermoFisherScientific Qubit® 3.0 Fluorometer. https://www.thermofisher.com/at/en/home/life-science/laboratory-instruments/fluorometers/qubit/qubit-fluorometer.html. Accessed 25 Dec 2015
35. Nan L, Huabang C (2010) An accurate and rapid PCR-based zygosity testing method for genetically modified maize. GMO Biosaf Res 1(1):1–4. doi:10.5376/gmo.2010.01.0001
36. Garritano S, Gemignani F, Voegele C, Nguyen-Dumont T, Le Calvez-Kelm F, De Silva D, Lesueur F, Landi S, Tavtigian SV (2009) Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. BMC Genet 10:5. doi:10.1186/1471-2156-10-5
37. Martino A, Mancuso T, Rossi AM (2010) Application of high-resolution melting to large-scale, high-throughput SNP genotyping: a comparison with the TaqMan method. J Biomol Screen 15(6):623–629. doi:10.1177/1087057110365900
38. Bush SM, Krysan PJ (2010) iTILLING: a personalized approach to the identification of induced mutations in Arabidopsis. Plant Physiol 154(1):25–35. doi:10.1104/pp.110.159897

39. Qiagen EN-Type-it-HRM-PCR-Handbook. http://www.qiagen.com/resources/resourcedetail?id=a800ede3-2322-4341-a5c1-a9cbb6f630a3&lang=en. Accessed 25 Dec 2015
40. Qiagen Guidelines for Successful HRM. https://www.qiagen.com/de/resources/technologies/hrm/guidelines%20for%20successful%20hrm/. Accessed 22 Dec 2016
41. Platt AR, Woodhall RW, George AL Jr (2007) Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. Biotechniques 43(1):58–62. doi:10.2144/000112499
42. Brown SM, Santino CG (1995) Enhanced expression in plants. US 5859347 A
43. JRC (2005) Event-specific method for the quantitation of maize line NK603 using real-time PCR. http://bookshop.europa.eu/en/event-specific-method-for-the-quantitation-of-maize-line-nk603-using-real-time-pcr-pbLBNA21825/. Accessed 14 June 2016
44. Heck GR, Armstrong CL, Astwood JD, Behr CF, Bookout JT, Brown SM, Cavato TA, DeBoer DL, Deng MY, George C, Hillyard JR, Hironaka CM, Howe AR, Jakse EH, Ledesma BE, Lee TC, Lirette RP, Mangano ML, Mutz JN, Qi Y, Rodriguez RE, Sidhu SR, Silvanovich A, Stoecker MA, Yingling RA, You J (2005) Development and characterization of a CP4 EPSPS-based. Glyphosate Toler Corn Event Crop Sci 45(1):329–339. doi:10.2135/cropsci2005.0329
45. Conner AJ (1997) Production of true-breeding transgenic seeds from plants heterozygous for transgene insertions. US 6057496 A
46. Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44(7):812–815. doi:10.1038/ng.2312
47. Lauer MJ, Leafgren RL, Nelson SM (2009) Chemical selection of resistant gametes of plants in the field. US 20090275473:A1
48. Kok EJ, Pedersen J, Onori R, Sowa S, Schauzu M, De Schrijver A, Teeri TH (2014) Plants with stacked genetically modified events: to assess or not to assess? Trends Biotechnol 32(2):70–73. doi:10.1016/j.tibtech.2013.12.001

European Food Research and Technology
https://doi.org/10.1007/s00217-018-3062-z

ORIGINAL PAPER

Genetic and epigenetic characterization of the *cry1Ab* coding region and its 3′ flanking genomic region in MON810 maize using next-generation sequencing

Sina-Elisabeth Ben Ali[1,2] · Alexandra Schamann[1,2] · Stefanie Dobrovolny[1,3] · Alexander Indra[4] · Sarah Zanon Agapito-Tenfen[5] · Rupert Hochegger[1] · Alexander G. Haslberger[2] · Christian Brandes[1]

Received: 22 December 2017 / Revised: 6 March 2018 / Accepted: 10 March 2018

Abstract
MON810 maize was first commercialized in 1997 and it is one of the most marketed genetically modified crops worldwide. Although MON810 maize has been studied extensively, its genetic stability and epigenetics have not been studied very well. We used next-generation sequencing to investigate the genetics and epigenetics of the *cry1Ab* coding region and its 3′ flanking genomic region in three different MON810 maize varieties. Genetic characterization of the *cry1Ab* coding region allowed us to identify and quantify several sequence variants. Samples from seeds containing a stacked MON810 event had more variants than MON810 single event varieties. Specifically, position 71 of the analyzed region varied in 15 of 600 samples tested and thus appears to be a mutational hotspot. In addition, position 71 varied at very different frequencies in the samples. Epigenetic analysis revealed a low degree of methylation, making it difficult to associate the coding region variants with methylation status. In conclusion, the variation in the coding region is either due to the increased age of the seeds from the tested maize varieties, which is known to increase the mutation rate, or due to the presence of a second (non-functional) *cry1Ab* fragment in the genome of the MON810 maize variety.

Keywords Genetic stability · GMO · MON810 · Amplicon sequencing · SNP · Bt maize · Methylation · Bisulfite sequencing

Introduction

A high percentage of economically important crops such as soybeans, maize, cotton, and canola are transgenic. Insect resistance is one of the most frequently marketed GM traits worldwide. In 2016, insect resistant maize hybrids were cultivated on an area of 53.7 million hectares, which corresponds to 88.6% of the farmland used for GM maize. The majority of GM maize varieties are stacked events, which contain mainly transgenes conferring herbicide resistance

✉ Christian Brandes
christian.brandes@ages.at

Sina-Elisabeth Ben Ali
sina-elisabeth.ben-ali@ages.at

Alexandra Schamann
alexandra.schamann@ages.at

Stefanie Dobrovolny
stefanie.dobrovolny@ages.at

Alexander Indra
alexander.indra@ages.at

Sarah Zanon Agapito-Tenfen
sarah.agapito@genok.no

Rupert Hochegger
rupert.hochegger@ages.at

Alexander G. Haslberger
alexander.haslberger@univie.ac.at

[1] Institute for Food Safety, Austrian Agency for Health and Food Safety, Spargelfeldstrasse 191, 1220 Vienna, Austria

[2] Department of Nutritional Sciences, University of Vienna, Althanstraße 14, 1090 Vienna, Austria

[3] Department of Analytical Chemistry, University of Vienna, Währinger Straße 38, 1090 Vienna, Austria

[4] Institute for Medical Microbiology and Hygiene, Austrian Agency for Health and Food Safety, Währingerstrasse 25a, 1096 Vienna, Austria

[5] Genøk-Centre for Biosafety, PB 6418 Science Park, 9294 Tromsø, Norway

Published online: 26 March 2018

and insect resistance [1]. In this work, the term single event is used for GM plants containing a single transgene, and the term stacked event is used for plants produced via conventional cross-breeding of single events, which results in plants with more than one transgene. Next to conventional breeding, re-transformation, or co-transformation can be used for gene stacking, but these are not addressed in the current work. The recombinant *cry1Ab*-expressing insect resistant maize variety MON810 (trade name: Yieldgard®) was first commercialized in the US in 1997. It was developed by Monsanto to protect maize plants from feeding damage caused by the European corn borer (*Ostrinia nubilalis*), the southwestern corn borer (*Diatraea grandiosella*), and the pink borer (*Sesamia cretica*) [2].

MON810 maize cultivars contain a transgene cassette consisting of a 35S promoter derived from the cauliflower mosaic virus, an hsp70 intron from a maize heat shock protein, and a *cry1Ab* element derived from *Bacillus thuringiensis* (see Fig. 1) [3]. The MON810 event was produced by microprojectile bombardment of embryogenic maize tissue with a plasmid [4]. This method of transformation can lead to multi-copy events [5] with complex integration patterns if whole plasmids are used for bombardment (instead of low doses of cassette DNA) [6]. Rearranged fragments of the full-length transgene with varying copy numbers are frequently detected after bombardment [7]. Rearrangements may also be caused by regenerating transformed plants through tissue culture [5, 8, 9]. In fact, MON810 shows a complex integration pattern because of a truncation event that led to the complete loss of the NOS terminator at the 3′ end of the recombinant *cry1Ab* gene [10]. The MON810 transgene also interrupts exon 8 of the putative maize HECT E3 ligase gene 8 (on chromosome 5) and has an additional insertion [11] with 100% homology to the *Zea mays* mitochondrial genome within the first 44 bp of its 3′ flanking region. However, Monsanto [2] stated in their technical report that "Southern blot analysis of corn event MON810 demonstrated that a single functional copy of the *cry1Ab* coding sequence was integrated into the corn genome". Thus, the integration of non-functional fragments of the recombinant *cry1Ab* coding sequence or other parts of the transformation vector at additional sites in the genome has not been explicitly excluded. Nevertheless, other studies assessing the MON810 copy number by real-time PCR and digital PCR showed that only one insert is present in the genome of different MON810 maize varieties [12, 13]. Figure 1 shows the MON810 transgene cassette.

Commercialization of a GM crop requires a regulatory approval that is usually contingent on a positive safety assessment (positive means that the GM crop is considered as safe) [14]. Molecular characterization is a crucial part of safety assessments and usually relies on the results of Southern blot analysis and Sanger sequencing. It is used to determine the copy number and function of the insert, transgene stability and integrity, the DNA sequence of the insert and the flanking genomic regions, as well as the general characteristics of the introduced trait and possible interactions among stacked transgenes [15, 16]. However, while genetic stability of the transgenic locus has to be verified in single events (over five generations), it is sufficient for stacked events to establish that the integrity of authorized single events has been preserved [15].

We know from earlier investigations of MON810 maize that the coding region of the MON810 *cry1Ab* gene has a

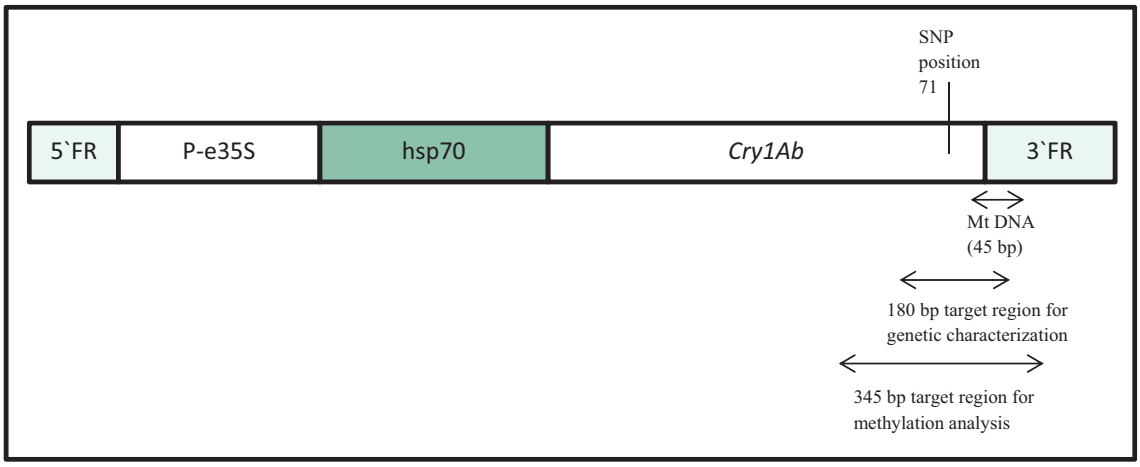


Fig. 1 Illustration of the genetic elements of the MON810 transgene cassette and the analyzed regions. The integrated transgene cassette consists of the constitutive 35S promoter from the cauliflower mosaic virus, the hsp70 intron, and a truncated *cry1Ab* coding sequence. The 180-bp target region (Amplicon-3m810fw + rev; marked with arrows), which was selected for genetic characterization, ranges from the *cry1Ab* element to the mitochondrial (Mt) DNA in the 3′ flanking region (3′FR). The frequently detected variant at position 71 lies within this 180-bp target region. The 345-bp target region (Meth F fw + rev) was used for methylation analysis

variant (a heterozygous C → T transition at position 71) in a few samples of the stacked MON810 variety 4421VT3, but investigations of a MON810 single event did not reveal any variants [17, 18]. The influence of gene stacking on the genetic stability of transgenic inserts remains unclear. Therefore, we have analyzed three other MON810 commercial varieties to improve the safety assessment of MON810 and to investigate the impact of transgene stacking in recently detected variants. In this study, we focused on the molecular characterization of the *cry1Ab* coding region by NGS. We employed amplicon sequencing for the detection and quantification of very low-frequency variants, as well as to more accurately characterize specific genomic regions [19]. In addition, we used conventional PCR to determine the zygosity of the MON810 insert and to gain a better understanding of its genetic structure. We also investigated the methylation patterns in the *cry1Ab* coding region. 5-methylcytosine resulting from the methylation of cytosine is an important epigenetic mechanism and very common in eukaryotes [20]. Therefore, we hypothesized that this epigenetic regulatory mechanism could be linked to the observed genetic instability in that region. We used bisulfite sequencing, which is considered to be the "gold standard" method for DNA methylation studies, to assess the link between DNA methylation and the frequent occurrence of C → T transitions in the *cry1Ab* coding region [21].

Materials and methods

Biological materials

Three maize varieties were analyzed: hybrid 631RR2/Bt, hybrid DKB 350YG, and a third MON810 hybrid (name not specified). Hybrid 631RR2/Bt (Croplan Genetics®) is a stacked event (single cross hybrid) with the unique identifier MON-ØØ6Ø3–6 x MON-ØØ81Ø-6. It contains two transgenes, a herbicide resistance transgene (event NK603, trade name: Roundup Ready), and an insect resistance transgene (event MON810, tradename: Yieldgard). NK603 enables resistance against the herbicide Roundup. MON810 produces a toxin (*Cry1Ab*, derived from Bacillus thuringiensis) against lepidopteran insects. Each transgene of the stacked event is located on a different chromosome. Hybrid 631RR2/Bt was obtained from the US market (Indiana, 2005) and the hybrid seeds were used for our analyses. DKB 350YG and the unspecified MON810 hybrid are single events, each containing the single transgene MON810 with the unique identifier MON-ØØ81Ø-6. DKB 350YG (Dekalb Brasil®) was obtained from the Brazilian market in 2013. The unspecified MON810 variety was obtained from Germany in 2008. Genomic DNA was extracted from 200 individual maize grains from the unspecified MON810 variety, the 631RR2/Bt variety, and the DKB 350YG variety, respectively.

Extraction of genomic DNA

Maize grains were individually homogenized with a household garlic crusher. Then, 150 mg of the homogenized grains was incubated (o/n) with a mixture of 820 µL of TNE buffer [10 mM Tris (pH 8), 150 mM NaCl, 2 mM EDTA (pH 8), 1% SDS], 150 µL of 5 M guanidine HCl, and 30 µL of proteinase K (600 µg/mL) at 60 °C under constant shaking. Next, the mixture was centrifuged for 5 min at 16,100×*g*. Then, 300 µL of chloroform (99%) was added to 600 µL of the supernatant and the mixture was vortexed for 20 s. Then, the mixture was centrifuged at 16,100×*g* for 8 min to separate the phases and 500 µL of the aqueous phase was transferred into a new tube. Two microliters of RNAse (8 µg/mL) was added to the aqueous phase and the mixture was incubated at 60 °C for 30 min under constant shaking. The extracted DNA was purified using the Wizard DNA Clean-up System (Promega). After a pre-elution with 20 µL of ddH_2O, the DNA was eluted with 10 mM Tris buffer (pH 7.4, 70 °C) for 10 min. The eluted DNA was immediately stored at − 20 °C.

DNA quality check

Agarose gel electrophoresis, as well as spectrophotometry, were used to assess DNA quality. DNA samples and Quantitas DNA Marker (Biozym) were loaded onto a 1% w/v agarose gel containing ethidium bromide and were then run for 23 min (140 V, 400 mA, 100 W). Afterwards, the gel was viewed under UV light using a Chemi XRS Gel documentation system (Bio-Rad). Samples with clear and sharp bands in the upper region of the gel were considered acceptable. DNA from these samples was then measured with a Nanophotometer (Implen). DNA sample purity was assessed using the 260/280 absorbance ratio. Samples with a ratio between 1.75 and 1.90 were used for additional analyses.

DNA quantitation

DNA concentrations were measured using a Qubit® 2.0 Fluorometer (Life Technologies) with a Qubit® dsDNA BR Assay Kit (Life Technologies). The instrument uses fluorescent dyes that bind to specific target molecules to determine the concentration of nucleic acids. Therefore, quantitation by the Qubit® 2.0 Fluorometer method is more sensitive than the widespread UV absorbance method [22].

PCR-based zygosity testing

MON810 zygosity was determined by adapting a PCR-based testing method developed by Liu and Chen [23]. Three primers (HECTExfwd, CRYfwd, and HECTupRev3, see Table 3) [11] were used to identify the degree of MON810 zygosity in the maize samples. The primer pair HECTExfwd/HECTupRev3 was used to detect the wild type (no MON810 transgene). CRYfwd and HECTupRev3 were used to identify the 3′-border region of MON810 in genomic and transgenic DNA (in a hemizygous or homozygous state). Maize samples that are homozygous for MON810 will produce only the MON810-specific CRYfwd/HECTupRev3 fragment and not the wild-type fragment. Samples without MON810 will produce only the wild-type HECTExfwd/HECTupRev3 fragment. Samples that are hemizygous for MON810 will produce both the wild type and MON810 fragments.

The PCR was performed using genomic DNA as a template. Each 20 µL reaction contained 4 µL of 5X Green GoTaq® Reaction Buffer (Promega), 250 nM of each forward primer, 500 nM of the reverse primer, 250 µM of each deoxynucleoside triphosphate (dNTP), 1.5 U of GoTaq® G2 DNA Polymerase (Promega), and 60 ng of DNA. A Mastercycler® (Eppendorf) was used for this analysis. The temperature profile is shown in Table 1.

The commercial hybrid NK603 AG8025 was used as a negative control (producing only the wild-type band) and the commercial hybrid PR33P67 (Pioneer Hi-Bred), which contains the MON810 trait in a hemizygous state, was used as a positive control (producing two bands, one for MON810 and one for the wild type).

SNP genotyping using high-resolution melting analysis

The first step of high-resolution melting (HRM) analysis uses real-time PCR to amplify the target sequence [24]. Each 16 µL reaction used the Type-it HRM PCR Kit (Qiagen), 8 µL master mix, 700 nM primers, and 1.6 µL of diluted DNA (40 ng/µL) template. HRM analysis was performed on a Rotor-Gene Q instrument (Qiagen) with a temperature ramping rate of 0.2 °C per 4 s. The initial and final temperatures were 80 and 92 °C. The thermocycling profile used is shown in Table 2. The melting curves and the temperature-shifted curves were normalized to enable sample-to-sample comparisons. Modified curves and HRM scores were obtained using the Rotor-Gene Q series software (version 2.0.2.4, Qiagen). After normalization, the normalized and temperature-shifted melting curves were used to derive the final curve. One sample (per run) was selected as a reference genotype (HRM score of 100%) for the difference plot. The software compared the melting curves between the samples and the reference and generated confidence intervals relative to the selected genotype.

To identify samples with sequence variations in the 180 bp target sequence (Fig. 1), 600 genomic samples containing the MON810 insert were screened by HRM analysis. Amplicons with sequence variations have different melting curve shapes and, therefore, lower HRM scores, which enables the identification of DNA samples containing these variations. Each variety was tested separately and a maximum of 70 genomic DNA samples was analyzed per run (one replicate per run). Every run was repeated on a different day. The mean of the HRM scores was calculated, and in each run, the seven samples with the lowest HRM scores and the reference sample (with an HRM score of 100%) were selected for amplicon sequencing. In total, 24 samples of each variety were sequenced. We have described the use of HRM analysis for variant detection several times [17, 25, 26].

Amplicon sequencing

Amplicon paired-end sequencing of the 180 bp target region (Fig. 1) was performed to characterize samples with low HRM scores in detail. Deep sequencing of amplicons (PCR products) enables the detection and quantification of very low-frequency variants (even frequencies of 1% are detectable), as well as accurate characterization of specific genomic regions [19].

The 16S Metagenomic Sequencing Library Preparation protocol [27] and the protocol of Dobrovolny et al. (2018, under review) were modified to sequence amplicons on an Illumina MiSeq sequencer. According to the 16S

Table 1 Thermal cycling profile of the PCR zygosity experiment

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	94	10 min	
Denaturation	94	30 s	30
Annealing	54	60 s	
Extension	72	60 s	
Final extension	72	8 min	

Table 2 Thermal cycling profile of the HRM analysis experiment

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	5 min	
Denaturation	95	10 s	40
Annealing	55	30 s	
Extension	72	20 s	
HRM	80–92	+0.2 °C every 4 s	

Table 3 Primers used in this work

Primer	Sequence
HECTExfwd	5′-TCAATCATCAAAGCATCATCG-3′ [11]
CRYfwd	5′-TCTTCACGTCCAGCAATCAG-3′ [11]
HECTupRev3	5′-TTTGGGAAGGAAAAGGTATC-3′ [11]
Amplicon-3m810fw	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAAGCACGAGACCGTCAA-3′ modified after [18]
Amplicon-3m810rev	5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTCGCAAGCAAATTCGGAA-3′ modified after [18]
Nextera XT Index 1 Primers	[28]
Nextera XT Index 2 Primers	[28]
Meth F fw	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGYGAYGATGTGTTYAAGGAGAA-3′
Meth F rev	5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAARCCAAACTCARATRAATCAAAA-3′
AmpliconATP1-1 fw	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAAYGAGATTYAAGYTGGGGAAATGGT-3′ modified after [30]
AmpliconATP1-1 rev	5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCTCTTCCATCAATARRTACTCCCA-3′ modified after [30]

Table 4 Thermal cycling profile of the amplicon PCR and the index PCR

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	15 min	
Denaturation	95	30 s	40 for amplicon PCR 8 for index PCR
Annealing	62 for amplicon PCR 55 for index PCR	30 s	
Extension	72	30 s	
Final extension	72	10 min	

Metagenomic Sequencing Library Preparation protocol [27], Illumina overhang adapter sequences were added to the MON810 coding region primer sequences [18] to generate amplicon primers (Amplicon-3m810fw and Amplicon-3m810rev). All primers used in this work are listed in Table 3. Amplicon PCR was performed in a thermal cycler (Eppendorf) to amplify the region of interest. Then, 2.5 µL of genomic DNA (5 ng/µL), 5 µL of each primer (1 µM) (Amplicon-3m810fw and rev, see Table 3), and 12.5 µL of HotStarTaq Master Mix (Qiagen) were used to set up the reaction. The temperature profile of the amplicon PCR is shown in Table 4.

Subsequently, amplicon PCR products were purified using AMPure XP beads to eliminate free primers and primer dimers. Index PCR was used to attach dual indices and sequencing adapters to the purified amplicons. Each index PCR reaction contained 5 µL of purified PCR amplicon product, 5 µL of each primer [Nextera XT Index Primers 1 and 2 (Illumina)] [28], 25 µL of HotStarTaq Master Mix (Qiagen), and 10 µL of ddH_2O. The thermal cycler profile used for index PCR is shown in Table 4. A second clean-up step was performed on the PCR products and the purified products were quantified using the Qubit® 2.0 Fluorometer (Life Technologies). Depending on the size of the DNA amplicon, products were diluted with 10 mM Tris (pH 8.5) to a 4 nM concentration. Five microliters of each diluted sample was pooled together, and 5 µL of this pool was combined with 5 µL 0.2 N NaOH and incubated for 5 min at room temperature to denature the DNA. The denatured DNA was diluted with 990 µL of pre-chilled HT1 buffer and the resulting 20 pM denatured library was stored at − 20 °C until sequencing. For amplicon sequencing, 2 µL of the 20 pM denatured library was spiked into a whole bacterial genome sequencing run (300 bp paired-end sequencing) using the MiSeq Reagent Kit v3 (Illumina). Sequencing data were evaluated using the CLC Genomics Workbench version 10.1.1 (Qiagen).

Bisulfite sequencing

For bisulfite sequencing, 24 selected DNA samples (see section "Methylation status in the coding region of the MON810 *cry1Ab gene*") were bisulfite-converted and purified using the EpiTect® Bisulfite Kit (Qiagen) according to the manufacturer's protocol. Next, amplicon sequencing was performed as described above (section "Amplicon sequencing"). Only the amplicon PCR primers, reagents, and PCR temperature profiles were different. For amplicon PCR with bisulfite-converted samples, a specific primer pair targeting the coding region of the MON810 *cry1Ab* gene was designed using Kismeth [29] and Illumina overhang adapter sequences [27] were added to the primer sequences. The resulting primers are depicted in Table 3 (Meth F fw and Meth F rev). Each 25 µL amplicon PCR reaction mixture contained 50 ng of bisulfite-converted DNA, 5 µL of 5X Colorless GoTaq® Buffer (Promega), 0.625 µL of each primer (10 µM) (Meth F fw and Meth F rev, see Table 3), 250 µM of each dNTP, and 1.875 U of GoTaq® Hot Start

Table 5 Thermal cycling profile of the amplicon PCR and the index PCR with bisulfite-converted samples

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	94	10 min	
Denaturation	94	30 s	40 for amplicon PCR
Annealing	50	30 s	8 for index PCR
Extension	72	45 s	
Final extension	72	10 min	

Table 6 Thermal cycling profile of the Amplicon PCR and the Index PCR for the endogenous gene control

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	15 min	
Denaturation	95	30 s	35 for amplicon PCR
Annealing	54	30 s	8 for index PCR
Extension	72	45 s	
Final extension	72	10 min	

DNA Polymerase (Promega). Amplicon PCR products were purified as described in the "Amplicon sequencing" section. The purified amplicons were used as templates for index PCR. Each 50 µL index PCR reaction contained 5 µL of purified amplicon PCR product, 5 µL of each primer [Nextera XT Index Primers 1 and 2 (Illumina)], 10 µL of 5X Colorless GoTaq® Buffer (Promega), 3.75 U of GoTaq® Hot Start DNA Polymerase (Promega), and 250 µM of each dNTP. Thermal cycling temperature profiles are shown in Table 5. Afterwards, the index PCR samples were treated as described in the "Amplicon sequencing" section. CG, CHG, and CHH methylation patterns were analyzed and displayed using the CLC Genomics Workbench version 10.1.1 (Qiagen).

Bisulfite conversion frequency

To determine the bisulfite conversion rate, a universal endogenous control gene developed by Wang et al. [30] was used. Illumina overhang adapter sequences were added to the sequences specific to the endogenous ATP1-1 control gene. The resulting primers are shown in Table 3 (AmpliconATP1-1 fw and rev). Amplicon PCR and index PCR were set up, as described in "Amplicon sequencing" section, and differed only by the specific primers used. The thermal cycler temperature profiles for amplicon PCR and index PCR are shown in Table 6. After index PCR, control samples were treated as described in "Amplicon sequencing" section. The bisulfite conversion frequency was calculated using the CLC Genomics Workbench version 10.1.1 (Qiagen).

Results

Genetic characterization of the MON810 *cry1Ab* coding region

First, we screened all samples by HRM analysis to detect mutations in the 180 bp target region. The mean of the HRM scores of two HRM runs was calculated and the samples with the lowest mean HRM scores as well as the reference samples were chosen for amplicon sequencing (a total of 72 samples, 24 of each variety). The mean HRM scores of these samples are presented in Table 7. Next, the 3′ flanking region of MON810 was analyzed for sequence variations by amplicon sequencing. Insert Table 7.

The 72 sequenced samples (Table 7) generated between 5308 and 26,990 sequencing reads. All sequenced samples passed a quality filter (trimming) to eliminate sequencing errors. Thereby, between 0 and 289 reads were removed for each sample. Reads were mapped to the reference sequence and the CLC Variant Detector was used for SNP calling. The minimum variant frequency was set to 1% to prevent false-positive variants due to sequencing signal noise. Table 7 gives an overview in which of selected samples variations were detected. Seventeen of the 72 sequenced samples had a variant in the coding region of MON810 *cry1Ab* with a minimum variant frequency of 1% (see Fig. 2). Fifteen of the 17 samples with a variant harbored the previously described variant (T instead of C) at position 71 of the analyzed region. However, the frequency of the variant at position 71 ranged from 1.2 to 53.2%. 631RR2/Bt samples 63, 68, 74, and 207 had the highest frequencies of variant 71 (31.7, 53.2, 52.2, and 26.8%). Low-frequency (<2%) transitions were detected at positions 39 (C → T), 50 (G → A), 53 (C → T), 71 (C → T), 79 (C → T), 83 (C → T), and 91 (G → A). Of the 17 samples with a variant, 14 were derived from the stacked MON810 event (631RR2/Bt), two were from the unspecified (single event) MON810 variety, and one was from the (single event) hybrid DKB 350YG. An overview of the results from the SNP analysis is depicted in Table 8.

Taken together, noticeably, more samples from the stacked MON810 event (631RR2/Bt) had a variant in the *cry1Ab* coding region and at higher frequencies compared to samples from the single event varieties (DKB 350YG and the unspecified MON810 variety). The variant frequencies of samples from sequenced single events were <2%, whereas the samples from the stacked event variety had variant frequencies up to 53.2% (see Table 8). Furthermore, the heterozygous variant at position 71 (C → T) was the most common variant in all of the tested samples and had

Table 7 Results of HRM analysis and amplicon sequencing

631RR2/Bt samples	Mean HRM score (%)	SNP detection*	DKB 350YG samples	Mean HRM score (%)	SNP detection*	Unspecified MON810 samples	Mean HRM score (%)	SNP detection*
82	4.07	✓	242	52.74	✗	22	23.51	✗
130	28.93	✓	136	56.24	✗	282	38.62	✗
68	51.05	✓	66	56.73	✗	296	49.56	✗
74	51.61	✓	185	66.18	✗	46	50.11	✗
117	54.28	✓	187	66.33	✗	41	53.12	✗
121	60.35	✓	190	69.01	✗	40	54.53	✗
35	64.80	✗	216	70.03	✗	34	55.82	✓
129	64.82	✓	191	71.34	✗	44	55.98	✗
207	64.92	✓	181	72.61	✗	5	57.31	✓
81	67.02	✓	194	74.73	✗	323	60.10	✗
72	68.68	✓	48	75.71	✓	234	61.94	✗
1	69.35	✗	163	75.83	✗	116	62.61	✗
120	69.82	✓	184	76.12	✗	206	63.07	✗
41	70.13	✗	186	76.39	✗	187	63.12	✗
63	70.49	✓	138	76.59	✗	269	66.82	✗
205	79.58	✗	141	76.70	✗	126	68.58	✗
182	80.22	✗	219	78.23	✗	319	68.79	✗
206	81.48	✗	155	78.29	✗	435	77.24	✗
158	83.35	✗	126	79.10	✗	444	77.57	✗
181	83.77	✓	189	80.50	✗	446	82.82	✗
201	84.12	✓	54	89.00	✗	402	83.26	✗
39	100.00	✗	123	100.00	✗	14	100.00	✗
108	100.00	✗	253	100.00	✗	276	100.00	✗
173	100.00	✗	225	100.00	✗	292	100.00	✗

The mean values of the obtained HRM scores are shown. Low HRM scores indicate mutations

*By amplicon sequencing

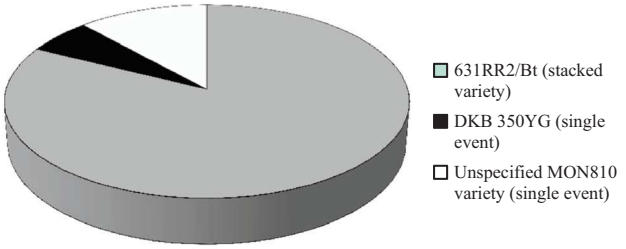


Fig. 2 Distribution of samples with a sequence variant in the MON810 *cry1Ab* coding region. The stacked variety 631RR2/Bt had the highest variant frequency

the highest variant frequencies (between 1.2 and 53.2%). The cytosine-to-thymine transition at position 71 does not lead to an amino acid change (silent mutation) [17] and the low-frequency variants at positions 50, 53, and 83 are also silent mutations. In contrast, the low-frequency variants at positions 39, 79, and 91 result in amino acid changes. The C → T transition at position 39 results in the translation of a phenylalanine (F) instead of a leucine (L), the C → T transition at position 79 leads to a valine (V) instead of an alanine (A), and the G → A transition at position 91 results in an arginine (R) instead of a histidine (H). It was striking that all variants were either C → T or G → A transitions. C → T transitions are very frequent at 5-methylcytosine sites. This is because deamination of 5-methylcytosine leads to thymine, which is not recognized by the enzyme uracil-DNA glycosylase and, therefore, not repaired [31]. This prompted us to investigate the methylation status of this region.

Methylation status of the MON810 *cry1Ab* coding region

Bisulfite sequencing was used to investigate the methylation status of the MON810 *cry1Ab* coding region in detail. The bisulfite-sequenced region was 345 bp, which included the 180 bp region used for SNP characterization (Fig. 1). Position 71 of the 180 bp fragment appears to be a mutational hotspot, which prompted our detailed

Table 8 Overview of detected variants in the three tested varieties

	SNP 39 (C → T)	SNP 50 (G → A)	SNP 53 (C → T)	SNP 71 (C → T)	SNP 79 (C → T)	SNP 83 (C → T)	SNP 91 (G → A)
631RR2/Bt samples							
63	–	–	–	31.71	–	–	–
68	–	–	–	53.15	–	–	–
72	–	–	–	6.15	–	–	–
74	–	–	–	52.24	–	–	–
81	–	–	–	4.27	–	–	–
82	–	–	1.42	1.16	–	–	–
117	–	–	–	2.61	–	–	–
120	–	–	–	1.28	–	–	–
121	1.11	1.24	–	1.68	–	–	–
129	–	–	–	1.74	–	–	–
130	–	–	–	12.68	1.26	–	–
181	–	–	–	–	–	1.17	–
201	–	–	–	1.78	–	–	–
207	–	–	–	26.80	1.25	–	–
DKB 350YG samples							
48	–	–	–	–	–	–	1.46
Unspecified MON810 samples							
5	–	–	–	1.60	–	–	–
34	–	–	–	1.87	–	–	–

The upper column shows the position and type of variant. Frequencies of the different variants are depicted in %

investigation of its methylation status. In total, eight samples of each variety were bisulfite-sequenced. For variety 631RR2/Bt with stacked transgenes, four samples with a high frequency of variants at position 71 and four samples without any nucleotide variations were selected for bisulfite sequencing. Only samples without variants were chosen for methylation analysis from the DKB 350YG and unspecified MON810 varieties. The analyzed samples generated between 4612 and 49,062 reads. A maximum of two reads was removed via quality trimming for each sample. Bisulfite reads were mapped to the reference sequence. Methylation levels were quantified using the Bisulfite Sequencing tool of the CLC Genomics Workbench (version 10.1.1). The analyzed region (excluding primer regions) included 25 cytosines in CpG motifs, 22 cytosines in CHG (where H is an A or T) motifs and 71 cytosines in CHH (where H is any base other than G) motifs. Overall, the average methylation level of the 24 bisulfite-sequenced samples was low (0.83% overall methylation, with 0.86% CpG, 0.78% CHG, and 0.86% CHH). The bisulfite conversion frequency averaged 97.9%. Figure 3 shows the methylation levels for each individual variety.

To determine if DNA methylation differed significantly, we compared the methylation levels in the different motifs (CpG, CHG, and CHH)

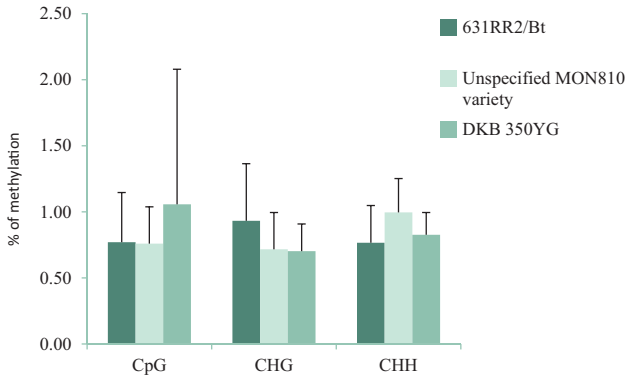


Fig. 3 Cytosine methylation levels in the *cry1Ab* coding region and the 3′ flanking region among the three tested MON810 maize varieties. The levels of cytosine methylation were measured in the CG, CHG, and CHH motifs. Vertical bars show the positive standard deviation

1. among the three different varieties (631RR2/Bt, DKB 350YG and the unspecified MON810 variety);
2. between 631RR2/Bt samples with a high-frequency variant and 631RR2/Bt samples without a variant.

In addition, we compared the methylation levels exclusively at position 71 (a CHH motif) between 631RR2/Bt

samples with a high-frequency variant and 631RR2/Bt samples without a variant.

There were no statistically significant differences among the three different varieties (ANOVA, $p \leq 0.05$). No significant differences were detected between the tested 631RR2/Bt samples either (ANOVA, $p \leq 0.05$). Interestingly, the methylation status of the analyzed region was not correlated with the presence or the absence of sequence variants. To gain a better understanding of the genetic structure of the MON810 locus, the insert zygosity was investigated in the three different varieties (631RR2/Bt, DKB 350YG, and the unspecified MON810 variety).

MON810 zygosity

MON810 hybrid maize grains can have different genetic structures: hemizygous (produced by a transgenic male or female parent) or homozygous at the transgenic locus [32]. If the genetic structure is homozygous at the transgene locus, it would theoretically be possible to detect two different alleles for every nucleotide position of MON810 (one maternal and one paternal). Since we discovered different alleles in our samples, MON810 zygosity was determined by conventional PCR as described in the "Materials and methods". Overall, 20 samples of each variety were investigated to identify the MON810 insert zygosity. All samples analyzed produced two bands on the agarose gel, the wild-type band, and the MON810-specific band, which means that they were hemizygous for the transgene locus. Therefore, we assumed that all of the analyzed F1 seeds from 631RR2/Bt, DKB350YG, and the unspecified MON810 variety were hemizygous for the MON810 insert.

Conclusions and discussion

Although there are already studies describing the molecular characterization of MON810 maize [10–12, 17, 18, 33–37], the complex integration pattern of the MON810 event [10, 11], recent SNP findings in commercial MON810 maize [17], and a fragile 45S rDNA phenotype in a stacked MON810 variety [36], as well as its frequent use worldwide [1], clearly indicate that MON810 maize varieties warrant further investigation. Currently, NGS is the method of choice for a detailed molecular characterization of transgenic events. NGS has many advantages over traditional Southern blot analysis; for example, NGS is much better at identifying incomplete and multiple integration events in complex regions [16, 38].

Deep sequencing of amplicons from the *cry1Ab* coding region and its flanking region verified the frequent presence of a heterozygous C → T transition at position 71, which was reported previously [17]. In contrast to our earlier study [17], NGS was used, and therefore, it was possible to quantify the frequencies of this variant. Additional maize varieties were also analyzed, allowing us to directly compare the variant frequencies between the different varieties and between single and stacked events. We observed more samples with coding region variants and much higher variant frequencies in the stacked MON810 event (631RR2/Bt) than in the single events (DKB 350YG and the unspecified MON810 variety). The occurrence of a T at position 71 was up to 53% in samples from variety 631RR2/Bt compared to < 2% in samples from the unspecified MON810 variety. Thus, the occurrence of sequence variants differs markedly between single and stacked events. Whether the stacking of transgenes alone is the reason for the increased mutation rate in the tested stacked event has to be clarified in further experiments. Possibly, other factors (e.g., age of the seeds) may contribute to genetic differences as well.

Interestingly, all identified variants were either C → T or G → A transitions. Furthermore, four of the seven identified variants depicted in Table 8 were silent mutations. These were located at the third position of the codon, which is known to be degenerated. Many amino acids are encoded by more than one codon. Such codons are called synonymous codons [39]. Although synonymous codons encode the same amino acid, codon usage is not random. Codon usage influences gene expression levels, protein folding, and protein cellular function [39, 40]. Thus, also silent mutations can have huge impact on proteins. The recombinant *cry1Ab* toxin of MON810 maize seems to be significantly stronger expressed compared to natural *cry1Ab* toxins due to modified codon usage prior to transgene insertion [41]. Protein folding and toxin activation may also be influenced in the recombinant *cry1Ab* toxin due to modified codon usage, but also because of major changes on the DNA level [42]. Synonymous codon usage patterns can vary markedly among genomes, but also within a genome. The evolutionary pressure on codon bias is explained by selection for translational efficiency, GC mutational bias, and by mRNA secondary-structure stability [40]. *Cry* genes are AT-rich compared to endogenous plant genes. This preference can lead to serious problems such as low expression if a *cry* gene becomes inserted into a plant organism. To enable stable and strong transgene expression, codon optimization is commonly performed prior to transgene insertion [43]. This was also performed for the *cry1Ab* gene; several A and T nucleotides were replaced by C and G nucleotides, which are preferred codon endings in *Zea maize* [41, 44]. In view of this, it was remarkable that the variants we have detected are only A and T variants, which does not reflect the distribution of nucleotides in the *Z. mays* background.

Another important result from our study is the determination of the zygosity of the MON810 transgene. We found that the transgenic locus was hemizygous in all three tested

varieties. This means that the analyzed maize grains inherited the transgene only from one parent (a transgenic male or female parent). We detected two alleles at single nucleotide positions of the MON810 transgene in individual maize grain samples. If the maize varieties were homozygous at the transgene locus, it would theoretically be possible to detect two different alleles at single nucleotide positions. Given the fact that the analyzed maize grains were hemizygous for the transgene locus, somatic mutations might explain the identification of two alleles at single positions in the MON810 transgene. Moreover, it has to be noted that every single seed of an ear is an independent event [45]. Thus, a mutation in pollen grain or ovule might also be the cause of the observed mutations. Moreover, in many plant species, genetic mutations and changes in chromosome structure in embryo meristems increase as the seeds age. The mutation rate can increase substantially in later developmental phases (seed senescence) [46]. Since we analyzed older maize grains, this may have influenced the sequence variations that we observed.

The identification of two alleles at single positions of the MON810 transgene may also be explained by the existence of a second (possibly only partial) copy of the recombinant *cry1Ab* gene, which contains a thymine instead of a cytosine at position 71. A BLAST search of the primer sequences (Amplicon-3m810fw and rev) used for SNP characterization revealed that the reverse primer can also bind to maize mitochondrial DNA. This is because the last base of the transgene and the first 44 bp of the 3′ flanking region of MON810 are 100% homologous to maize mitochondrial DNA (see Fig. 1). This additional, most-likely unintended, insertion of mitochondrial DNA has become a functional part of the *cry1Ab* coding region since it contains the stop codon for the *cry1Ab* gene. However, the forward primer can only bind to the *cry1Ab* coding region, so it can be assumed that this primer pair is specific to only one locus in the MON810 varieties' genomes (providing that no other copies are present). Monsanto states that only one functional copy of the *cry1Ab* gene is present in the maize genome and other studies confirmed the presence of only one MON810 insert in the MON810 maize genome [12, 13]. Nevertheless, since particle bombardment frequently results in variable copy numbers of rearranged transgene fragments [7], it is possible that small non-functional fragments of the *cry1Ab* coding sequence coupled to the mitochondrial DNA sequence are also contained in the MON810 maize genome and that these fragment(s) have been overlooked until now. Mandatory transgene characterization might be complicated by these sequence variations. Consequently, transgene presence and frequency may be underestimated [47].

Most remarkably, position 71 of the *cry1Ab* sequence seems to be a mutational hotspot in at least three MON810 maize varieties (631RR2/Bt, the unspecified MON810 variety, and variety 4421VT3 [17]). Mutational hotspots can arise through the pro-mutagenic activities of cytosine modifications such as methylation, deamination, and halogenation [20]. DNA methylation of cytosine in symmetrical CG motifs is the most common methylation pattern in plants. However, in plants cytosine can also be methylated at symmetrical CHG (H is A or T) or asymmetrical CHH (H is A, C or T) sites [30]. Moreover, C → T transitions frequently occur at 5-methylcytosine sites because deamination of 5-methylcytosine leads to thymine, which is not recognized by the enzyme uracil-DNA glycosylase and, therefore, not repaired [31]. Since all detected variants were either C → T or the complementary G → A transitions, we analyzed the methylation status of the *cry1Ab* coding region and its 3′ flanking genomic region by bisulfite sequencing. However, the methylation status was very low (0.83% overall methylation, of which 0.86% were CpG, 0.78% were CHG, and 0.86% were CHH) and no significant differences were detected among the varieties or between samples with or without variants. Therefore, we conclude that the presence of variants in the *cry1Ab* locus is not influenced by methylation status.

The methylation status of stacked versus single events (MON-89Ø34-3 and MON-ØØ6Ø3-6) was previously investigated [48] and differences in cytosine methylation levels were found in the FMV promoter and *cry2Ab2* transgene in four Bt-expressing hybrid varieties. Comparing single and stacked events in the same genetic background also revealed differences in the 35S promoter sequence and the results of transgene transcript accumulation levels showed differences in both *cry1A.105* and *cry2Ab2* transgenes among the four Bt-expressing hybrid varieties. However, only the methylation level of the *cry2Ab2* transgene differed between single and stacked events. Such differences are important for the ongoing discussion about whether stacked events should be regulated as conventional hybrids or as new GM plants [49]. If they are regulated as conventional hybrids, it is sufficient to refer to the validation parameters for each individual single event. In contrast, if they are regulated as new GM plants, the risk assessment has to be repeated with plants materials from stacked events.

Spontaneous mutations can occur in all living organisms during DNA replication, homologous or intrachromosomal recombination, mitosis and meiosis [9, 18]. Furthermore, hemizygous DNA architectures show considerably more ectopic recombination events compared to homozygous DNA architectures, which can lead to severe chromosomal rearrangements [50]. In addition, transposable elements and repetitive DNA sequences have the ability to alter genomes, e.g., create or reverse mutations. Transposable elements can also increase gene copy number or even lead to the creation of new genes through exon shuffling [51, 52]. Moreover, DNA repeats can cause instabilities via homologous

recombination. Mobile elements and repetitive DNA strongly influence genome diversity [53]. Due to plasticity, plant genomes have the ability to reorganize themselves at the DNA level, but also at the chromatin level [54]. It needs to be clarified, if transgene stacking increases genetic reorganization, which would mean an increased mutation rate compared to the natural mutation rate.

The results of our study support the statement by Waminal et al. [36]: "In the context of genetic engineering, it may be useful to note that transformation itself is known to be mutagenic, and during plant transformation, each transformation event receives a unique transgene integration pattern, and thus, a unique accompanying genomic rearrangement. Thus, it is logical to treat each event individually and uniquely". We also share the opinion that every event should be examined and evaluated separately. Since in GM risk assessment, toxicological tests are usually performed with *cry* proteins purified from laboratory strains of bacteria engineered to express *cry* protein instead of using *cry* proteins extracted from the assessed GM plant [42], toxicological relevant issues may be overlooked. This is especially relevant when evaluating stacked events, which may contain genetic differences in the transgene compared to single events. In view of this, we recommend a separate assessment for stacked events, which includes DNA-sequencing data of the transgenes.

Given our results, we have to clarify, if transgene stacking has an influence on the genetic stability, which would justify a separate evaluation of single and stacked events in the authorization of GM plants. Furthermore, molecular mechanisms responsible for the differences between single and stacked events have to be identified. Nevertheless, transgene integration processes can have unpredictable plant-wide effects on different processes and signaling pathways due to the interactions of biological networks [55]. These effects can be observed on the genomic, epigenomic, transcriptomic, proteomic, or metabolomic levels. To better understand interactions in plant biological networks, to improve breeding techniques, and to perform adequate risk assessments, it is important to identify and investigate such unintended effects.

Acknowledgements We would like to express our gratitude to all three anonymous reviewers for their helpful and detailed comments. Material costs for this work were partially supported by the Hochschuljubiläumsstiftung (HJST) of the city of Vienna.

Compliance with ethical standards

Conflict of interest All the authors declare that they have no competing interest.

Ethics requirements This article does not contain any studies with human or animal subjects.

References

1. ISAAA (2016) Global Status of Commercialized Biotech/GM Crops: 2016. ISAAA. http://www.isaaa.org/purchasepublications/itemdescription.asp?ItemType=BRIEFS&Control=IB052-2016. Accessed 26 Sept 2017
2. Monsanto (2017) Safety Assessment of YieldGard Insect-Protected Corn Event MON 810. Monsanto, https://monsanto.com/company/media/statements/mon810-bans/. Accessed 30 Oct 2017
3. Trtikova M, Wikmark OG, Zemp N, Widmer A, Hilbeck A (2015) Transgene expression and Bt protein content in transgenic Bt maize (MON810) under optimal and stressful environmental conditions. PLoS One 10(4):e0123011. https://doi.org/10.1371/journal.pone.0123011
4. Armstrong CL, Green CE, Phillips RL (1991) Development and availability of germplasm with high Type II culture formation response. Maize Genet Cooper Newsl 65:92–93
5. Haslberger AG (2003) Codex guidelines for GM foods include the analysis of unintended effects. Nat Biotech 21(7):739–741
6. Lowe BA, Shiva Prakash N, Way M, Mann MT, Spencer TM, Boddupalli RS (2009) Enhanced single copy integration events in corn via particle bombardment using low quantities of DNA. Transgenic Res 18(6):831–840. https://doi.org/10.1007/s11248-009-9265-0
7. Yin Z, Plader W, Malepszy S (2004) Transgene inheritance in plants. J Appl Genet 45(2):127–144
8. Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P (1999) Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. Plant J Cell Mol Biol 17(6):591–601
9. Schnell J, Steele M, Bean J, Neuspiel M, Girard C, Dormann N, Pearson C, Savoie A, Bourbonnière L, Macdonald P (2015) A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. Transgenic Res 24(1):1–17. https://doi.org/10.1007/s11248-014-9843-7
10. Hernandez M, Pla M, Esteve T, Prat S, Puigdomenech P, Ferrando A (2003) A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard based on the 3′-transgene integration sequence. Transgenic Res 12(2):179–189
11. Rosati A, Bogani P, Santarlasci A, Buiatti M (2008) Characterisation of 3′transgene insertion site and derived mRNAs in MON810 yield gard maize. Plant Mol Biol. https://doi.org/10.1007/s11103-008-9315-7
12. Aguilera M, Querci M, Pastor S, Bellocchi G, Milcamps A, Van den Eede G (2009) Assessing copy number of MON 810 integrations in commercial seed maize varieties by 5′ event-specific real-time PCR validated method coupled to 2-ΔΔCT analysis. Food Anal Methods 2(1):73–79. https://doi.org/10.1007/s12161-008-9036-1
13. Corbisier P, Bhat S, Partis L, Xie VR, Emslie KR (2010) Absolute quantification of genetically modified MON810 maize (*Zea mays* L.) by digital polymerase chain reaction. Anal Bioanal Chem 396(6):2143–2150. https://doi.org/10.1007/s00216-009-3200-3
14. Privalle LS, Chen J, Clapper G, Hunst P, Spiegelhalter F, Zhong CX (2012) Development of an agricultural biotechnology crop product: testing from discovery to commercialization. J Agric Food Chem 60(41):10179–10187. https://doi.org/10.1021/jf302706e
15. EFSA (2011) Scientific Opinion on Guidance for risk assessment for food and feed from genetically modified plants. EFSA J. https://doi.org/10.2903/j.efsa.2011.2150
16. Guttikonda SK, Marri P, Mammadov J, Ye L, Soe K, Richey K, Cruse J, Zhuang M, Gao Z, Evans C, Rounsley S, Kumpatla SP

(2016) Molecular characterization of transgenic events using next generation sequencing approach. PLoS One 11(2):e0149515. https://doi.org/10.1371/journal.pone.0149515
17. Ben Ali SE, Madi ZE, Hochegger R, Quist D, Prewein B, Haslberger AG, Brandes C (2014) Mutation scanning in a single and a stacked genetically modified (GM) event by real-time PCR and high resolution melting (HRM) analysis. Int J Mol Sci 15(11):19898–19923. https://doi.org/10.3390/ijms151119898
18. Neumann G, Brandes C, Joachimsthaler A, Hochegger R (2011) Assessment of the genetic stability of GMOs with a detailed examination of MON810 using Scorpion probes. Eur Food Res Technol 233(1):19–30. https://doi.org/10.1007/s00217-011-1487-8
19. Illumina (2015) Highly targeted resequencing of regions of interest. https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/amplicon-sequencing.html. Accessed 20 Jul 2017
20. Sassa A, Kanemaru Y, Kamoshita N, Honma M, Yasui M (2016) Mutagenic consequences of cytosine alterations site-specifically embedded in the human genome. Genes Environ 38(1):17. https://doi.org/10.1186/s41021-016-0045-9
21. Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. Biology 5(1):3. https://doi.org/10.3390/biology5010003
22. ThermoFisherScientific Qubit® 3.0 Fluorometer. https://www.thermofisher.com/at/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-fluorometer.html. Accessed 22 Jun 2017
23. Liu N, Chen H (2010) An accurate and rapid PCR-based zygosity testing method for genetically modified maize. Mol Plant Breed 7(3):619–623
24. Druml B, Cichna-Markl M (2014) High resolution melting (HRM) analysis of DNA—its role and potential in food analysis. Food Chem 158:245–254. https://doi.org/10.1016/j.foodchem.2014.02.111
25. Madi ZE, Brandes C, Neumann G, Quist D, Ruppitsch W, Hochegger R (2013) Evaluation of Adh1 alleles and transgenic soybean seeds using Scorpion PCR and HRM analysis. Eur Food Res Technol 237(2):125–135. https://doi.org/10.1007/s00217-013-1969-y
26. Castan M, Ben Ali S-E, Hochegger R, Ruppitsch W, Haslberger AG, Brandes C (2017) Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing. Eur Food Res Technol 243(3):353–365. https://doi.org/10.1007/s00217-016-2749-2
27. Illumina (2013) 16S metagenomic sequencing library preparation. https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html. Accessed 22 Jun 2017
28. Illumina (2016) Illumina adapter sequences. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/experiment-design/illumina-adapter-sequences_1000000002694-01.pdf. Accessed 28 Jul 2017
29. Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R (2008) Kismeth: Analyzer of plant methylation states through bisulfite sequencing. BMC Bioinf 9(1):371. https://doi.org/10.1186/1471-2105-9-371
30. Wang J, Wang C, Long Y, Hopkins C, Kurup S, Liu K, King GJ, Meng J (2011) Universal endogenous gene controls for bisulphite conversion in analysis of plant DNA methylation. Plant Methods 7:39–39. https://doi.org/10.1186/1746-4811-7-39
31. Griffiths A, Gelbart W, Miller J (1999) Modern genetic analysis. W. H. Freeman, New York
32. Zhang D, Corlet A, Fouilloux S (2008) 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 Res 17(3):393–402. https://doi.org/10.1007/s11248-007-9114-y
33. la Paz JL, Pla M, Papazova N, Puigdomènech P, Vicient CM (2010) Stability of the MON 810 transgene in maize. Plant Mol Biol 74(6):563–571. https://doi.org/10.1007/s11103-010-9696-2
34. Aguilera M, Querci M, Balla B, Prospero A, Ermolli M, Van den Eede G (2008) A qualitative approach for the assessment of the genetic stability of the MON 810 trait in commercial seed maize varieties. Food Anal Methods 1(4):252–258. https://doi.org/10.1007/s12161-008-9035-2
35. la Paz JL, Pla M, Centeno E, Vicient CM, Puigdomènech P (2014) The use of massive sequencing to detect differences between immature embryos of MON810 and a comparable non-GM maize variety. PLoS One. https://doi.org/10.1371/journal.pone.0100895
36. Waminal NE, Ryu KH, Choi S-H, Kim HH (2013) Randomly detected genetically modified (GM) maize (*Zea mays* L.) near a transport route revealed a fragile 45S rDNA phenotype. PLoS One 8(9):e74060. https://doi.org/10.1371/journal.pone.0074060
37. Singh CK, Ojha A, Kamle S, Kachru DN (2007) Assessment of cry1Ab transgene cassette in commercial Bt corn MON810: gene, event, construct and GMO specific concurrent characterization. Protocol Exch. https://doi.org/10.1038/nprot.2007.440
38. Zhang R, Yin Y, Zhang Y, Li K, Zhu H, Gong Q, Wang J, Hu X, Li N (2012) Molecular characterization of transgene integration by next-generation sequencing in transgenic cattle. PLoS One 7(11):e50348. https://doi.org/10.1371/journal.pone.0050348
39. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH (2004) Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci USA 101(10):3480–3485. https://doi.org/10.1073/pnas.0307827100
40. Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J 6(6):650–659. https://doi.org/10.1002/biot.201000332
41. van der Salm T, Bosch D, Honée G, Feng L, Munsterman E, Bakker P, Stiekema WJ, Visser B (1994) Insect resistance of transgenic plants that express modified *Bacillus thuringiensis* cryIA(b) and cryIC genes: a resistance management strategy. Plant Mol Biol 26(1):51–59. https://doi.org/10.1007/bf00039519
42. Latham JR, Love M, Hilbeck A (2017) The distinct properties of natural and GM cry insecticidal proteins. Biotechnol Genet Eng Rev 33(1):62–96. https://doi.org/10.1080/02648725.2017.1357295
43. Khan MS, Musatafa G, Nazir S, Joyia FA (2016) Applied molecular biotechnology: the next generation of genetic engineering. Plant molecular biotechnology: applications of transgenics. Taylor & Francis Group, New York
44. Liu H, He R, Zhang H, Huang Y, Tian M, Zhang J (2010) Analysis of synonymous codon usage in *Zea mays*. Mol Biol Rep 37(2):677–684. https://doi.org/10.1007/s11033-009-9521-7
45. Giroux MJ, Shaw J, Barry G, Cobb BG, Greene T, Okita T, Hannah LC (1996) A single mutation that increases maize seed weight. Proc Natl Acad Sci 93(12):5824–5829
46. D'Amato F (1997) Role of somatic mutations in the evolution of higher plants. Caryologia 50(1):1–15. https://doi.org/10.1080/00087114.1997.10797380
47. Agapito-Tenfen SZ, Wickson F (2017) Challenges for transgene detection in landraces and wild relatives: learning from 15 years of debate over GM maize in Mexico. Biodivers Conserv. https://doi.org/10.1007/s10531-017-1471-0
48. Vilperte V, Agapito-Tenfen SZ, Wikmark O-G, Nodari RO (2016) Levels of DNA methylation and transcript accumulation in leaves of transgenic maize varieties. Environ Sci Eur 28(1):29. https://doi.org/10.1186/s12302-016-0097-2
49. Rocca E, Andersen F (2017) How biological background assumptions influence scientific risk evaluation of stacked genetically modified plants: an analysis of research hypotheses and argumentations. Life Sci Soc Policy 13(1):11. https://doi.org/10.1186/s40504-017-0057-7

50. Sun XQ, Li DH, Xue JY, Yang SH, Zhang YM, Li MM, Hang YY (2016) Insertion DNA accelerates meiotic interchromosomal recombination in *Arabidopsis thaliana*. Mol Biol Evol 33(8):2044–2053. https://doi.org/10.1093/molbev/msw087
51. Ayarpadikannan S, Kim H-S (2014) The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. Genom Inf 12(3):98–104. https://doi.org/10.5808/GI.2014.12.3.98
52. Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15(6):621–627. https://doi.org/10.1016/j.gde.2005.09.010
53. Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genom Proteom Bio 12(4):164–171. https://doi.org/10.1016/j.gpb.2014.07.003
54. Weber N, Halpin C, Hannah LC, Jez JM, Kough J, Parrott W (2012) Editor's choice: crop genome plasticity and its relevance to food and feed safety of genetically engineered breeding stacks. Plant Physiol 160(4):1842–1853. https://doi.org/10.1104/pp.112.204271
55. Buiatti M, Christou P, Pastore G (2013) The application of GMOs in agriculture and in food production for a better nutrition: two different scientific points of view. Genes Nutr 8(3):255–270. https://doi.org/10.1007/s12263-012-0316-4

Analysis of transcriptomic differences between NK603 maize and near-isogenic maize using RNA sequencing and RT-qPCR

Sina-Elisabeth Ben Ali[1,2], Agnes Draxler[1,2], Alexandra Schamann[1,2], Sarah Agapito-Tenfen[3], Rupert Hochegger[1], Alexander G. Haslberger[2] and Christian Brandes[1,*]

[1] Institute for Food Safety, Austrian Agency for Health and Food Safety, Spargelfeldstrasse 191, 1220 Vienna, Austria. E-Mails: sina-elisabeth.ben-ali@ages.at; rupert.hochegger@ages.at; christian.brandes@ages.at

[2] Department of Nutritional Sciences, University of Vienna, Althanstraße 14, 1090 Vienna, Austria; E-Mail: alexander.haslberger@univie.ac.at

[3] Genøk- Centre for Biosafety, PB 6418 Science Park, 9294 Tromsø, Norway. E-Mail: sarah.agapito@genok.no

***** Author to whom correspondence should be addressed; E-Mail: christian.brandes@ages.at; Tel.: +43-5-0555-32202; Fax: +43-5-0555-32219.

Abstract

The aim of the present study was to compare the transcriptome profiles of NK603 maize and conventional near-isogenic varieties to identify differences between the crops as first step in the identification of unintended effects due to the NK603 transgene. High-throughput RNA sequencing was employed for this purpose and allowed a very deep sequencing analysis of the maize transcriptome. Differential gene expression analysis revealed 286 genes differentially expressed between the tested NK603 maize variety AG8025RR2 and a near-isogenic, non-transgenic variety. These genes could be assigned to different biological functions; especially those related to stress response, response to different stimuli, and signaling pathways were strongly represented. We also identified some gene candidates, which were consistently differentially expressed between a further NK603 variety (AG9045RR2) and its near-isogenic variety. However, differential expression of genes related to stress response was not confirmed in the second variety pair, indicating that these differences were most likely not due to the NK603 transgene, but rather due to external influences or genetic background. Moreover, our as well as other study results indicate an influence of the NK603 transgene on the fatty acid metabolism, but the resulting effects are within the natural variability.

Introduction

In the field of genetic engineering, there is still a lack of clarity about transgene integration mechanisms and about which biological processes take place in the plant after successful transgene integration [1,2]. DNA integration occurs randomly in the genome, however, is preferred in gene-rich regions and, thus, can lead to gene disruptions, DNA rearrangements, and the production of new proteins [3]. Thus, it is not surprising that there are also

unintended effects in genetically modified (GM) plants: the insertion of a transgene into a plant organism can, additionally to the intended effects (e.g. herbicide resistance), also lead to unintended effects of the plants [4,2,5]. Other processes which may lead to unintended effects are integration-associated processes, such as *in vitro* culture [4,6]. Unintended effects need to be identified and evaluated to detect and prevent possible adverse effects [7], but also to enable a deeper understanding of plant gene networking.

Currently, unintended effects are primarily investigated by targeted approaches, in which GM crops are compared to their near-isogenic non-GM counterparts. Molecular, compositional, phenotypic, and agronomic analyses are performed in order to identify similarities and differences between the crops. Significant differences between a GM crop and its conventional counterpart indicate an unintended effect, which requires further investigation [8]. Targeted approaches have the disadvantage that important differences can be overlooked, as only selected features or chemical/nutritional compounds are investigated. Untargeted OMICS-approaches, e.g. the analysis of gene expression in its entirety, enable the identification of broad coordinated trends which cannot be discerned by targeted approaches [9]. Thus, new non-targeted profiling approaches are an appropriate tool to complement current investigations regarding unintended effects of transgenic plants [3,10].

There are already several studies investigating unintended effects in MON810 maize, which produces a toxin against lepidopteran insects. Coll et al. [11,12] compared grown MON810 maize varieties with some comparable varieties using microarrays under in-vitro conditions, and conventional agricultural field conditions. In the study no gene was found to be significant differentially expressed in all variety pairs tested under in-vitro or conventional field conditions. Further, Coll et al. [13] analyzed the transcriptome of MON810 maize and comparators under low-nitrogen fertilization on the field. Major differences in gene expression were detected in one variety pair, but natural variation explained most of the variability. Furthermore, a proteome analysis was conducted by Coll et al. [14], which revealed virtually identical proteomic patterns between MON810 and comparators, with a very low number of spots showing fold-variations in the range of 1-1.8.

Differences in lignin content between MON810 maize and comparators were reported by Saxena and Stotzky, and Poerschmann et al. [15,16]. Herrero et al. [17] detected significant differences in enantiomeric amino acid composition between two MON810 varieties and comparators but not in a third tested variety pair (MON810 vs. non-GM). Ben Ali et al. [18,19] identified a variant in the coding region of the *cry1ab* gene in different MON810 maize varieties using High Resolution Melting (HRM) analysis, Sanger sequencing, and

amplicon sequencing. The detected variant was especially present in stacked MON810 varieties and did not lead to a modified protein since it was a silent mutation.

Agapito-Tenfen et al. [20] conducted a comparative analysis on MON810 maize and comparators grown under different agroecosystem conditions in Brazil using two-dimensional gel electrophoresis combined with mass spectrometry which identified 32 differentially expressed proteins. La Paz et al. [21] performed RNA sequencing on MON810 maize and comparators, which revealed 140 differentially expressed genes. A slight, but significant delay in seed and plant maturation of MON810 maize plants was also observed and used as explanation for the detected differences.

Gullì et al. [22] compared drought stress response and gene expression between MON810 maize and a near-isogenic non-GM variety. The *cry1ab* expression level itself was not influenced by the provoked drought stress. However, many stress-responsive genes were upregulated under drought-stress conditions in both, the GM and the non-GM variety, but more genes were upregulated in the GM variety, which indicates a more efficient stress response.

Also NK603 (Roundup Ready) maize, which is tolerant to the broad spectrum herbicide Roundup, was studied for unintended effects [23-25]. Agapito-Tenfen et al. [23] analyzed, additionally to MON89034 maize and NK603 maize, stacked maize varieties, which contained a transgene for herbicide resistance as well as one for insect resistance. The group used proteomics for the identification of unintended effects and detected 22 proteins to be differentially expressed in stacked and single GM events versus near-isogenic non-GM maize and a landrace variety. Mesnage et al. [24] generated proteome profiles of Roundup Ready maize and near-isogenic maize varieties using molecular profiling. A comparison of GM maize vs. the near-isogenic maize revealed alterations in the levels of enzymes of glycolysis and TCA cycle pathways, which can be interpreted as an imbalance in energy metabolism. Benevenuto et al. [25] performed proteomic and metabolomic analyses on NK603 maize plants under abiotic stress conditions. 20 differentially modulated proteins were identified between GM and non-GM hybrids under water deficiency conditions and herbicide sprays.

Barros et al. [26] compared two Roundup Ready maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics, and metabolomics. The results showed that the environment had a major effect on protein and gene expression and metabolite production. Harrigan et al. [27] analyzed differences in the metabolome between NK603 hybrids, corresponding negative segregants, and conventional comparator hybrids. It was

concluded that the largest effects on metabolomic variation were due to growing location and genomic differences associated with back-crossing practices.

Castan et al. [28] performed a molecular analysis of NK603 maize and identified two insertions, which are not present in the NK603 patent sequence. These insertions were located in the transgenic promoter region, therefore, they may have an influence on the promoter activity and consequently also on transgene expression [29].

Since 20 years, herbicide tolerant crops (e.g. Roundup Ready maize) are the most cultivated GM crops worldwide. 47% of the global biotech crop area belongs to herbicide tolerant crops [30]. Roundup Ready maize expresses the well-known and characterized protein 5-enolpyruvylshikimate-3-phosphate synthase [31]. Thus, the evaluation of this protein is relative straight-forward: toxicity and allergenicity as well as metabolites, which are associated with the protein´s function, are checked for possible hazards. On the other hand, some unpredictable, unintended effects, "which could materialize as a consequence of gene insertion, from random mutations that take place during the transformation and tissue culture process, or from pleiotropic effects of the introduced protein" [5], may be overlooked by targeted approaches. In other words, transgene insertion may influence the expression of hundreds or thousands of neighboring and downstream genes from the insertion site [103]. Beside this, conventional crossing is used to incorporate proven, transgenic traits into adapted crop varieties with different genetic backgrounds [21]. Therefore, the question arises, whether there are differences between transgenic maize and near-isogenic varieties, which can be detected in several varieties, so to speak, are consistent.

In this study, we focused on a non-targeted transcriptomics approach for the identification of possible unintended effects in herbicide-resistant maize (NK603 maize). The aim was to identify differences in the entire transcriptome between Roundup Ready maize and near-isogenic varieties, to identify gene candidates which may be consistently, differently expressed between Roundup Ready maize and near-isogenic varieties, independent of the genetic background and external influences, and to analyze, which biological functions are particularly affected by the modification of the gene expression.

Materials and Methods

Plant Material

Maize hybrids of the commercial GM varieties AG8025RR2 and AG9045RR2 (unique identifier MON-ØØ6Ø3-6 from Monsanto Company, carrying the NK603 insert which enables glyphosate herbicide tolerance) and the near-isogenic non-GM varieties AG8025 (conven-

tional counterpart of AG8025RR2) and AG9045 (conventional counterpart of AG9045RR2) were obtained from the Brazilian market (Sementes Agroceres). AG8025RR2 and its conventional counterpart AG8025 have the same genetic background since they are produced from the same endogamic parental lines. The same is for AG9045RR2 and its conventional counterpart AG9045. The AG8025 variety is the hybrid progeny of the single-cross between maternal endogamous line “A” with the paternal endogamous line “B”. Thus, the tested hybrid variety grains have high genetic similarity (AB genotype). The AG9045 variety is the hybrid progeny of the single-cross between maternal endogamous line “C” with the paternal endogamous line “D” resulting in CD genotype. Presence/absence of the transgenic NK603 insert was confirmed by qPCR in both, GM and conventional maize grains.

RNA extraction

Total RNA from pools of ten maize grains (of each variety) was isolated using a protocol of van Dijk (unpublished) which is based on [26,33]. Maize grains were homogenized in liquid nitrogen using mortar and pestle. 500mg of the homogenized grains was transferred into a 15ml tube and 5ml of 60°C warm extraction buffer (2% CTAB, 2% PVP [12], 2M NaCl, 0.9mM DEPC, 0.5mM spermidine, 100mM Tris [pH=8]) and 100µl of mercaptoethanol were added. After vortexing, 5ml of chloroform-isoamylalcohol (24:1; CIA) was added. The mixture was vortexed and centrifuged (15 min, 15°C, 3184 × g). The upper aqueous phase was mixed with 5ml of CIA and centrifuged again (15 min, 15°C, 3184 × g). This CIA extraction was repeated once. After the third centrifugation, 900µl of 10M lithium chloride with 1mM DEPC was added to the upper phase (in a new tube). The mixture was vortexed and stored in the refrigerator (4°C, o/n). After o/n incubation, the mixture was centrifuged (30 min, 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 new tube and 500µl of CIA was added. After vortexing, the mixture was centrifuged (10 min, 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 (approx. 1.5ml). Then, the samples were incubated on ice (5 min) and centrifuged (4°C, 35 minutes, 16363 × g). After eliminating the supernatant, 250µl of 75% ethanol was added and the mixture was vortexed shortly. The mixture was centrifuged (4°C, 10 minutes, 16363 × g) and the supernatant was removed. For drying the pellet, the tubes were incubated with open lids (10 min, rt), and 150µl of 10mM Tris (pH>7) was added. After an incubation step (65°C, 10 minutes), the RNA was dissolved and extracted RNA was purified (including DNA digestion) using the

RNeasy Mini Kit (Qiagen) and the RNase-Free DNase Set (Qiagen) following the manufacturer's instructions. Purified RNA was quantified by a fluorometer (Qubit). The integrity of the purified RNA intended for RNA sequencing was verified with an Agilent 2100 bioanalyzer (Agilent Technologies). RNA was stored at −80°C until use.

RNA sequencing and data analysis

Three pools of variety AG8025 conventional and three pools of variety AG8025RR2 were subjected to standard Illumina library preparation using the NEB Next Ultra RNA library prep kit according to the manufacturer's instructions. The resulting six cDNA libraries were paired end sequenced (125 bp) using an Illumina HiSeq2500 machine at the Vienna Biocenter Core Facilities (VBCF) NGS unit (http://www.vbcf.ac.at). Reads, which passed basic quality control (Illumina chastity filter), were preprocessed: Adapters were removed with cutadapt (http://cutadapt.readthedocs.io/en/stable/guide.html) and the reads were filtered against the rDNA using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), which is a very sensitive contaminants database. The remaining reads were paired end aligned with STAR (open source software) [34] against the B73 reference genome of *Zea mays* (AGPv3). Differential gene expression (DGE) analysis was performed with the statistic packages DESeq2, EdgeR, and CLC Genomics Workbench [35-37]. All genes with a $\log_2$fold change ≥ 1 and a FDR corrected p-value below 0.05 were considered to be significantly modulated. A $\log_2$fold change ≥ 1 means a $\log_2$fold value ≥ 1 (2-fold increase) or ≤ -1 (2-fold decrease), respectively. We selected this threshold for a better comparability to similar studies [21]. Only genes which showed a significant modulation (according to the above mentioned criteria) in all three statistic programs were considered for further data analysis. Functional classification of differentially expressed genes was based on gene ontology (GO) terms. To identify classes of genes that are over-represented in our gene set of differentially expressed genes, a singular enrichment analysis (SEA) was performed using the agriGO tool [38].

Confirmation of differentially expressed genes with RT-qPCR

The gene expression of 21 differentially regulated genes was assayed by reverse transcription coupled to real-time PCR (RT-qPCR) to validate the RNA sequencing results. A new set of three RNA pools, which was not used for RNA sequencing, was produced from AG8025 conventional and AG8025RR2, respectively, to test the biological and the technical reproducibility of the RNA sequencing results. In addition, three pools of varieties AG9045RR2 and its conventional counterpart AG9045, respectively, were generated to

test if the identified gene expression differences between AG8025 conventional and AG8025RR2 were also detectable between AG9045 conventional and AG9045RR2.

RT-qPCR was performed on a Rotor Gene Q (Qiagen) using the GoTaq® 1-Step RT-qPCR System (Promega) according to the manufacturer's instructions. Each reaction was performed in a 20µL final volume containing 10ng total RNA and 0.2µM of each target-specific primer. Ubiquitin was used as reference gene (F: 5'-AGACCCTGACTGGAAAAACC-3', R: 5'-CGACCCATGACTTACTGACC-3' [39]). Primers targeting differentially expressed genes were designed using Primer-BLAST [40]. Table 1 shows primer sequences, amplicon lengths, and gene IDs of primers targeting selected differentially expressed genes. Depicted primers were purchased from Sigma-Aldrich (Vienna, Austria).

Table 1. Primer sequences (forward and reverse) and Gene IDs used for the verification of differentially expressed genes. Amplicon lengths of all primer pairs are depicted in the right column.

Gene ID	Primer sequences for gene expression testing (forward and reverse)	Amplicon length
GRMZM2G026780	5`-ACCAGCGAACCAGAGATT-3'	292
	5`-CTTGCTTTCCAGCTCCTTATC-3'	
AC203989,4_FG001	5`-TACTCCATAGAGGAGGACAGGG-3'	161
	5`-CTGCACATCTTGAGCTCTCTCG-3'	
GRMZM2G092137	5`-GTTCTAGATCCGAGGATGACCTA-3'	175
	5`-ATGCTGAGCTGCTTGTTCT-3'	
GRMZM2G112538	5`-AGCGTCAGGCAGTTCAA-3'	109
	5`-CCTCGATGAGCGTGTTCT-3'	
GRMZM2G300424	5`-CTACCGTGCTCGATTCTTC-3'	364
	5`-GAGCGACAGGGAGAAATG-3'	
GRMZM2G472852	5`-GTCGCTCAAGAACTTCCTC-3'	257
	5`-ACTCGAACTCCATTTCTTAGAC-3'	
GRMZM2G005499	5`-GCGTCTCTGGTGTTCAT-3'	185
	5`-CAGAAGTCAAACTCCGTCTC-3'	
GRMZM2G112488	5`-GTGATGGACTGGCACAC-3'	182
	5`-GACTTGCACTCGCACTT-3'	
GRMZM2G316362	5`-TGTCGCTAGCTGTCAGTGTC-3'	149
	5`-CCAATCTGGGTTCCAAATCGTA-3'	
AC204711.3_FG003	5`-ACTCAACGTGCCCTATCT-3'	141
	5`-CATGCTAATTCAGACCTCACT-3'	
GRMZM2G117989	5`-GCGACGTACCACCTGTA-3'	190
	5`-CACGGTTCTTCACCCTTATG-3'	
GRMZM2G117971	5`-GAACAACTGGGACCTCAAC-3'	198
	5`-TGGTCCACGATCCTCAC-3'	
GRMZM2G112238	5`-ACTTCAGCAGACTCGCCTTC-3'	292
	5`-GATCCATCCGTCACCACTCC-3'	
GRMZM2G106445	5`-TTCTCTGCACCTCAGCACAA -3'	149

	5`-GACGCTCATCGCCACCA-3’	
GRMZM2G024527	5`-CAGCAGATACCGTGCATAC-3’	252
	5`-GAAAGCCCTCACTTTCTTTAATC-3’	
GRMZM2G386674	5`-AACCCCTCAAGGGAAAGGCTA-3’	123
	5`-GCTCAAGCTCGACGACCATC-3’	
GRMZM2G127948	5`-ACATGCTCATCAAGCTCATC-3’	162
	5`-TCGTGGTAGTTGAGGTAGTT-3’	
GRMZM2G018786	5`-AAGGATGGATGCTTGGATTT-3’	209
	5`-GCTTTCCATCTGCTCTCATAG-3’	
GRMZM2G425482	5`-ACGGGAGGGACATCTTCAT -3’	265
	5`-CACTAGACCAGAGAAGACGGAT-3’	
GRMZM2G034724	5`-CCAAGGTCGCCTACGTC-3’	231
	5`-CCCGTGAGCTGGAAGTT-3’	
GRMZM2G085964	5`-CGCAGCTGTTTTGGATCGAG-3’	201
	5`-AGCCGGCAATTAACAGACCA-3’	

All reactions were performed with a hold step at 37°C (15 min) followed by an initial denaturation step at 95°C (10 min), and 45 cycles of denaturation (95°C for 10 s), annealing (60°C for 30 s), and extension (72°C for 30 s) with a fluorescence measurement at the last step of each cycle. After every RT-qPCR the temperature was increased from 60°C to 99°C with fluorescence measurements at 1°C intervals to obtain a melting curve in order to determine the specificity of the reaction. The differential expression of selected genes was measured by comparing the RNA of three independent pools (“independent” means that each pool was created from ten different maize kernels) of GM maize grains with three independent pools of conventional maize grains. Each sample pool was tested in triplicate. For inhibition testing and for the evaluation of the efficiencies of the specific and reference gene PCR assays, standard curves were included in every run. Negative-control and reverse transcriptase (RT)-minus controls (reverse transcription reaction without addition of reverse transcriptase) were also checked in every run. The RT-qPCR with each specific primer was repeated on another day and the mean of the values was calculated.

Data were evaluated using the Rotor Gene Q Series software. Linearity (R2) and efficiency ($E = 10^{[-1/slope]}$) of every reaction were within the accepted values (PCR efficiencies between 90% and 110%). Relative quantitation was calculated using the $2^{-\Delta\Delta Ct}$ method [41]. Values were normalized using the reference gene ubiquitin and efficiency correction was performed as described by Pfaffl [42].

Results

Experimental design

The main objective of this study was to compare gene expression profiles of two different NK603 varieties with the profiles of the corresponding near-isogenic varieties which do not contain any transgenes, AG8025RR2 vs. AG8025 conventional and AG9045RR2 vs. AG9045 conventional. For this purpose we used high-throughput RNA sequencing as well as RT-qPCR. AG8025RR2 and AG8025 conventional were analyzed by RNA sequencing and the obtained transcriptome profiles were compared. DGE analysis was performed by three different statistic packages (DeSeq2, EdgeR, and CLC Genomics workbench [35-37]) and the results were further examined by RT-qPCR. In addition, we performed RT-qPCR to test, if some of the detected differences in gene expression are also present in the AG9045 pair (AG9045RR2 vs. AG9045 conventional). Finally, we identified biological functions of differentially expressed genes using GO annotations and performed a SEA to identify if special gene classes are overrepresented among the differentially expressed genes.

RNA sequencing and identification of differentially expressed genes

The HiSeq 2500 System (Illumina) was used for the transcriptome comparison of AG8025RR2 maize and its near-isogenic maize variety AG8025. Three cDNA libraries per variety were constructed. In average, the libraries yielded 93.6 M raw reads. 14.6% of the raw reads were removed due to adapter removing and cleaning (quality filtering and elimination of rRNA). 73.8% of the raw reads could be mapped uniquely to the B73 maize reference genome version AGPv3. 5.7% of the raw reads represented repetitive sequences, which aligned at multiple sites. 5.9% of the reads could not be mapped because they had too many errors to match the target sequence. Uniquely mapped reads corresponded to 31,333 and 31,073 unigenes in the transgenic and near-isogenic variety, respectively. Table 2 gives an overview of the run statistics and the bioinformatics data processing.

Table 2. Run statistics and data processing

Run statistics	AG8025RR2	AG8025 conv.
Raw reads	95.5 M	91.7 M
Uniquely mapped reads	71.7 M	65.9 M
Repetitive sequences	5.4 M	5.2 M
Not aligned	5.0 M	6.1 M
Median insert size (nt) (from aligned paired end reads)	107	128

Genes showing differential expression between AG8025RR2 and AG8025 conventional (log_2fold change ≥ 1 and a FDR-corrected p-value below 0.05) were identified using the R bioconductor packages DESeq2, EdgeR, and the CLC Genomics Workbench [35-37]. A total of 440 genes were identified as differentially expressed by DESeq2. 735 genes were calculated as differentially expressed by EdgeR and the CLC genomic workbench tool found 566 genes as differentially expressed. 286 genes were identified to have a significant difference in the gene expression by all the three programs and, therefore, these genes were used for further data analysis. Of these, 236 genes were up-regulated and 50 genes were down-regulated in the transgenic variety AG8025RR2. Table S1 of the supplementary material lists gene IDs and relative expression values of the 286 differentially expressed genes. Figure 1 shows the consistency of the DGE analysis using the three different statistic packages for data evaluation.

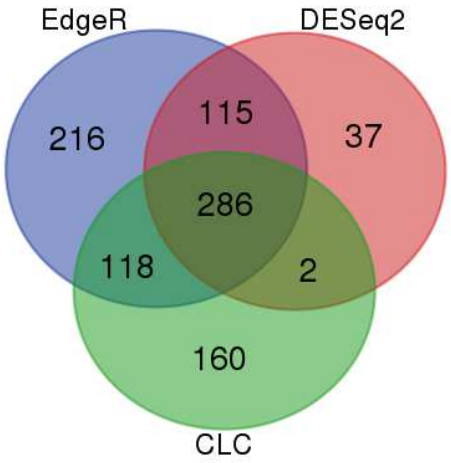


Figure 1. Venn diagram showing the number of differentially expressed genes identified by the statistic packages EdgeR (blue), DESeq2 (red), and CLC Genomics Workbench (green). 286 genes were consistently identified as differentially expressed.

Singular enrichment analysis

In order to identify if differentially expressed genes preferentially cluster into specific GO categories, we performed a SEA with the 286 differentially expressed genes using the agriGO tool [38]. A large part of the 286 genes was not characterized and for these genes, no GO category could be assigned. However, 188 of the 286 differentially expressed genes could be annotated by agriGO. The agriGO tool identified 52 significant GO terms which belong to biological processes and seven significant GO terms which belong to molecular functions (Table S2 of the supplementary material). We focused our analysis on biological processes of the differentially expressed genes. SEA results were imported to REVIGO (http://revigo.irb.hr/), a computational approach that summarizes long GO lists by reducing functional redundancies and visualizes the remaining GO terms [43]. By using REVIGO, the list of significant GO terms which belong to biological processes could be reduced to 34 GO terms, which are presented in Table 3.

Table 3. List of significantly overrepresented GO terms in the set of differentially expressed genes. Only GO terms from the category biological process were considered. The GO terms are sorted by the $\log_{10}$ p-value. Frequency means the proportion of this GO term in the underlying protein annotation database (http://www.uniprot.org/). Higher frequencies indicate more general terms and lower frequencies show more specific ones.

Go term ID	Description	Frequency [%]	Log_{10} p-value
GO:0009266	response to temperature stimulus	0.21	-3.70
GO:0006950	response to stress	4.58	-3.70
GO:0009628	response to abiotic stimulus	0.57	-3.02

GO:0009968	negative regulation of signal transduction	0.28	-2.92
GO:0050896	response to stimulus	12.21	-2.89
GO:0010033	response to organic substance	0.90	-2.72
GO:0065007	biological regulation	20.50	-2.34
GO:0010646	regulation of cell communication	0.93	-2.03
GO:0023051	regulation of signaling	0.93	-2.03
GO:0016311	dephosphorylation	1.25	-2.00
GO:0050794	regulation of cellular process	18.84	-2.00
GO:0010118	stomatal movement	0.01	-1.89
GO:0009719	response to endogenous stimulus	0.53	-1.89
GO:0006470	protein dephosphorylation	0.59	-1.64
GO:0044282	small molecule catabolic process	1.27	-1.55
GO:0019318	hexose metabolic process	0.64	-1.41
GO:0048519	negative regulation of biological process	1.98	-1.02
GO:0007154	cell communication	7.22	-1.02
GO:0019748	secondary metabolic process	0.14	-1.00
GO:0006066	alcohol metabolic process	0.42	-0.92
GO:0008152	metabolic process	75.39	-0.89
GO:0042398	cellular modified amino acid biosynthetic process	0.54	-0.89
GO:0005975	carbohydrate metabolic process	5.26	-0.89
GO:0055114	oxidation-reduction process	15.06	-0.77
GO:0006952	defense response	0.57	-0.77
GO:0048856	anatomical structure development	2.54	-0.74
GO:0016070	RNA metabolic process	15.95	-0.74
GO:0048583	regulation of response to stimulus	1.12	-0.74
GO:0032502	developmental process	2.81	-0.64
GO:0009607	response to biotic stimulus	0.34	-0.55
GO:0042742	defense response to bacterium	0.10	-0.39
GO:0065009	regulation of molecular function	1.73	-0.38
GO:0044237	cellular metabolic process	53.06	-0.36
GO:0006575	cellular modified amino acid metabolic process	0.79	-0.32

A treemap was generated for a better visualization and interpretation of the significantly overrepresented GO terms (Figure 2).

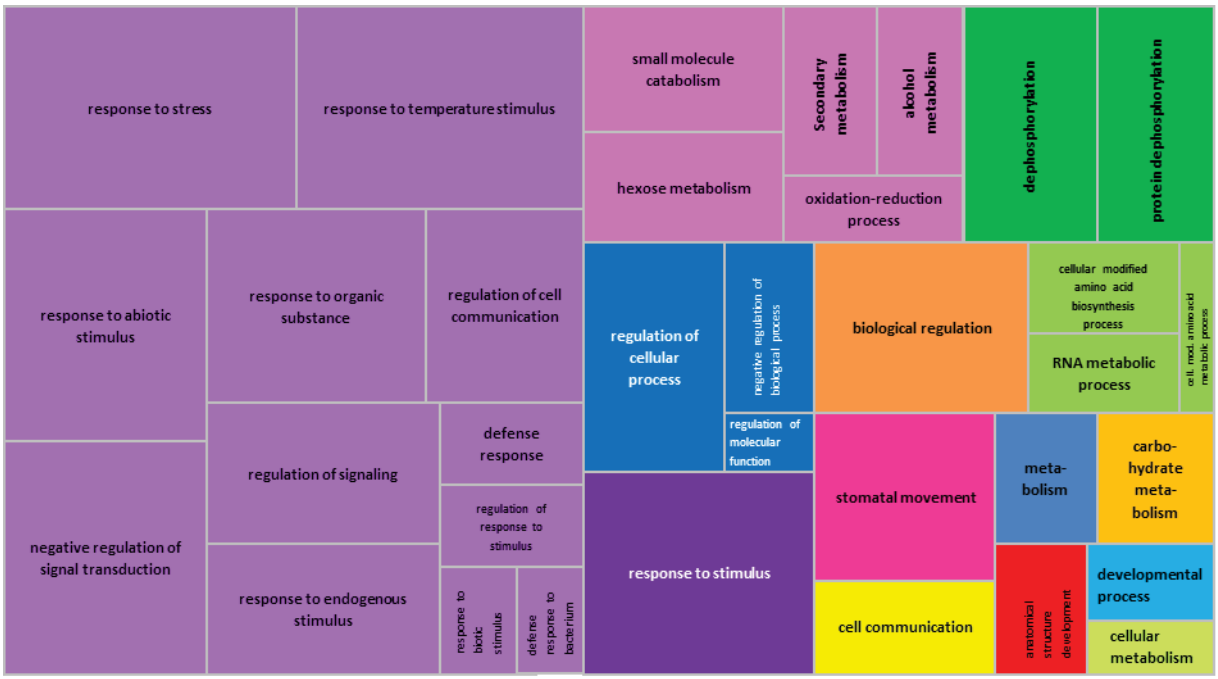


Figure 2: Each rectangle represents one significant GO term and semantically similar [44] GO terms have the same color. The size of a rectangle depends on the significance of the corresponding GO term. This treemap was generated using REVIGO.

The biological processes “response to abiotic stimulus” (GO:0009628), its subcategory “temperature stimulus” (GO:0009266), and the related category “response to stress” (GO:0006950), were significantly overrepresented among the 286 genes (see Table 3). Also signaling pathways, such as “negative regulation of signal transduction” (GO:0009968) and “regulation of signaling” (GO:0023051) were strongly represented. These processes are also clearly visible, when looking at the tree map (Figure 2): All five mentioned biological processes are located on the left side of the tree map and are colored purple, which means that they are semantically related (semantic similarity applied to GO annotations means functional similarity). On the left side of the tree map, the biggest rectangles (most significant) belong to ”response to stimulus” (GO:0050896) and “biological regulation” (GO:0065007). However, these biological processes are very frequent and broad, which makes interpretation difficult. Further noticeable are the pink (can be summarized as catabolic processes) and green area (phosphate metabolic processes) on the left side of the treemap, which can be summarized under “metabolic processes” (GO:0008152).

Verification of RNA sequencing results by RT-qPCR

For the verification of the RNA sequencing results, 14 up-regulated and seven down-regulated genes were tested by RT-qPCR. The RT-qPCR results of these 21 genes are

depicted in Table 4. The differences in gene expression could be verified for 16 of these 21 genes for AG8025RR2 vs. AG8025 conventional. For two genes (GRMZM2G106445 and GRMZM2G127948) the log_2FC was not ≥ 1, but very close to this value (0.91 and 0.81, repectively). For three down-regulated genes (GRMZM2G005499, GRMZM2G018786, and GRMZM2G034724), differential expression could not be verified.

Table 4. Verification of differentially expressed genes between varieties AG8025RR2 and AG8025 conventional and results of differentially expressed genes between varieties AG9045RR2 and AG9045 conventional by RT-qPCR. On the left side of the Table, gene IDs and corresponding GO term IDs including description are shown. On the right side of the Table, the log_2FC of the RNA sequencing experiment and the log_2FC of the RT-qPCR experiments are depicted. Results were generally counted as verified if the log_2FC was ≥ 1. RT-qPCR results were marked green if the log_2FC was verified, yellow if the log_2FC was ≥ 0.5 or ≤ -0.5 into the same direction, and red if the RNA sequencing results were not verified.

Upregulated Genes					
Gene ID	GO term ID	Description	log_2FC RNA seq AG8025RR2 vs AG8025 conv.	log_2FC RT-qPCR AG8025RR2 vs AG8025 conv.	log_2FC RT-qPCR AG9045RR2 vs AG9045 conv.
GRMZM2G316362	GO:0006631 (P)	fatty acid metabolic process	5.35	2.59	2.76
	GO:0016491 (F)	oxidoreductase activity			
	GO:0046914 (F)	transition metal ion binding			
	GO:0045300 (F)	acyl-[acyl-carrier-protein] desaturase activity			
	GO:0055114 (P)	oxidation-reduction process			
GRMZM2G112238	-	-	3.82	1.85	1.72
GRMZM2G026780	-	-	4.03	1.77	0.55
AC203989.4_FG001	GO:0055114 (P)	oxidation-reduction process	2.89	3.39	0.48
	GO:0047800 (F)	cysteamine dio-xygenase activity			
GRMZM2G092137	GO:0003700 (F)	DNA binding tran-scription factor activity	2.31	1.92	-0.06
	GO:0006355 (P)	regulation of tran-scription, DNA-templated			
	GO:0043565 (F)	sequence-specific DNA binding			
GRMZM2G112538	GO:0006952 (P)	defense response	2.37	2.45	-0.18

	GO:0009607 (P)	response to biotic stimulus			
AC204711.3_FG003	-	-	2.72	2.32	2.96
GRMZM2G112488	GO:0006952 (P)	defense response	1.98	2.57	-0.23
	GO:0009607 (P)	response to biotic stimulus			
GRMZM2G117989	GO:0080027 (P)	response to herbivore	1.33	1.80	-1.97
	GO:0009817 (P)	defense response to fungus, incompatible interaction			
	GO:0009627 (P)	systemic acquired resistance			
	GO:0009723 (P)	response to ethylene			
	GO:0042742 (P)	defense response to bacterium			
	GO:0009615 (P)	response to virus			
	GO:0009651 (P)	response to salt stress			
GRMZM2G117971	Same GO IDs as GRMZM2G117989		3.60	2.55	-2,67
GRMZM2G386674	-	-	2.83	2.23	-0.22
GRMZM2G425482	-	-	2.17	1.74	-0.34
GRMZM2G085964	GO:0003700 (F)	DNA binding transcription factor activity	1.89	1.44	1.89
	GO:0006355 (P)	regulation of transcription, DNA-templated			
	GO:0005634 (C)	nucleus			
GRMZM2G106445	-	-	3.72	0.91	-0.40

Downregulated Genes					
Gene ID	GO ID	Description	log_2FC RNA seq AG8025RR2 vs AG8025 conv.	log_2FC RT-qPCR AG8025RR2 vs AG8025 conv.	log_2FC RT-qPCR AG9045RR2 vs AG9045 conv.
GRMZM2G472852	GO:0009535 (C)	chloroplast thylakoid membrane	-1.30	-1.12	-0.01
	GO:0016020 (C)	membrane			
	GO:0009523 (C)	photosystem II			
	GO:0015979 (P)	photosynthesis			
GRMZM2G024527	GO:0005515 (F)	protein binding	-1.99	-1.37	-0.68
GRMZM2G300424	-	-	-2.24	-1.62	0.82

GRMZM2G005499	GO:0016021 (C)	integral component of membrane	-1.28	-0.27	-1.28
	GO:0016020 (C)	membrane			
GRMZM2G127948	GO:0032259 (P)	methylation	-1.88	-0.81	-1.28
	GO:0009809 (P)	lignin biosynthetic process			
	GO:0009805 (P)	coumarin biosynthetic process			
	GO:0016740 (F)	transferase activity			
	GO:0046686 (P)	response to cadmium ion			
	GO:0042409 (F)	caffeoyl-CoA O-methyltransferase activity			
	GO:0046872 (F)	metal ion binding			
GRMZM2G018786	GO:0016310 (P)	phosphorylation	-1.59	-0.14	0.07
	GO:0008152 (P)	metabolic process			
	GO:0016740 (F)	transferase activity			
	GO:0000166 (F)	nucleotide binding			
	GO:0016301 (F)	kinase activity			
	GO:0005524 (F)	ATP binding			
GRMZM2G034724	GO:0005507 (F)	copper ion binding	-1.43	-0.14	0.29
	GO:0008270 (F)	zinc ion binding			
	GO:0045735 (F)	nutrient reservoir activity			

(P) Biological Process, (F) Molecular Function

Identification of genes with differential expression between variety AG9045RR2 and AG9045

To check whether some of the identified significant differences in gene expression can also be detected in another NK603 variety vs. its comparator, the selected 21 genes were also tested for AG9045RR2 and AG9045 conventional by RT-qPCR (see Table 4). A comparable difference in the gene expression to the variety pair AG8025RR2 vs. AG8025 conventional could be measured in four genes (GRMZM2G316362, GRMZM2G112238,

AC204711.3_FG003, and GRMZM2G085964). Two genes (GRMZM2G026780 and GRMZM2G024527) did not meet the criteria for a $\log_2$FC ≥ 1, but showed clearly differential regulation into the right direction. 15 genes showed no, too low, or contrary differential expression. GO categories could be assigned for two genes of the four consistent differentially regulated genes, whereas no GO categories could be assigned for the other two genes. Interestingly, differential expression for genes, which could be assigned to defense/stress response and biotic stimulus were not confirmed for varieties AG9045RR2 and AG9045 conventional, whereas expression for genes, which could be assigned to fatty acid metabolic process, methylation, and lignin biosynthetic process were verified.

Measurement of NK603 grains vs. grains of the near-isogenic variety

30 grains of each variety were weighed to determine if there are significant weight differences between transgenic and conventional maize varieties. AG8025RR2 maize grains had an average weight of 0.321g (+/-0.027) and AG8025 conventional maize grains of 0.374g (+/-0.035g). The weight of the maize grains was significantly different (two-tailed t-test; $p < 0.05$). Also between AG9045RR2 maize grains (av. 0.342g +/- 0.020) and AG9045 conventional maize grains (av. 0.232g +/- 0.021), a significant weight difference was observed (two-tailed t-test; $p < 0.05$). However, the data did not match, in case of the AG8025 comparison, the conventional variety was heavier whereas in case of the AG9045 comparison, the transgenic variety was heavier. Figure 3 shows the average dry weight of the maize grains of all tested varieties.

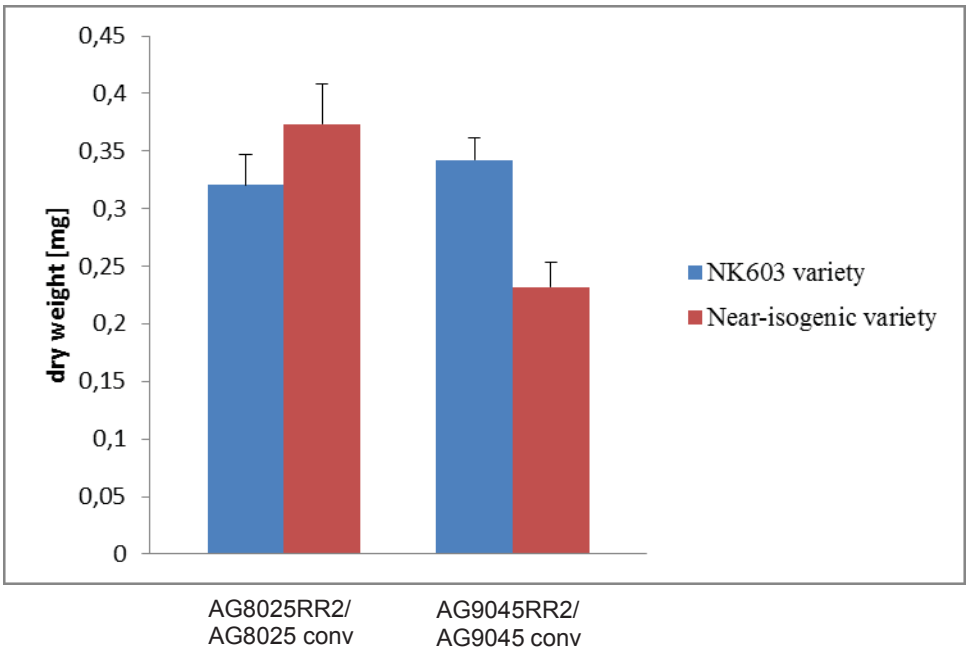


Figure 3. Dry weight of single maize grains of the four tested varieties. Vertical bars show the positive standard deviation.

Discussion and Conclusion

The employment of a non-targeted high-throughput RNA sequencing approach allowed a very deep analysis of an enormous amount of valuable information of the sequenced maize transcriptomes. Differentially expressed genes were detected with high stringency and false positives were limited by employment of three different statistic packages for the DGE analysis. We detected 286 genes, which were differentially expressed between the variety pair AG8025RR2 and AG8025 conventional (82.5% were up-regulated and 17.5% were down-regulated). This corresponded to ~0.91% (for AG8025RR2) and ~0.92% (for AG8025 conventional) of all detected maize genes, which is very similar to the results of la Paz et al. [21], who compared MON810 maize and near-isogenic maize by RNA sequencing. We also detected statistically significant differences in dry weight of the maize grains. However, these differences were negligible, because in variety pair AG8025RR2/AG8025 conventional, the maize grains of the conventional variety had significant more weight, whereas in variety pair AG9045RR2/AG9045 conventional, the maize kernels of the transgenic variety had significant more weight. Moreover, the dry weight of the transgenic maize grains was in the range of the conventional ones, and thus within natural variation.

21 differentially expressed genes in AG8025RR2 and AG8025 conventional were further analyzed by RT-qPCR and RNA sequencing results were confirmed for 16 of the 21 tested genes. RT-qPCR is very costly and time consuming. Thus, we focused mainly on high dif-

ferentially expressed genes as well as on genes from over-represented gene ontologies (see results of singular enrichment analysis). However, in some cases we could not find suitable primers for gene expression analysis. As we used a new set of RNA samples for validation with RT-qPCR, we confirmed technical and biological replication for these 16 genes. In addition, gene expression was tested in AG9045RR2 and AG9045 conventional by RT-qPCR and four of 21 tested genes showed significant differentially expression. These four genes showed significant differentially expression in both tested variety pairs (AG8025RR2/AG8025 conventional and AG9045RR2/AG9045 conventional). However, we are aware that there may be much more genes, which show consistent differentially expression in both variety pairs, and perhaps even in further variety pairs with different genetic background. For instance, in the aforementioned study of la Paz et al [21], ~75% of the significant differentially regulated genes between MON810 and near-isogenic maize were verified in two further variety pairs. To better understand the details of differentially expressed genes in NK603 maize, it would be interesting to perform RNA sequencing with AG9045RR2 and AG9045 conventional and perhaps further variety pairs as well, which will be one of our next steps.

In this study, we also analyzed the gene ontologies of the 286 differentially expressed genes. Over-represented gene classes among the 286 genes were identified by SEA. SEA analysis revealed 52 significantly over-represented biological processes and seven significantly over-represented molecular functions among the 286 genes. REVIGO reduced the list of 52 affected biological processes to 34 GO biological processes by reducing functional redundancies and visualized them as a tree map. In our transcriptomics study response to different stimuli, stress response, signaling pathways, and metabolic processes (which includes carbohydrate and protein metabolism) were significantly overrepresented, whereas in the MON810 transcriptomics study of la Paz et al. [21] carbohydrate metabolism, protein metabolism, and chromatin organization were significantly overrepresented. In the comparative proteomics and metabolomics studies of Agapito-Tenfen et al. [20,23] and Benevenuto et al. [25], differentially expressed proteins were also mainly assigned to carbohydrate and energy metabolism. Three genes, which were assigned to different stress responses, were also tested by RT-qPCR. Up-regulation of genes GRMZM2G112538, GRMZM2G112488, and GRMZM2G117989 was confirmed by RT-qPCR for AG8025RR2 vs. AG8025 conventional, but not for AG9045RR2 vs. AG9045 conventional. This implies that the differences in stress response are most likely not due to the NK603 transgene, but rather due to external influences.

If differences between a GM variety and its near-isogenic variety are not conserved among different variety pairs (which have different genetic backgrounds), it cannot be concluded that they are a direct consequence of the transgene [12]. It has been demonstrated that the majority of transcriptomic and proteomic differences are due to genotypic, environmental and crop treatment factors [23,26,27]. However, there are several studies where an impact of transgenes on gene expression is concluded [32,21,26]. Moreover, the impact on gene expression differs between transformation events. There is no consistent, functional pattern, on how gene expression is influenced by a transgene [32].

Genes which showed differential expression in both variety pairs were assigned to fatty acid metabolic process (GO:000663), lignin biosynthetic process (GO:0009809), oxidation-reduction process (GO:0055114), regulation of transcription, DNA-templated (GO:0006355), methylation (GO:0032259), response to cadmium ion (GO:0046686), and others. The gene GRMZM2G316362, for which significant up-regulation in the GM varieties was confirmed in both variety pairs, is involved in the fatty acid metabolic process (GO:000663). It codes for the enzyme stearoyl-acyl-carrier-protein desaturase9 (SAD9). Stearoyl-acyl-carrier-protein desaturase (SAD) genes affect the conversion of stearic acid (C18:0) to oleic acid (C18:1) in maize grains, and represent, therefore, a key-enzyme for improving maize oil quality [45]. Interestingly, the gene GRMZM2G316362 was also up-regulated in the comparison of MON810 maize and near-isogenic maize by la Paz et al [21]. Moreover, Agapito-Tenfen et al. [23] identified significant differences in stearoyl-acyl-carrier-protein desaturase1 (SAD1) between AG8025PRO (MON89034), which showed the highest expression level of SAD1, non-transgenic AG8025 conventional, AG8025RR2 (NK603) and AG8025PRO2 (MON89034 × NK603), using proteomic profiling. Taken together, the modulations in SAD1 and SAD9 indicate modifications of the fatty acid composition in some NK603, MON89034, and MON810 maize varieties.

Fatty acids are key nutrients, which are generally measured in the composition analysis of GMO comparative safety assessments of the European Union, which are performed by GMO producing companies and assessed by the European Food Safety Authority [46]. In this frame, statistically significant differences between NK603 maize and non-transgenic control maize were reported for oleic acid (C18:1), palmitic acid (C16:0), stearic acid (C18:0), and arachidic acid (C20:0) in two different field trials. The differences in fatty acid composition were within the natural variability [47]. For MON810, statistically significant differences were reported for palmitic acid (C16:0) content in one field trial, which was higher in MON810 maize compared to the control. These findings were not confirmed in other field trials [46]. It should be noted that in the mentioned NK603 and MON810 field

trials, in each single trial, samples were pooled from different field sites, which may mask differences between field sites.

A large number of different biological functions could be assigned to the 286 differentially expressed genes, which might be modified due to the NK603 transgene, but also due to external factors or genetic background. To confirm that these functions are affected due to the NK603 transgene, it would be necessary to test further NK603/non-NK603 variety pairs with different genetic backgrounds for differential expression of genes. In addition, the consistence of these effects needs to be tested under different environmental conditions.

The results of our comparative gene expression study can be utilized to identify further unintended effects resulting from transgene integration and associated processes, and to study identified changes in additional NK603 varieties. More research employing new profiling techniques is necessary to get a better impression of unintended effects and the underlying mechanisms. This will enable a better understanding of the whole biological plant gene network and the impact of transgenes on plant gene organization.

Acknowledgements

The material costs for this work were partially supported by the Hochschuljubiläumsstiftung (HJST) of the city of Vienna. The staff costs for this work were partially supported by the Austrian Research Promotion Agency (FFG) through FEMtech internships for the students.

Conflict of Interest

None.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

References

1. Kohli A, Twyman R, Abranches R, Wegel E, Stoger E, Christou P (2003) Transgene integration, organization and interaction in plants. Plant Mol Biol 52 (2):247-258. doi:10.1023/A:1023941407376
2. Schnell J, Steele M, Bean J, Neuspiel M, Girard C, Dormann N, Pearson C, Savoie A, Bourbonnière L, Macdonald P (2015) A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. Transgenic Res 24 (1):1-17. doi:10.1007/s11248-014-9843-7
3. Rischer H, Oksman-Caldentey K-M (2006) Unintended effects in genetically modified crops: revealed by metabolomics? Trends in biotechnology 24 (3):102-104. doi:https://doi.org/10.1016/j.tibtech.2006.01.009
4. Haslberger AG (2003) Codex guidelines for GM foods include the analysis of unintended effects. Nat Biotech 21 (7):739-741

5. Ladics GS, Bartholomaeus A, Bregitzer P, Doerrer NG, Gray A, Holzhauser T, Jordan M, Keese P, Kok E, Macdonald P, Parrott W, Privalle L, Raybould A, Rhee SY, Rice E, Romeis J, Vaughn J, Wal JM, Glenn K (2015) Genetic basis and detection of unintended effects in genetically modified crop plants. Transgenic Res 24 (4):587-603. doi:10.1007/s11248-015-9867-7
6. Fonseca C, Planchon S, Serra T, Chander S, Saibo NJ, Renaut J, Oliveira MM, Batista R (2015) In vitro culture may be the major contributing factor for transgenic versus nontransgenic proteomic plant differences. Proteomics 15 (1):124-134. doi:10.1002/pmic.201400018
7. Eckerstorfer M, Miklau M, Gaugitsch H (2014) Biosafety considerations for New Plant Breeding Techniques. Environment Agency Austria, Vienna
8. EFSA (2010) Guidance on the environmental risk assessment of genetically modified plants. EFSA Journal 8(11):1879. doi:10.2903/j.efsa.2010.1879
9. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. PLoS computational biology 13 (5):e1005457. doi:10.1371/journal.pcbi.1005457
10. OECD (2015) Molecular characterisation of plants derived from modern biotechnology. Safety Assessment of Foods and Feeds Derived from Transgenic Crops, vol 2. Paris
11. Coll A, Nadal A, Palaudelmas M, Messeguer J, Mele E, Puigdomenech P, Pla M (2008) Lack of repeatable differential expression patterns between MON810 and comparable commercial varieties of maize. Plant Mol Biol 68 (1-2):105-117. doi:10.1007/s11103-008-9355-z
12. Coll A, Nadal A, Collado R, Capellades G, Messeguer J, Mele E, Palaudelmas M, Pla M (2009) 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 Res 18 (5):801-808. doi:10.1007/s11248-009-9266-z
13. Coll A, Nadal A, Collado R, Capellades G, Kubista M, Messeguer J, Pla M (2010) Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. Plant Mol Biol 73 (3):349-362
14. Coll A, Nadal A, Rossignol M, Puigdomenech P, Pla M (2011) Proteomic analysis of MON810 and comparable non-GM maize varieties grown in agricultural fields. Transgenic Res 4:939 - 949
15. Poerschmann J, Gathmann A, Augustin J, Langer U, Gorecki T (2005) Molecular composition of leaves and stems of genetically modified Bt and near-isogenic non-Bt maize--characterization of lignin patterns. Journal of environmental quality 34 (5):1508-1518. doi:10.2134/jeq2005.0070
16. Saxena D, Stotzky G (2001) Bt corn has a higher lignin content than non-Bt corn. American journal of botany 88 (9):1704-1706
17. Herrero M, Ibanez E, Martin-Alvarez PJ, Cifuentes A (2007) Analysis of chiral amino acids in conventional and transgenic maize. Anal Chem 79 (13):5071-5077. doi:10.1021/ac070454f
18. Ben Ali SE, Madi ZE, Hochegger R, Quist D, Prewein B, Haslberger AG, Brandes C (2014) 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 15 (11):19898-19923. doi:10.3390/ijms151119898
19. Ben Ali S-E, Schamann A, Dobrovolny S, Indra A, Agapito-Tenfen SZ, Hochegger R, Haslberger AG, Brandes C (2018) Genetic and epigenetic characterization of the cry1Ab coding region and its 3′ flanking genomic region in MON810 maize using next-generation sequencing. Eur Food Res Technol. doi:10.1007/s00217-018-3062-z
20. Agapito-Tenfen SZ, Guerra MP, Wikmark OG, Nodari RO (2013) Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. Proteome Sci 11 (1). doi:10.1186/1477-5956-11-46

21. la Paz JL, Pla M, Centeno E, Vicient CM, Puigdomènech P (2014) The use of massive sequencing to detect differences between immature embryos of MON810 and a comparable non-GM maize variety. PLoS ONE 9 (6). doi:10.1371/journal.pone.0100895
22. Gullì M, Salvatori E, Fusaro L, Pellacani C, Manes F, Marmiroli N (2015) Comparison of Drought Stress Response and Gene Expression between a GM Maize Variety and a Near-Isogenic Non-GM Variety. PLoS ONE 10 (2):e0117073. doi:10.1371/journal.pone.0117073
23. Agapito-Tenfen S, Vilperte V, Benevenuto R, Rover C, Traavik T, Nodari R (2014) Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biology 14 (1):346
24. Mesnage R, Agapito-Tenfen SZ, Vilperte V, Renney G, Ward M, Séralini G-E, Nodari RO, Antoniou MN (2016) An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process. Scientific reports 6:37855. doi:10.1038/srep37855
25. Benevenuto RF, Agapito-Tenfen SZ, Vilperte V, Wikmark O-G, van Rensburg PJ, Nodari RO (2017) Molecular responses of genetically modified maize to abiotic stresses as determined through proteomic and metabolomic analyses. PLoS ONE 12 (2):e0173069. doi:10.1371/journal.pone.0173069
26. Barros E, Lezar S, Anttonen M, van Dijk J, Rohlig R, Kok E, Engel K (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnol J 8:436 - 451
27. Harrigan GG, Venkatesh TV, Leibman M, Blankenship J, Perez T, Halls S, Chassy AW, Fiehn O, Xu Y, Goodacre R (2016) Evaluation of metabolomics profiles of grain from maize hybrids derived from near-isogenic GM positive and negative segregant inbreds demonstrates that observed differences cannot be attributed unequivocally to the GM trait. Metabolomics 12:82. doi:10.1007/s11306-016-1017-6
28. Castan M, Ben Ali S-E, Hochegger R, Ruppitsch W, Haslberger AG, Brandes C (2017) Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing. Eur Food Res Technol 243 (3):353-365. doi:10.1007/s00217-016-2749-2
29. Shastry BS (2009) SNPs: impact on gene function and phenotype. Methods Mol Biol 578:3-22. doi:10.1007/978-1-60327-411-1_1
30. ISAAA (2016) Global Status of Commercialized Biotech/GM Crops: 2016. ISAAA. http://www.isaaa.org/purchasepublications/itemdescription.asp?ItemType=BRIEFS&Control=IB052-2016. Accessed 26.09. 2017
31. Heck GR, Armstrong CL, Astwood JD, Behr CF, Bookout JT, Brown SM, Cavato TA, DeBoer DL, Deng MY, George C, Hillyard JR, Hironaka CM, Howe AR, Jakse EH, Ledesma BE, Lee TC, Lirette RP, Mangano ML, Mutz JN, Qi Y, Rodriguez RE, Sidhu SR, Silvanovich A, Stoecker MA, Yingling RA, You J (2005) Development and Characterization of a CP4 EPSPS-Based, Glyphosate-Tolerant Corn Event. Crop Science 45 (1):329-339. doi:10.2135/cropsci2005.0329
32. Lambirth KC, Whaley AM, Blakley IC, Schlueter JA, Bost KL, Loraine AE, Piller KJ (2015) A Comparison of transgenic and wild type soybean seeds: analysis of transcriptome profiles using RNA-Seq. BMC biotechnology 15:89. doi:10.1186/s12896-015-0207-z
33. Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter 11 (2):113-116. doi:10.1007/BF02670468
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29 (1):15-21. doi:10.1093/bioinformatics/bts635
35. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15 (12):550. doi:10.1186/s13059-014-0550-8

36. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) 26 (1):139-140. doi:10.1093/bioinformatics/btp616
37. CLC Genomics Workbench. RNA seq analysis. http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/601/index.php?manual=RNA_Seq_analysis.html. Accessed 05.02. 2018
38. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res 45 (W1):W122-w129. doi:10.1093/nar/gkx382
39. Thiebaut F, Rojas CA, Grativol C, Motta MR, Vieira T, Regulski M, Martienssen RA, Farinelli L, Hemerly AS, Ferreira PC (2014) Genome-wide identification of microRNA and siRNA responsive to endophytic beneficial diazotrophic bacteria in maize. BMC genomics 15:766. doi:10.1186/1471-2164-15-766
40. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134. doi:10.1186/1471-2105-13-134
41. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif) 25 (4):402-408. doi:10.1006/meth.2001.1262
42. Pfaffl MW (2004) Quantification strategies in real-time PCR. In: Bustin SA (ed) A-Z of quantitative PCR. International University Line (IUL), La Jolla, CA, USA, pp 87 - 112
43. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE 6 (7):e21800. doi:10.1371/journal.pone.0021800
44. Pesquita C, Faria D, Falcao AO, Lord P, Couto FM (2009) Semantic similarity in biomedical ontologies. PLoS computational biology 5 (7):e1000443. doi:10.1371/journal.pcbi.1000443
45. Han Y, Xu G, Du H, Hu J, Liu Z, Li H, Li J, Yang X (2017) Natural variations in stearoyl-acp desaturase genes affect the conversion of stearic to oleic acid in maize kernel. Theor Appl Genet 130 (1):151-161. doi:10.1007/s00122-016-2800-5
46. EFSA (2009) Applications (EFSA-GMO-RX-MON810) for renewal of authorisation for the continued marketing of (1) existing food and food ingredients produced from genetically modified insect resistant maize MON810; (2) feed consisting of and/or containing maize MON810, including the use of seed for cultivation; and of food and feed additives, and feed materials produced from maize MON810, all under Regulation (EC) No 1829/2003 from Monsanto. The EFSA Journal 1149:1-85
47. Ridley WP, Sidhu RS, Pyla PD, Nemeth MA, Breeze ML, Astwood JD (2002) Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (Zea mays L.). J Agric Food Chem 50 (25):7235-7243

4. DISCUSSION

4.1. Genetic stability

Stacked events are gaining more and more importance in crop cultivation due to improved pest management, whereas single events often cause resistance of the target organisms [104]. Consequently, the use of single events is decreasing [6]. Due to this development, it is very important to perform an adequate risk assessment for stacked events. Also several scientists argue that stacked events should be regulated as conventional hybrids [105,10,106], which need no risk assessment; there are fundamental reasons against this. One important reason is that small DNA changes in the transgenes of stacked events, which might arise during conventional breeding, will be overlooked by the current approval process. Another safety risk is the potential interaction of combined GM traits [47].

In the EU, stacked events are considered as new GM plants and applicants have to perform a risk assessment for the approval of stacked events. However, in these risk assessments, applicants often refer to the data of the corresponding single events; especially for the verification of the genetic stability of transgenes present in stacked events, but also for the toxicological assessment of newly expressed proteins [107-109]. Thus, toxicological relevant issues, such as modified amino acids in the newly expressed protein due to point mutations in the coding region, might be overlooked. In view of this, it is of great importance to sequence the transgenes of stacked events, especially the coding regions and the flanking regions. We addressed this and investigated two of the most commonly used commercial transgenes (MON810 and NK603) in stacked and single events using HRM analysis and DNA sequencing.

4.1.1. MON810 analyses

The 3´flanking region of insect-resistant maize was studied in seeds of four different varieties, two stacked event varieties (MON88017 × MON810 and NK603 × MON810) and two single event varieties (MON810) [110,111]. The analyzed region included a part of the *cry1ab* coding region and the transition from the MON810 transgene to the genomic maize domain. Interestingly, DNA sequencing revealed some variants in the recombinant *cry1ab* gene. Most remarkable was a heterozygous C→T transition in the coding region (at position 71 of the analyzed amplicon), which was detected in individual samples of both, stacked event varieties and in one single event variety. Moreover, the tested stacked

events harbored markedly more variants as well as markedly higher frequencies of the variants than the analyzed single events. Hence, position 71 appeared to be a mutational hotspot. Further variants were detected, but these had very low frequencies (<2%). Thus, it cannot be excluded that these further variants were due to PCR or sequencing error.

To date, the theory of “instabilities through sequence homologies” [10] is the only explanation for an increased mutation rate in stacked transgenes. Possible other factors (e.g. age of the seeds) may contribute to genetic differences as well. In many plant species, genetic mutations and changes in chromosome structure in embryo meristems increase with the seeds age. The mutation rate can increase substantially in later developmental phases (seed senescence) [112]. Since the analyzed stacked events were older maize grains, this may explain the sequence variations that we observed. Further, since mutational hotspots can also arise through the pro-mutagenic activities of cytosine modifications such as methylation, deamination, and halogenation [89], we analyzed cytosine methylation patterns in the coding region. The results of these experiments are discussed in chapter 4.2.

The mutation frequency observed in our two MON810 studies exceeds the natural background mutation rate for maize endogenous genes. Jiao et al. [82] estimated the average single-base substitution rate for genic maize regions to be 4.79×10^{-8} per site per year. Clark et al. [113] reported a nucleotide substitution rate between $\sim 2.9 \times 10^{-8}$ and 3.3×10^{-8} for an endogenous maize gene, which is very similar to Jiao et al. [82]. Our variant frequencies markedly exceed the above mentioned mutation rates of Jiao et al. [82] and Clark et al. [113]. Further experiments with the investigated GM maize material, in which the background mutation frequency of endogenous maize genes is measured, would enable an appropriate comparison with the observed mutation frequency in the MON810 transgene.

Whether stacked transgenes influence genetic reorganization of plant genomes cannot be answered based on our results. However, it is clear that the variant frequencies are markedly different between single and stacked events. This observation is supported by the study of Agapito-Tenfen et al. [47], which reported that transgene transcript levels of stacked events had a significant reduction of about 34% when compared to single events. Moreover, there was some evidence of synergistic and antagonistic interactions in the maize genome due to transgene stacking, which might have influenced the gene expression profile.

The majority of the detected variants were transition mutations (C↔T or A↔G), which often create synonymous substitutions without altering the encoded amino acid sequence. Also the variant at position 71 was a synonymous substitution. This variant was located at the degenerated third position of the codon for glycine, hence resulting in no amino acid change of the coded protein (the codons GGC and GGU both code for the amino acid glycine). Many amino acids are encoded by more than one codon. Such codons are called synonymous codons [114]. Although synonymous codons encode the same amino acid, codon usage is not random. Codon usage influences gene expression levels, protein folding, and protein cellular function [115]. Thus, also silent mutations can have huge impact on proteins.

The recombinant *cry1ab* toxin of MON810 maize seems to be significantly stronger expressed compared to natural *cry1ab* toxins due to modified codon usage prior to transgene insertion [116]. Protein folding and toxin activation may also be influenced in the recombinant *cry1ab* toxin due to modified codon usage, but also because of major changes on the DNA level [79]. Synonymous codon usage patterns can vary markedly among genomes, but also within a genome. The evolutionary pressure on codon bias is explained by selection for translational efficiency, GC mutational bias, and by mRNA secondary-structure stability [115]. *Cry* genes are AT-rich compared to endogenous plant genes. This preference can lead to serious problems such as low expression, if a *cry* gene becomes inserted into a plant organism. To enable stable and strong transgene expression, codon optimization is commonly performed prior to transgene insertion [65]. This was also performed for the *cry1ab* gene; several A and T nucleotides were replaced by C and G nucleotides, which are preferred codon endings in *Zea mays* [116,117]. In view of this, it was remarkable that the variants which we have detected were mostly A and T variants, which does not reflect the distribution of nucleotides in the *Zea mays* background.

To my best knowledge, there are no other studies, which investigated the genetic stability of the MON810 transgene in stacked events. Zaulet et al. [74] analyzed the F2 generation of an unknown MON810 event and detected five point mutations in the *cry1ab* coding region. The analyzed region does not overlap with our analyzed region. Hence, it was not possible to compare our detected variants directly to the results of Zaulet et al. La Paz et al. [76] investigated the genetic stability of the MON810 event in 28 different single event varieties and detected no changes in the transgene. However, la Paz and colleagues investigated only three biological replicates (seed material) of each variety, while we analyzed at least 100 individual maize seeds per variety. Neumann et al. [29] tested 567 indi-

vidual maize seeds of a single event MON810 variety using real-time PCR with Scorpion primers in combination with Sanger sequencing. No changes were identified in the MON810 transgene and its flanking regions.

We analyzed the same single event variety as Neumann et al. [29] using another DNA sequencing method and identified the above mentioned C→T transition at position 71 of the coding region in two individual maize seed samples of this variety [111]. Instead of Sanger sequencing, we employed deep amplicon sequencing. Although Sanger sequencing is still considered to be the gold standard for the detection of DNA mutations, comparisons of the two methods clearly demonstrate that amplicon sequencing is much more sensitive. Low frequency variants (<10%) are usually not detected by Sanger sequencing, whereas amplicon sequencing enables already the detection of 1% frequencies [118,119]. Therefore, amplicon sequencing is more suitable for the analysis of genetic instabilities in GM plants.

It has to be clarified in further studies whether the stacking of transgenes is the only reason for the increased mutation rate in the tested stacked events. Further possible mechanisms have to be examined. Interestingly, the analysis of the genetic stability of stacked NK603 maize (see 4.1.2.) revealed no differences to the DNA sequence of the certified reference material. Thus, the observed difference in mutation frequency between single events and stacked events may be specific for MON810.

4.1.2. NK603 analyses

In contrast to MON810 maize, no studies about the genetic stability of NK603 maize were published after its authorization for import and processing in the EU in 2003. In order to fill this gap, the NK603 transgene and its flanking regions were screened for mutations using HRM analysis and Sanger sequencing. Two stacked NK603 events (NK603 x MON810) were investigated and the obtained DNA sequence was compared to the NK603 patent sequence [120]. Interestingly, we observed two nucleotides in the rice actin 1 promoter in all sequenced samples, which are not present in the published NK603 patent [120]. Moreover, these additional nucleotides were also present in certified NK603 reference material. These results might be explained by a sequencing error or a manual copy/paste error of the company that determined the patent sequence. Furthermore, it cannot be excluded that the identified insertions were introduced during crop breeding. However, an error in the patent sequence is more likely, because the reference material was also affected by

these insertions. DNA sequencing data which are submitted by GMO developers for the approval of new GM events often contain errors. New sequencing data of soybean event 305423 revealed a four base pair difference compared to the sequencing data originally submitted for soybean event 305423 [121]. The originally submitted sequencing data are from 2007 and the new sequencing data are from 2016. Thus, it took nine years for this mistake to be discovered. Moreover, the sequencing data of soybean event 40-3-2 were corrected four times, the last time in 2017 [122-125]. A similar incident is reported for GM maize GA21 [126]. These incidents, as well as our findings emphasize once again, how important the performance of additional sequencing analyses of transgenic constructs is, since the sequencing data provided by the biotech companies are not always reliable.

As with the nucleotide differences we detected in NK603 maize, none of the nucleotide differences reported for soybean event 305423 were located in the transgene coding region [121]. Nucleotide differences which are located in promoter regions (as in our case) may have an influence on the promoter activity and consequently also on transgene expression [127]. However, since transgene expression has already been reviewed as part of the risk assessment, it is not necessary to repeat this step.

In our NK603 study, the result of the NK603 transgene zygosity, which was determined by conventional PCR, was surprising. While the NK603 transgene was tested hemizygous in all samples of the stacked event 631RR2/Bt, most samples of the progenies of the stacked event DKC 26-79 were tested homozygous for the transgenic locus and none was tested null (absence of the NK603 transgene). The tested varieties belong to different generations. The stacked event 631RR2/Bt was directly tested for its zygosity at the transgenic locus (thus the F1 generation was tested). In case of stacked event DKC 26-79 only the progenies were tested for NK603 transgene zygosity (thus the F2 generation was analyzed). Assumed that the hemizygous DKC 26-79 individuals were self-pollinated, the last mentioned result does not match with Monsanto´s assertion that NK603 × MON810 progeny segregate according to the Mendelian genetics [128]. According to Mendelian genetics, a 3:1 ratio for herbicide-resistant and herbicide-sensitive progenies would be expected. Interestingly, US patent 6,057,496 reports similar findings, which are explained by a preferential selection for transgene-positive gametes in herbicide-resistant plants after herbicide treatment [129]. However, the frequency of herbicide-resistant progeny was much higher in our study (96%) compared to the frequency observed by Lauer et al. (68%) [129]. To elucidate this difference, it would be necessary to investigate further NK603 progenies from different varieties which were treated with different amounts of herbicide.

Our findings clearly demonstrate that additional (independent from the data which are submitted by the developers) molecular characterization of transgenic inserts is of major importance for adequate risk assessment and for correct proof of GMOs in food and feed. However, without public access to transgene sequence data and appropriate GMO materials, independent research in GMO biosafety will remain rare [130].

4.2. DNA methylation patterns in the MON810 transgene

Due to the detection of a mutational hotspot in the *cry1ab* gene of the MON810 transgene described in chapter 4.1.1., we analyzed cytosine methylation patterns in this region in a stacked MON810 event and in two MON810 single events. The observed methylation status was very low in all tested varieties. Therefore, a connection with the mutational hotspot is unlikely. Moreover, no significant differences were identified among the tested varieties or between samples with or without SNPs. These results indicate that the partial instability of the *cry1ab* locus is most likely not influenced by the cytosine methylation status.

Our observed cytosine methylation level (average total methylation: 0.83%) is slightly lower compared to the results of Vilperte et al. [90] and la Paz et al. [76]. La Paz et al. reported average total methylation of 1.2% in the *cry1ab* coding region. Unfortunately, la Paz and colleagues did not specify the distribution of symmetrical and asymmetrical methylation patterns. Vilperte et al. [90] did not analyze MON810 maize, but other Bt maize varieties (single MON-89034-3 and stacked MON-89034-3 × MON-00603-6). They announced the distribution of the different methylation motifs (symmetrical CpG and CHG and asymmetrical CHH), but did not report average total methylation. However, a rough estimate based on the methylation levels depicted in a figure of the mentioned study is possible. The average total methylation of the analyzed *cry2ab2* region is in the range of ~1.1% and for the analyzed *cry1A.105* region ~1.5%. In our study, the distribution of the different methylation motifs was 0.86% CpG, 0.78% CHG, and 0.86% CHH. Thus, CpG methylation and CHH methylation were the most frequent patterns. In the study of Vilperte et al., CpG methylation is most frequent (1.8% in the *cry2ab2* region and ~2% in the *cry1A.105* region). In contrast to our results, where no significant differences were identified, Vilperte and colleagues reported significant differences for CpG methylation levels of the *cry2ab2* region and the FMV promoter region among tested single event varieties. Further, significant differences in CpG methylation levels of the 35S promoter were reported between single and stacked events. Though, all reported methylation levels are low and, therefore,

it is unlikely that they cause genetic instability. Moreover, transgene silencing is also very unlikely because of the low methylation levels.

Vilperte et al. [90] and la Paz et al. [76] also used bisulfite sequencing for the analysis of cytosine methylation levels. In contrast to our high-throughput bisulfite sequencing method, they used cloning of PCR products coupled to conventional Sanger sequencing, which is much more time-consuming and gives a lower output. La Paz et al. [76] analyzed only one plant per variety and five bacteria colonies of every plant sample. Vilperte et al. [90] investigated at least three plants per variety and eight bacteria colonies of every plant sample. We analyzed eight samples per variety and obtained between 4612 and 49062 sequencing reads per sample. Cloning is superfluous, when high-throughput next generation sequencing is employed. Thus, our bisulfite sequencing approach, which included targeted next generation sequencing, enabled a markedly larger output.

One limitation of our methylation analysis was that the CLC Genomics Workbench, which was used for data evaluation, cannot differentiate between bisulfite-induced changes and genetic variants. Usually, unmethylated cytosines are deaminated to uracile and methylated cytosine (5-mC) residues remain unaffected. In our case, we frequently detected a C→T variant [110,111] at the same position and the program was not able to recognize this variant. Instead, a thymine was counted as unmethylated cytosine at this position. However, since we analyzed these samples before (without prior bisulfite conversion), we were able to determine all genetic variants present in our samples. Nevertheless, for future studies, a program which can distinguish bisulfite-induced changes from genetic variants would be preferred. Since bisulfite-induced C-to-T variants exhibit a guanine on the opposing strand, whereas genetic C-to-T variants exhibit an adenine instead, a distinction is generally possible [131]. Up to date, the only program which considers this difference is Bis-SNP [132].

Our MON810 methylation study provides valuable information on the methylation status of commercial transgenes and demonstrates the advantages of high-throughput bisulfite sequencing in this context. Although, we did not identify the cause of the genetic instability of the *cry1ab* gene, our results may be valuable for plant breeders and GM plant regulators as hardly any data on DNA methylation patterns in commercial GM plants are available.

4.3. Investigation of possible unintended effects in NK603 maize

Possible unintended effects of the commercial NK603 transgene were studied on the transcriptomic level using high-throughput RNA sequencing and RT-qPCR. The use of non-targeted new profiling approaches (such as RNA sequencing) in the comparative assessment of GM plants can be a valuable addition to usual targeted approaches, such as compositional, phenotypic and agronomic analyses (see chapter 1.2.1.). Moreover, to my best knowledge, there are no public studies in which commercial NK603 maize was studied by RNA sequencing.

RNA sequencing of the Brazilian NK603 variety AG8025RR2 and its near-isogenic variety AG8025 conventional resulted in 95.5 million and 91.7 million raw reads, respectively. Of these, 73.8% could be mapped uniquely to the B73 maize reference genome version AGPv3. Compared to other transcriptomics studies, our RNA sequencing was a very deep analysis. RNA sequencing yielded about 80 times more raw reads than in the transcriptomics study of la Paz et al. [43] and about eight times more raw reads than in the RNA sequencing study of Lambirth et al. [133]. Since the maize genome has about 2.3 Gb [92], it is important to perform a deep sequencing analysis. Otherwise, a large part of the maize genes (~40,000 protein-coding genes [134]) would not be captured. Our mapped reads corresponded to 31,333 and 31,073 unigenes in the transgenic and near-isogenic variety, respectively. La Paz et al. [43] could map their reads to ~14,800 maize genes. Further, we detected 286 genes which were differentially expressed between the variety pair AG8025RR2 and AG8025 conventional (82.5% were up-regulated and 17.5% were down-regulated). This corresponded to ~0.91% (for AG8025RR2) and ~0.92% (for AG8025 conventional) of total detected maize genes, which is very similar to the frequencies reported by la Paz et al. [43] (0.94% for the GM variety and 0.95% for the non-GM variety).

Three different programs (DeSeq2, EdgeR, and CLC Genomics Workbench) were employed for the differential gene expression analysis, since their statistical calculation methods differ [135-137]. We only considered genes as significantly different regulated, if all three programs confirmed differential expression. Thus, our results are based on reliable statistics. La Paz et al. [43] and Lambirth et al. [133] used two different statistic packages for the differential gene expression analysis. Moreover, we employed RT-qPCR for the verification of the RNA sequencing results, which further supports our results.

Singular enrichment analysis of the 286 differentially regulated genes identified 52 significantly overrepresented gene ontology (GO) terms which belong to biological processes and seven significantly overrepresented GO terms which belong to molecular functions. While response to different stimuli, stress response, signaling pathways, but also metabolic processes (which includes carbohydrate and protein metabolism) were significantly overrepresented in our transcriptomics study, carbohydrate metabolism, protein metabolism, and chromatin organization were significantly overrepresented in the MON810 transcriptomics study of la Paz et al. [43]. In the comparative proteomics and metabolomics studies of Agapito-Tenfen et al. [46,47] and Benevenuto et al. [48], differentially expressed proteins were mainly assigned to carbohydrate and energy metabolism. Although Agapito-Tenfen et al. [46,47], Benevenuto et al. [48], and la Paz et al. [43] studied other varieties (and partially other transgenes) and performed their investigations on different molecular levels, there are still commonalities: A high percentage of differences between GM plants and non-GM plants can be assigned to metabolic processes. Nevertheless, it has to be stated that metabolic process of plants is a very broad concept. It includes the chemical reactions and pathways (anabolism and catabolism) by which living plants transform chemical substances. Further, it involves small molecules as well as macromolecular processes [138].

Moreover, our study results and the results of la Paz et al. [43] and Agapito-Tenfen et al. [47] indicate an influence of transgenes on stearoyl-acyl-carrier-protein desaturase enzymes, which are involved in fatty acid metabolic processes. However, composition analyses in the frame of EU GMO risk assessment showed that differences in fatty acid composition were within the natural variability [139,140]. Furthermore, our detected differences in stress response genes were most likely not due to the NK603 transgene, but rather due to external factors. Generally, it can be assumed that environmental factors have a stronger impact on the transcriptome, proteome, and metabolome than the genetic modification [49,141,142].

The results from our comparative study demonstrate that the transcriptome profiles of transgenic plants can be significantly different from near-isogenic non-transgenic plants. This is not surprising, since the genotypes of transgenic plants can also be conceptually different from those obtained by classical breeding [7]. Whether the 286 differentially expressed genes are linked to the NK603 transgene or more to external factors (such as temperature or farming practice) needs to be elucidated in further experiments. Moreover, the natural variability of the differentially expressed genes needs to be determined. For

substantial differences which are outside the range of natural variation, biological relevance needs to be clarified. Future experiments should also consider that *in vitro* culture stress can cause transcriptomic alterations. The analysis of negative segregants (in addition to GM and non-GM plants) would provide valuable information on this detail.

Our results can be utilized to identify further unintended effects resulting from transgene integration and associated processes, and to study identified changes in additional NK603 varieties. In order to get a better impression of unintended effects and underlying mechanisms, it would be appropriate to continue using new profiling techniques in GM plant risk assessment. If several OMIC levels are covered by new profiling technologies, the results are even more meaningful. This would enable a better understanding of the whole biological plant gene network and the impact of transgenes on plant gene organization.

5. CONCLUSION

Our study results confirm that genetic engineering can lead to unintended effects in plants, thereby changing plant organization on different molecular levels. Although we did not identify large mutations in the analyzed transgenes, our studies on genetic stability clearly demonstrate that single nucleotide polymorphisms are present. Especially stacked MON810 maize varieties seem to be affected by a mutational hotspot, which, however, is most likely not related to cytosine methylation patterns. As long as it is not clear whether stacked events are genetically more unstable than single events, they should be examined separately in GM plant risk assessments. Further, sequencing data submitted by biotech companies for authorization of GM plants are not always reliable. In order to prevent serious errors in risk assessments (especially in toxicological assessments), additional analyses of transgenic inserts and their flanking regions are desirable.

The transcriptomic comparison of NK603 maize with near-isogenic conventional maize revealed several significantly different expressed genes. Since some of these were also verified in an additional NK603 variety and its near-isogenic non-GM counterpart, there may be some gene candidates, which are consistent differentially expressed in NK603 maize varieties in general. The identification and study of such genes enables a better understanding of plant gene organization and additionally provides supplement information for GM plant risk assessments.

Over the last years, many new molecular biology techniques, such as next generation sequencing, were developed. Although several techniques have huge potential for GM plant risk assessment, only few of these techniques are employed for this purpose. Our studies on genetic stability, epigenetics, and transcriptomics clearly demonstrated that next generation sequencing is well suited for the characterization of GM plants. Targeted next generation sequencing, high-throughput bisulfite sequencing as well as RNA sequencing provide valuable information, which expands the current knowledge of GM plants.

The studies performed in the frame of this PhD project enable a deeper inside into the genetics, epigenetics, and transcriptomics of frequently marketed GM maize varieties, which are important for GM plant risk assessment and plant breeders.

6. INDICES

6.1. List of Abbreviations

bp	base pairs
Bt	*Bacillus thuringiensis*
cDNA	copy DNA
CRISPR-Cas9	Clustered Regulatory Interspaced Short Palindromic Repeats – Cas9
DNA	deoxyribonucleic acid
DSB	double-strand break
EC	European Commission
EFSA	European Food Safety Authority
ERA	Environmental risk assessment
EU	European Union
FMV	Figwort mosaic virus
Gb	giga base pairs
GM	genetically modified
GMO	genetically modified organism
GO	Gene ontology
HDR	homology directed repair
HRM	High Resolution Melting
mRNA	messenger RNA
NHEJ	non-homologous end-joining
NOS	Nopaline synthase
NPBT	New Plant Breeding Techniques
PCR	Polymerase chain reaction
RNA	ribonucleic acid
RT-qPCR	reverse transcriptase quantitative PCR
SDN	site-directed nuclease
SEA	Singular enrichment analysis
SNP	single nucleotide polymorphism
Ti	Tumor-inducing

6.2. List of Figures

6.3. List of References

1. Jackson DA, Symons RH, Berg P (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc Natl Acad Sci U S A 69 (10):2904-2909
2. Cohen SN, Chang ACY (1973) Recircularization and Autonomous Replication of a Sheared R-Factor DNA Segment in Escherichia coli Transformants. Proceedings of the National Academy of Sciences of the United States of America 70 (5):1293-1297
3. Jaenisch R, Mintz B (1974) Simian Virus 40 DNA Sequences in DNA of Healthy Adult Mice Derived from Preimplantation Blastocysts Injected with Viral DNA. Proceedings of the National Academy of Sciences of the United States of America 71 (4):1250-1254
4. Bevan MW, Flavell RB, Chilton M-D (1983) A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. Nature 304:184. doi:10.1038/304184a0
5. Horsch RB, Fry JE, Hoffmann NL, Eichholtz DA, Rogers SG, Fraley RT (1985) A simple and general method for transferring genes into plants. Science 227 (4691):1229-1231. doi:10.1126/science.227.4691.1229
6. ISAAA (2016) Global Status of Commercialized Biotech/GM Crops: 2016. ISAAA. http://www.isaaa.org/purchasepublications/itemdescription.asp?ItemType=BRIEFS&Control=IB052-2016. Accessed 26.09. 2017
7. Nielsen KM (2003) Transgenic organisms—time for conceptual diversification? Nature Biotechnology 21:227. doi:10.1038/nbt0303-227
8. Transgen (2016) MON810: Anbau noch in elf EU-Ländern erlaubt. http://www.transgen.de/zulassung/2226.mon810.html. Accessed 10.04. 2018
9. Taverniers I, Papazova N, Bertheau Y, De Loose M, Holst-Jensen A (2008) Gene stacking in transgenic plants: Towards compliance between definitions, terminology, and detection within the EU regulatory framework. Environ Biosaf Res 7 (4):197-218. doi:10.1051/ebr:2008018
10. Weber N, Halpin C, Hannah LC, Jez JM, Kough J, Parrott W (2012) Editor's Choice: Crop Genome Plasticity and Its Relevance to Food and Feed Safety of Genetically Engineered Breeding Stacks. Plant Physiol 160 (4):1842-1853. doi:10.1104/pp.112.204271
11. Khan F (2012) Biotechnology Fundamentals. Taylor & Francis Group, London, New York
12. Gao C, Nielsen KK (2013) Comparison between Agrobacterium-mediated and direct gene transfer using the gene gun. Methods Mol Biol 940:3-16. doi:10.1007/978-1-62703-110-3_1
13. Schnell J, Steele M, Bean J, Neuspiel M, Girard C, Dormann N, Pearson C, Savoie A, Bourbonnière L, Macdonald P (2015) A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. Transgenic Res 24 (1):1-17. doi:10.1007/s11248-014-9843-7
14. Park SY, Vaghchhipawala Z, Vasudevan B, Lee LY, Shen Y, Singer K, Waterworth WM, Zhang ZJ, West CE, Mysore KS, Gelvin SB (2015) Agrobacterium T-DNA integration into the plant genome can occur without the activity of key non-homologous end-joining proteins. The Plant journal : for cell and molecular biology 81 (6):934-946. doi:10.1111/tpj.12779
15. Griffiths A, Gelbart W, Miller J (1999) Modern Genetic Analysis. In. W. H. Freeman, New York,
16. Mumm RH (2013) A look at product development with genetically modified crops: Examples from maize. J Agric Food Chem 61 (35):8254-8259. doi:10.1021/jf400685y
17. Nannas NJ, Dawe RK (2015) Genetic and Genomic Toolbox of Zea mays. Genetics 199 (3):655-669. doi:10.1534/genetics.114.165183

18. Wang K, B. Frame, Y. Ishida, Komari T (2009) Maize transformation. In: Maize Handbook: Genetics and Genomics. Springer, New York, pp 609–639
19. Somers DA, Makarevitch I (2004) Transgene integration in plants: poking or patching holes in promiscuous genomes? Curr Opin Biotechnol 15 (2):126-131. doi:10.1016/j.copbio.2004.02.007
20. Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. The EMBO journal 17 (20):6086-6095. doi:10.1093/emboj/17.20.6086
21. Davis AJ, Chen DJ (2013) DNA double strand break repair via non-homologous end-joining. Translational cancer research 2 (3):130-143. doi:10.3978/j.issn.2218-676X.2013.04.02
22. Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S, Bechtold N, Cruaud C, DeRose R, Pelletier G, Lepiniec L, Caboche M, Lecharny A (2002) T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites. EMBO Reports 3 (12):1152-1157. doi:10.1093/embo-reports/kvf237
23. Li Y, Rosso MG, Ülker B, Weisshaar B (2006) Analysis of T-DNA insertion site distribution patterns in Arabidopsis thaliana reveals special features of genes without insertions. Genomics 87 (5):645-652. doi:https://doi.org/10.1016/j.ygeno.2005.12.010
24. Schneeberger RG, Zhang K, Tatarinova T, Troukhan M, Kwok SF, Drais J, Klinger K, Orejudos F, Macy K, Bhakta A, Burns J, Subramanian G, Donson J, Flavell R, Feldmann KA (2005) Agrobacterium T-DNA integration in Arabidopsis is correlated with DNA sequence compositions that occur frequently in gene promoter regions. Functional & Integrative Genomics 5 (4):240-253. doi:10.1007/s10142-005-0138-1
25. Kim SI, Veena, Gelvin SB (2007) Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. The Plant journal : for cell and molecular biology 51 (5):779-791. doi:10.1111/j.1365-313X.2007.03183.x
26. Lambirth KC, Whaley AM, Schlueter JA, Bost KL, Piller KJ (2015) CONTRAILS: A tool for rapid identification of transgene integration sites in complex, repetitive genomes using low-coverage paired-end sequencing. Genomics data 6:175-181. doi:10.1016/j.gdata.2015.09.001
27. De Buck S, De Wilde C, Van Montagu M, Depicker A (2000) T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by Agrobacterium-mediated transformation. Mol Breed 6 (5):459-468. doi:10.1023/A:1026575524345
28. Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R (2003) A comprehensive characterization of single-copy T-DNA insertions in the Arabidopsis thaliana genome. Plant Mol Biol 52 (1):161-176. doi:10.1023/A:1023929630687
29. Neumann G, Brandes C, Joachimsthaler A, Hochegger R (2011) Assessment of the genetic stability of GMOs with a detailed examination of MON810 using Scorpion probes. Eur Food Res Technol 233 (1):19-30. doi:10.1007/s00217-011-1487-8
30. Haslberger AG (2003) Codex guidelines for GM foods include the analysis of unintended effects. Nat Biotech 21 (7):739-741
31. Yin Z, Plader W, Malepszy S (2004) Transgene inheritance in plants. Journal of applied genetics 45 (2):127-144
32. Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P (1999) Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. The Plant journal : for cell and molecular biology 17 (6):591-601
33. Kohli A, Twyman R, Abranches R, Wegel E, Stoger E, Christou P (2003) Transgene integration, organization and interaction in plants. Plant Mol Biol 52 (2):247-258. doi:10.1023/A:1023941407376

34. Shou H, Frame BR, Whitham SA, Wang K (2004) Assessment of transgenic maize events produced by particle bombardment or Agrobacterium-mediated transformation. Mol Breed 13 (2):201-208. doi:10.1023/b:molb.0000018767.64586.53
35. Fagard M, Vaucheret H (2000) (Trans)Gene Silencing in Plants: How many mechanisms? Annu Rev Plant Physiol Plant Mol Biol 51:167 - 194
36. Vaucheret H, Fagard M (2001) Transcriptional gene silencing in plants: targets, inducers and regulators. Trends in genetics : TIG 17 (1):29-35
37. Hernandez M, Pla M, Esteve T, Prat S, Puigdomenech P, Ferrando A (2003) A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard based on the 3'-transgene integration sequence. Transgenic Res 12 (2):179-189
38. Rosati A, Bogani P, Santarlasci A, Buiatti M (2008) Characterisation of 3'transgene insertion site and derived mRNAs in MON810 yield gard maize. Plant Mol Biol 67. doi:10.1007/s11103-008-9315-7
39. EFSA (2003) Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the safety of foods and food ingredients derived from herbicide-tolerant genetically modified maize NK603, for which a request for placing on the market was submitted under Article 4 of the Novel Food Regulation (EC) No 258/97 by Monsanto1 (QUESTION NO EFSA-Q-2003-002). The EFSA Journal 9:1-14
40. EFSA (2010) Scientific Opinion on an application (EFSA-GMO-RX-MON863) for renewal of the authorisation for continued marketing of existing feed materials, feed additives and food additives produced from maize MON863, under Regulation (EC) No 1829/2003 from Monsanto. The EFSA Journal 8 (3)
41. ISAAA (2018) Trait Introduction Method: Microparticle bombardment of plant cells or tissue.http://www.isaaa.org/gmapprovaldatabase/traitintro/default.asp?IntroMethodID=8&IntroMethod=Microparticle%20bombardment%20of%20plant%20cells%20or%20tissue. Accessed 23.02 2018
42. EFSA (2010) Guidance on the environmental risk assessment of genetically modified plants. EFSA Journal 8(11):1879. doi:10.2903/j.efsa.2010.1879
43. la Paz JL, Pla M, Centeno E, Vicient CM, Puigdomènech P (2014) The use of massive sequencing to detect differences between immature embryos of MON810 and a comparable non-GM maize variety. PLoS ONE 9 (6). doi:10.1371/journal.pone.0100895
44. Gullì M, Salvatori E, Fusaro L, Pellacani C, Manes F, Marmiroli N (2015) Comparison of Drought Stress Response and Gene Expression between a GM Maize Variety and a Near-Isogenic Non-GM Variety. PLoS ONE 10 (2):e0117073. doi:10.1371/journal.pone.0117073
45. Coll A, Nadal A, Rossignol M, Puigdomenech P, Pla M (2011) Proteomic analysis of MON810 and comparable non-GM maize varieties grown in agricultural fields. Transgenic Res 4:939 - 949
46. Agapito-Tenfen SZ, Guerra MP, Wikmark OG, Nodari RO (2013) Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. Proteome Sci 11 (1). doi:10.1186/1477-5956-11-46
47. Agapito-Tenfen S, Vilperte V, Benevenuto R, Rover C, Traavik T, Nodari R (2014) Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biology 14 (1):346
48. Benevenuto RF, Agapito-Tenfen SZ, Vilperte V, Wikmark O-G, van Rensburg PJ, Nodari RO (2017) Molecular responses of genetically modified maize to abiotic stresses as determined through proteomic and metabolomic analyses. PLoS ONE 12 (2):e0173069. doi:10.1371/journal.pone.0173069
49. Barros E, Lezar S, Anttonen M, van Dijk J, Rohlig R, Kok E, Engel K (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnol J 8:436 - 451
50. Sorochinskii BV, Burlaka OM, Naumenko VD, Sekan AS (2011) Unintended effects of genetic modifications and methods of their analysis in plants. Cytology and Genetics 45 (5):324. doi:10.3103/s0095452711050124

51. Bujnicki J, Dykstra P, Fortunato E, Heuer R-D, Keskitalo C, Nurse P, Villani C (2017) New techniques in Agricultural Biotechnology. Scientific Advice Mechanism (SAM). Brussels
52. Ladics GS, Bartholomaeus A, Bregitzer P, Doerrer NG, Gray A, Holzhauser T, Jordan M, Keese P, Kok E, Macdonald P, Parrott W, Privalle L, Raybould A, Rhee SY, Rice E, Romeis J, Vaughn J, Wal JM, Glenn K (2015) Genetic basis and detection of unintended effects in genetically modified crop plants. Transgenic Res 24 (4):587-603. doi:10.1007/s11248-015-9867-7
53. Eckerstorfer M, Miklau M, Gaugitsch H (2014) Biosafety considerations for New Plant Breeding Techniques. Environment Agency Austria, Vienna
54. EFSA (2011) Scientific Opinion on Guidance for risk assessment for food and feed from genetically modified plants. EFSA journal 9 (5). doi:10.2903/j.efsa.2011.2150
55. OECD (1993) Guidelines for the Testing of Chemicals. Organisation for Economic Cooperation and Development. OECD, PAris
56. FAO (1996) Biotechnology and Food Safety. Report of a Joint FAO/WHO Consultation. FAO Food and Nutrition Paper, vol 61. FAO,
57. Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Gatehouse AM, Karenlampi S, Kok EJ, Leguay JJ, Lehesranta S, Noteborn HP, Pedersen J, Smith M (2004) Unintended effects and their detection in genetically modified crops. Food Chem Toxicol 42 (7):1089-1125. doi:10.1016/j.fct.2004.02.003
58. OECD (2015) Molecular characterisation of plants derived from modern biotechnology. Safety Assessment of Foods and Feeds Derived from Transgenic Crops, vol 2. Paris
59. Kok EJ, Dijk JPv, Prins TW, Kleter GA (2010) Safety Assessment of GMO-derived Foods. Collection of Biosafety Reviews 5:82-111
60. EFSA (2018) EFSA scientific colloquium “Omics in risk assessment: state-of-the-art and next steps” https://www.efsa.europa.eu/en/events/event/180424-0. Accessed 23.04.2018
61. Horgan RP, Kenny LC (2011) ‘Omic’ technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist 13 (3):189-195. doi:10.1576/toag.13.3.189.27672
62. Ricroch A, Berge J, Kuntz M (2011) Evaluation of Genetically Engineered Crops Using Transcriptomic, Proteomic, and Metabolomic Profiling Techniques. Plant Physiol 155:1752 - 1761
63. EFSA (2014) Modern methodologies and tools for human hazard assessment of chemicals The EFSA Journal 12 (4)
64. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. PLoS computational biology 13 (5):e1005457. doi:10.1371/journal.pcbi.1005457
65. Muhammad Sarwar Khan, Ghulam Musatafa, Shahid Nazir, Joyia FA (2016) Applied Molecular Biotechnology: The Next Generation of Genetic Engineering. Plant molecular biotechnology: Applications of transgenics. Taylor & Francis Group, New York
66. De Schrijver A, Devos Y, Van den Bulcke M, Cadot P, De Loosed M, Reheul D, Sneyers M (2007) Risk assessment of GM stacked events obtained from crosses between GM events. Trends Food Sci Tech 18:101 - 109
67. Moch K (2006) Epigenetische Effekte bei transgenen Pflanzen: Auswirkungen auf die Risikobewertung. . BfN-Skripten 187:1-80
68. Broothaerts W, Corbisier P, Schimmel H, Trapmann S, Vincent S, Emons H (2008) A single nucleotide polymorphism (SNP839) in the adh1 reference gene affects the quantitation of genetically modified maize (Zea mays L.). J Agric Food Chem 56 (19):8825-8831. doi:10.1021/jf801636d
69. Windels P, De Buck S, Van Bockstaele E, De Loose M, Depicker A (2003) T-DNA Integration in Arabidopsis Chromosomes. Presence and Origin of Filler DNA Sequences. Plant Physiol 133 (4):2061-2068. doi:10.1104/pp.103.027532

70. Wilson AK, Latham JR, Steinbrecher RA (2006) Transformation-induced mutations in transgenic plants: Analysis and biosafety implications. Biotechnol Genet Eng Rev 23:209-237
71. Choffnes DS, Philip R, Vodkin LO (2001) A transgenic locus in soybean exhibits a high level of recombination. In Vitro CellDevBiol-Plant 37 (6):756-762. doi:10.1007/s11627-001-0125-8
72. Papazova N, Windels P, Depicker A, Taverniers I, Roldan-Ruiz I, Milcamps A, Van Bockstaele E, Van Den Eede G, De Loose M (2006) Sequence stability of the T-DNA - plant junctions in tissue culture in Arabidopsis transgenic lines. Plant cell reports 25 (12):1362-1368. doi:10.1007/s00299-006-0192-8
73. Morisset D, Demsar T, Gruden K, Vojvoda J, Stebih D, Zel J (2009) Detection of genetically modified organisms - closing the gaps. Nat Biotech 27 (8):700-701. doi:http://www.nature.com/nbt/journal/v27/n8/suppinfo/nbt0809-700_S1.html
74. Zaulet M, Rusu L, Kevorkian S, Luca C, Mihacea S, Badea EM, Costache M (2009) Detection and quantification of GMO and sequencing of the DNA amplified products. Romanian Biotechnological Letters 14 (5):4733-4746
75. Ogasawara T, Chikagawa Y, Arakawa F, Nozaki A, Itoh Y, Sasaki K, Umetsu H, Watanabe T, Akiyama H, Maitani T, Toyoda M, Kamada H, Goda Y, Ozeki Y (2005) 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 51 (2):197-201. doi:10.1248/jhs.51.197
76. la Paz JL, Pla M, Papazova N, Puigdomènech P, Vicient CM (2010) Stability of the MON 810 transgene in maize. Plant Mol Biol 74 (6):563-571. doi:10.1007/s11103-010-9696-2
77. Madi ZE, Brandes C, Neumann G, Quist D, Ruppitsch W, Hochegger R (2013) Evaluation of Adh1 alleles and transgenic soybean seeds using Scorpion PCR and HRM analysis. Eur Food Res Technol 237 (2):125-135. doi:10.1007/s00217-013-1969-y
78. BAC (2001) Advice on the reply of Syngenta of 28 July concerning the application Bt11 sweet maize under Regulation (EC) No. 258/97.
79. Latham JR, Love M, Hilbeck A (2017) The distinct properties of natural and GM cry insecticidal proteins. Biotechnology & genetic engineering reviews 33 (1):62-96. doi:10.1080/02648725.2017.1357295
80. R. Hull SNCPD (2000) Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate. Microbial Ecology in Health and Disease 12 (1):1-5. doi:10.1080/089106000435527
81. Mehrotra S, Goyal V (2014) Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. Genomics, Proteomics & Bioinformatics 12 (4):164-171. doi:10.1016/j.gpb.2014.07.003
82. Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44 (7):812-815. doi:10.1038/ng.2312
83. Ayarpadikannan S, Kim H-S (2014) The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. Genomics & Informatics 12 (3):98-104. doi:10.5808/GI.2014.12.3.98
84. Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Current opinion in genetics & development 15 (6):621-627. doi:10.1016/j.gde.2005.09.010
85. Rajeevkumar S, Anunanthini P, Sathishkumar R (2015) Epigenetic silencing in transgenic plants. Frontiers in Plant Science 6:693. doi:10.3389/fpls.2015.00693
86. Dupont C, Armant DR, Brenner CA (2009) Epigenetics: definition, mechanisms and clinical perspective. Seminars in reproductive medicine 27 (5):351-357. doi:10.1055/s-0029-1237423
87. Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, Myers CL, Vaughn MW, Springer NM (2013) Epigenetic and

genetic influences on DNA methylation variation in maize populations. Plant Cell 25 (8):2783-2797. doi:10.1105/tpc.113.114793
88. Vanyushin BF (2006) DNA methylation in plants. Current topics in microbiology and immunology 301:67-122
89. Sassa A, Kanemaru Y, Kamoshita N, Honma M, Yasui M (2016) Mutagenic consequences of cytosine alterations site-specifically embedded in the human genome. Genes and Environment 38 (1):17. doi:10.1186/s41021-016-0045-9
90. Vilperte V, Agapito-Tenfen SZ, Wikmark O-G, Nodari RO (2016) Levels of DNA methylation and transcript accumulation in leaves of transgenic maize varieties. Environmental Sciences Europe 28 (1):29. doi:10.1186/s12302-016-0097-2
91. Flint-Garcia SA (2017) Maize kernel development. CABI, Boston
92. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326 (5956):1112-1115. doi:10.1126/science.1178534
93. Giroux MJ, Shaw J, Barry G, Cobb BG, Greene T, Okita T, Hannah LC (1996) A single mutation that increases maize seed weight. Proceedings of the National Academy of Sciences 93 (12):5824-5829
94. Lusser M, Davies HV (2013) Comparative regulatory approaches for groups of new plant breeding techniques. New Biotechnology 30 (5):437-446. doi:https://doi.org/10.1016/j.nbt.2013.02.004
95. Maria Lusser, Claudia Parisi, Damien Plan, Rodríguez-Cerezo E (2011) New plant breeding techniques. State-of-the-art and prospects for commercial development. JRC Reference Reports.
96. Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346 (6213). doi:10.1126/science.1258096
97. Peng R, Lin G, Li J (2016) Potential pitfalls of CRISPR/Cas9-mediated genome editing. The FEBS journal 283 (7):1218-1231. doi:10.1111/febs.13586
98. Wang H, La Russa M, Qi LS (2016) CRISPR/Cas9 in Genome Editing and Beyond. Annu Rev Biochem 85:227-264. doi:10.1146/annurev-biochem-060815-014607
99. Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnology Advances 33 (1):41-52. doi:https://doi.org/10.1016/j.biotechadv.2014.12.006
100. Char SN, Neelakandan AK, Nahampun H, Frame B, Main M, Spalding MH, Becraft PW, Meyers BC, Walbot V, Wang K, Yang B (2017) An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. Plant Biotechnol J 15 (2):257-268. doi:10.1111/pbi.12611
101. Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K, Qiu JL, Gao C (2016) Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nature communications 7:12617. doi:10.1038/ncomms12617
102. Liang C, van Dijk JP, Scholtens IMJ, Staats M, Prins TW, Voorhuijzen MM, da Silva AM, Arisi ACM, den Dunnen JT, Kok EJ (2014) Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing. Analytical and Bioanalytical Chemistry 406 (11):2603-2611. doi:10.1007/s00216-014-7667-1
103. Qiagen Type-it HRM PCR Handbook. http://www.qiagen.com/resources/resourcedetail?id=a800ede3-2322-4341-a5c1-a9cbb6f630a3&lang=en. Accessed 23.04.2018
104. Storer NP, Thompson GD, Head GP (2012) Application of pyramided traits against Lepidoptera in insect resistance management for Bt crops. GM crops & food 3 (3):154-162. doi:10.4161/gmcr.20945
105. Kok EJ, Pedersen J, Onori R, Sowa S, Schauzu M, De Schrijver A, Teeri TH (2014) Plants with stacked genetically modified events: to assess or not to assess? Trends in biotechnology 32 (2):70-73. doi:10.1016/j.tibtech.2013.12.001

106. Steiner H-Y, Halpin C, Jez JM, Kough J, Parrott W, Underhill L, Weber N, Hannah LC (2013) Evaluating the Potential for Adverse Interactions within Genetically Engineered Breeding Stacks. Plant Physiol. doi:10.1104/pp.112.209817
107. EFSA (2013) Scientific Opinion on application EFSA-GMO-NL-2011-97 for the placing on the market of insect-resistant and herbicide-tolerant genetically modified cotton T304-40 for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Bayer CropScience AG. EFSA Journal 11 (6). doi:
10.2903/j.efsa.2013.3251
108. EFSA (2017) Assessment of genetically modified maize 1507 × 59122 × MON810 × NK603 and subcombinations, for food and feed uses, under Regulation (EC) No 1829/2003 (application EFSA-GMO-NL-2011-92). EFSA Journal 15 (11). doi:10.2903/j.efsa.2017.5000
109. EFSA (2017) Scientific Opinion on application EFSA-GMO-BE-2013-117 for authorisation of genetically modified maize MON 874273 x MON 890343 x NK603 and subcombinations independently of their origin, for food and feed uses, import and processing submitted under Regulation (EC) No 1829/2003 by Monsanto Company. EFSA Journal 15 (8). doi:
10.2903/j.efsa.2017.4922
110. Ben Ali SE, Madi ZE, Hochegger R, Quist D, Prewein B, Haslberger AG, Brandes C (2014) 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 15 (11):19898-19923. doi:10.3390/ijms151119898
111. Ben Ali S-E, Schamann A, Dobrovolny S, Indra A, Agapito-Tenfen SZ, Hochegger R, Haslberger AG, Brandes C (2018) Genetic and epigenetic characterization of the cry1Ab coding region and its 3′ flanking genomic region in MON810 maize using next-generation sequencing. Eur Food Res Technol. doi:10.1007/s00217-018-3062-z
112. D'Amato F (1997) Role of somatic mutations in the evolution of higher plants. Caryologia 50 (1):1-15. doi:10.1080/00087114.1997.10797380
113. Clark RM, Tavaré S, Doebley J (2005) Estimating a Nucleotide Substitution Rate for Maize from Polymorphism at a Major Domestication Locus. Molecular Biology and Evolution 22 (11):2304-2312. doi:10.1093/molbev/msi228
114. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH (2004) Codon usage between genomes is constrained by genome-wide mutational processes. Proceedings of the National Academy of Sciences of the United States of America 101 (10):3480-3485. doi:10.1073/pnas.0307827100
115. Angov E (2011) Codon usage: Nature's roadmap to expression and folding of proteins. Biotechnology Journal 6 (6):650-659. doi:10.1002/biot.201000332
116. van der Salm T, Bosch D, Honée G, Feng L, Munsterman E, Bakker P, Stiekema WJ, Visser B (1994) Insect resistance of transgenic plants that express modified Bacillus thuringiensis cryIA(b) and cryIC genes: a resistance management strategy. Plant Mol Biol 26 (1):51-59. doi:10.1007/bf00039519
117. Liu H, He R, Zhang H, Huang Y, Tian M, Zhang J (2010) Analysis of synonymous codon usage in Zea mays. Molecular biology reports 37 (2):677-684. doi:10.1007/s11033-009-9521-7
118. Arsenic R, Treue D, Lehmann A, Hummel M, Dietel M, Denkert C, Budczies J (2015) Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer. BMC Clinical Pathology 15:20. doi:10.1186/s12907-015-0020-6
119. Xu C, Nezami Ranjbar MR, Wu Z, DiCarlo J, Wang Y (2017) Detecting very low allele fraction variants using targeted DNA sequencing and a novel molecular barcode-aware variant caller. BMC genomics 18 (1):5. doi:10.1186/s12864-016-3425-4
120. Behr CF, Heck GR, Hironaka CH, You J (2012) Corn event PV-ZMGT32(nk603) and compositions and methods for detection thereof. US 8,273,959 B2,

121. EFSA (2017) Risk assessment of new sequencing information on genetically modified soybean event 305423. EFSA Journal 15 (8). doi:10.2903/j.efsa.2017.4967
122. Astwood JD, Hammond BG, Dobert RC, Fuchs RL (2000) Updated molecular characterization and safety assessment of Roundup Ready soybean event 40-3-2. Monsanto Technical Report MSL-16712. Monsanto, St. Louis, Missouri. USA
123. Lirette RP, Cavato TA, Reiser SE, Beazley KA, Borovkova IG, Higgins MR, Albee LE, Dixon DA (2000) Further characterization of glyphosate tolerant soybean event 40-3-2. Monsanto Technical Report: MSL-16646. Monsanto, St. Louis, Missouri
124. Windels P, Taverniers I, Depicker A, Van Bockstaele E, De Loose M (2001) Characterisation of the Roundup Ready soybean insert. Eur Food Res Technol 213 (2):107-112. doi:10.1007/s002170100336
125. EFSA (2017) Risk assessment of new sequencing information on genetically modified soybean event 40-3-2. EFSA Journal 15 (8):6. doi:10.2903/j.efsa.2017.4968
126. EFSA (2015) Risk assessment of new sequencing data of GM maize event GA21. EFSA Journal 13 (11). doi:doi:10.2903/j.efsa.2015.4296
127. Shastry BS (2009) SNPs: impact on gene function and phenotype. Methods Mol Biol 578:3-22. doi:10.1007/978-1-60327-411-1_1
128. EFSA (2005) Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the notification (Reference C/GB/02/M3/3) for the placing on the market of glyphosate-tolerant and insect-resistant genetically modified maize NK603 x MON810, for import and processing, under Part C of Directive 2001/18/EC from Monsanto1. The EFSA Journal 308:1-22
129. Lauer MJ, Leafgren RL, SM N (2009) Chemical selection
of resistant gametes of plants in the field.
130. Nielsen KM (2013) Biosafety Data as Confidential Business Information. PLOS Biology 11 (3):e1001499. doi:10.1371/journal.pbio.1001499
131. Bock C (2012) Analysing and interpreting DNA methylation data. Nat Rev Genet 13 (10):705-719. doi:10.1038/nrg3273
132. Liu Y, Siegmund KD, Laird PW, Berman BP (2012) Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. Genome Biology 13 (7):R61-R61. doi:10.1186/gb-2012-13-7-r61
133. Lambirth KC, Whaley AM, Blakley IC, Schlueter JA, Bost KL, Loraine AE, Piller KJ (2015) A Comparison of transgenic and wild type soybean seeds: analysis of transcriptome profiles using RNA-Seq. BMC biotechnology 15:89. doi:10.1186/s12896-015-0207-z
134. Law M, Childs KL, Campbell MS, Stein JC, Olson AJ, Holt C, Panchy N, Lei J, Jiao D, Andorf CM, Lawrence CJ, Ware D, Shiu S-H, Sun Y, Jiang N, Yandell M (2015) 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 Physiol 167 (1):25-39. doi:10.1104/pp.114.245027
135. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15 (12):550. doi:10.1186/s13059-014-0550-8
136. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) 26 (1):139-140. doi:10.1093/bioinformatics/btp616
137. CLC Genomics Workbench. RNA seq analysis. http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/601/index.php?manual=RNA_Seq_analysis.html. Accessed 05.02. 2018
138. Consortium TGO (2018) AmiGO2. http://amigo.geneontology.org/amigo/term/GO:0008152. Accessed 14.03. 2018
139. Ridley WP, Sidhu RS, Pyla PD, Nemeth MA, Breeze ML, Astwood JD (2002) Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (Zea mays L.). J Agric Food Chem 50 (25):7235-7243

140. EFSA (2009) Applications (EFSA-GMO-RX-MON810) for renewal of authorisation for the continued marketing of (1) existing food and food ingredients produced from genetically modified insect resistant maize MON810; (2) feed consisting of and/or containing maize MON810, including the use of seed for cultivation; and of food and feed additives, and feed materials produced from maize MON810, all under Regulation (EC) No 1829/2003 from Monsanto. The EFSA Journal 1149:1-85
141. Batista R, Fonseca C, Planchon S, Negrao S, Renaut J, Oliveira MM (2017) Environmental stress is the major cause of transcriptomic and proteomic changes in GM and non-GM plants. Scientific reports 7 (1):10624. doi:10.1038/s41598-017-09646-8
142. Harrigan GG, Venkatesh TV, Leibman M, Blankenship J, Perez T, Halls S, Chassy AW, Fiehn O, Xu Y, Goodacre R (2016) Evaluation of metabolomics profiles of grain from maize hybrids derived from near-isogenic GM positive and negative segregant inbreds demonstrates that observed differences cannot be attributed unequivocally to the GM trait. Metabolomics 12:82. doi:10.1007/s11306-016-1017-6

7. APPENDIX

7.1. Abstract

The cultivation of genetically modified (GM) plants, especially stacked events (containing at least two transgenes), is substantially increasing worldwide. Recombinant DNA techniques, such as genetic engineering, are very common in plant breeding. Genetic engineering can, like other breeding techniques, lead to unintended effects in the plant. The performance of adequate risk assessments is a very important issue to ensure biosafety. Since the first release of GM plants, several new molecular biology methods, such as next generation sequencing, were developed. These methods have the ability to provide new scientific information for GM plant risk assessments. In the era of next generation sequencing, we used this powerful tool, in addition to other appropriate molecular biology methods, to provide a deeper inside into the genetics, epigenetics, and transcriptomics of frequently marketed GM maize varieties. Insect-resistant and herbicide-resistant maize was investigated on different molecular levels.

We tested the genetic stability of transgenic inserts and their flanking regions, which is an essential safety criterion for the admission of GM plants in the EU. The 3´flanking region of insect-resistant maize was studied in seeds of four different MON810 maize varieties, two stacked event varieties and two single event varieties. The analyzed region included a part of the *cry1ab* coding region and the transition from the MON810 transgene to the genomic maize domain. Interestingly, DNA sequencing revealed some genetic variants in the recombinant *cry1ab* gene. The analyzed stacked events harbored markedly more variants as well as markedly higher frequencies of the variants than the analyzed single events. Since mutational hotspots can also arise through the pro-mutagenic activities of cytosine modifications such as methylation, deamination, and halogenation, we analyzed cytosine methylation patterns in the coding region. The observed methylation status was very low in all tested varieties. Therefore, a connection with the mutational hotspot was unlikely. Moreover, no significant differences were identified in the methylation status among the tested varieties or between samples with or without the mutational hotspot. These results indicate that the partial instability of the *cry1ab* locus is most likely not influenced by the cytosine methylation status.

We also studied herbicide-resistant maize for its genetic stability. Two stacked NK603 maize varieties were screened for mutations in the NK603 transgene and its flanking re-

gions. Interestingly, we found two nucleotides in all sequenced samples in the rice actin 1 promoter, which were not present in the published NK603 patent. Moreover, these additional nucleotides were also present in certified NK603 reference material. These results indicated a sequencing error or a manual copy/paste error of the company that determined the patent sequence. However, it cannot be excluded that the identified insertions were introduced during crop breeding.

Furthermore, possible unintended effects of the commercial NK603 transgene were studied on transcriptomic level using high-throughput RNA sequencing and RT-qPCR. Mapped reads corresponded to more than 30,000 unigenes. Transcriptome profiles of NK603 maize and near-isogenic maize were compared and differential gene expression analysis revealed 286 differentially expressed genes. Singular enrichment analysis (SEA) with the 286 differentially regulated genes was performed to filter out significantly overrepresented gene functions. SEA revealed that especially response to different stimuli, stress response, signaling pathways, and metabolic processes were significantly overrepresented among the 286 genes. The results of this study show that the transcriptomic profiles of GM plants can be significantly different from their near-isogenic non-GM comparators. Whether the 286 differentially expressed genes are mainly modified due to the NK603 transgene or more due to external factors (such as temperature or farming practice) needs to be elucidated in further experiments.

The studies performed in the frame of this PhD project addressed different biosafety issues of GM plants. The results provide valuable information for GM plant risk assessment and plant breeding companies. Given the rapid development of the biotechnology industry, the results are of vehement importance. Moreover, the approaches used for the characterization of GM plants can also be applied for the investigation of plants produced by new plant breeding techniques, which will play a crucial role in the future of agriculture.

7.2. Zusammenfassung

Der Anbau von gentechnisch veränderten (GV) Pflanzen, vor allem von sogenannten „Stacked Events“ (enthalten mindestens 2 Transgene), nimmt weltweit stark zu. Rekombinante DNA-Techniken, wie die Gentechnik, sind in der Pflanzenzüchtung sehr verbreitet. Gentechnische Veränderungen können, genauso wie andere Züchtungstechniken, zu unbeabsichtigten Effekten in der Pflanze führen. Die Durchführung angemessener Risikobewertungen ist ein sehr wichtiges Thema um die biologische Sicherheit zu gewährleisten. Seit der ersten Freisetzung von GV-Pflanzen wurden mehrere neue molekularbiologische Methoden, wie Next Generation Sequencing, entwickelt. Diese Methoden ermöglichen die Bereitstellung neuer wissenschaftlicher Informationen für Risikobewertungen von GV-Pflanzen. Wir nutzten diese vielversprechende Methode, neben anderen geeigneten molekularbiologischen Methoden, um die Genetik, Epigenetik und das Transkriptom von häufig vermarkteten GV-Maissorten näher zu erforschen. Insektenresistenter und herbizidresistenter Mais wurde auf verschiedenen molekularen Ebenen untersucht.

Im Rahmen der Dissertation haben wir die genetische Stabilität von transgenen Inserts und deren flankierenden Regionen getestet, was ein essentielles Sicherheitskriterium für die Zulassung von GV-Pflanzen in der EU darstellt. Die 3'-flankierende Region von insektenresistentem Mais wurde in Samen von vier verschiedenen MON810-Maissorten, zwei „Stacked Events“ und zwei „Single Events“, untersucht. Die analysierte Region umfasste einen Teil der *cry1ab*-codierenden Region und den Übergang vom MON810-Transgen zum natürlichen Maisgenom. Interessanterweise zeigte die DNA-Sequenzierung einige genetische Varianten im rekombinanten *cry1ab*-Gen. Die analysierten „Stacked Events“ enthielten deutlich mehr Varianten sowie deutlich höhere Frequenzen der Varianten als die analysierten „Single Events“. Da Mutations-Hotspots auch durch die pro-mutagenen Aktivitäten von Cytosin-Modifikationen wie Methylierung, Desaminierung und Halogenierung entstehen können, haben wir Cytosin-Methylierungsmuster in der kodierenden Region untersucht. Der gemessene Methylierungsstatus war jedoch in allen getesteten Sorten sehr niedrig. Daher war eine Verbindung mit dem Mutations-Hotspot unwahrscheinlich. Darüber hinaus wurden keine signifikanten Unterschiede im Methylierungsstatus bei den getesteten Sorten oder zwischen Proben mit oder ohne Mutations-Hotspot festgestellt. Diese Ergebnisse zeigen, dass die partielle Instabilität des *cry1ab*-Locus höchstwahrscheinlich nicht vom Cytosin-Methylierungsstatus beeinflusst wird.

Des Weiteren wurde auch herbizidresistenter Mais auf genetische Stabilität untersucht. Zwei „stacked“ NK603-Maissorten wurden auf Mutationen im NK603-Transgen und seinen flankierenden Regionen gescreent. Wir fanden interessanterweise zwei Nukleotide im Reis-Actin-1-Promotor in allen sequenzierten Proben, die im veröffentlichten NK603-Patent nicht vorhanden waren. Darüber hinaus waren diese zusätzlichen Nukleotide auch in zertifiziertem NK603-Referenzmaterial vorhanden. Diese Ergebnisse deuten auf einen Sequenzierungsfehler oder einen manuellen Kopier-/Einfügefehler der Firma hin, die die Patentsequenz bestimmte. Es kann jedoch nicht ausgeschlossen werden, dass die identifizierten Insertionen während der Pflanzenzüchtung entstanden sind.

Im Rahmen der Dissertation wurden auch mögliche unbeabsichtigte Wirkungen des kommerziellen NK603-Transgens auf Transkriptom-Ebene untersucht. Hierfür wurden Hochdurchsatz-RNA-Sequenzierung und RT-qPCR angewandt. Der beim Sequenzieren erhaltenen Nukleotidabfolge konnten mehr als 30.000 Maisgenen zugeordnet werden. Transkriptom-Profile von NK603-Mais und der Vergleichsmaissorte wurden verglichen und eine differentielle Genexpressionsanalyse identifizierte 286 unterschiedlich exprimierte Gene. Eine Singular Enrichment Analyse (SEA) wurde mit den 286 unterschiedlich regulierten Genen durchgeführt, um signifikant überrepräsentierte Genfunktionen herauszufiltern. Die SEA zeigte, dass insbesondere die Reaktion auf verschiedene Stimuli, Stressreaktionen, Signalwege und Stoffwechselprozesse unter den 286 Genen signifikant überrepräsentiert sind. Die Ergebnisse dieser Studie zeigen, dass die Transkriptom-Profile von GM-Pflanzen sich signifikant von ihren nahezu isogenen Vergleichspflanzen unterscheiden können. Ob die 286 unterschiedlich exprimierten Gene hauptsächlich aufgrund des NK603-Transgens oder eher aufgrund externer Faktoren (wie Temperatur oder landwirtschaftliche Praxis) verändert sind, muss in weiteren Experimenten geklärt werden.

Die im Rahmen dieses Promotionsvorhabens durchgeführten Studien befassten sich mit verschiedenen Fragen der biologischen Sicherheit von GV-Pflanzen. Die Ergebnisse liefern wertvolle Informationen für die Risikobewertung von GV-Pflanzen und Pflanzenzüchtungsunternehmen. Angesichts der rasanten Entwicklung der Biotechnologiebranche sind die Ergebnisse von großer Bedeutung. Darüber hinaus können die Ansätze zur Charakterisierung von GV-Pflanzen auch für die Untersuchung von Pflanzen genutzt werden, die durch neue Pflanzenzüchtungstechniken erzeugt werden und die in Zukunft eine entscheidende Rolle in der Landwirtschaft spielen werden.

7.3. Supplementary material for: Mutation Scanning in a Single and a Stacked Genetically Modified (GM) Event by Real-Time PCR and High Resolution Melting (HRM) Analysis

Figure S1. Amplification curves of the primary 5´ GT73 Scorpion PCR screening are presented. Genomic samples 14 (blue), 18 (red), and the no template control (ntc) (grey) are shown. The Ct values can be found in Table 1.

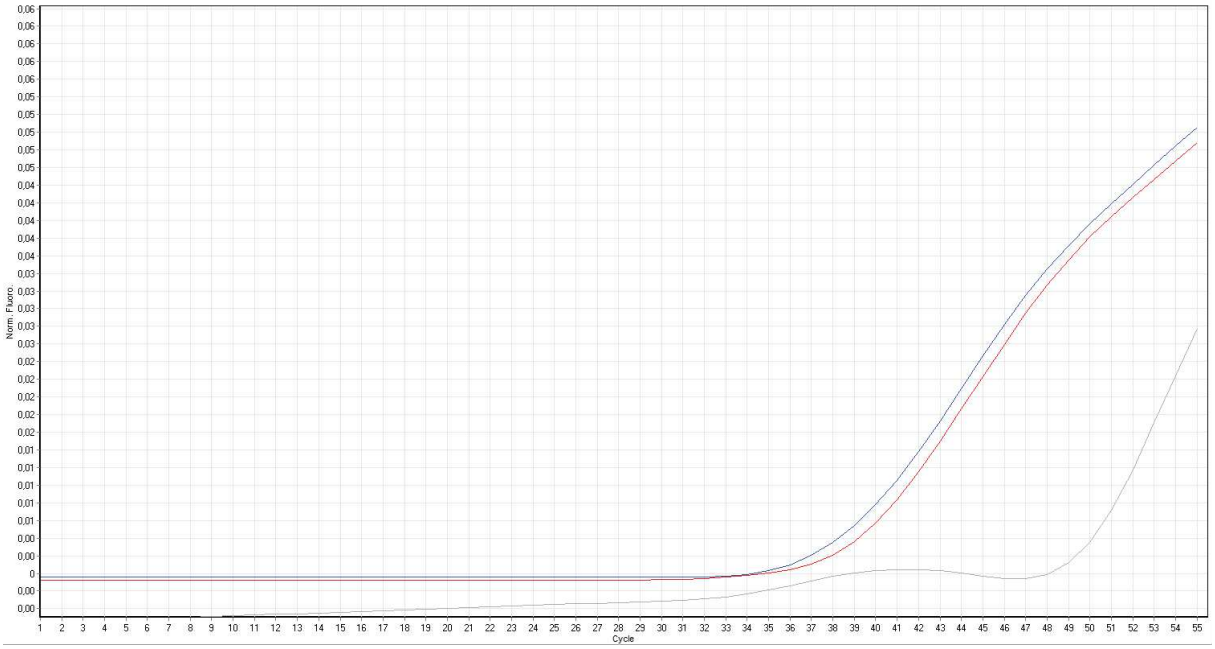


Figure S2. Amplification curves of the primary 3´ MON810 real-time PCR screening with subsequent HRM analysis are presented. Genomic samples 264 (green), 323 (blue), the reference = VT3 pool (red), and the ntc (grey) are shown. The Ct values are 29.26 (sample 264), 30.90 (sample 323), and 30.81 (VT3 pool).

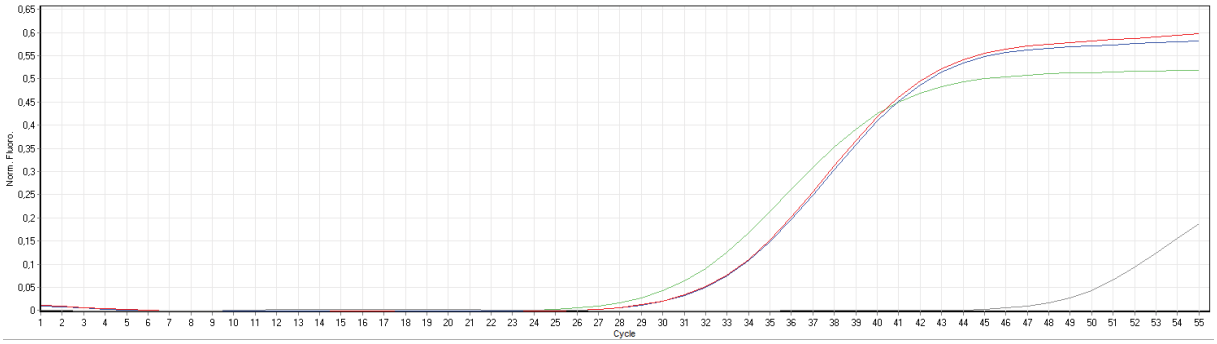


Figure S3. The normalized HRM graph of the 3´ MON810 HRM analysis is presented. Genomic samples 264 (green), 323 (blue), and the reference = VT3 pool (red) are shown. The confidence values [%] are 1.23 (sample 264), 97.03 (sample 323), and 100.00 (reference).

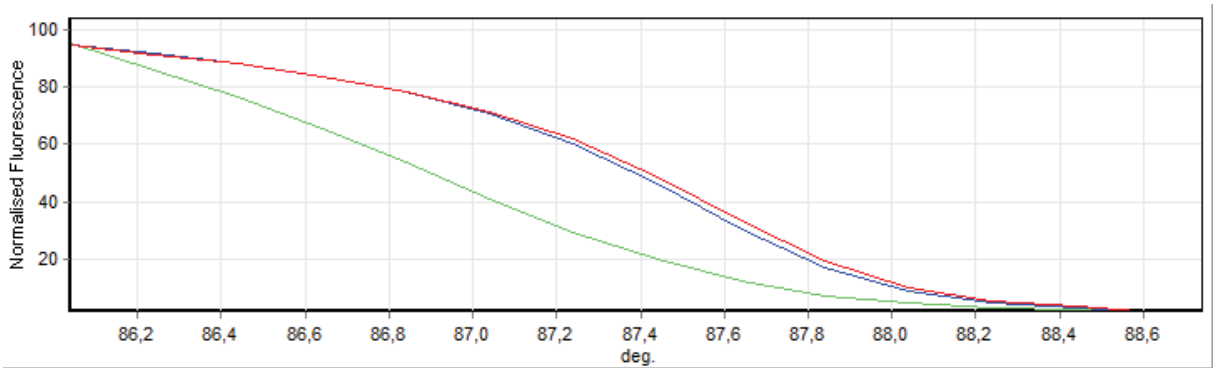


Figure S4. The difference HRM graph of the 3´ MON810 HRM analysis is presented. Genomic samples 264 (green), 323 (blue), and the reference = VT3 pool (red) are shown. The confidence values [%] are 1.23 (sample 264), 97.03 (sample 323), and 100.00 (reference).

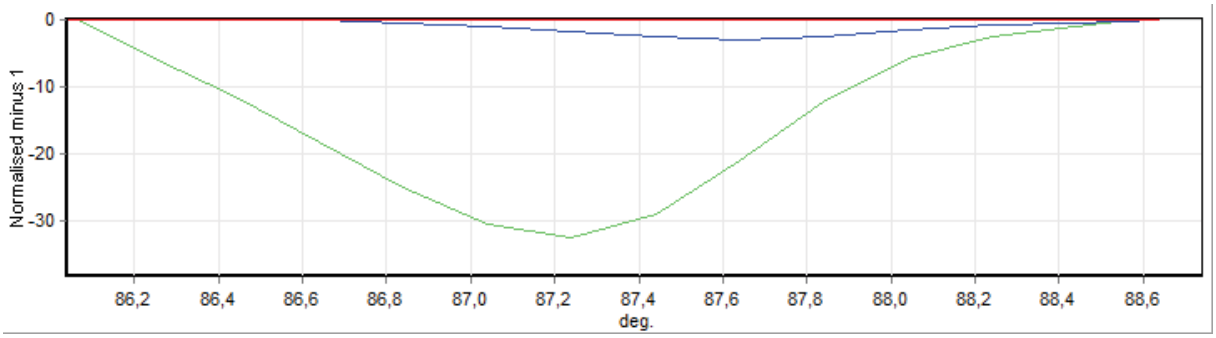


Table S4. All Ct and confidence values of selected samples of the primary screening of the 5′-GT73 region in oilseed rape samples by real-time PCR employing Scorpion primers and by HRM analysis are presented. This table is a complete version of Table 1.

Sample	Ct value	Confidence value
13	51.84	79.97
14	47.64	65.92
18	48.30	30.23
20	45.77	47.56
30	45.93	73.56

40	46.67	24.87
118	40.54	35.13
137	43.81	11.80
152	46.57	2.86
153	37.69	9.17
156	39.48	9.11
157	41.92	17.55
170	45.11	39.61
175	41.40	37.49
333	42.67	10.15

7.4. Supplementary material for: Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing

Table S1. Results of the first screening section for DKC 26-79 (F2) samples. Samples on the left (samples 1 to 183; first run) and on the right (samples 5 to 176; second run) were analyzed separately (two different runs which were repeated on a different day). We selected six samples (first run) and eleven samples (second run) which had confidence values below 10 respectively 32.

DKC 26-79 (F2) Samples	Mean Ct value	Mean Confidence value [%]	DKC 26-79 (F2) samples	Mean Ct value	Mean Confidence value [%]
1	19.5	37.3	5	18.0	91.1
3	19.8	56.2	9	20.8	82.7
4	19.6	84.5	10	29.7	39.4
6	19.2	40.8	15	18.9	72.2
7	28.3	50.1	20*	30.5	30.0
8	19.1	57.5	22*	28.1	3.4
11	20.8	35.0	57	19.4	74.6
12	20.3	72.8	64	19.5	99.2
13	19.4	64.5	65	18.8	66.9
14	20.1	52.7	67	19.6	93.8
16	19.5	95.5	68	18.4	72.2
17	20.1	81.5	69	19.0	59.5
18	19.7	92.6	70	19.2	72.8
19	20.2	93.9	71	19.1	97.9
21	20.3	94.9	72*	20.8	1.3
23	20.8	27.5	73	19.3	86.4
24	20.9	49.4	74	18.2	94.5
25	20.3	99.5	75*	25.9	3.8
26	20.2	44.9	77	18.9	89.2
27	19.0	98.8	78	20.0	83.7
29	20.5	12.4	79	20.4	54.7

31	19.4	71.4	80	19.2	94.3
32	20.5	96.0	81	18.5	98.6
33	20.8	33.5	82	17.9	86.0
34*	20.9	3.0	83*	27.8	1.5
35	21.0	21.5	84	19.2	92.4
36	21.2	51.8	85	18.5	88.8
37	19.1	25.8	86	18.3	68.4
38*	20.8	0.0	87	18.3	87.4
39	21.1	95.9	89	19.4	80.4
40	20.7	89.0	90	19.3	89.4
41	20.0	48.7	91	19.5	71.4
42	20.1	47.6	92	19.3	88.1
43	19.6	62.4	93	18.6	74.7
44	20.5	84.4	95	20.0	93.7
45	20.1	11.1	97*	27.5	0.0
46	20.4	81.3	104	19.4	96.9
49	19.8	74.9	105**	19.3	100.0
50	20.7	82.4	106	18.0	72.5
51	19.6	98.3	107	18.2	42.3
52	20.6	82.8	108	17.8	65.1
53	20.3	89.4	109	19.6	80.4
54	19.8	82.2	110*	19.5	17.4
55	21.2	46.7	112*	19.3	0.3
58	20.0	97.1	113	19.2	51.7
59*	20.6	2.9	114	19.4	89.1
60	29.9	10.2	115	19.1	64.0
61	19.7	7.7	116	19.5	86.4
62	20.2	7.6	117	19.4	97.3
63	19.3	9.8	118	18.2	86.8
96**	19.6	100.0	119	19.3	75.1
98*	25.1	1.8	138	19.0	98.1
99	19.9	98.5	155	19.0	62.0
100*	19.2	6.0	156	18.1	98.9
101	19.6	43.2	157	18.8	74.5
102	28.3	15.0	158	17.7	94.0
103	20.7	22.0	159	17.8	68.1
127	19.4	63.2	160	19.4	93.3
128	20.9	17.1	161*	32.2	2.3
129	20.9	98.3	162	18.7	46.6
130	19.2	41.0	163	18.8	99.5
131	20.7	92.3	164	18.7	39.0
133	20.9	8.6	165	18.5	77.5
177	20.6	14.8	166*	29.7	31.9
178	20.9	29.9	167	20.0	70.7
179	21.0	36.4	170*	18.0	31.0
180	19.7	9.9	171	18.7	42.3
181	19.5	56.5	174	19.9	85.3
182	20.7	36.4	175	18.4	52.0

183*	21.0	1.5	176	19.5	59.2

* Selected for sequencing; ** Reference

Table S2. Results of the first screening section for 631RR2/Bt (F1) samples. Samples on the left (samples 1 to 80; first run), in the middle (samples 81 to 152; second run) and on the right (samples 153 to 214; third run) were analyzed separately (three different runs which were repeated on a different day). We selected three samples (first and second run) and one sample (third run) which had confidence values below 76, 43 respectively 73.

631 RR2/Bt samples	Mean Ct value	Mean Confidence value [%]	631 RR2/Bt samples	Mean Ct value	Mean Confidence value [%]	631 RR2/Bt samples	Mean Ct value	Mean Confidence value [%]
1	18.1	88.3	81	18.5	87.3	153	18.7	91.5
2	19.2	96.9	82*	19.1	35.7	154	18.6	97.8
3	18.2	98.1	83	18.5	83.8	155	19.3	95.8
4	19.0	97.2	84	18.0	69.9	156	18.5	94.5
5	18.2	99.2	85	18.3	86.2	157	18.3	97.9
6	18.0	99.6	86	18.3	99.7	158	17.7	95.2
7	19.0	83.3	87	18.6	88.5	159	19.6	94.8
8	18.3	98.4	88	18.7	76.2	160	18.2	95.1
9	18.1	98.6	89	18.5	66.8	161	19.0	96.8
10	18.8	90.6	90	17.8	87.9	162	18.2	98.9
11	18.3	94.0	91	18.5	77.5	163	18.4	97.8
12	18.4	90.4	92	17.7	84.6	165	18.3	99.2
14	18.5	97.2	93	18.5	49.8	166	18.0	89.4
15	18.1	90.1	94	17.8	83.3	167	17.8	98.1
16	18.2	93.1	95	18.6	87.2	168	17.8	96.9
18	18.1	97.9	96	17.9	81.9	169	17.8	98.7
19	17.9	97.5	97	17.9	76.5	170	17.8	99.0
20	18.4	98.6	98	17.7	92.0	171	18.1	91.0
21	19.1	89.8	99	18.9	70.6	172	18.5	82.5
22	18.4	81.2	100	17.6	84.4	173	18.4	90.9
23	18.0	90.4	101	17.6	78.4	174	17.9	97.2
24	18.0	76.2	102	17.6	98.8	175	18.3	93.8
25	18.0	82.9	103	17.5	95.2	176	18.4	99.9
26*	18.1	75.6	104	18.2	86.8	177	18.3	98.9
29	18.1	95.4	105	17.5	51.3	178	18.4	97.3
30	18.1	95.4	106	17.4	94.3	179	18.3	90.0
31	18.0	99.3	107**	17.8	100.0	180	19.0	94.5
32	18.4	85.4	108	18.0	83.3	181	18.3	90.3
33	18.0	96.2	109	17.7	64.4	182	18.4	94.9
34	18.3	85.6	110	17.4	75.5	183	18.2	80.5
35	18.3	78.9	111	17.7	61.7	184	18.2	95.1
36	18.0	94.6	112	17.4	64.3	185	18.1	93.2
37	18.2	96.5	113	17.7	55.5	186	18.7	97.6
38	18.2	97.5	114	17.7	66.2	187	18.5	97.0

39	18.5	98.5	115*	17.7	42.9	188	18.0	99.7
40	18.3	98.3	116	17.5	62.0	189	18.4	99.1
41	18.2	81.4	117*	17.6	38.7	190	18.3	99.3
42*	18.2	72.5	119	17.5	55.9	191	18.2	99.3
43	18.3	93.4	120	17.4	53.1	192	18.3	98.8
44	17.9	94.4	121	17.7	51.3	193	20.1	87.8
45	18.0	93.3	122	17.5	75.5	194	18.2	99.7
46	18.0	89.8	123	17.7	68.7	195	18.3	99.5
48	17.8	76.1	125	17.4	66.2	196	19.8	97.2
50	17.8	91.5	126	17.6	59.9	197	18.2	95.5
51	18.1	77.1	127	17.4	71.1	198	18.5	86.5
55	17.9	83.3	128	17.8	99.1	199	18.1	93.3
56*	17.9	71.9	129	17.4	79.3	200	18.2	97.3
58	18.4	94.1	130	19.6	94.3	201	18.4	84.9
59	18.2	94.1	131	17.7	72.2	202	18.4	98.0
60	18.4	96.7	132	17.7	59.1	203	18.4	87.3
61	18.4	98.0	133	18.1	81.5	204	18.3	97.4
62	18.3	93.3	134	17.6	82.0	205*	18.4	72.9
63	18.2	94.3	135	17.6	62.4	206	18.9	95.7
64	18.4	83.5	136	17.8	68.5	207	18.4	98.6
65	18.4	96.3	137	18.0	64.4	208	17.9	98.8
66	18.3	97.1	138	17.5	59.8	209	18.3	98.0
67**	18.2	100.0	139	18.2	57.7	211	18.4	94.5
68	18.2	98.8	140	18.0	66.4	212	18.3	98.2
69	18.4	98.7	141	18.4	47.9	213	18.3	97.0
70	18.9	91.1	142	18.7	51.8	214**	18.2	100.0
71	18.2	93.1	143	18.4	55.4			
72	17.8	76.9	144	17.8	65.7			
73	18.4	85.5	145	17.5	57.7			
74	18.2	91.0	146	18.5	48.2			
75	18.8	94.2	147	17.6	52.2			
76	18.1	84.6	148	17.5	74.0			
77	18.0	86.4	149	18.2	57.1			
78	18.1	88.6	150	18.1	77.0			
79	18.1	98.1	151	18.0	86.1			
80	18.2	87.3	152	19.5	90.8			

* Selected for sequencing; ** Reference

Table S3. Results of the screening section 23 for DKC 26-79 (F2) samples. Samples on the left (samples 1 to 183; first run) and on the right (samples 5 to 176; second run) were analyzed separately (two different runs which were repeated on a different day). We selected five samples (first run) and three samples (second run) which had confidence values below 12 respectively 23.

DKC 26-79 (F2) Samples	Mean Ct value	Mean Confidence value [%]	DKC 26-79 (F2) samples	Mean Ct value	Mean Confidence value [%]
1	21.1	82.9	5	21.1	94.0
3	21.1	92.5	9	22.7	91.0
4	20.8	88.7	10	30.4	87.7
6	20.9	96.5	15	22.4	99.7
7	28.8	97.0	20	30.5	67.3
8	20.8	94.9	22	29.9	50.8
11	23.2	42.8	57	22.4	97.6
12	22.1	95.8	64	22.6	99.1
13	21.0	88.2	65	23.1	89.4
14	22.2	89.4	67	22.5	98.6
16	20.9	90.6	68	21.3	95.8
17	21.2	88.9	69	21.9	84.6
18	20.8	85.6	70	22.1	97.8
19	20.8	90.4	71	22.4	98.3
21	22.1	97.9	72*	22.9	2.5
23	22.6	49.2	73	23.0	99.6
24	22.1	52.5	74	21.1	98.5
25	22.1	99.0	75	27.5	92.1
26	22.0	67.7	77	22.0	96.5
27	20.8	89.1	78	23.7	95.7
29	23.3	27.0	79	22.1	84.3
31	20.8	74.6	80	21.9	98.2
32	21.2	94.7	81	21.4	95.9
33	22.4	20.2	82	20.9	99.5
34	22.4	16.0	83	29.5	82.7
35	22.3	71.4	84	21.9	98.4
36	22.3	86.7	85	21.8	94.9
37	20.9	62.5	86	21.4	99.7
38*	22.4	0.0	87	21.7	91.4
39	22.6	82.6	89	22.7	88.5
40	21.6	98.3	90	22.9	96.1
41	21.0	51.0	91	23.6	93.4
42	22.1	61.2	92	22.4	99.5
43	21.0	84.6	93**	21.3	100.0
44	22.1	83.3	95	22.6	97.0
45	21.8	38.2	97	24.1	97.6
46	22.3	94.2	104	22.6	99.5
49	21.1	84.1	105	22.8	99.5
50	22.2	98.2	106	21.0	93.7

51	21.0	95.9	107	20.9	94.5
52	22.0	96.2	108	21.0	97.0
53	21.4	67.1	109	23.1	99.7
54	21.0	99.0	110*	22.0	14.6
55	22.6	38.2	112*	22.4	22.4
58	21.0	86.0	113	22.3	89.7
59	22.4	25.9	114	22.6	90.1
60	28.3	64.6	115	22.3	95.8
61	22.5	27.9	116	22.7	99.5
62	22.0	53.1	117	22.4	98.9
63	21.3	44.6	118	21.6	99.3
96**	21.1	100.0	119	22.7	98.6
98	28.5	43.2	138	22.4	99.5
99	21.2	88.2	155	22.2	99.8
100*	21.0	9.6	156	21.0	90.0
101	21.2	52.4	157	22.5	91.5
102	28.9	48.4	158	21.2	98.2
103	22.7	42.9	159	21.2	91.9
127	21.3	42.1	160	22.3	96.1
128	22.4	20.1	161	32.0	70.0
129	23.2	96.6	162	21.3	98.9
130	21.3	38.9	163	22.6	88.2
131	22.4	95.6	164	22.6	94.3
133	23.1	31.8	165	22.1	91.4
177	22.3	41.1	166	31.1	64.8
178	22.5	22.2	167	22.7	98.6
179	22.6	21.4	170	21.0	88.2
180*	21.1	9.2	171	21.2	81.3
181	21.1	33.1	174	22.6	95.9
182*	22.5	11.6	175	21.5	99.3
183*	22.5	0.3	176	22.7	98.9

* Selected for sequencing; ** Reference

Table S4. Results of screening section 23 for 631RR2/Bt (F1) samples. Samples on the left (samples 1 to 80; first run), in the middle (samples 81 to 152; second run) and on the right (samples 153 to 214; third run) were analyzed separately (three different runs which were repeated on a different day). We selected three samples (first run), one sample (second run) and no sample (third run) which had confidence values below 67, 47 respectively 80.

631 RR2/Bt samples	**Mean Ct value**	**Mean Confidence value [%]**	**631 RR2/Bt samples**	**Mean Ct value**	**Mean Confidence value [%]**	**631 RR2/Bt samples**	**Mean Ct value**	**Mean Confidence value [%]**
1	20.4	99.3	81	19.3	98.8	153	19.1	98.6
2	21.4	87.8	82*	20.0	46.3	154	18.9	98.8
3	20.5	95.7	83	19.0	99.1	155	20.5	95.7
4	21.1	95.1	84	19.0	97.4	156	19.1	94.7

5	20.5	95.6	85	19.4	99.9	157	18.8	99.4
6	20.6	90.2	86	19.7	98.1	158	18.4	96.4
7	20.9	99.9	87	19.5	98.4	159	20.8	93.9
8	20.5	99.7	88	19.6	99.5	160	18.6	98.9
9	20.6	98.5	89	19.5	99.3	161	19.5	99.2
10	20.9	99.5	90	19.2	98.9	162	18.5	99.6
11	20.7	79.1	91	19.8	98.8	163	18.7	98.8
12	20.8	99.5	92	19.0	99.8	165	18.6	99.7
14	21.0	99.9	93	19.5	94.1	166	18.5	99.5
15	20.5	80.9	94	19.1	98.4	167	18.5	97.2
16*	20.5	66.2	95	20.0	98.8	168	18.5	99.4
18	20.6	97.1	96	18.9	96.1	169	18.4	98.6
19	20.6	96.1	97	19.0	91.8	170	18.3	96.4
20	20.7	87.5	98	19.1	99.3	171	18.9	99.6
21	21.3	66.8	99	19.6	98.5	172	19.0	88.3
22	20.7	93.1	100	19.0	99.2	173	18.8	99.2
23	20.4	75.6	101	18.5	99.6	174	18.4	98.0
24	20.2	74.7	102	18.9	96.4	175	18.8	95.9
25*	20.2	61.2	103	18.9	99.6	176	18.9	99.5
26*	20.4	49.3	104	19.0	97.0	177	18.9	97.8
29	20.5	94.5	105	18.7	88.0	178	19.0	90.8
30	20.6	99.6	106	18.8	98.8	179	18.9	97.7
31	20.3	99.2	107	18.9	99.6	180	19.2	98.8
32	20.9	99.0	108**	18.9	100.0	181	19.0	92.9
33	20.4	99.4	109	18.8	92.7	182	18.9	87.4
34	20.7	93.0	110	18.6	95.9	183	18.5	98.3
35	20.6	97.9	111	18.9	94.5	184	18.8	98.7
36	20.8	98.2	112	18.9	93.5	185	18.4	99.1
37	20.6	91.2	113	19.1	89.8	186	19.0	97.8
38	20.6	87.1	114	18.9	93.8	187	18.7	99.3
39	20.9	82.1	115	18.8	93.9	188	18.5	99.2
40	20.2	95.4	116	19.0	96.1	189	19.0	99.7
41	20.2	91.6	117	18.9	88.0	190	18.9	99.8
42	20.2	81.7	119	19.1	96.1	191	18.8	99.2
43	20.9	92.0	120	18.7	93.2	192	18.9	97.5
44	19.8	86.2	121	19.0	89.6	193	19.5	91.8
45	19.8	83.5	122	19.0	91.4	194	18.9	99.3
46	19.9	99.7	123	19.0	97.8	195	19.0	99.7
48	19.9	78.6	125	18.8	96.9	196	19.6	95.4
50	19.9	88.8	126	18.8	93.8	197	18.7	98.8
51	20.0	79.0	127	18.8	98.3	198	19.2	93.1
55	19.7	77.7	128	19.0	99.0	199	18.6	97.1
56	19.6	77.5	129	18.9	91.3	200	18.7	99.2
58	20.2	94.4	130	19.9	84.4	201	19.1	92.2
59	20.2	99.8	131	18.9	97.3	202	19.0	92.1
60	20.5	97.0	132	19.0	92.1	203	18.5	95.8
61**	20.4	100.0	133	19.0	98.3	204	18.7	96.5
62	20.3	98.8	134	18.7	98.2	205	18.1	82.7

63	19.9	95.4	135	19.1	96.0	206	19.5	84.6
64	20.3	94.3	136	18.9	99.8	207	19.1	98.4
65	20.3	96.4	137	18.9	98.2	208	18.5	99.7
66	20.3	97.8	138	18.9	97.8	209	18.7	99.7
67	20.2	97.2	139	19.3	97.3	211	19.1	99.5
68	20.2	98.8	140	19.4	97.5	212	19.1	99.0
69	20.5	99.7	141	19.5	93.2	213	18.7	97.9
70	21.0	98.0	142	19.9	91.8	214**	18.7	100.0
71	20.2	97.9	143	19.4	94.0			
72	19.8	95.3	144	19.1	97.7			
73	20.4	99.0	145	18.8	94.4			
74	20.2	99.4	146	19.5	91.1			
75	20.5	99.9	147	19.2	83.4			
76	20.0	98.2	148	18.9	98.2			
77	20.0	87.2	149	19.3	89.4			
78	20.1	96.5	150	19.4	99.0			
79	19.9	91.4	151	19.1	97.9			
80	19.6	96.7	152	20.3	98.9			

* Selected for sequencing; ** Reference

Table S5. Results of screening section 2 for DKC 26-79 (F2) samples. The Ct values ranged from 18.4 to 28.5 with an average Ct value of 19.8 and a standard deviation of 2.8 (CV = 14.1%). The sample with the lowest confidence value (83.6) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	19.6	99.9
56	28.5	99.5
59	19.6	99.2
61	27.5	99.1
63	18.6	98.1
98	19.5	99.9
99	18.5	98.2
100	18.6	97.5
101	18.6	99.6
103*	19.3	83.6
127	18.4	99.9
128	18.5	100.0
129	19.4	99.2
130	18.6	99.8
132	18.5	95.3
133	19.3	91.6
137	18.6	99.6
139	19.2	99.9
140**	18.6	100.0

141	19.4	94.8

* Selected for sequencing; ** Reference

Table S6. Results of screening section 3 for DKC 26-79 (F2) samples. The Ct values ranged from 19.3 to 21.2 with an average Ct value of 20.3 and a standard deviation of 0.7 (CV = 3.4%). The samples with the lowest confidence values (49.1 respectively 61.7) were selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.0	83.1
56	21.0	93.5
59**	20.8	100.0
61	20.1	97.7
63	19.4	89.5
98	21.2	87.9
99	19.8	91.7
100	19.8	76.7
101	20.1	79.1
103*	21.1	49.1
127	19.7	99.7
128	21.1	97.4
129	20.5	95.0
130	19.8	97.5
132	19.3	96.7
133	20.7	95.7
137	19.3	99.5
139*	20.7	61.7
140	19.5	95.3
141	20.9	77.3

* Selected for sequencing; ** Reference

Table S7. Results of screening section 4 for DKC 26-79 (F2) samples. The Ct values ranged from 21.2 to 24.6 with an average Ct value of 22.1 and a standard deviation of 0.9 (CV = 3.9%). The sample with the lowest confidence value (88.8) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	22.8	98.5
56**	22.8	100.0
59	22.6	99.9
61	21.4	99.3
63	21.3	100.0

98*	24.6	88.8
99	21.6	99.0
100	21.5	94.9
101	21.4	99.2
103	22.7	94.1
127	21.2	99.4
128	21.4	99.0
129	22.5	99.2
130	21.6	96.1
132	21.3	98.0
133	22.3	99.9
137	21.3	96.5
139	22.4	97.5
140	21.4	90.8
141	22.8	96.3

* Selected for sequencing; ** Reference

Table S8. Results of screening section 5 for DKC 26-79 (F2) samples. The Ct values ranged from 22.9 to 30.1 with an average Ct value of 25.7 and a standard deviation of 1.7 (CV = 6.4%). The sample with the lowest confidence value (89.4) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55**	26.5	100.0
56	29.1	96.0
59	26.0	99.3
61	25.1	99.8
63	25.1	98.0
98	26.7	99.1
99	24.5	99.9
100	24.3	99.6
101	25.1	98.4
103	26.4	99.7
127	24.6	99.7
128	26.1	98.4
129	26.0	99.5
130	24.4	97.7
132	26.9	99.6
133	25.6	99.7
139*	24.7	89.4
140	24.6	99.4
141	30.1	98.7
142	22.9	97.4

* Selected for sequencing; ** Reference

Table S9. Results of screening section 7 for DKC 26-79 (F2) samples. The Ct values ranged from 19.5 to 22.0 with an average Ct value of 20.8 and a standard deviation of 0.7 (CV = 3.4%). The sample with the lowest confidence value (77.5) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.3	99.8
56	22.0	98.9
59	21.4	99.9
61**	20.5	100.0
63	20.5	98.7
98	21.9	97.2
99	20.3	99.9
100	20.4	92.1
101	20.5	98.2
103*	21.3	77.5
127	19.8	98.5
128	21.1	93.4
129	21.1	100.0
130	20.3	98.5
132	19.8	96.9
133	21.0	94.6
137	19.5	99.6
139	20.9	91.0
140	20.3	99.7
141	21.8	96.5

* Selected for sequencing; ** Reference

Table S10. Results of screening section 8 for DKC 26-79 (F2) samples. The Ct values ranged from 18.6 to 20.6 with an average Ct value of 19.5 and a standard deviation of 0.6 (CV = 3.3%). The sample with the lowest confidence value (74.6) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	19,9	85,1
56	20,2	91,2
59	19,9	81,3
61	19,0	91,2
63	18,6	90,6
98 *	19,7	74,6
99	19,3	91,5
100	18,8	77,9
101	19,1	90,9

103	19,9	77,9
127	19,1	99,3
128	19,8	98,5
129	19,8	99,9
130	18,7	99,8
132	19,6	99,8
133**	20,5	100,0
137	18,8	99,1
139	20,0	90,2
140	20,6	97,0
141	18,6	96,0

* Selected for sequencing; ** Reference

Table S11. Results of screening section 9 for DKC 26-79 (F2) samples. The Ct values ranged from 19.5 to 21.7 with an average Ct value of 20.4 and a standard deviation of 0.7 (CV = 3.2%). The sample with the lowest confidence value (73.5) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	**Ct value**	**Confidence value [%]**
55	20.9	99.3
56**	20.9	100.0
59	20.9	99.1
61	20.1	95.3
63	19.8	98.9
98	21.1	99.7
99	19.9	97.4
100*	19.7	73.5
101	19.9	95.9
103	20.9	87.6
127	19.6	99.3
128	21.1	97.3
129	20.5	96.1
130	19.6	98.9
132	19.5	94.0
133	20.6	98.7
137	19.5	89.6
139	20.9	74.6
140	20.3	90.5
141	21.7	74.9

* Selected for sequencing; ** Reference

Table S12. Results of screening section 10 for DKC 26-79 (F2) samples. The Ct values ranged from 19.2 to 22.0 with an average Ct value of 20.8 and a standard deviation of 0.8 (CV = 3.7%). The sample with the lowest confidence value (55.0) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.9	95.1
56	22.0	81.5
59	21.7	93.2
61	20.5	99.7
63	20.6	99.6
98	21.8	99.7
99	20.6	97.7
100	20.4	85.1
101	20.5	83.3
103	21.6	89.4
127	20.3	94.8
128	21.5	99.7
129**	21.3	100.0
130	20.1	94.7
132	20.3	96.6
133	21.4	99.6
137	20.2	98.3
138	20.1	93.7
140	20.8	86.6
142*	19.2	55.0

* Selected for sequencing; ** Reference

Table S13. Results of screening section 11 for DKC 26-79 (F2) samples. The Ct values ranged from 18.6 to 21.0 with an average Ct value of 19.6 and a standard deviation of 0.7 (CV = 3.3%). The sample with the lowest confidence value (81.4) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	20.0	99.8
56**	20.3	100.0
59	20.0	97.6
61	19.0	96.3
63	19.0	99.8
98	20.3	97.8
99	19.0	97.6
100	19.0	89.1
101	19.1	93.1

103	20.0	90.9
127	19.1	86.9
128	20.2	99.3
129	19.9	99.5
130	18.9	99.8
132	18.8	97.6
133	19.7	99.1
137	18.6	88.9
139	20.0	87.4
140	19.4	96.6
141*	21.0	81.4

* Selected for sequencing; ** Reference

Table S14. Results of screening section 12 for DKC 26-79 (F2) samples. The Ct values ranged from 18.6 to 20.4 with an average Ct value of 19.4 and a standard deviation of 0.7 (CV = 3.4%). The sample with the lowest confidence value (79.8) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	19.9	99.3
56	20.4	98.6
59	20.0	94.8
61	19.1	91.4
63	18.9	100.0
98	20.4	98.5
99	18.9	99.5
100	18.9	99.3
101	19.0	97.9
103	20.2	87.7
127	18.7	97.9
128	20.3	99.9
129	19.7	99.3
130	18.7	99.0
132**	18.6	100.0
133	19.7	97.3
137	18.6	96.1
139*	19.9	79.8
140	18.7	98.4
141	19.9	89.7

* Selected for sequencing; ** Reference

Table S15. Results of screening section 13 for DKC 26-79 (F2) samples. The Ct values ranged from 17.9 to 20.4 with an average Ct value of 19.4 and a standard deviation of 0.7 (CV = 3.6%). The sample with the lowest confidence value (62.1) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	20.2	90.6
56*	20.4	62.1
59	20.1	92.9
61	19.0	99.4
63	19.0	93.0
98	20.0	99.5
99	19.1	99.7
100	18.9	80.6
101	19.1	90.9
103	20.3	99.7
127	18.8	99.8
128	20.3	94.6
129	20.0	95.8
130	18.8	83.9
132**	18.8	100.0
133	19.7	99.4
137	18.7	91.2
138	19.7	98.3
140	19.1	89.6
142	17.9	65.8

* Selected for sequencing; ** Reference

Table S16. Results of screening section 14 for DKC 26-79 (F2) samples. The Ct values ranged from 19.2 to 21.8 with an average Ct value of 20.3 and a standard deviation of 0.7 (CV = 3.7%). The sample with the lowest confidence value (86.5) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.1	99.3
56	20.8	98.8
59	20.9	99.9
61	19.9	99.9
63	19.8	99.7
98	21.1	98.7
99	20.0	98.5
100	19.7	92.9
101	19.9	98.9

103	20.9	95.8
127	19.2	99.9
128	21.0	97.6
129	20.7	99.6
130	19.4	99.9
132**	19.3	100.0
133	20.7	97.2
137	19.4	95.7
139	20.9	86.6
140	20.0	98.2
141*	21.8	86.5

* Selected for sequencing; ** Reference

Table S17. Results of screening section 15 for DKC 26-79 (F2) samples. The Ct values ranged from 18.6 to 21.3 with an average Ct value of 19.6 and a standard deviation of 0.7 (CV = 3.5%). The sample with the lowest confidence value (69.4) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	**Ct value**	**Confidence value [%]**
55	20,3	99,8
56	20,5	91,1
59	20,2	99,9
61	19,1	80,9
63	18,9	94,0
98	19,8	89,9
99	19,1	99,9
100*	19,0	69,4
101	19,3	92,5
103	20,2	94,8
127	19,0	96,4
128	19,7	90,5
129	20,0	95,1
130	18,9	72,6
132	18,9	93,5
133**	19,9	100,0
137	18,6	94,3
139	20,1	70,0
140	19,6	70,0
141	21,3	77,2

* Selected for sequencing; ** Reference

Table S18. Results of screening section 16 for DKC 26-79 (F2) samples. The Ct values ranged from 20.1 to 22.4 with an average Ct value of 21.3 and a standard deviation of 0.7 (CV = 3.2%). The samples with the lowest confidence values (58.5 respectively 3.7) were selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	22.0	90.3
56	21.9	94.7
59	21.7	67.5
61	20.8	71.3
63*	20.5	58.5
98	22.2	95.9
99	20.9	97.9
100	20.8	64.1
101	21.1	83.4
103	22.0	94.5
127	20.5	97.5
128	22.4	97.5
129	21.8	94.1
130*	21.2	3.7
132	20.4	99.1
133**	21.5	100.0
137	20.1	74.2
139	21.6	75.3
140	20.4	99.1
141	21.7	76.0

* Selected for sequencing; ** Reference

Table S19. Results of screening section 17 for DKC 26-79 (F2) samples. The Ct values ranged from 19.0 to 22.3 with an average Ct value of 20.1 and a standard deviation of 0.8 (CV = 4.2%). The samples with the lowest confidence values (50.8 respectively 27.7) were selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	20,7	77,5
56	20,8	86,4
59	20,5	99,4
61**	19,5	100,0
63	19,5	92,3
98	21,1	74,3
99*	19,6	50,8

100*	19,4	27,7
101	19,6	70,8
103	20,5	64,3
127	19,3	84,3
128	20,8	70,2
129	20,5	74,6
130	19,4	55,4
132	19,4	89,7
133	20,2	64,0
137	19,0	82,1
139	20,6	99,6
140	19,3	49,5
141	22,3	59,9

* Selected for sequencing; ** Reference

Table S20. Results of screening section 18 for DKC 26-79 (F2) samples. The Ct values ranged from 19.5 to 21.8 with an average Ct value of 20.5 and a standard deviation of 0.6 (CV = 2.8%). The sample with the lowest confidence value (94.7) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	20.3	99.4
56	20.3	98.5
59	20.2	99.8
61	20.4	95.5
63	20.3	95.3
98	21.4	97.5
99	20.4	97.4
100	20.0	99.9
101**	20.4	100.0
103	20.7	97.6
127	19.5	99.6
128	21.5	98.4
129	20.7	99.8
130	19.7	99.9
132	20.2	100.0
133	20.6	98.6
137*	19.9	94.7
139	20.6	96.5
140	20.4	97.3
141	21.8	95.1

* Selected for sequencing; ** Reference

Table S21. Results of screening section 19 for DKC 26-79 (F2) samples. The Ct values ranged from 20.4 to 23.3 with an average Ct value of 21.4 and a standard deviation of 0.7 (CV = 3.0%). The sample with the lowest confidence value (69.8) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.3	97.5
56	21.2	95.9
59	21.1	99.4
61	21.2	99.4
63	21.2	93.5
98*	22.2	69.8
99	21.2	88.5
100	20.8	92.4
101	21.4	87.7
103	21.8	88.3
127	20.5	99.9
128	22.2	92.8
129	21.5	98.8
130	20.4	79.0
132**	21.0	100.0
133	21.6	93.0
137	20.7	98.9
139	21.6	75.7
140	21.2	97.3
141	23.3	79.6

* Selected for sequencing; ** Reference

Table S22. Results of screening section 20 for DKC 26-79 (F2) samples. The Ct values ranged from 19.3 to 21.3 with an average Ct value of 20.2 and a standard deviation of 0.7 (CV = 3.3%). The sample with the lowest confidence value (49.1) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	20.8	91.3
56	21.0	88.4
59**	20.8	100.0
61	19.8	98.7
63	19.3	99.8
98	21.0	92.7
99	19.9	93.5
100	19.7	72.1
101	20.0	87.4
103	20.9	54.6
127	19.6	91.7

128	21.3	91.9
129	20.6	99.5
130	19.7	82.6
132	19.5	97.4
133	20.5	96.6
137	19.3	98.5
139	20.5	67.0
140	19.5	95.6
141*	20.9	49.1

* Selected for sequencing; ** Reference

Table S23. Results of screening section 21 for DKC 26-79 (F2) samples. The Ct values ranged from 18.7 to 20.8 with an average Ct value of 19.6 and a standard deviation of 0.7 (CV = 3.6%). The sample with the lowest confidence value (79.6) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	20.2	98.6
56	20.3	99.7
59	20.2	99.9
61	19.3	99.9
63	19.0	98.3
98	20.5	99.8
99**	19.3	100.0
100	18.8	96.1
101	19.3	99.5
103	20.2	95.2
127	18.8	95.5
128	20.3	97.8
129	20.2	93.7
130	18.8	96.9
132	18.8	92.3
133	20.0	94.0
137	18.7	86.3
139*	20.2	79.6
140	18.8	93.5
141	20.8	80.7

* Selected for sequencing; ** Reference

Table S24. Results of screening section 22 for DKC 26-79 (F2) samples. The Ct values ranged from 19.6 to 21.9 with an average Ct value of 20.6 and a standard deviation of 0.7 (CV = 3.4%). The sample with the lowest confidence value (94.5) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.3	99.5
56	21.4	98.9
59	21.2	99.8
61	20.2	99.8
63	19.9	98.0
98	21.4	99.6
99**	20.1	100.0
100*	19.9	94.5
101	20.2	99.5
103	21.2	97.7
127	19.8	99.8
128	20.9	98.8
129	20.8	99.2
130	19.7	99.2
132	19.7	99.9
133	20.7	98.2
137	19.6	99.5
139	21.0	97.4
140	20.2	99.4
141	21.9	94.9

* Selected for sequencing; ** Reference

Table S25. Results of screening section 24 for DKC 26-79 (F2) samples. The Ct values ranged from 16.9 to 19.0 with an average Ct value of 17.6 and a standard deviation of 0.5 (CV = 3.1%). The sample with the lowest confidence value (86.5) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	18,1	98,6
56	17,8	99,7
59	17,2	98,0
61*	19,0	86,5
63**	17,7	100,0
98	17,7	90,4
99	17,2	99,6
100	17,4	97,9
101	17,5	98,7

103	16,9	92,8
127	17,2	96,3
128	17,2	99,0
129	17,4	97,3
130	18,2	99,9
132	18,7	98,1
133	17,9	97,2
137	17,5	92,0
139	17,5	87,9
140	17,3	91,5
141	17,1	92,8

* Selected for sequencing; ** Reference

Table S26. Results of screening section 25 for DKC 26-79 (F2) samples. The Ct values ranged from 19.4 to 21.2 with an average Ct value of 20.3 and a standard deviation of 0.6 (CV = 3.0%). The sample with the lowest confidence value (84.8) was selected for Sanger sequencing.

DKC 26-79 (F2) samples	Ct value	Confidence value [%]
55	21.2	96.0
56	21.0	100.0
59	21.1	89.1
61	20.1	99.8
63	19.6	99.9
98	21.2	98.8
99	20.0	99.7
100	20.0	96.6
101	20.1	100.0
103*	21.2	84.8
127	19.8	98.0
128	20.3	97.4
129	20.8	98.3
130	19.7	99.5
132**	19.7	100.0
133	20.8	99.9
137	19.4	97.0
139	20.7	89.7
140	19.8	93.1
141	20.5	94.9

* Selected for sequencing; ** Reference

Table S27. Sequencing results of all 25 sections. All sequence data were assembled and submitted to GenBank (Banklt1939702 Seq1 KX640115).

Screening section	DNA sequence	Length (bp)
1	AGA GCCTCACGTT TCCAGGGTGA AGTATCAGAG GATTTACCGC CCATGCCTTT TATGGAGACA AGAAGGGGAG GAGGTAAACA GATCAGCATC AGCGCTCGAA AGTTTCGTCA AAGGATGCGG AACTGTTTCC AGCCGCCGTC GCCATTCGGC CAGACTCCTC CTCTCTCGGC ATGAGCCGAT CTTTTCTCTG GCATTTCCAA CCCTAGAGAC GTGCGTCCCT GGTGGGCTGC TCGGCCAGCA AGCCTTGTAG CGGCCCACGC GTGGTACCAA GCTTGATATC CCTAGGGCGG C	294
2	TATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAAC AAAGGTAAGA TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAAGGTGGTA TAAAGTAAAA TATCGGTAAT AAAAGGTGGC CCAA	123
3	TCGGTAAT AAAAGGTGGC CCAAAGTGAA ATTTACTCTT TTCTACTATT ATAAAAATTG AGGATGTTTT TGTCGGTACT TTGATACGTC ATTTTTGTAT GAATTGGTTT TTAAGTTTAT TCGCTTTTGG AAATGCATAT CTGTATTTGA GTCGGGTTTT AAGTTCGTTT GCTTTTGTAA ATACAGAGGG ATTTGTATAA GAAATATCTT TAGAAAAACC CATATGCTAA TTTGACATAA TTTTTGAGAA AAATATATAT TCAGGCGAAT TCTCACAATG AACAATAATA AGATTAAAAT AGCTTTCCCC CGTTGCAGCG CATGGGTATT TTTTCTAGTA AAAATAAAAG ATAAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTTCCTAAA GCCCAAAGTG CT	410
4	TAAAAT AGCTTTCCCC CGTTGCAGCG CATGGGTATT TTTTCTAGTA AAAATAAAAG ATAAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTTCCTAAA GCCCAAAGTG CT	118
5	CGTTGCAGCG CATGGGTATT TTTTCTAGTA AAAATAAAAG ATAAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTTCCTAAA GCCCAAAGTG CTATCCACGA TCCATAGCAA GCCCAGCCCA ACCCAACCCA ACCCAACCCA CCCCAGTCCA GCCAACTGGA CAATAGTCTC CACACCCCCC CACTATCACC GTGAGTTGTC CGCACGCACC GCACGTCTCG CAGCCAAAAA AAAAAAGAAA GAAAAAAAAG AAAAAGAAAA AACAGCAGGT GGGTCCGGGT CGTGGGGGCC GGAAACGCGA GGAGGATCGC GAGCCAGCGA CGAGGCCGGC CCTCCCTCCG CTTCCAAAGA AACGC	365
6	CCTCCG CTTCCAAAGA AACGCCCCCC ATCGCCACTA TATACATACC CCCCCCTCTC CTCCCATCCC CCCAACCCTA CCACCACCAC CACCACCACC TCCACCTCCT CCCCCCTCGC TGCCGGACGA CGAGCTCCTC CCCCCTCCCC CTTCGCCGCC GCCGCGCCGG TAACCACCCC GCCCCTCTCC TCTTTCTTTC TCCGTTTTTT TTTCCGTCTC GGTCTCGATC TTTGGCCTTG GTAGTTTGGG TGGGCGAGAG GCGGCTTCGT GCGCGCCCAG ATCGGTGCGC GGGAGGGGCG GGATCTCGCG GCTGGGGCTC TCGCCGGCGT GGATCCGGCC CGGATCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTGC GATCCGCCGT TGTTGGGGGA GATGATGGGG GGTTTAAAAT TTCCGCCGTG CTAAACAAGA TCAGG	441
7	GAATGGG GCTCTCGGAT GTAGATCTGC GATCCGCCGT TGTTGGGGGA GATGATGGGG GGTTTAAAAT TTCCGCCGTG CTAAACAAGA TCAGGAAGAG GGGAAAAGGG CACTATGGTT TATATTTTTA TATATTTCTG CTGCTTCGTC AGGCTTAGAT GTGCTAGATC TTTCTTTCTT CTTTTTGTGG GTAGAATTTG AATCCCTCAG CATTGTTCAT CGGTAGTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGCGGAGC TTTTTTGTAG GTAGAAGTGA TCAACCATGG CGCAAGTTAG CAGAATCTGC AATGGTGTGC AGAA	331
8	TGACA AATGCAGCCT CGTGCGGAGC TTTTTTGTAG GTAGAAGTGA TCAACCATGG CGCAAGTTAG CAGAATCTGC AATGGTGTGC AGAACCCATC TCTTATCTCC AATCTCTCGA AATCCAGTCA ACGCAAATCT CCCTTATCGG TTTCTCTGAA GACGCAGCAG CATCCACGAG CTTATCCGAT TTCGTCGTCG TGGGGATTGA AGAAGAGTGG GATGACGTTA ATTGGCTCTG	374

	AGCTTCGTCC TCTTAAGGTC ATGTCTTCTG TTTCCACGGC GTGCATGCTT CACGGTGCAA GCAGCCGGCC CGCAACCGCC CGCAAATCCT CTGGCCTTTC CGGAACCGTC CGCATTCCCG GCGACAAGTC GATCTCCCA	
9	CGTCGTCG TGGGGATTGA AGAAGAGTGG GATGACGTTA ATTGGCTCTG AGCTTCGTCC TCTTAAGGTC ATGTCTTCTG TTTCCACGGC GTGCATGCTT CACGGTGCAA GCAGCCGGCC CGCAACCGCC CGCAAATCCT CTGGCCTTTC CGGAACCGTC CGCATTCCCG GCGACAAGTC GATCTCCCAC CGGTCCTTCA TGTTCGGCGG TCTCGCGAGC GGTGAAACGC GCATCACCGG CCTTCTGGAA GGCGAGGACG TCATCAATAC GGGCAAGGCC ATGCAGGCGA TGGGCGCCCG CATCCGTAAG GAAGGCGACA CCTGGATCAT CGATGGCGTC GGCAATGGCG GCCTCCTGGC GCCTGAGGCG CCGCTCGATT TCGGCAATGC C	389
10	AGGCGACA CCTGGATCAT CGATGGCGTC GGCAATGGCG GCCTCCTGGC GCCTGAGGCG CCGCTCGATT TCGGCAATGC CGCCACGGGC TGCCGCCTGA CGATGGGCCT CGTCGGGGTC TACGATTTCG ACAGCACCTT CATCGGCGAC GCCTCGCTCA CAAAGCGCCC GATGGGCCGC GTGTTGAACC CGCTGCGCGA AATGGGCGTG CAGGTGAAAT CGGAAGACGG TGACCGTCTY CCCGTTACCT TGCGCGGGCC GAAGACGCCG ACGCCGATCA CCTACCGCGT GCCGATGGCC TCCGCACAGG TGAAGTCCGC CGTGCTGCTC GCCGGCCTCA ACACGCCCGG CATCACGACG GTCATCGAG	367
11	TC TACGATTTCG ACAGCACCTT CATCGGCGAC GCCTCGCTCA CAAAGCGCCC GATGGGCCGC GTGTTGAACC CGCTGCGCGA AATGGGCGTG CAGGTGAAAT CGGAAGACGG TGACCGTCTY CCCGTTACCT TGCGCGGGCC GAAGACGCCG ACGCCGATCA CCTACCGCGT GCCGATGGCC TCCGCACAGG TGAAGTCCGC CGTGCTGCTC GCCGGCCTCA ACACGCCCGG CATCACGACG GTCATCGAGC CGATCATGAC GCGCGATCAT ACGGAAAAGA TGCTGCAGGG CTTTGGCGCC AACCYTACCG TCGAGACGGA TGCGGACGGC GTGCGCACCA TCCGCCTG	350
12	C TCCGCACAGG TGAAGTCCGC CGTGCTGCTC GCCGGCCTCA ACACGCCCGG CATCACGACG GTCATCGAGC CGATCATGAC GCGCGATCAT ACGGAAAAGA TGCTGCAGGG CTTTGGCGCC AACCYTACCG TCGAGACGGA TGCGGACGGC GTGCGCACCA TCCGCCTGGA AGGCCGCGGC AAGCTCACCG GCCAAGTCAT CGACGTGCCG GGCGACCCGT CCTCGACGGC CTTCCCGCTG GTTGCGGCCC TGCTTGTTCC GGGCTCCGAC GTCACCATCC TCAACGTGCT GATGAACCCC ACCCGCACCG GCCTCATCCT GACGCTGCAG GAAATGGGCG CCGACATCGA AGTCATCAAC CCGCGC	367
13	CG TCGAGACGGA TGCGGACGGC GTGCGCACCA TCCGCCTGGA AGGCCGCGGC AA-GCTCACCG GCCAAGTCAT CGACGTGCCG GGCGACCCGT CCTCGACGGC CTTCCCGCTG GTTGCGGCCC TGCTTGTTCC GGGCTCCGAC GTCACCATCC TCAACGTGCT GATGAACCCC ACCCGCACCG GCCTCATCCT GACGCTGCAG GAAATGGGCG CCGACATCGA AGTCATCAAC CCGCGCCTTG CCGGCGGCGA AGACGTGGCG GACCTGCGCG TTCGCTCCTC CACGCTGAAG GGCGTCACGG TGCCGGAAGA CCGCGCGCCT TCGATGATCG ACGAATATCC GATTCTCGCT GTCGCCGCCG CCTTCGCGGA AGGGGCGACC G	383
14	ATATCC GATTCTCGCT GTCGCCGCCG CCTTCGCGGA AGGGGCGACC GTGATGAACG GTCTGGAAGA ACTCCGCGTC AAGGAAAGCG ACCGCCTCTC GGCCGTCGCC AATGGCCTCA AGCTCAATGG CGTGGATTGC GATGAGGGCG AGACGTCGCT CGTCGTGCGT GGCCGCCCTG ACGGCAAGGG GCTCGGCAAC GCCTCGGGCG CCGCCGTCGC CACCCATCTC GATCACCGCA TCGCCATGAG CTTCCTCGTC ATGGGCCTCG TGTCGGAAAA CCCTGTCACG GTGGACGATG CCACGATGAT CGCCACGAGC TTCCCGGAGT TCATGGACCT GATGGCCGGG CTGGGCGCGA AGATCGAACT CTC	369
15	TCGC CACCCATCTC GATCACCGCA TCGCCATGAG CTTCCTCGTC ATGGGCCTCG TGTCG-GAAAA CCCTGTCACG GTGGACGATG CCACGATGAT CGCCACGAGC TTCCCGGAGT TCATGGACCT GATGGCCGGG CTGGGCGCGA AGATCGAACT CTCCGATACG AAGGCTGCCT GATGAGCTCG AATTCCCGAT CGTTCAAACA TTTGGCAATA AAGTTTCTTA AGATTGAATC CTGTTGCCGG TCTTGCGATG ATTA	258

16	GTTGCCGG TCTTGCGATG ATTATCATAT AATTTCTGTT GAATTACGTT AAGCATGTAA TAATTAACAT GTAATGCATG ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCCGC AATTATACAT TTAATACGCG ATAGAAAACA AAATATAGCG CGCAAACTAG GATAAATTAT CGCGCGCGGT GTCATCTATG TTACTAGATC GGGGATAGCT TCTGCAGGTC CGATTGAGAC TTTTCAACAA AGGGTAATAT CCGGAAACCT CCTCGGATTC CATTGCCCAG CTATCTGTCA CTTTATTGTG AAGATAGTGG AAAAGGAAGG TGGCTCCTAC AAATGCCATC ATTGCGATAA AGGAAAGGCC ATCGTTGAAG ATGCCTCTGC CGACAGTGGT CCCAAAGATG GACCCCCACC CACGAGGAGC ATCGTGGAAA AAGAAGACGT TCCAACCACG TCTTCAAAGC AAGTGGATTG ATGTGATGGT CCGATTGAGA C	499
17	AAGTGGATTG ATGTGATGGT CCGATTGAGA CTTTTCAACA AAGGGTAATA TCCGGAAACC TCCTAGGATT CCATTGCCCA GCTATCTGTC ACTTTATTGT GAAGATAGTG GAAAAGGAAG GTGGCTCCTA CAAATGCCAT CATTGCGATA AAGGAAAGGC CATCGTTGAA GATGCCTCTG CCGACAGTGG TCCCAAAGAT GGACCCCCAC CCACGAGGAG CATCGTGGAA AAAGAAGACG TTCCAACCAC GTCTTCAAAG CAAGTGGATT GATGTGATAT CTCCACTGAC GTAAGGGATG ACGCACAATC CCACTATCCT TCGCAAGACC CTTCCTCTAT ATAAGGAAGT TCATTTCATT TGGAGAGGAC ACGCTGACAA GCTGACTCTA GCAGATCTAC CGTCTTCGGT ACGCGCTCAC TCCGCCCTCT GCCT	434
18	ACGCGCTCAC TCCGCCCTCT GCCTTTGTTA CTGCCACGTT TCTCTGAATG CTCTCTTGTG TGGTGATTGC TGAGAGTGGT TTAGCTGGAT CTAGAATTAC ACTCTGAAAT CGTGTTCTGC CTGTGCTGAT TACTTGCCGT CCTTTGTAGC AGCAAAATAT AGGGACATGG TAGTACGAAA CGAAGATAGA ACCTACACAG CAATACGAGA AATGTGTAAT TTGGTGCTTA GCGGTATTTA TTTAAGCACA TGTTGGTGTT ATAGGGCACT TGGATTCAGA AGTTTGCTGT TAATTTAGGC ACAGGCTTCA TACTACATGG GTCAATAGTA TAGGGATTCA TATTATAGGC GATACTATAA TAATTTGTTC GTCTGCAGAG CTTATT	386
19	TAATTTGTTC GTCTGCAGAG CTTATTATTT GCCAAAATTA GATATTCCTA TTCTGTTTTT GTTTGTGTGC TGTTAAATTG TTAACGCCTG AAGGAATAAA TATAAATGAC GAAATTTTGA TGTTTATCTC TGCTCCTTTA TTGTGACCAT AAGTCAAGAT CAGATGCACT TGTTTTAAAT ATTGTTGTCT GAAGAAATAA GTACTGACAG TATTTTGATG CATTGATCTG CTTGTTTGTT GTAACAAAAT TTAAAAATAA AGAGTTTCCT TTTTGTTGCT CTCCTTACCT CCTGATGGTA TCTAGTATCT ACCAACTGAC ACTATATTGC TTCTCTTTAC ATACGTATCT TGCTCGATGC CTTCT	365
20	GAGTTTCCT TTTTGTTGCT CTCCTTACCT CCTGATGGTA TCTAGTATCT ACCAACTGAC AC-TATATTGC TTCTCTTTAC ATACGTATCT TGCTCGATGC CTTCTCCCTA GTGTTGACCA GTGTTACTCA CATAGTCTTT GCTCATTTCA TTGTAATGCA GATACCAAGC GGCCTCTAGA GGATCCAGGA GCAACCATGG CGCAAGTTAG CAGAATCTGC AATGGTGTGC AGAACCCATC TCTTATCTCC AATCTCTCGA AATCCAGTCA ACGCAAATCT CCCTTATCGG TTTCTCTGAA GACGCAGCAG CATCCACGAG CTTATCCGAT TTCGTCGTCG TGGGGATTGA AGAAGAGTGG GATGACGTTA ATTGGCTCTG AGCTTCGTCC TCTTAAGGTC ATGTCTTCTG TTTCCACGGC GTGCATGCTT CACGGTGCAA GCAGC	444
21	ACGAGC TTCCCGGAGT TCATGGACCT GATGGCCGGG CTGGGCGCGA AGATCGAACT CTCCGATACG AAGGCTGCCT GATGAGCTCG AATTCCCGAT CGTTCAAACA TTTGGCAATA AAGTTTCTTA AGATTGAATC CTGTTGCCGG TCTTGCGATGA TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT AATTAACATG TAATGCATGA CGTTATTTAT GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAAACAA AATATAGCGC GCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT TACTAGATCG GGGATATCCC CGGGGAATTC GGTACCAAGC TT	379
22	AATTATC GCGCGCGGTG TCATCTATGT TACTAGATCG GGGATATCCC CGGGGAATTC GGTACCAAGC TTTTATAATA GTAGAAAAGA GTAAATTTCA CTTTGGGCCA CCTTTTATTA CCGATATTTT ACTTTATACC ACCTTTTAAC TGATGTTTTC ACTTTTGACC AGGTAATCTT	305

	ACCTTTGTTT TATTTTGGAC TATCCCGACT CTCTTCTCAA GCATATGAAT GACCTCGAGT AAGCTTGTTA ACGCGGCCGC CCTAGGGATA TCAAGCTTGG TACCACGCGA CACACTTCCA CTCTAGTG	
23	ATGAAT GACCTCGAGT AAGCTTGTTA ACGCGGCCGC CCTAGGGATA TCAAGCTTGG TAC-CACGCGA CACACTTCCA CTCTAGTGTT TGAGTGGATC CTGTTATCTC TT	108
24	A CACACTTCCA CTCTAGTGTT TGAGTGGATC CTGTTATCTC TTCTCGAACC ATAACAGACT AGTATTATTT GATCATTGAA TCGTTTATTT CTCTTGAAAG CGGTTTCATT TTTTTTTACA GACGTCTTTT TTTAGGAGGT CGACATCCAT TATGCGGCAT AGGTGTTACA TCGCGTATAC AACTTAACCG TACACCACTT	201
25	TTAGCAATGG CTCGTAATGC GGCATCTCTT CCGCTACCAG CACCTTTTAC CATAACTTCT GCTCGTTGCA AACCCACTGT ACGAATAGCA TCTACTGCTG TTCTGCTGAC TTTATTTTTT TTAATAAAGT GAAAAACCAT AAAATGGACA ACAACACCCT GCCCTTCACT ACCGGTCGGA GCGACGCCGA AGATGGGGTT	200

7.5. Supplementary material for: Analysis of transcriptomic differences between NK603 maize and near-isogenic maize using RNA sequencing and RT-qPCR

Table S1. Gene IDs, FDR corrected p-values, and relative expression values of the 286 differentially expressed genes. Evaluations are highlighted in orange for DeSeq2, in green for EdgeR, and in blue for CLC Genomics. The average expression value (Log_2FC) of every differentially expressed gene is shown in the right column (highlighted in purple).

Gene ID	Log_2FC DeSeq2	FDR corr. p-value	Log_2FC EdgeR	FDR corr. p-value	Log_2FC CLC Genomics	FDR corr. p-value	MW Log_2FC
GRMZM2G446253	3.92	4.86E-28	10.64	6.98E-22	3.61	2.35E-05	6.06
GRMZM2G316362	4.91	4.62E-118	5.50	5.02E-100	5.63	1.57E-101	5.35
GRMZM2G013448	1.77	1.82E-05	5.59	1.67E-06	5.62	3.15E-04	4.33
GRMZM2G026780	3.65	4.16E-49	4.25	5.64E-49	4.20	4.03E-40	4.03
GRMZM2G131421	2.38	1.84E-09	4.74	3.22E-32	4.62	1.30E-18	3.91
GRMZM2G112238	3.69	3.46E-128	3.90	9.22E-91	3.88	1.04E-102	3.82
GRMZM2G106445	3.54	3.88E-88	3.81	2.50E-72	3.81	6.96E-86	3.72
GRMZM2G106393	2.55	1.04E-11	4.18	1.60E-37	4.20	2.08E-20	3.64
GRMZM2G117971	1.48	5.59E-04	4.18	4.74E-04	5.15	3.37E-05	3.60
GRMZM2G106413	3.33	1.88E-72	3.61	2.52E-63	3.75	5.19E-59	3.56
GRMZM2G171096	1.44	8.33E-04	4.28	4.04E-04	4.13	1.00E-02	3.28
GRMZM2G479038	1.84	7.72E-06	3.45	3.66E-05	3.64	8.99E-04	2.98
GRMZM2G074543	1.86	7.38E-06	3.83	1.46E-04	3.17	4.00E-02	2.95
AC203989.4_FG001	2.73	3.69E-52	2.95	1.61E-46	3.00	1.46E-44	2.89
GRMZM2G028677	1.88	4.18E-06	3.35	5.04E-06	3.43	2.14E-04	2.89
AC197705.4_FG001	1.56	3.19E-04	3.47	1.97E-13	3.62	9.50E-08	2.88
GRMZM2G170017	1.42	1.21E-03	3.58	1.55E-03	3.55	2.00E-02	2.85
GRMZM2G386674	2.69	3.92E-44	2.93	1.86E-39	2.88	4.57E-28	2.83
GRMZM2G051619	2.48	8.63E-25	2.84	1.42E-24	2.87	1.01E-19	2.73
AC204711.3_FG003	2.20	3.98E-11	2.94	1.99E-12	3.01	4.17E-08	2.72

GRMZM2G300965	2.34	5.76E-17	2.81	2.08E-14	2.91	2.99E-11	2.69
GRMZM2G075333	2.49	1.10E-23	2.87	1.76E-16	2.67	2.19E-10	2.68
GRMZM2G396402	2.14	6.17E-13	2.62	1.65E-14	2.93	9.59E-09	2.56
GRMZM2G106511	1.61	1.51E-04	2.98	5.65E-16	2.98	1.20E-07	2.52
GRMZM2G419994	2.34	1.07E-29	2.57	2.64E-29	2.55	2.80E-24	2.49
GRMZM2G057483	1.55	2.94E-04	2.81	2.69E-03	3.05	6.85E-03	2.47
GRMZM2G075459	2.28	9.41E-27	2.53	1.43E-26	2.56	9.25E-18	2.46
GRMZM2G112538	2.02	3.14E-11	2.50	7.09E-13	2.59	5.21E-13	2.37
GRMZM2G041714	1.45	9.46E-04	2.73	3.58E-12	2.76	1.26E-06	2.31
GRMZM2G092137	2.12	2.32E-16	2.46	2.22E-16	2.35	1.31E-12	2.31
GRMZM2G060544	1.98	1.89E-11	2.42	6.69E-12	2.48	9.08E-08	2.29
GRMZM2G043893	1.48	7.47E-04	2.91	3.40E-03	2.48	2.00E-02	2.29
GRMZM5G860761	1.36	2.14E-03	2.73	4.58E-03	2.70	4.00E-02	2.27
GRMZM5G843555	1.94	6.99E-09	2.55	2.52E-09	2.16	1.97E-03	2.22
GRMZM2G011523	2.09	7.62E-32	2.25	9.86E-29	2.27	1.81E-23	2.21
GRMZM2G117164	1.87	1.79E-08	2.41	4.69E-09	2.34	2.19E-06	2.21
GRMZM2G082387	1.99	3.10E-16	2.27	1.62E-14	2.32	9.89E-13	2.19
GRMZM2G425482	2.03	1.45E-24	2.22	1.63E-24	2.26	4.48E-21	2.17
GRMZM2G147399	1.98	6.46E-19	2.21	5.20E-20	2.28	2.67E-17	2.16
GRMZM2G306643	1.26	5.05E-03	2.44	1.73E-08	2.76	3.62E-03	2.15
GRMZM2G139300	1.79	4.86E-08	2.30	7.19E-08	2.31	1.02E-04	2.13
GRMZM2G026956	1.80	6.99E-09	2.24	4.44E-06	2.24	9.95E-05	2.09
GRMZM2G104626	1.62	7.59E-06	2.21	5.46E-05	2.34	7.55E-04	2.06
GRMZM2G075244	1.70	8.83E-08	2.12	6.82E-07	2.27	3.22E-05	2.03
GRMZM2G339327	1.89	4.15E-20	2.08	5.86E-20	2.07	9.96E-15	2.02
GRMZM2G094165	1.44	6.01E-04	2.28	4.05E-03	2.28	3.00E-02	2.00
GRMZM2G112488	1.54	1.32E-05	2.05	2.62E-06	2.34	4.49E-04	1.98
GRMZM2G023847	1.39	8.64E-04	2.17	7.92E-03	2.33	2.00E-02	1.97
GRMZM2G360234	1.76	8.99E-13	2.00	2.55E-14	1.97	6.55E-11	1.91
GRMZM2G308615	1.79	2.92E-20	1.94	1.01E-19	1.95	8.01E-15	1.90
GRMZM2G085964	1.77	1.13E-20	1.91	7.45E-18	1.99	8.02E-15	1.89
GRMZM2G011893	1.60	2.44E-06	2.07	1.43E-05	2.00	2.01E-03	1.89
GRMZM2G056335	1.62	4.82E-07	2.03	4.71E-07	1.99	4.21E-04	1.88
GRMZM2G168984	1.52	3.18E-05	2.05	5.35E-04	1.98	1.00E-02	1.85
GRMZM2G074604	1.62	1.24E-08	1.93	4.58E-07	1.96	3.37E-05	1.84
GRMZM2G396553	1.64	1.53E-07	2.03	1.49E-06	1.84	2.33E-04	1.84
GRMZM2G043921	1.76	6.27E-28	1.87	1.92E-23	1.84	1.00E-17	1.82
GRMZM2G088166	1.26	4.43E-03	2.22	1.02E-02	1.96	2.00E-02	1.81
GRMZM2G054224	1.70	2.19E-12	1.94	1.81E-12	1.79	1.38E-08	1.81
GRMZM2G131409	1.68	3.98E-15	1.86	1.12E-15	1.88	7.30E-13	1.81
GRMZM2G171912	1.48	5.18E-07	1.77	7.46E-06	2.16	6.11E-04	1.80
GRMZM2G136663	1.75	1.18E-17	1.92	3.76E-16	1.73	1.24E-10	1.80
GRMZM2G105844	1.36	8.12E-04	2.00	7.43E-04	2.01	7.45E-03	1.79
GRMZM5G818655	1.44	7.81E-05	1.94	1.31E-04	1.98	1.68E-03	1.79

GRMZM2G421033	1.43	1.49E-06	1.71	7.47E-05	2.21	2.87E-05	1.78
GRMZM2G156632	1.41	2.19E-04	1.94	2.05E-13	1.98	5.42E-10	1.78
GRMZM2G112629	1.52	8.76E-07	1.85	3.67E-06	1.95	1.30E-04	1.77
GRMZM2G091578	1.48	4.13E-06	1.84	6.59E-05	1.98	3.47E-04	1.77
GRMZM2G128219	1.41	9.31E-05	1.87	9.38E-04	2.02	4.24E-03	1.77
GRMZM2G413652	1.68	1.21E-21	1.80	3.58E-18	1.81	2.69E-15	1.76
GRMZM2G125813	1.69	2.88E-22	1.81	6.69E-20	1.77	1.53E-16	1.76
GRMZM2G174347	1.46	2.18E-05	1.88	3.94E-06	1.92	4.41E-04	1.75
AC225718.2_FG009	1.43	2.15E-05	1.82	2.88E-05	1.99	1.00E-02	1.74
GRMZM2G044670	1.54	2.48E-06	1.93	3.66E-05	1.74	1.62E-03	1.74
GRMZM2G428216	1.61	1.17E-13	1.78	3.03E-14	1.79	7.07E-11	1.73
GRMZM2G413897	1.48	1.25E-05	1.90	9.84E-05	1.79	2.11E-03	1.72
GRMZM2G129018	1.43	9.85E-05	1.93	1.09E-03	1.80	9.97E-03	1.72
GRMZM5G800586	1.64	1.53E-21	1.75	1.77E-17	1.77	6.00E-13	1.72
GRMZM2G066413	1.28	1.51E-03	1.86	4.06E-03	2.02	2.00E-02	1.72
GRMZM2G108991	1.30	1.57E-03	1.94	8.76E-04	1.91	1.00E-02	1.72
GRMZM5G822947	1.01	2.90E-02	1.79	1.38E-05	2.29	1.32E-06	1.70
GRMZM2G007252	1.46	2.41E-05	1.87	7.41E-05	1.75	8.96E-04	1.69
GRMZM2G090043	1.42	2.89E-05	1.82	2.87E-04	1.83	4.71E-03	1.69
GRMZM2G040559	1.50	1.61E-09	1.70	1.49E-07	1.82	2.87E-05	1.67
GRMZM2G354319	1.61	6.63E-13	1.79	1.70E-11	1.61	1.45E-05	1.67
GRMZM2G095892	1.28	1.43E-03	1.81	2.89E-03	1.89	2.00E-02	1.66
GRMZM2G445634	1.57	3.33E-14	1.72	7.64E-13	1.67	3.52E-09	1.65
GRMZM2G112524	1.54	1.63E-12	1.71	4.37E-13	1.69	5.03E-11	1.65
GRMZM2G050705	1.47	2.52E-06	1.80	2.03E-06	1.67	3.68E-03	1.65
GRMZM2G145444	1.46	3.75E-06	1.80	2.69E-04	1.63	1.00E-02	1.63
GRMZM2G084491	1.16	2.93E-03	1.58	2.51E-02	2.11	4.87E-03	1.62
GRMZM2G073044	1.51	5.44E-11	1.69	1.22E-10	1.65	1.84E-09	1.62
GRMZM2G172574	1.54	1.19E-23	1.63	1.67E-18	1.66	7.38E-14	1.61
GRMZM2G393062	1.34	1.51E-04	1.73	4.08E-05	1.74	4.46E-03	1.61
GRMZM2G147687	1.34	3.46E-04	1.80	1.93E-04	1.64	7.35E-03	1.60
GRMZM2G169013	1.43	1.58E-06	1.72	1.59E-07	1.62	3.22E-05	1.59
GRMZM2G008287	1.41	1.83E-05	1.77	6.86E-05	1.58	7.77E-04	1.59
GRMZM2G119175	1.38	1.45E-06	1.64	4.33E-06	1.73	4.15E-05	1.58
GRMZM2G057466	1.44	5.78E-09	1.63	4.01E-07	1.68	6.73E-06	1.58
GRMZM2G168365	1.32	3.03E-05	1.62	1.82E-04	1.81	2.04E-04	1.58
GRMZM2G700969	1.40	5.94E-04	2.07	3.31E-04	1.28	2.32E-03	1.58
GRMZM2G036351	1.35	8.16E-05	1.71	1.70E-03	1.68	2.00E-02	1.58
GRMZM2G031613	1.35	1.54E-05	1.64	7.04E-04	1.73	3.03E-03	1.57
GRMZM2G025611	1.54	7.40E-18	1.65	2.26E-14	1.48	2.68E-10	1.55
GRMZM2G085078	1.43	2.89E-05	1.83	7.17E-05	1.39	1.00E-02	1.55
GRMZM2G047347	1.17	1.03E-03	1.49	6.20E-03	1.98	1.00E-02	1.55
GRMZM2G080524	1.40	1.04E-05	1.71	2.26E-04	1.53	2.00E-02	1.55
GRMZM2G377613	1.42	1.56E-08	1.61	1.93E-09	1.60	1.29E-06	1.54

GRMZM2G063069	1.33	6.85E-05	1.67	2.33E-05	1.60	3.33E-03	1.53
GRMZM2G077607	1.40	6.89E-07	1.64	5.64E-07	1.56	5.39E-07	1.53
GRMZM2G438202	1.49	3.85E-09	1.70	2.69E-10	1.41	9.56E-06	1.53
GRMZM2G176307	1.37	2.11E-06	1.62	1.41E-07	1.61	1.36E-04	1.53
GRMZM2G095826	1.32	1.20E-05	1.58	6.35E-05	1.66	1.02E-03	1.52
GRMZM2G051806	1.41	4.27E-13	1.53	7.51E-13	1.62	2.23E-12	1.52
GRMZM5G884137	1.17	4.20E-03	1.69	1.88E-02	1.63	4.00E-02	1.50
GRMZM5G831577	1.36	2.40E-10	1.50	8.36E-10	1.62	9.00E-09	1.49
GRMZM2G053503	1.46	1.12E-31	1.53	4.74E-19	1.46	3.65E-14	1.49
GRMZM2G099337	1.30	5.22E-06	1.52	2.53E-05	1.62	1.25E-04	1.48
GRMZM5G858197	1.34	1.12E-04	1.72	1.48E-04	1.36	3.02E-03	1.47
GRMZM2G121573	1.24	7.46E-04	1.61	1.75E-03	1.57	3.00E-02	1.47
GRMZM2G425206	1.33	2.52E-06	1.56	4.97E-07	1.50	2.22E-06	1.46
AC149829.2_FG002	1.36	3.39E-15	1.46	1.26E-13	1.56	1.50E-11	1.46
GRMZM2G062042	1.02	4.32E-03	1.29	6.41E-03	2.04	1.29E-05	1.45
GRMZM2G084173	1.26	8.90E-05	1.54	2.00E-04	1.54	1.77E-10	1.45
AC206901.3_FG005	1.22	2.66E-03	1.73	5.83E-03	1.39	5.00E-02	1.45
AC206425.3_FG002	1.34	2.89E-14	1.44	1.08E-11	1.57	8.38E-11	1.45
GRMZM2G005954	1.31	2.26E-07	1.49	2.09E-06	1.54	1.37E-05	1.45
GRMZM2G078723	1.34	3.92E-11	1.46	3.72E-11	1.51	1.02E-08	1.44
GRMZM2G323470	1.19	7.04E-04	1.52	4.85E-03	1.58	1.00E-02	1.43
GRMZM2G104269	1.21	2.95E-04	1.50	8.66E-04	1.59	5.40E-03	1.43
GRMZM2G084005	1.36	2.74E-19	1.44	7.45E-15	1.50	2.26E-12	1.43
GRMZM2G329181	1.17	4.88E-04	1.45	7.13E-04	1.68	4.45E-04	1.43
GRMZM2G099023	1.14	3.79E-03	1.55	6.16E-03	1.60	5.00E-02	1.43
GRMZM2G112122	1.03	1.45E-02	1.49	3.09E-02	1.77	3.00E-02	1.43
GRMZM2G142315	1.30	9.04E-11	1.42	7.57E-09	1.51	1.04E-07	1.41
GRMZM2G129674	1.27	6.81E-07	1.45	1.85E-07	1.47	1.07E-06	1.40
GRMZM2G061487	1.35	1.21E-21	1.42	2.31E-15	1.41	7.99E-12	1.39
GRMZM2G036826	1.16	1.18E-03	1.48	3.93E-03	1.54	3.00E-02	1.39
GRMZM2G066902	1.15	2.77E-04	1.39	4.35E-03	1.64	1.00E-02	1.39
GRMZM2G013957	1.06	1.27E-02	1.56	8.75E-07	1.54	9.56E-06	1.39
GRMZM5G829304	1.15	9.81E-04	1.45	7.30E-03	1.56	9.21E-03	1.38
AC199526.5_FG002	1.33	3.66E-11	1.45	3.20E-11	1.36	1.28E-06	1.38
GRMZM5G839518	1.28	7.30E-10	1.41	9.73E-09	1.45	6.99E-08	1.38
GRMZM5G898768	1.20	5.51E-05	1.43	5.63E-05	1.50	1.45E-03	1.38
GRMZM2G142709	1.21	6.69E-06	1.39	4.33E-06	1.52	2.05E-04	1.37
GRMZM2G701063	1.21	6.76E-05	1.44	2.43E-05	1.46	1.30E-04	1.37
GRMZM2G135381	1.23	2.63E-06	1.41	4.14E-05	1.48	7.68E-04	1.37
GRMZM2G003489	1.29	4.86E-08	1.45	6.16E-08	1.37	1.77E-04	1.37
GRMZM2G106928	1.07	5.73E-03	1.44	7.77E-03	1.58	3.00E-02	1.36
GRMZM5G871463	1.25	8.51E-06	1.46	1.43E-06	1.38	3.93E-05	1.36
GRMZM2G113771	1.21	6.64E-06	1.39	3.63E-05	1.48	2.11E-05	1.36
GRMZM2G406099	1.30	3.75E-15	1.39	1.08E-12	1.38	3.28E-10	1.36

GRMZM2G070375	1.15	8.34E-04	1.44	9.77E-04	1.49	2.00E-02	1.36
GRMZM2G133836	1.25	5.20E-07	1.42	1.43E-06	1.40	5.78E-05	1.36
GRMZM2G467907	1.17	1.57E-03	1.52	8.64E-04	1.37	3.15E-03	1.35
GRMZM2G113355	1.29	2.26E-16	1.37	9.20E-13	1.38	7.27E-10	1.35
GRMZM2G021406	1.28	2.76E-19	1.34	1.07E-13	1.39	3.03E-10	1.34
GRMZM2G117989	1.19	1.06E-05	1.38	1.88E-06	1.42	2.08E-04	1.33
GRMZM2G433137	1.21	4.40E-06	1.39	3.45E-06	1.39	6.34E-04	1.33
GRMZM2G135044	1.29	2.05E-07	1.46	3.42E-08	1.24	1.25E-04	1.33
GRMZM2G102938	1.16	1.47E-04	1.38	1.22E-03	1.41	1.00E-02	1.31
GRMZM2G110333	1.24	2.48E-06	1.42	2.64E-06	1.27	2.81E-04	1.31
GRMZM2G103771	1.25	2.83E-13	1.33	2.36E-11	1.35	2.67E-08	1.31
GRMZM2G180246	1.13	5.62E-04	1.38	2.02E-03	1.42	2.00E-02	1.31
GRMZM2G169201	1.20	6.20E-06	1.37	6.90E-05	1.35	6.65E-04	1.31
GRMZM2G334574	1.19	1.41E-04	1.44	3.58E-04	1.28	3.00E-02	1.30
GRMZM2G166944	1.18	7.96E-05	1.40	2.39E-04	1.33	4.23E-03	1.30
GRMZM2G152175	1.15	1.80E-04	1.37	1.51E-04	1.40	3.39E-03	1.30
GRMZM2G095964	1.18	1.47E-04	1.42	7.00E-05	1.30	3.06E-03	1.30
GRMZM2G009223	1.15	4.39E-04	1.41	2.50E-03	1.32	3.00E-02	1.29
GRMZM2G157422	1.15	1.94E-06	1.29	5.54E-05	1.44	3.74E-05	1.29
GRMZM2G056600	1.24	2.35E-10	1.34	8.87E-10	1.28	4.39E-07	1.29
GRMZM2G422576	1.08	1.05E-03	1.32	7.02E-03	1.45	1.00E-02	1.28
GRMZM2G144997	1.19	3.14E-07	1.33	9.94E-08	1.32	2.77E-06	1.28
GRMZM2G303465	1.22	6.44E-12	1.32	7.54E-10	1.30	4.33E-08	1.28
GRMZM2G156486	1.13	1.22E-04	1.33	7.19E-04	1.38	2.76E-03	1.28
GRMZM2G001200	1.19	6.31E-06	1.36	8.28E-06	1.28	3.21E-04	1.27
GRMZM2G024958	1.20	7.10E-09	1.31	3.61E-07	1.29	9.78E-08	1.27
GRMZM2G477236	1.17	1.30E-06	1.31	2.78E-06	1.32	8.36E-06	1.26
GRMZM2G039828	1.12	1.39E-05	1.27	7.77E-06	1.38	4.41E-04	1.25
GRMZM2G083347	1.14	1.02E-09	1.23	1.51E-08	1.38	4.35E-06	1.25
GRMZM2G134389	1.05	1.43E-03	1.28	1.27E-03	1.42	3.00E-02	1.25
GRMZM2G001334	1.15	5.07E-06	1.30	7.41E-05	1.30	5.28E-04	1.25
GRMZM2G152908	1.07	1.05E-03	1.31	1.36E-03	1.37	1.00E-02	1.25
GRMZM5G838098	1.20	5.23E-09	1.31	1.38E-08	1.23	7.69E-06	1.25
GRMZM2G049057	1.10	8.77E-05	1.27	4.32E-04	1.34	2.37E-03	1.24
GRMZM2G080183	1.13	1.25E-07	1.24	7.20E-08	1.32	1.29E-06	1.23
GRMZM2G000397	1.05	2.32E-03	1.31	3.31E-03	1.31	3.00E-02	1.23
GRMZM2G028521	1.07	9.16E-04	1.29	3.58E-03	1.32	1.00E-02	1.23
GRMZM2G180262	1.09	1.31E-04	1.27	1.40E-04	1.31	9.27E-04	1.22
GRMZM2G004795	1.17	1.74E-15	1.24	1.90E-11	1.24	1.71E-08	1.22
GRMZM2G082924	1.13	7.09E-05	1.32	6.78E-05	1.19	6.33E-03	1.21
GRMZM2G159200	1.12	2.30E-05	1.29	4.55E-05	1.23	4.10E-03	1.21
GRMZM2G120320	1.12	5.78E-09	1.22	2.13E-08	1.29	3.58E-07	1.21
GRMZM2G088689	1.16	3.76E-12	1.23	3.91E-10	1.23	4.33E-08	1.21
GRMZM2G042870	1.08	1.35E-03	1.33	2.02E-03	1.20	2.00E-02	1.20

GRMZM2G021784	1.07	4.19E-04	1.26	1.88E-03	1.27	9.80E-03	1.20
GRMZM2G001645	1.18	1.73E-11	1.27	5.26E-10	1.14	5.07E-07	1.20
GRMZM2G029587	1.15	2.65E-18	1.21	7.51E-12	1.23	3.32E-09	1.20
GRMZM2G005066	1.13	1.97E-15	1.19	4.94E-11	1.26	4.18E-08	1.20
GRMZM2G425629	1.15	4.15E-20	1.20	3.78E-12	1.24	1.88E-09	1.20
GRMZM2G404973	1.10	7.01E-05	1.27	2.76E-05	1.21	7.39E-04	1.19
GRMZM2G140041	1.13	1.73E-11	1.21	2.05E-09	1.24	7.54E-08	1.19
GRMZM2G025833	1.14	3.55E-06	1.29	2.87E-06	1.14	1.46E-04	1.19
GRMZM2G327692	1.08	1.51E-04	1.25	1.31E-04	1.24	2.00E-02	1.19
GRMZM2G155512	1.10	1.14E-05	1.24	1.82E-04	1.23	4.41E-04	1.19
GRMZM2G092877	1.10	4.60E-07	1.21	2.93E-07	1.21	5.73E-07	1.18
GRMZM5G833389	1.04	4.83E-05	1.18	2.18E-05	1.30	1.54E-05	1.17
GRMZM2G471357	1.14	1.59E-13	1.21	8.40E-10	1.16	2.22E-07	1.17
GRMZM2G096352	1.03	3.61E-04	1.19	2.21E-03	1.28	1.00E-02	1.17
GRMZM2G052336	1.09	7.72E-07	1.20	3.78E-07	1.20	2.35E-05	1.16
GRMZM2G434514	1.05	2.74E-04	1.22	3.70E-03	1.21	3.00E-02	1.16
GRMZM2G120971	1.09	2.92E-07	1.20	2.82E-07	1.20	2.19E-06	1.16
GRMZM2G301860	1.16	7.13E-12	1.24	3.94E-10	1.07	3.51E-06	1.16
GRMZM2G082487	1.10	1.59E-13	1.16	4.95E-10	1.21	6.78E-08	1.16
GRMZM2G166297	1.09	4.70E-11	1.17	3.10E-09	1.20	2.14E-07	1.16
GRMZM2G442555	1.10	4.44E-08	1.20	5.63E-08	1.16	7.69E-06	1.15
GRMZM2G409104	1.06	1.16E-07	1.15	1.98E-07	1.21	3.50E-06	1.14
GRMZM2G108997	1.06	4.00E-07	1.16	7.52E-07	1.20	1.91E-05	1.14
GRMZM2G000812	1.09	4.23E-07	1.20	3.15E-07	1.12	5.26E-04	1.14
GRMZM2G134753	1.08	8.46E-05	1.24	1.23E-04	1.06	4.94E-04	1.13
GRMZM2G396541	1.00	5.09E-03	1.27	5.74E-03	1.11	3.00E-02	1.13
GRMZM2G347280	1.06	2.89E-06	1.17	1.67E-06	1.15	1.15E-05	1.13
GRMZM2G154735	1.07	3.82E-06	1.20	1.62E-06	1.10	4.51E-04	1.12
GRMZM2G000739	1.02	4.23E-07	1.12	3.09E-06	1.22	2.39E-04	1.12
GRMZM2G019056	1.02	1.60E-04	1.16	4.37E-04	1.18	1.00E-02	1.12
GRMZM2G303149	1.03	3.75E-06	1.14	2.77E-06	1.18	2.76E-05	1.12
GRMZM2G071582	1.00	1.06E-05	1.11	6.42E-05	1.21	1.40E-04	1.11
GRMZM2G368886	1.05	6.56E-11	1.12	1.35E-08	1.13	3.38E-06	1.10
GRMZM2G311187	1.05	1.85E-09	1.13	3.51E-08	1.10	1.65E-05	1.09
GRMZM2G162359	1.05	3.66E-11	1.12	1.19E-08	1.10	1.82E-06	1.09
GRMZM2G425863	1.02	7.62E-06	1.13	9.65E-06	1.10	2.05E-04	1.08
GRMZM2G407223	1.03	1.58E-10	1.09	2.42E-08	1.11	3.13E-06	1.08
GRMZM2G059453	1.02	3.92E-11	1.08	1.73E-08	1.12	1.07E-06	1.08
GRMZM2G053206	1.03	2.40E-09	1.11	5.65E-08	1.09	3.76E-06	1.08
GRMZM2G089557	1.01	8.80E-07	1.11	2.83E-05	1.05	7.42E-04	1.06
GRMZM2G130127	-1.04	1.15E-08	-1.07	1.88E-07	-1.14	1.23E-05	-1.09
GRMZM2G417859	-1.02	1.23E-03	-1.17	8.16E-04	-1.09	2.00E-02	-1.09
GRMZM2G345055	-1.03	1.23E-05	-1.11	1.51E-05	-1.17	2.82E-04	-1.10
GRMZM2G497928	-1.07	9.57E-04	-1.26	2.36E-03	-1.05	3.00E-02	-1.13

GRMZM5G896756	-1.01	4.08E-03	-1.23	4.26E-03	-1.21	3.00E-02	-1.15
GRMZM2G152447	-1.11	8.16E-08	-1.17	2.09E-07	-1.20	1.19E-06	-1.16
AC195587.4_FG004	-1.01	2.30E-03	-1.19	7.51E-03	-1.28	2.00E-02	-1.16
GRMZM2G306597	-1.11	8.65E-04	-1.32	4.65E-04	-1.06	2.00E-02	-1.16
GRMZM2G040750	-1.13	9.29E-07	-1.21	1.85E-06	-1.15	3.69E-05	-1.16
GRMZM2G335146	-1.07	1.70E-04	-1.20	7.41E-05	-1.23	8.35E-05	-1.17
GRMZM2G073351	-1.12	8.01E-07	-1.20	9.22E-07	-1.26	1.04E-05	-1.19
GRMZM2G386273	-1.05	2.16E-03	-1.26	2.90E-03	-1.30	4.00E-02	-1.20
GRMZM2G015599	-1.15	3.38E-06	-1.25	6.97E-06	-1.24	3.66E-06	-1.21
GRMZM5G899800	-1.13	9.28E-04	-1.36	1.27E-03	-1.16	5.00E-02	-1.21
GRMZM2G104942	-1.01	6.56E-03	-1.27	1.37E-02	-1.38	3.00E-02	-1.22
GRMZM2G064145	-1.10	1.90E-03	-1.34	4.01E-03	-1.23	5.00E-02	-1.22
GRMZM2G143367	-1.07	6.43E-03	-1.40	7.66E-03	-1.25	1.68E-03	-1.24
GRMZM2G141355	-1.07	3.22E-03	-1.33	6.47E-03	-1.37	5.00E-02	-1.26
GRMZM2G055578	-1.06	1.41E-03	-1.26	2.36E-03	-1.49	1.00E-02	-1.27
GRMZM2G005499	-1.15	3.12E-04	-1.35	4.64E-03	-1.34	3.00E-02	-1.28
GRMZM2G472852	-1.24	4.28E-09	-1.32	9.19E-08	-1.33	1.22E-06	-1.30
GRMZM2G331316	-1.21	2.15E-03	-1.63	3.99E-02	-1.05	5.74E-04	-1.30
GRMZM5G845532	-1.11	8.25E-04	-1.32	1.82E-03	-1.47	1.29E-06	-1.30
GRMZM2G046430	-1.25	3.12E-05	-1.45	4.56E-05	-1.23	2.00E-02	-1.31
GRMZM2G052206	-1.04	7.55E-03	-1.35	2.66E-02	-1.56	4.00E-02	-1.31
GRMZM2G020411	-1.06	1.79E-02	-1.70	3.93E-02	-1.23	3.00E-02	-1.33
GRMZM2G123459	-1.33	8.80E-07	-1.50	3.75E-05	-1.31	2.20E-03	-1.38
GRMZM2G003711	-1.12	4.04E-03	-1.48	9.97E-03	-1.56	2.11E-03	-1.39
GRMZM5G836174	-1.21	2.97E-04	-1.47	1.70E-04	-1.49	3.39E-03	-1.39
GRMZM2G034724	-1.38	8.04E-23	-1.41	4.75E-15	-1.50	8.48E-13	-1.43
GRMZM2G357737	-1.18	2.10E-03	-1.56	4.28E-03	-1.61	2.00E-02	-1.45
GRMZM5G850924	-1.09	1.84E-02	-2.21	4.18E-02	-1.16	2.00E-02	-1.49
GRMZM2G106092	-1.42	5.18E-05	-1.80	3.57E-05	-1.29	1.79E-06	-1.50
GRMZM2G145632	-1.50	4.07E-06	-1.83	4.48E-06	-1.34	4.07E-05	-1.56
GRMZM2G022945	-1.31	7.77E-04	-1.78	1.55E-03	-1.60	1.00E-02	-1.57
GRMZM2G018786	-1.37	2.68E-05	-1.67	2.38E-03	-1.72	6.46E-03	-1.59
GRMZM2G075018	-1.23	6.18E-03	-2.19	3.91E-02	-1.41	3.64E-04	-1.61
GRMZM2G062752	-1.43	7.11E-04	-2.28	2.82E-03	-1.27	3.00E-02	-1.66
GRMZM2G128865	-1.24	3.80E-03	-1.88	1.33E-02	-1.92	2.00E-02	-1.68
GRMZM2G312110	-1.41	9.16E-05	-1.81	3.83E-05	-1.91	1.95E-03	-1.71
GRMZM2G089750	-1.59	3.00E-08	-1.85	1.12E-07	-1.78	2.49E-05	-1.74
GRMZM2G109327	-1.12	1.44E-02	-2.13	3.91E-02	-2.07	1.01E-06	-1.78
GRMZM2G127948	-1.81	1.74E-16	-1.96	1.75E-13	-1.88	6.51E-11	-1.88
GRMZM2G024527	-1.53	1.59E-04	-2.31	7.37E-04	-2.13	3.00E-02	-1.99
GRMZM5G857998	-1.44	1.07E-03	-2.95	3.52E-03	-1.84	1.50E-03	-2.07
GRMZM2G085921	-2.05	6.34E-10	-2.64	1.87E-07	-1.86	2.17E-05	-2.18
GRMZM2G300424	-2.35	1.19E-15	-2.86	7.97E-10	-1.51	1.75E-08	-2.24
GRMZM2G177771	-1.48	7.11E-04	-2.87	5.01E-03	-2.54	9.92E-03	-2.30

GRMZM2G103033	-2.15	8.64E-12	-2.69	3.16E-09	-2.51	9.17E-05	-2.45

Table S2. Result of the Singular Enrichment Analysis using the agriGO tool: http://bioinfo.cau.edu.cn/agriGO/

Species: *Zea mays*

GO type: Completed GO

Background/Reference: Zea mays AGPv3.30

Annotated number in query list: 188

Annotated number in background/reference: 25864

Significant GO terms: 59

Detail information

GO term	Ontology*	Description	Number in input list	Number in BG/Ref	p-value	FDR
GO:0009266	P	response to temperature stimulus	16	464	4.6e-07	0.0002
GO:0006950	P	response to stress	39	2243	3.2e-07	0.0002
GO:0009788	P	negative regulation of abscisic acid mediated signaling pathway	5	29	4.4e-06	0.00096
GO:0009409	P	response to cold	12	318	5.4e-06	0.00096
GO:0009628	P	response to abiotic stimulus	27	1397	4.2e-06	0.00096
GO:0009968	P	negative regulation of signal transduction	5	36	1.1e-05	0.0012
GO:0023057	P	negative regulation of signaling process	5	36	1.1e-05	0.0012
GO:0010648	P	negative regulation of cell communication	5	36	1.1e-05	0.0012
GO:0050896	P	response to stimulus	48	3551	1.3e-05	0.0013
GO:0042221	P	response to chemical stimulus	33	2052	1.5e-05	0.0013
GO:0010033	P	response to organic substance	21	1030	2.5e-05	0.0019
GO:0009414	P	response to water deprivation	10	254	2.4e-05	0.0019
GO:0009415	P	response to water	10	264	3.2e-05	0.0022
GO:0009787	P	regulation of abscisic acid mediated signaling pathway	5	50	4.8e-05	0.003
GO:0065007	P	biological regulation	47	3697	7.9e-05	0.0046

☐	GO:0071215	P	cellular response to abscisic acid stimulus	6	97	0.0001	0.0057
☐	GO:0048585	P	negative regulation of response to stimulus	5	68	0.00019	0.0094
☐	GO:0009966	P	regulation of signal transduction	7	158	0.0002	0.0094
☐	GO:0023051	P	regulation of signaling process	7	158	0.0002	0.0094
☐	GO:0050789	P	regulation of biological process	41	3207	0.00023	0.0094
☐	GO:0010646	P	regulation of cell communication	7	162	0.00023	0.0094
☐	GO:0010468	P	regulation of gene expression	28	1863	0.00023	0.0094
☐	GO:0009725	P	response to hormone stimulus	16	787	0.00025	0.0096
☐	GO:0060255	P	regulation of macromolecule metabolic process	28	1909	0.00033	0.01
☐	GO:0050794	P	regulation of cellular process	38	2928	0.0003	0.01
☐	GO:0006355	P	regulation of transcription, DNA-dependent	26	1706	0.00031	0.01
☐	GO:0006350	P	transcription	28	1916	0.00035	0.01
☐	GO:0006351	P	transcription, DNA-dependent	28	1915	0.00035	0.01
☐	GO:0016311	P	dephosphorylation	7	172	0.00033	0.01
☐	GO:0051252	P	regulation of RNA metabolic process	26	1719	0.00035	0.01
☐	GO:0045449	P	regulation of transcription	26	1708	0.00032	0.01
☐	GO:0032774	P	RNA biosynthetic process	28	1922	0.00037	0.01
☐	GO:0009738	P	abscisic acid mediated signaling pathway	5	83	0.00045	0.012
☐	GO:0019219	P	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	26	1755	0.00048	0.012
☐	GO:0010118	P	stomatal movement	5	86	0.00052	0.013
☐	GO:0009719	P	response to endogenous stimulus	16	848	0.00056	0.013
☐	GO:0051171	P	regulation of nitrogen compound metabolic process	26	1773	0.00055	0.013
☐	GO:0019222	P	regulation of metabolic process	29	2085	0.00061	0.013
☐	GO:0010556	P	regulation of macromolecule biosynthetic process	26	1783	0.0006	0.013
☐	GO:0009737	P	response to abscisic acid stimulus	10	383	0.0006	0.013
☐	GO:0009743	P	response to carbohydrate stimulus	6	142	0.00074	0.016

☐	GO:0031326	P	regulation of cellular biosynthetic process	26	1823	0.00083	0.017
☐	GO:0009889	P	regulation of biosynthetic process	26	1828	0.00086	0.018
☐	GO:0006470	P	protein amino acid dephosphorylation	5	103	0.0011	0.023
☐	GO:0080090	P	regulation of primary metabolic process	26	1868	0.0012	0.023
☐	GO:0031323	P	regulation of cellular metabolic process	27	1976	0.0012	0.023
☐	GO:0044282	P	small molecule catabolic process	8	289	0.0015	0.028
☐	GO:0006096	P	glycolysis	5	113	0.0017	0.031
☐	GO:0009723	P	response to ethylene stimulus	5	119	0.0021	0.037
☐	GO:0019318	P	hexose metabolic process	7	242	0.0023	0.039
☐	GO:0006970	P	response to osmotic stress	11	538	0.0022	0.039
☐	GO:0044275	P	cellular carbohydrate catabolic process	6	186	0.0028	0.047
☐	GO:0003700	F	transcription factor activity	20	891	1e-05	0.0022
☐	GO:0030528	F	transcription regulator activity	20	974	3.6e-05	0.0039
☐	GO:0016791	F	phosphatase activity	8	242	0.00049	0.035
☐	GO:0043169	F	cation binding	47	4135	0.001	0.035
☐	GO:0046872	F	metal ion binding	47	4124	0.00097	0.035
☐	GO:0043565	F	sequence-specific DNA binding	12	571	0.0011	0.035
☐	GO:0043167	F	ion binding	47	4136	0.001	0.035

* P = Biological Process, F = Molecular Function

„Ich habe mich bemüht sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.“