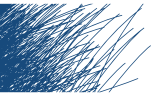


UNIVERSITEIT GENT

FACULTY OF **PHARMACEUTICAL SCIENCES**

Strategies to detect unauthorized GMO in the food and feed chain

Marie-Alice FRAITURE

Dissertation presented in partial
fulfilment of the requirements for the
degree of Doctor in Pharmaceutical Sciences

June 2016

Strategies to detect unauthorized GMO in the food and feed chain

Marie-Alice FRAITURE

Supervisors:
Prof. D. Deforce, UGent
Prof. M. De Loose, ILVO
Dr. N. Roosens, WIV-ISP

Members of the Examination Committee:
Prof. T. Coenye, chair, UGent
Prof. S. De Saeger, secretary, UGent
Prof. A. Depicker, UGent
Dr. P. Herman, WIV-ISP
Dr. E. Kok, RIKILT Wageningen

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Pharmaceutical Sciences

June 2016

Uitgegeven in eigen beheer, Marie-Alice Fraiture, Ottergemsesteenweg 460, B-9000 Ghent, Belgium

ISBN 9789059898790

This PhD was incorporated within the framework of the UGMMONITOR project, which is funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment (convention RF 11/6242). This project, aiming to develop a platform in molecular biology and a database (biosafety information) for the detection of EU unauthorized GMO in food and feed chain, was a consortium between three partners that are WIV-ISP (Nancy Roosens and Philippe Herman), ILVO (Marc De Loose and Isabel Taverniers) and CRA-W (Gilbert Berben, Frédéric Debode and Eric Janssen).

During this PhD, Marie-Alice Fraiture was a member of both WIV-ISP, ILVO and UGent. At the practical level, even if this PhD was made in close collaboration with Prof. Dr. Dieter Deforce (UGent), Dr. Ir. Nancy Roosens and Prof. Dr. Marc De Loose, Marie-Alice Fraiture worked for two years in WIV-ISP (2012-2014) and for two years in ILVO (2014-2016).

Abstract

To guarantee the traceability on the market and the freedom of choice for consumers, genetically modified organisms (GMO) legislations have been established in many countries, including in Europe (EU). However, the implementation of these legislations by the enforcement laboratories is becoming complex due mainly to the increasing number and diversity of GMO. To cope with the problematic of EU unauthorized GMO, this PhD aims to improve and strengthen the existing GMO detection system using high-tech approaches.

First, as a study case, an overview of genetically modified (GM) rice, developed around the world was carried out to collect information related *inter alia* on elements found in their transgenic cassette. Second, according to this information, key targets, frequently found in GMO (p35S and tNOS) or exclusively observed in EU unauthorized GMO (t35S pCAMBIA), were selected to develop a strategy allowing to detect and characterize a broad range of GMO. This strategy, fully integrated in the GMO routine analysis, consists to characterize sequences surrounding detected key transgenic elements using a DNA walking approach. By this way, the acquisition of sequences from the junction between the transgenic cassette and the plant genome as well as the associations of elements typically found in transgenic constructs allow to confirm the presence of GMO in food/feed matrices. Due to its good performance thoroughly assessed via several unprocessed and processed food/feed matrices, this strategy represents a key tool, easily implementable by the enforcement laboratories. With the aim to even more simplify the workflow and increase the throughput of this strategy, the sequencing step was performed using the Next Generation Sequencing (NGS) technology instead of the Sanger technology.

In parallel, the detection of GMO in alimentary matrices using exclusively the NGS technology, through a whole genome sequencing (WGS) approach, was also investigated. As this last approach does not theoretically require any prior information about the targeted sequences, GMO composed only of unknown transgenic elements could be detected.

This work has thus allowed to provide additional strategies to the current GMO detection system in order to characterize a larger spectrum of GMO, both authorized or not.

Abbreviations

C_q	quantification cycle
C_t	threshold cycle
D	sequenced paired-end distance
H	length of the non-GM host genome
I	length of the insert
M	minimum overlap of nucleotides
N	number of reads
O	overlap
P	probability
R	sequenced read length
R^2	linearity of the PCR
T_m	melting temperature
T_p	total of different theoretical paired-end reads
T_s	total of different mates
p	ratio of targeted reads over total reads
r	ratio of the genome reference length over the actual genome length
δC_t	difference of C_t values
$\delta\delta C_\mathrm{t}$	difference between the observed δC_t value and the theoretical δC_t value
aadA	aminoglycoside 3'-adenylyltransferase
ACC	acetyl-CoA-carboxylase gene from colza
acp1	acyl carrier protein 1 gene from cotton
ADH	alcohol dehydrogenase I gene from maize
adhC	alcohol dehydrogenase C gene from cotton
AHAS	AHAS fragment unique recombination from BPS-CV-127
Als	acetolactate synthase
Amp	ampicillin resistance gene

Bar	bialophos resistance
bar	phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*
BLAST	basic local alignment search tool
bp	base pairs
BS	bispyribac-sodium
BWA	Burrows-Wheeler alignment
CaMV	ORFIII from CaMV
CaMV	cauliflower mosaic virus
CCS	consensus circular sequencing
cdPCR	chamber dPCR
CGE	capillary gel electrophoresis
CGH	comparative genomic hybridisation
Chy	chymopapain gene from papaya
CoSYPS	Combinatory SYBR®Green qPCR Screening
CP4-EPSPS	5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain
CpTI	trypsin inhibitor in cowpea *Vigna unguiculata*
CRA-W	Centre wallon de Recherches Agronomiques
CRISPR	clustered regularly interspaced short palindromic repeats
CRM	certified reference material
CRT	reverse transcriptase gene from the cauliflower mosaic virus
crt	carotene desaturase
CRU	cruciferin gene from colza
Cry	gene encoding the *Bacillus thuringiensis* δ-endotoxin
Cry1B	synthetic Cry1B gene
Cry3Bb	gene encoding the *Bacillus thuringiensis* δ-endotoxin 3Bb
Cry3Bb1	synthetic Cry3Bb1 gene
CSHL	Cold Spring Harbor Laboratory
CTAB	C-hexadecyl-trimethyl-ammonium-bromide
CTP2/CP4-EPSP	junction region between the chloroplast transit peptide 2 (CTP2) sequence from the *Arabidopsis thaliana* epsps gene and the CP4 epsps gene from *Agrobacterium tumefaciens* (CP4-EPSPS)
CTP4	chloroplast transit peptide 4 from the *Arabidopsis thaliana* epsps gene

ddNTP	dideoxynucleotides
ddPCR	droplet dPCR
DNA	deoxyribonucleic acid
dNTP	deoxynucleotides
dPCR	digital PCR
DRT	degenerated random tagging
DSS	decision support system
EC	European Community
EEC	European Economic Community
EFSA	European Food Safety Authority
ELISA	enzyme-linked immunosorbent assay
ENGL	European Network of GMO Laboratories
EPSPS	5-enolphyruvylshikimate-3-phosphate synthase
EU	European Union
EU-RL GMFF	European Union Reference Laboratory for GM Food and Feed
F	forward
FASFC	Federal Agency for the Safety of the Food Chain (Federaal Agentschap voor de Veiligheid van de Voedselketen (FAVV) - Agence Fédérale pour la sécurité de la Chaîne Alimentaire (AFSCA))
FRUp	β-fructosidase gene from potato
FRUt	β-fructosidase gene from tomato
gDNA	genomic DNA
GE	genome editing
GLU	glutamine synthetase gene from sugar beet
GM	genetically modified
GMO	genetically modified organism
Gox	glyphosate oxidoreductase gene
HGE	haploid genome equivalent
HMG	major high mobility group protein gene from maize
hpt	hygromycin phosphotransferase gene
iAct	Rice Actin intron
ILVO	Institute for Agricultural and Fisheries Research (Instituut voor Landbouw- en Visserijonderzoek)

IPC	internal positive control
IVR	invertase gene from maize
JRC	Joint Research Centre
Kbp	kilo base pairs
KL	knowledge level
lacZ	LacZ alpha fragment
LAMP	loop-mediated isothermal amplification
LB	left border
LEC	lectin gene from soybean
LLP	low level presence
LOD	limit of detection
LS28	choline kinase
LT-RADE	long template-rapid amplification of genomic DNA ends
LY	transition from *Zea mays* chloroplast transit peptide sequence for dihydrodipicolinate synthase to *Corynebacterium glutamicum* dihydrodipicolinate synthase (cordapA) gene encoding for a lysine-insensitive dihydrodipicolinate synthase enzyme
MACRO	multiplex amplification on a chip with readout on an oligo microarray
mas1	mannopine synthase region
Mbp	mega base pairs
MLSEB	multi-labelling system based on electrochemical biosensor
MQDA-PCR	multiplex quantitative DNA array-based PCR
NAIMA	nucleic acid sequence based amplification implemented microarray analysis
nBLAST	nucleotide BLAST
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
NPBT	new plant breeding techniques
nptII	neomycin phosphotransferase II gene
NRL	national reference laboratory
NTC	no template control

ODM	oligonucleotide directed mutagenesis
p35S	cauliflower mosaic virus 35S promoter
p4-AS1	modified cauliflower mosaic virus 35S promoter
PacBio	Pacific Biosciences
pAct	promoter region of rice actin gene
pAct1	rice actin 1 gene promoter
PAT	phosphinotricin-N-acetyltransferase
pat	phosphinothricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*
PBB	Platform Biotechnology and Molecular Biology
PC	phosphoenolpyruvate carboxylase gene from wheat
PCR	polymerase chain reaction
pFMV	promoter of the figwort mosaic virus
phy	phytase gene from maize
PINII	potato proteinase inhibitor II
PLD	phospholipase D gene from rice
pNOS	promoter of the nopaline synthase gene
PPLMD	padlock probe ligation-microarray detection
PPO	protoporphyrinogen oxidase
pRbcS4	ribulose 1,5-bisphosphate carboxylase small subunit promoter from *Arabidopsis thaliana*
Pro	prolamin gene from rice
PSY	phytoene synthase
pUBI	maize ubiquitin promoter
qPCR	quantitative real-time PCR
R	reverse
r.f. CP4/EPSPS	repeated fragment of CP4/EPSPS
R&D	research and development
rAct	rice actin intron
RB	right border
RBCL	ribulose-1,5-biphosphate carboxylase oxygenase
RdDM	RNA-dependent DNA methylation
RNA	ribonucleic acid
RPA	recombinase polymerase amplification
RSD_r	relative standard deviation of the test results
RT	room temperature

SAD1	stearoyl-acyl-carrier protein desaturase gene from cotton
SAM	sequence alignment/map
SAMS	transition from S-adenosyl-L-methionine synthetase (SAMS) promoter to *Glycine max* acetolactate synthase (gm-hra) gene
SBB	Biosafety and Biotechnology
SMRT	single-molecule real-time
SPS	sucrose phosphate synthase gene from rice
ssDNA	single strand DNA
ssIIb	starch synthase IIb gene from maize
T-DNA	transfer DNA
t35S	cauliflower mosaic virus 35S terminator
t35S pCAMBIA	terminator of the cauliflower mosaic virus from pCAMBIA vector
T&V	Technology and Food Science Unit (Technologie en Voeding)
tAHASL	acetohydroxy acid synthase large subunit terminator from *Arabidopsis thaliana*
tahsp17	wheat heat shock protein terminator
TAIL-PCR	thermal asymmetric interlaced-PCR
TALEN	transcription-activator like effector nucleases
tE9	ribulose-1,5-bisphosphate carboxylase terminator E9 from *Pisum sativum*
Ti	tumor-inducing
TLC	tRNA-Leu chloroplastic gene
tNOS	*Agrobacterium tumefaciens* terminator of the nopaline synthase gene
tORF23	open reading frame 23 terminator from *Agrobacterium tumefaciens*
tpinII	inhibitor II terminator from potato
UGent	Ghent University
UGM	unauthorised GMO
uidA	β-glucuronidase
USA	United Stated of America
VIP3A	vegetative insecticidal protein 3A
WGS	whole genome sequencing

WIV-ISP	Scientific Institute of Public Health (Wetenschappelijk Instituut Volksgezondheid - Institut Scientifique de Santé Publique)
WT	wild-type
wtCAB	wheat major chlorophyll a/b binding protein gene
Xa21	Xa21 gene from *Oryza longistaminata*
Zein	Zein gene from maize
ZFN	zinc finger nucleases
ZMW	zero-mode-waveguide

Contents

List of figures

List of tables

Chapter 1

General introduction

1.1 What is a GMO?

A genetically modified organism (GMO) is an organism in which the genetic material has been modified through the use of modern biotechnology. Each instance of a GMO is referred to the term event, which is characterized by an unique insertion of a certain DNA combination into a specific location on the host genome. Most of genetically modified (GM) crops were produced via the biolistic or *Agrobacterium* approaches (Figure 1.1).
With the biolistic transformation, DNA is coated to microprojectilles to be then propelled into plant cells by acceleration. The delivered DNA is eventually integrated into the plant genome. In the case of the *Agrobacterium*-mediated transformation, T-DNA of the binary vector within the gram negative bacterial *Agrobacterium tumefaciens* species is transferred into the plant cell with the help of virulence genes. The T-DNA, defined by left and right borders, is then randomly integrated into the plant genome.
In this way, genes of interest from one living organism could be introduced into the genome of another. Following to its regeneration from a single transformed cell, the transgenic organism could present new agricultural, nutritional or therapeutic properties (Nill 2002; Mirkov 2003; Parisi *et al.* 2016; SCBD 2016).
The first transgenic plant was obtained in 1983. This transgenic tobacco was transformed with the *Agrobacterium* transformation approaches in order to present a resistance to kanamycin and methotrexate (Herrera-Estrella *et al.* 1983;

Schell *et al.* 1983). The same year, transgenic petunia resistant to kanamycin and transgenic sunflower with bean phasoelin gene were also created (Fraley *et al.* 1983; Murai *et al.* 1983). In 1986, the first field trial was organized in the USA and France with transgenic tobacco plants (James and Krattiger 1996). As the present work was mainly focused on transgenic rice, as a study case, the key transformation steps of its history are briefly recounted here. The first transgenic rice were generated in 1988 by electroporation-mediated or polyethylene glycol-mediated protoplast transformation methods. In 1991, transgenic rice plants were then produced following to a biolistic transformation. This approach presents the advantage to be effective regardless of species. However, the delivered DNA is usually integrated in multiple copies, often fragmented and rearranged. Moreover, the delivery of long fragments could induce damages to the naked DNA (Hiei *et al.* 1997; Kathuria *et al.* 2007). To overcome this problem, the *Agrobacterium*-mediated transformation was used on rice in 1990. Given that monocotyledonous species, as rice, are not the natural host of this soil bacteria, some modifications were thus brought on the initial transformation protocol to successfully generate transgenic rice (Hiei *et al.* 1997; Hoque *et al.* 2005). For instance, the addition of phenolic compounds, such as acetosyringone, promotes the activation of the T-DNA transfer process through the stimulation of the expression of virulence genes. In addition, the kind of plant tissues used, such as with a high cell division activity, is determining. For the antibiotic selection of the transformed rice cells, the hygromycin phosphotransferase gene is more convenient than the genes resistant to kanamycin and G418 since this hygromycin B antibiotic is not damaging for the rice regeneration and fertility. This is due to the fact that rice present no innate resistance to the hygromycin B antibiotic, conversely to the kanamycin and G418. In addition, even if the phosphinothricin antibiotic is less observed in GM rice, its use was shown to be at least as effective as hygromycin (Hiei *et al.* 1997; Twyman *et al.* 2002; Kathuria *et al.* 2007).

1.2 GMO legislations

To guarantee the traceability in the food/feed chain as well as the freedom of choice for consumers, legislations have been established in many countries regarding to the introduction and the control of GMO in the food/feed chain. However, regarding the meaning of GMO, these legislations present some variations to the scientific definition indicated in the previous section. Indeed, two main kind of legislations, based on either the finished product (e.*g.*, USA, Canada and Japan) or the production process (*e.g.*, EU, China and Brazil), are distinguishable. In the first case, products are thus considered as GMO if they present new associations of genetic elements that do not occur in nature (product-

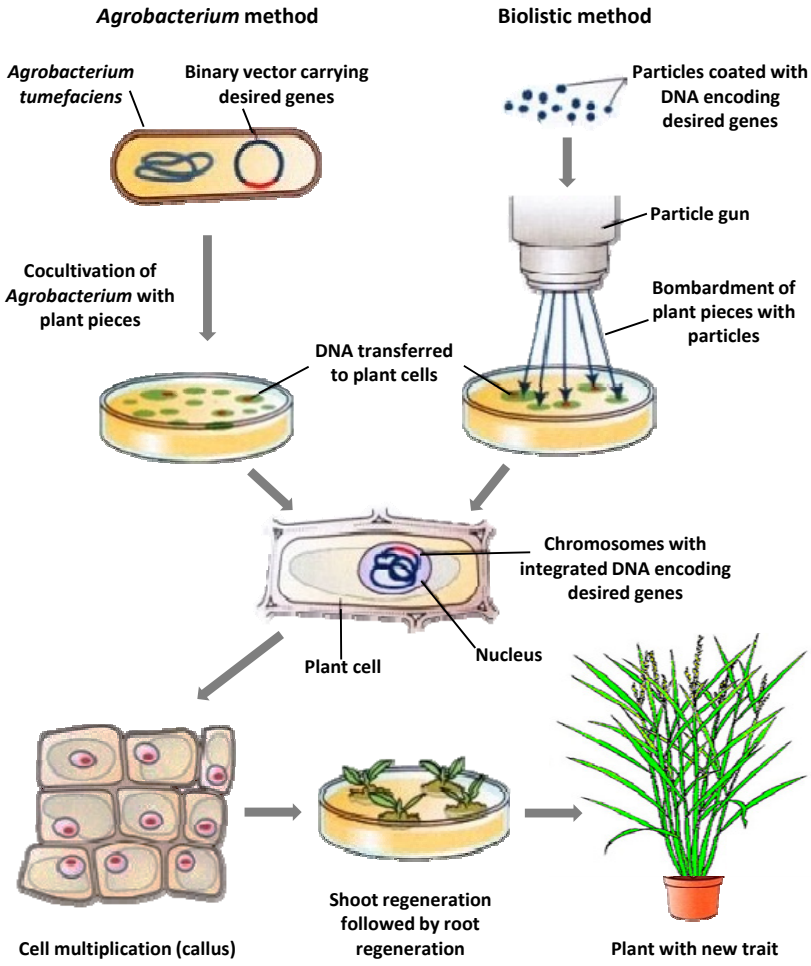


Figure 1.1: Schematic representation of classical processes to generate transgenic plants (adapted from Mirkov 2003).

based regulation). In the second case, GMO were generated by using certain biotechnology tools (process-based regulation). The EU legislation, where GMO are defined as organisms "in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination" (The European Parliament and The Council Of The European Union 2001; The European Parliament and The Council Of The European Union 2003a; The European Parliament and The Council Of The European Union 2003b), belongs to the second case. Among all the existing legislations, the labelling regulations differ in terms of practical application (mandatory or voluntary) and threshold of tolerance (varying from 0% to 5%). Moreover, some countries do not or partially implement their labelling policy in spite of their adoption, as notably observed in Brazil where the labelling law was voted since 2003 (Davison and Bertheau 2007; Gruère and Rao 2007; Broeders *et al.* 2012a; Katovich 2012; Andersen *et al.* 2015). On the EU market, a series of regulations, directives and recommendations have been implemented and updated following to a continuous collaboration between the research experts, including the National Reference Laboratories (NRL), and the competent authorities via the European Network of GMO laboratories (ENGL), hosted by the Joint Research Centre (JRC) of the European Commission (Table 1.1).

Table 1.1: EU legislations related to GMOs.

Official documents	Purposes
Directive 90/219/EEC	Contained use of GM micro-organisms.
Directive 90/220/EEC	The deliberate release into the environment of GMOs.
Regulation (EC) 258/97	Concerning novel foods and novel food ingredients.
Directive 2009/41/EC	Repeals Directive 90/219/EEC and its successive amendments (Directive 98/81/EC, Decision 2001/204/EC, Directive 94/51/EC).
Directive 2001/18/EC	The deliberate release into the environment of GMOs and repealing Council Directive 90/220/EEC.
Regulation (EC) 178/2002	Laying down the general principles and requirements of food law, establishing the European Food Safety Authority (EFSA) and laying down procedures in matters of food safety.
Regulation (EC) 1946/2003	Covering trans-boundary movements of GMOs.

Table 1.1 *continued*

Official documents	Purposes
Regulation (EC) 1829/2003	Covering commercialized GMOs in food and feed.
Regulation (EC) 1830/2003	The traceability and labelling of GMOs and the traceability of food and feed products produced from GMOs amending Directive 2001/18/EC.
Regulation (EC) 65/2004	Establishing of a system for the development and assignment of unique identifiers for GMOs.
Decision 2004/204/EC	Laying down detailed arrangements for the operation of the registers for recording information on genetic modifications in GMOs, provided for in Directive 2001/18/EC of the European Parliament and of the Council.
Regulation (EC) 641/2004	Detailed rules for the implementation of Regulation (EC) No 1829/2003 of the European Parliament and of the Council as regards the application for the authorization of new GM food and feed, the notification of existing products and adventitious or technically unavoidable presence of GM material which has benefited from a favorable risk evaluation.
Recommendation 2004/787/EC	Technical guidance for sampling and detection of GMOs and material produced from GMOs as or in products in the context of Regulation (EC) No 1830/2003.
Regulation (EC) 882/2004	Official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules.
Regulation (EC) 1981/2006	Detailed rules for the implementation of Article 32 of Regulation (EC) No 1829/2003 of the European Parliament and of the Council as regards the Community reference laboratory for GMOs.
Recommendation 2010/01/EC	Guidelines for the development of national coexistence measures to avoid the unintended presence of GMOs in conventional and organic crops.
Regulation (EC) 619/2011	Laying down the methods of sampling and analysis for the official control of feed as regards presence of GM material for which an authorization procedure is pending or the authorization of which has expired.
Regulation (EC) 503/2013	Applications for authorization of GM food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006.

To introduce GM crops on the EU market, a complex approval procedure should be followed to get the required authorization. A dossier from the applicant is assessed only by the relevant national competent authority for GM crop cultivation while, in case of food/feed use application, a dossier is submitted to the EFSA that takes into account the scientific and the member state competent authorities points of views (Devos *et al.* 2014; Hartung and Schiemann 2014). All these dossiers, including an event-specific method, allowing to identify each event by targeting specifically its junction between the transgenic cassette and the host genome, and a risk assessment, are evaluated by the EFSA in term of Biosafety assessment, including a molecular characterization of the insert(s) such as the transgene flanking regions, a toxic and allergen risk assessment of the inserted genes, a comparison to the non-GM counterpart at the agronomic, phenotypic and compositional levels and a nutritional assessment of GMO-derived food/feed products using the 90-days animal feeding trials. By this way, GMOs are considered as equivalent to their non-GM counterparts, apart from the new traits. In case of staked GM events, generated by conventional breeding of single GM events that are individually authorized or not, a safety assessment is also required by the Regulation (EC) 503/2013 in terms of genetic stability, genetic expression and potential synergistic or antagonistic effects (Devos *et al.* 2014; Kok *et al.* 2014; Ricroch *et al.* 2014). Regarding the submitted event-specific methods, the European Union Reference Laboratory for GM Food and Feed (EU-RL GMFF) is responsible for the validation. After reception of the documentations and materials, the documentations and data are submitted to a scientific assessment. In case of a positive assessment, the method is then experimentally tested to verify if the method fulfils the method acceptance criteria defined by the ENGL. The methods succeeding to this step are subsequently submitted to a collaborative trial, composed of at least twelve EU laboratories, for the method validation. Finally, the resulting evaluation reports, including the validation report and the validated protocols, are provided to the EFSA. Based on all these information, the EFSA give an opinion to the EU commission that will then take the final decision. A list of all EU authorized GMO for food and feed is publically available on the EU website (The European Commission 2016). The validity of each accepted dossier is however of only ten years and can be extended following to a new evaluation (Devos *et al.* 2014).

The labelling and the traceability of GMO depend essentially on the Regulations (EC) 1829/2003 and 1830/2003. All products with at least 0.9% of EU authorized GM ingredients have to be labelled as "containing GMO" whereas a policy of "zero tolerance" is applied for the EU unauthorized GMO. Due to the worldwide asynchronous approvals, GM events already authorized in some countries are still unauthorized in other countries. In reaction, the Regulation (EC) 619/2011 has been adopted to regulate the adventitious presence of GM events for which

the authorizations are pending or are subject to be withdrawn. However, the low level presence of these unauthorized GM events is only tolerated to a maximum of 0.1%. Regarding GM events only authorized in non-EU countries and for which no dossier is pending or subject to withdraw, they are considered as unauthorized in EU and a policy of "zero tolerance" is thus applied. In addition, GM events unauthorized worldwide are, of course, considered as EU unauthorized GMO. GM events accidentally/deliberately escaped from field trials constitute also a source of EU unauthorized GMO. The respect of all these legislations is controlled by NRL, accredited under ISO 17025 or equivalent international standards, through a GMO detection, identification and quantification system. Moreover, the certified reference materials (CRM), required by this GMO detection system, are produced by the JRC's Institute for Reference Materials and Measurements (Broeders *et al.* 2012a; Hartung and Schiemann 2014). This control of the legislations is supervised by the national competent authorities, such as the Federal Agency for the Safety of the Food Chain (FASFC) in Belgium, which carry out the sampling on the market that is then sent for analysis by NRL. The results are delivered to the competent authority that will take the appropriate decisions (AFSCA 2016).

1.3 Current and new approaches in GMO detection

This section is composed of two publications. The first one summarizes the current and new approaches in GMO detection, including their benefits and drawbacks. According to this critical view, some solutions were proposed to strengthen the current GMO detection strategy used by the enforcement laboratories. The second one illustrates the application on study cases of the current GMO detection system helped by new approaches, such as DNA walking and Next-Generation-Sequencing (NGS).

1.3.1 Current and new approaches in GMO detection: challenges and solutions

The following subsection was previously published as:
M.-A. Fraiture, P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2015c). "Current and new approaches in GMO detection: challenges and solutions". *BioMed Research International*, article ID 392872. DOI: 10.1155/2015/392872

Authors' contributions

M.-A. Fraiture collected all information and drafted the manuscript. P. Herman, I. Taverniers, M. De Loose, D. Deforce and N. H. Roosens evaluated critically the manuscript. All authors read and approved the final manuscript.

Abstract

In many countries, genetically modified organisms (GMO) legislations have been established in order to guarantee the traceability of food/feed products on the market and to protect the consumer freedom of choice. Therefore, several GMO detection strategies, mainly based on DNA, have been developed to implement these legislations. Due to its numerous advantages, the quantitative PCR (qPCR) is the method of choice for the enforcement laboratories in GMO routine analysis. However, given the increasing number and diversity of GMO developed and put on the market around the world, some technical hurdles could be encountered with the qPCR technology, mainly owing to its inherent properties. To address these challenges, alternative GMO detection methods have been developed, allowing faster detections of single GM target (*e.g.*, loop-mediated isothermal amplification), simultaneous detections of multiple GM targets (*e.g.*, PCR capillary gel electrophoresis, microarray, and Luminex), more accurate quantification of GM targets (*e.g.*, digital PCR), or characterization of partially known (*e.g.*, DNA walking and Next Generation Sequencing (NGS)) or unknown (*e.g.*, NGS) GMO. The benefits and drawbacks of these methods are discussed in this review.

1.3.1.1 Introduction

With the aim to improve the agricultural practices and nutritional quality, plant breeding techniques have been developed to produce genetically modified (GM) crops expressing interesting traits such as herbicide tolerance, insect resistance, and abiotic stress resistance (James 2014). To this end, new combinations of their genetic material are created through the use of modern biotechnology (SCBD 2015). The first genetically modified organism (GMO) approved for the commercialization was the Flavr-Savr tomato in 1994. From that time, 181.5 million hectares of planted GM plants in 28 countries were reported in 2014 (James 2014). Given that the "right to know" for the consumers, GMO labeling policies have been established in several countries around the world with a threshold of tolerance varying between 0 and 5%. Therefore, the presence of GMO in the food/feed chain is controlled by the competent authorities (Kamle and Ali 2013). To guarantee the GMO traceability, a key factor in the

implementation of these regulations, several strategies, categorized as indirect (protein-based methods) or direct (DNA-based methods), have been developed to detect GMO in food/feed samples by using different technologies. Among the protein-based approaches, which target proteins encoded by the transgenes, several methods depend on the Enzyme-Linked Immunosorbent Assay (ELISA) technique (Table 1.2) (Rogan *et al.* 1999; Lipp *et al.* 2000; Lipton *et al.* 2000; McKenzie *et al.* 2000; Fagan *et al.* 2001; Anklam *et al.* 2002; Stave 2002; Xu *et al.* 2005; Ermolli *et al.* 2006; Emslie *et al.* 2007; Shan *et al.* 2007; Giovannoli *et al.* 2008; Smith *et al.* 2008; Xu *et al.* 2009; Székács *et al.* 2010; Kamle *et al.* 2011; Zhu *et al.* 2011; Tan *et al.* 2013). A portable immunoassay system was also proposed (Table 1.2) (Jang *et al.* 2011). As an alternative, the immuno-PCR method was used to identify GMO (Table 1.2) (Allen *et al.* 2006; Santiago-Felipe *et al.* 2014).

Furthermore, protein-based methods include the use of the mass spectrometry-based technology as a tool allowing characterizing GM crops (García-Cañas *et al.* 2011). However, although they present several advantages such as the rapidity and simplicity, the protein-based methods depend on the expression level of targeted proteins, which is variable according to the plant tissues and the plant developmental status. Moreover, the proteins are highly degraded or denatured by food processing. Any modification in the targeted proteins could indeed alter the specificity and sensitivity of the assay. In addition, this strategy is not applicable if the genetic modification has no impact at the protein level (Morisset *et al.* 2008b; Mazzara *et al.* 2013). To overcome these issues, many DNA-based methods, targeting straightforward transgenic integrated sequences, have been widely developed. Even if quantitative PCR (qPCR) is the method of choice in GMO routine analysis, its inherent PCR properties imply some limitations. Therefore, to address these challenges, some alternative approaches have been developed, allowing notably providing faster detection of GM targets individually amplified in both routine laboratory and field (*e.g.*, loop-mediated isothermal amplification (LAMP)), simultaneous detection of several GM targets (*e.g.*, PCR capillary gel electrophoresis (CGE), microarray, and Luminex), more accurate quantification of GM targets (*e.g.*, digital PCR (dPCR)), or characterization of partially known (*e.g.*, DNA walking and Next Generation Sequencing (NGS)) or unknown (*e.g.*, NGS) GMO (Figure 1.2). These DNA-based approaches and their targets are described in this review. In addition, the most appropriate uses of these approaches are discussed according to the adopted strategy of GMO detection as well as the available information about the sequences of tested GMO.

Table 1.2: Representative examples illustrating protein-based methods targeting GMO.

Technologies	Targets	References
ELISA	CP4-EPSPS	Rogan *et al.* (1999)
	Cry1Ab	Stave (2002), Ermolli *et al.* (2006), Giovannoli *et al.* (2008), Székács *et al.* (2010), and Zhu *et al.* (2011)
	Cry1Ac	Stave (2002) and Shan *et al.* (2007)
	Cry2A	Stave (2002)
	Cry2Ab	Kamle *et al.* (2011)
	Cry3A	Stave (2002) and Smith *et al.* (2008)
	Cry9C	Stave (2002)
	nptII	McKenzie *et al.* (2000), Smith *et al.* (2008), and Jang *et al.* (2011)
	CP4-EPSPS	Lipp *et al.* (2000), Stave (2002), Emslie *et al.* (2007), and Jang *et al.* (2011)
	pat	Stave (2002), Xu *et al.* (2005), Emslie *et al.* (2007), and Jang *et al.* (2011)
	Gox	Xu *et al.* (2009)
	CpTI	Tan *et al.* (2013)
Immuno-PCR	Cry1Ac	Allen *et al.* (2006)
	p35S	Santiago-Felipe *et al.* (2014)
	tNOS	Santiago-Felipe *et al.* (2014)

CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; CpTI: trypsin inhibitor in cowpea *Vigna unguiculata*; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; Gox: glyphosate oxidoreductase gene; nptII: neomycin phosphotransferase II gene; p35S: promoter of the 35 S cauliflower mosaic virus; tNOS: terminator of the nopaline synthase gene.

1.3.1.2 GMO detection approaches

qPCR technology

The qPCR system, which is the most common strategy, allows detecting, identifying, and quantifying GMO via the SYBR Green or TaqMan chemistries (Figure 1.2) (Angers-Loustau *et al.* 2014). Using a primer pair specific to the target, these qPCR chemistries are both based on PCR amplification recorded in real time with the fluorescence originated either from the asymmetrical cyanine dye binding to double-stranded DNA (SYBR Green) or from the fluorogenic probe specific to the targeted sequence (TaqMan) (Navarro *et al.* 2015). This technology is suitable for both unprocessed and processed food/feed matrices since amplicons of around 100 bp are usually amplified. Even if numerous

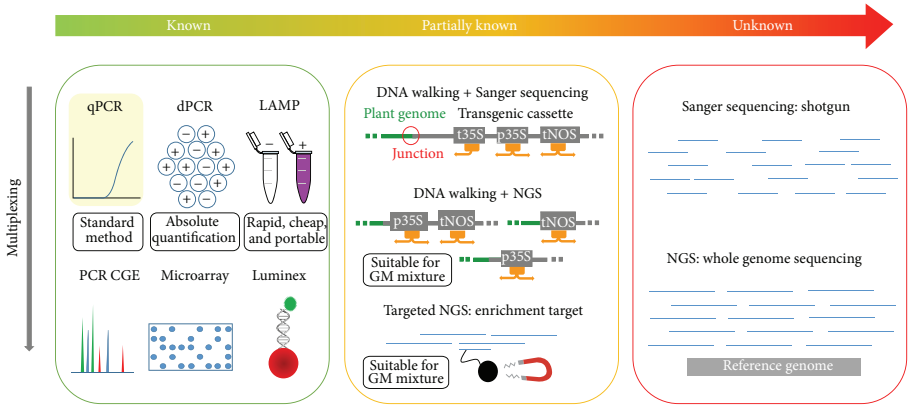


Figure 1.2: Suitable application of GMO detection approaches regarding the adopted strategy as well as the available information about the sequences of tested GMO.

qPCR methods have been reported, three main steps are typically followed in GMO routine analysis (Broeders *et al.* 2012b). First, the potential presence of GMO is assessed via a screening approach targeting the most common transgenic elements found in GMO, such as p35S (35S promoter from cauliflower mosaic virus) and tNOS (nopaline synthase terminator from *Agrobacterium tumefaciens*). In addition, some markers more discriminative, such as Cry3Bb, gat-tpinII, and t35S pCAMBIA, and taxon-specific markers could also be used. This step allows establishing a list of the potential GMO present in the tested samples and preventing further unnecessary assays in the subsequent steps (Table 1.3) (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2012a; Broeders *et al.* 2012b; Angers-Loustau *et al.* 2014; Fraiture *et al.* 2014; Broeders *et al.* 2015). Several of these screening markers are validated, based on minimum performance requirements, at the EU level following ring trials and are included in the Compendium of reference methods for GMO analysis (Joint Research Centre 2015a). According to the positive and negative signals observed for the different screening markers tested, GM events potentially detected are in a second step identified using construct specific or event-specific markers targeting, respectively, the junction between two elements inside the transgenic cassette or the junction between the transgenic cassette and the plant genome. In order to properly discriminate each GM event, the event-specific markers are currently favoured since the unique transgenic integration sites are targeted. Finally, the amount of identified GM events present in the tested food/feed samples is determined. Using event-specific and taxon-specific markers, this quantification step is carried out on the basis of the number of copies belonging

to the transgene and to the endogen (Table 1.3) (Broeders *et al.* 2012b). All the methods used to identify the EU-authorized GMO as well as the GMO for which the authorization is pending or is subjected to be withdrawn in the case of low level presence (LLP) have been provided by the applicants and are reported in the Compendium of reference methods for GMO analysis (Joint Research Centre 2015a). In combining several taxon-specific, event-specific, and construct-specific TaqMan markers in a 96-well prespotted plate, a real-time PCR based ready-to-use multitarget analytical system has been developed to allow the simultaneous identification of thirty-nine GM events (Querci *et al.* 2010).

In spite of its flexibility, simplicity, rapidity, and high analytical sensitivity, especially crucial to detect a low amount of GM targets, the success of the qPCR strategy depends however on some factors. For instance, the throughput of the qPCR strategy is usually limited to one marker per reaction. Due to the increasing number of GMO, additional markers have continually to be developed and used to fully cover their detection, which could thus make the laboratory work and the analysis of the results quite complex and laborious (Broeders *et al.* 2012a). In addition, this *a priori* approach targets only known sequences. Therefore, negative signals guarantee only the absence of known GMO in the tested food/feed samples. Similarly, in case of unexplained signals, in other words, the obtaining of positive and negative signals that found no correspondence with known GM events, the presence of unknown GMO could only be suspected. Indeed, the detection of GMO by qPCR is notably based on transgenic elements originated from natural organisms, such as p35S from CaMV and tNOS from *Agrobacterium*. For this reason, the qPCR system provides merely an indirect proof of the presence of GMO in a food/feed matrix since it could only be confirmed by the sequence of their transgene flanking regions. Concerning the quantification step, its achievement depends on the availability of Certified Reference Materials (CRM) (Broeders *et al.* 2012b; Holst-Jensen *et al.* 2012; Fraiture *et al.* 2014). Finally, the presence of inhibitors, such as polysaccharides, polyphenols, pectin, xylan, or fat, could alter the efficiency of the PCR reaction. Consequently, a later qPCR signal than theoretically expected will be observed, inducing an underestimation or even concealing the amount of GMO present in the tested sample (Demeke and Jenkins 2010; Opel *et al.* 2010; Schrader *et al.* 2012).

qPCR analysis tools In order to facilitate the interpretation of results, rapid and cost-efficient systems have been developed via analytical tools integrating simultaneously several targets. To this end, the CoSYPS platform (Combinatory SYBR Green qPCR Screening), which is a decision support system (DSS) at the screening level, has been successfully developed. For each tested food/feed matrix, this DSS combines immediately the experimental C_t and T_m values

Table 1.3: Representative examples illustrating simplex qPCR methods targeting GMO. Those validated at the EU level are indicated by an asterisk. Screening markers used in the CoSYPS are indicated by ∼.

Methods	Chemistries	Targets	References
		Screening markers	
Plant-specific	SYBR Green	RBCL∼	Mbongolo Mbella *et al.* (2011)
Taxon-specific	SYBR Green	LEC*∼	Joint Research Centre (2015a)
	SYBR Green	ADH*∼	Mbongolo Mbella *et al.* (2011)
	SYBR Green	CRU*∼	Mbongolo Mbella *et al.* (2011)
	SYBR Green	PLD∼	Mbongolo Mbella *et al.* (2011)
	SYBR Green	SAD1∼	Joint Research Centre (2015a)
	SYBR Green	GLU∼	Joint Research Centre (2015a)
Element-specific	SYBR Green	p35S*∼	Barbau-Piednoir *et al.* (2010)
	TaqMan	p35S*	Waiblinger *et al.* (2008)
	SYBR Green	tNOS*∼	Barbau-Piednoir *et al.* (2010)
	TaqMan	tNOS*	Waiblinger *et al.* (2008)
	SYBR Green	pFMV∼	Broeders *et al.* (2013)
	TaqMan	pFMV*	Joint Research Centre (2015a)
	SYBR Green	pNOS∼	Broeders *et al.* (2013)
	SYBR Green	t35S∼	In-house
	SYBR Green	Cry1Ab/Ac∼	Barbau-Piednoir *et al.* (2014)

Table 1.3 *continued*

Methods	Chemistries	Targets		References
	TaqMan	Cry1A(b)*		Joint Research Centre (2015a)
	SYBR Green	Cry3Bb~		Broeders *et al.* (2015)
	SYBR Green	pat*~		Barbau-Piednoir *et al.* (2014)
	TaqMan	pat*		Joint Research Centre (2015a)
	SYBR Green	bar*~		Barbau-Piednoir *et al.* (2014)
	TaqMan	bar*		Joint Research Centre (2015a)
	SYBR Green	CP4-EPSPS~		Barbau-Piednoir *et al.* (2014)
	SYBR Green	t35S pCAMBIA~		Fraiture *et al.* (2014)
	SYBR Green	nptII		Joint Research Centre (2015a)
Construct-specific	SYBR Green	gat-tpinII~		Broeders *et al.* (2015)
Virus-specific	SYBR Green	CRT~		In-House
		Event-specific methods		
GM-specific	TaqMan	Maize (*Zea mays*)	3272*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	5307*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	98140*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	Bt11*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	Bt176*	Joint Research Centre (2015a)

Table 1.3 *continued*

Methods	Chemistries	Targets		References
	TaqMan	Maize (*Zea mays*)	DAS-40278-9*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	DAS-59122-7*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	GA21*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	LY038*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MIR162*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MIR604*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MON810*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MON863*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MON87460*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MON88017*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	MON89034*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	NK603*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	T25*	Joint Research Centre (2015a)
	TaqMan	Maize (*Zea mays*)	TC1507*	Joint Research Centre (2015a)

Table 1.3 *continued*

Methods	Chemistries	Targets		References
	TaqMan	Soybean (*Glycine max*)	A2704-12*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	A5547-127*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	BPS-CV-127*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	DAS68416-4*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	DP-305423-1*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	DP-356043-5*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	FG72*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	GTS40-3-2*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	MON87701*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	MON87705*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	MON87708*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	MON87769*	Joint Research Centre (2015a)
	TaqMan	Soybean (*Glycine max*)	MON89788*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	281-24-236*	Joint Research Centre (2015a)

Table 1.3 *continued*

Methods	Chemistries	Targets		References
	TaqMan	Cotton (*Gossypium hirsutum*)	3006-210-23*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	GHB119*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	GHB614*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	LLCOTTON25*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	MON531*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	MON1445*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	MON15985*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	MON88913*	Joint Research Centre (2015a)
	TaqMan	Cotton (*Gossypium hirsutum*)	T304-40*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	73496*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	GT73*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	MON88302*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	Ms1*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	Ms8*	Joint Research Centre (2015a)

Table 1.3 *continued*

Methods	Chemistries	Targets		References
	TaqMan	Oilseed rape (*Brassica napus*)	Rf1*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	Rf2*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	Rf3*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	T45*	Joint Research Centre (2015a)
	TaqMan	Oilseed rape (*Brassica napus*)	Topas 19/2*	Joint Research Centre (2015a)
	TaqMan	Potato (*Solanum tuberosum*)	EH92-527-1*	Joint Research Centre (2015a)
	TaqMan	Rice (*Oryza sativa*)	LLRICE62*	Joint Research Centre (2015a)
	TaqMan	Sugar beet (*Beta vulgaris*)	H7-1*	Joint Research Centre (2015a)

ADH: alcohol dehydrogenase I gene from maize; bar: phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; CRT: reverse transcriptase gene from the cauliflower mosaic virus; CRU: cruciferin gene from colza; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; gat-tpinII: junction sequence between the glyphosate N-acetyltransferase of *Bacillus licheniformis* and the terminator of the *Solanum tuberosum* proteinase inhibitor; GLU: glutamine synthetase gene from sugar beet; LEC: lectin gene from soybean; nptII: neomycin phosphotransferase II gene; p35S: promoter of the 35 S cauliflower mosaic virus; pat: phosphinothricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*; pFMV: promoter of the figwort mosaic virus; phy: phytase gene from maize; PLD: phospholipase D gene from rice; pNOS: promoter of the nopaline synthase gene; RBCL: ribulose-1,5-biphosphate carboxylase oxygenase; SAD1: stearoylacyl carrier protein desaturase gene from cotton; t35S: terminator of the cauliflower mosaic virus; t35S pCAMBIA: terminator of the cauliflower mosaic virus from pCAMBIA vector; tNOS: terminator of the nopaline synthase gene.

obtained with the twenty SYBR Green methods, running in a single 96-well plate and targeting plant gene, taxon genes, and transgenic elements (Table 1.3). This selection of screening markers allows both covering at least all the EU-authorized GMO and LLP cases (*e.g.*, with p35S and tNOS) and, as far as possible, discriminating between themselves and some EU-unauthorized GMO (*e.g.*, with t35S pCAMBIA and gat-tpinII) in order to reduce the number of identifications/quantifications to carry out downstream (Van den Bulcke *et al.* 2010; Broeders *et al.* 2012b; Fraiture *et al.* 2014; Broeders *et al.* 2015). An alternative to interpret qPCR results is provided by the GMOseek and GMOfinder databases, containing reliable information on GMO. Following the interpretation of the experimental results, obtained with in-house or EU reference methods, the names of positive elements are introduced in the databases to provide a list of potentially detected GMO that will be then experimentally verified (Gerdes *et al.* 2012; Block *et al.* 2013). The truthfulness of these predictions is however diminished since elements identically named can possess different sequences and the detection methods used are not taken into account. Indeed, to target the same element, several methods could exist and could present different PCR efficiencies which could generate variation in the results. Most recently, the JRC-GMO-Matrix platform, combining information from the GMOMETHODS database (all reference methods for GMO analysis) and the Central Core DNA Sequences Information System (several annotated GMO sequences), was also proposed for the same purpose. This platform integrates the positive and negative signals experimentally observed with EU validated taxon-specific, element-specific, construct-specific, and event-specific methods for any tested food/feed matrix in order to predict more reliably the potential amplified GM events (Angers-Loustau *et al.* 2014). The JRC-GMO-Matrix platform is also strengthened by the JRC GMO-Amplicons database which contains publically available putative GMO-related sequences (Joint Research Centre 2015b).

Multiplex qPCR strategy With multiplex PCR-based methods, several DNA targets can be detected in a single reaction. It presents the advantage to decrease the number of reactions necessary to test the potential presence of GMO in a sample. Several multiplex qPCR TaqMan strategies have thus been investigated, including mainly the screening markers p35S and tNOS (Table 1.4) (Waiblinger *et al.* 2008; Bahrdt *et al.* 2010; Dörries *et al.* 2010; Huber *et al.* 2013; Samson *et al.* 2013; Chaouachi *et al.* 2014; Köppel *et al.* 2014; Köppel *et al.* 2015; Park *et al.* 2015). To provide a system with a high GMO coverage, twenty-three triplex and one duplex PCR were gathered on a 384-well plate to identify forty-seven targets (Table 1.4) (Cottenet *et al.* 2013).
However, compared to simplex qPCR, the development of optimal multiplex assays could be more challenging notably in terms of primers and probes design

as well as sensitivity and reproducibility. Moreover, the throughput of this strategy is relatively limited by the availability of dyes with an emission and absorption spectrum of fluorescence sufficiently distinct to avoid overlaps of signals. The combination of different dyes risks also increases the fluorescent background. Therefore, the majority of the reported multiplex qPCR assays amplify simultaneously only two or three targets. To date, a maximum of six markers have been successfully combined in one reaction to detect GMO (Bahrdt *et al.* 2010; Joint Research Centre 2015a).

Alternative multiplex strategies

Still with the aim of going further in the development of multiplex assays, several methods not based on qPCR have been also developed using notably the CGE, microarray, and Luminex technologies. Two main steps are generally followed. First, to guarantee a sufficient sensitivity, the samples are amplified by PCR since the GM targets are potentially at trace level in food/feed matrices. In a second step, the PCR products are analyzed using the CGE, microarray, or Luminex platforms. Despite the fact that these technologies present a higher throughput than qPCR, their multiplexing level is still influenced by the inherent properties of PCR which limit the number of reactions at commonly ten targets per PCR assay (Pla *et al.* 2012; Vega and Marina 2014).

PCR capillary gel electrophoresis technology In order to detect simultaneously several targets, the use of the PCR multiplex CGE, where fluorescently labelled primers allow discriminating different amplicons of the same size, has been also suggested to be applied in the GMO detection field (Figure 1.2 and Table 1.5). Compared to the electrophoresis gel, the resolution power of the CGE system to detect PCR products from a multiplex assay is clearly higher (Vega and Marina 2014). However, the sensitivity of CGE system is weaker than the qPCR technology (Milavec *et al.* 2014). Using the PCR CGE system, eight GM maize were identified via a nonaplex PCR including event-specific, construct-specific, and taxon-specific methods (Table 1.5) (Heide *et al.* 2008a; Heide *et al.* 2008b). Similarly, one pentaplex PCR and two hexaplex PCR were also developed to, respectively, detect specifically four GM maize, five GM cotton, and five GM maize (Table 1.5) (Nadal *et al.* 2006; Nadal *et al.* 2009; Holck *et al.* 2010; Holck and Pedersen 2011). Recently, a tetraplex targeting transgenic elements and cotton-specific gene was also reported (Table 1.5) (Basak *et al.* 2014). In addition, Guo *et al.* (2011) developed three octaplex PCR using universally tailed primers to preamplify GM targets under a short number of cycles. To increase the yield and PCR efficiency, these amplicons, earlier submitted to a PCR emulsion, are then enriched with universal primers.

Table 1.4: Representative examples illustrating multiplex qPCR TaqMan methods targeting GMO. Those validated at the EU level are indicated by an asterisk.

Multiplexing	Methods	Targets	References
Duplex	Element-specific	p35S* and tNOS*	Waiblinger *et al.* (2008)
Duplex	Element-specific	bar and pat	Huber *et al.* (2013)
Duplex	Plant-specific Other	TLC IPC	Cottenet *et al.* (2013)
Duplex	Taxon-specific Event-specific	ADH Bt11	Chaouachi *et al.* (2014)
Duplex	Taxon-specific Event-specific	ADH Bt176	Chaouachi *et al.* (2014)
Duplex	Taxon-specific Event-specific	ADH MON810	Chaouachi *et al.* (2014)
Duplex	Taxon-specific Event-specific	ADH T25	Chaouachi *et al.* (2014)
Triplex	Element-specific	p35S, tNOS, and CTP2/CP4-EPSPS	Huber *et al.* (2013)
Triplex	Taxon-specific Other	LEC and Zein ICP	Cottenet *et al.* (2013)
Triplex	Taxon-specific Other	Pro and PC ICP	Cottenet *et al.* (2013)
Triplex	Taxon-specific Other	ACC and FRUp ICP	Cottenet *et al.* (2013)

Table 1.4 *continued*

Multiplexing	Methods	Targets	References
Triplex	Taxon-specific Other	SAD1 and FRUt ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Other	p35S and pFMV ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Other	tE9 and tNOS ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Other	bar and CP4-EPSPS ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Other	hpt and pat ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Other	nptII and Cry1Ab/Ac ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	CBH351 and Bt176 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	MON810 and T25 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	Bt11 and MON863 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	NK603 and GA21 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	TC1507 and DAS-59122-7 ICP	Cottenet *et al.* (2013)

Table 1.4 *continued*

Multiplexing	Methods	Targets	References
Triplex	Construct-specific Other	MIR604 and MON88017 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	98140 and MON89034 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	3272 and MIR162 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	A2704-12 and GTS40-3-2 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	DP-305423-1 and DP-356043-5 ICP	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	MON87701 and MON89788 ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Construct-specific Other	AHAS FG72 IPC	Cottenet *et al.* (2013)
Triplex	Construct-specific Other	Bt63 and A5547-127 ICP	Cottenet *et al.* (2013)
Triplex	Element-specific Construct-specific Other	Xa21 KMD1 IPC	Cottenet *et al.* (2013)
Triplex	Taxon-specific Construct-specific	Zein MON810 and GA21	Samson *et al.* (2013)

Table 1.4 *continued*

Multiplexing	Methods	Targets	References
Triplex	Taxon-specific Construct-specific	ADH MON810 and GA21	Samson *et al.* (2013)
Triplex	Element-specific	p35s, tNOS, and t35S	Park *et al.* (2015)
Triplex	Element-specific	tE9, pRbcS4, and tORF23	Park *et al.* (2015)
Triplex	Element-specific Event-specific	tpinII and tAHASL DP-305423-1	Park *et al.* (2015)
Tetraplex	Element-specific	pFMV, bar, pat, and CTP2/CP4-EPSPS	Köppel *et al.* (2014)
Tetraplex	Element-specific	p35S, tNOS, pFMV, and bar	Dörries *et al.* (2010)
Pentaplex	Taxon-specific Element-specific Virus-specific	HMG and LEC p35S and tNOS CaMV	Köppel *et al.* (2014)
Pentaplex	Element-specific	p35S, tNOS, bar, pat, and CTP2/CP4-EPSPS	Huber *et al.* (2013)
Pentaplex	Taxon-specific Event-specific	LEC MON87769, MON87708, MON87705, and FG72	Köppel *et al.* (2015)
Hexaplex	Element-specific Construct-specific Other	p35S, tNOS, and pFMV SAMS and LY IPC	Bahrdt *et al.* (2010)

ACC: acetyl-CoA-carboxylase gene from colza; ADH: alcohol dehydrogenase I gene from maize; AHAS: AHAS fragment unique recombination from BPS-CV-127; bar: phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; CaMV: ORFIII from CaMV; CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; CTP2/CP4-EPSP: junction region between the chloroplast transit peptide 2 (CTP2) sequence from the *Arabidopsis*

thaliana epsps gene and the CP4 epsps gene from *Agrobacterium tumefaciens* (CP4-EPSPS); FRUp: β-fructosidase gene from potato; FRUt: β-fructosidase gene from tomato; HMG: major high mobility group protein gene from maize; hpt: hygromycin phosphotransferase gene; IPC: internal positive control; LEC: lectin gene from soybean; LS28: choline kinase; LY: transition from *Zea mays* chloroplast transit peptide sequence for dihydrodipicolinate synthase to *Corynebacterium glutamicum* dihydrodipicolinate synthase (cordapA) gene encoding for a lysine-insensitive dihydrodipicolinate synthase enzyme; nptII: neomycin phosphotransferase II gene; p35S: promoter of the 35 S cauliflower mosaic virus; pat: phosphinothricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*; PC: phosphoenolpyruvate carboxylase gene from wheat; pFMV: promoter of the figwort mosaic virus; pRbcS4: ribulose 1,5-bisphosphate carboxylase small subunit promoter from *A. thaliana*); Pro: prolamin gene from rice; SAD1: stearoyl-acyl carrier protein desaturase gene from cotton; SAMS: transition from S-adenosyl-L-methionine synthetase (SAMS) promoter to *Glycine max* acetolactate synthase (gm-hra) gene; t35S: terminator of the cauliflower mosaic virus); tAHASL: acetohydroxy acid synthase large subunit terminator from *A. thaliana*; tE9: ribulose-1,5-bisphosphate carboxylase terminator E9 from *Pisum sativum*; TLC: tRNA-Leu chloroplastic gene; tNOS: terminator of the nopaline synthase gene; tORF23: open reading frame 23 terminator from *A. tumefaciens*); tpinII: inhibitor II terminator from potato; Zein: Zein gene from maize; Xa21: Xa21 gene from *Oryza longistaminata*.

By this way, twenty-four targets from fourteen GM events were identified by the CGE system (Table 1.5) (Guo *et al.* 2011). A variant of this technique, which implies no fluorescent labels on primers, is reported by Burrell *et al.* (2011). This study proposed a tetraplex PCR composed of two species-specific methods and two screening markers allowing detecting the presence of Bt11 maize and GTS40-3-2 soybean events using commercialized electrophoresis instruments (Table 1.5) (Burrell *et al.* 2011).

Microarrays technology With the microarray technology applied to GMO detection, GM targets are amplified by PCR, using target-specific and/or universal primers, prior to being hybridized on the array, allowing the simultaneous detection of more than 250 000 targets in one assay (Figure 1.2 and Table 1.6) (Nakaya *et al.* 2007). Compared to the qPCR, the microarray strategy presents thus a well higher throughput but a slightly weaker sensitivity (Dobnik *et al.* 2010; Pla *et al.* 2012). One approach, called multiplex quantitative DNA array-based PCR (MQDA-PCR), tested on transgenic maize events, consists of a first PCR using target-specific primers that harbor a universal tail allowing using universal primers in the second PCR. The signal is then detected after the hybridization of the PCR products with the fluorescently labelled probes on the DNA array (Table 1.6) (Rudi *et al.* 2003). Furthermore, using a padlock probe ligation-microarray detection system (PPLMD), some GM maize, cotton, and soybean events were detected. With the PPLMD system, the targets are initially hybridized to linear padlock probes harboring target-specific and universal sequences to be then amplified by PCR with universal primers (Table 1.6) (Prins *et al.* 2008). In addition, a nucleic acid sequence based amplification implemented microarray (NAIMA) approach, using universal primers, has been tested on transgenic maize (Table 1.6) (Morisset *et al.* 2008a; Dobnik *et al.* 2010). As an alternative to the potential issue related to the use of fluorescent label, the DualChip GMO system was proposed. So, after PCR amplification with biotinylated target-specific primers, the amplicons hybridized on the arrays are detected by a colorimetric reaction, allowing identifying simultaneously some GM maize, soybean, and rapeseed events. The performance of the DualChip GMO system, targeting fourteen elements, was also validated through an EU collaborative ring trial. An upgraded version of this system (DualChip GMO V2.0) presents a higher GMO coverage in targeting thirty elements (Table 1.6) (Leimanis *et al.* 2006; Leimanis *et al.* 2008; Hamels *et al.* 2009; von Götz 2010; Pla *et al.* 2012). Most recently, a multiplex amplification on a chip with readout on an oligo microarray (MACRO) system, targeting ninety-one targets to cover a broad spectrum of GMO, was also reported (Shao *et al.* 2014).

Table 1.5: Representative examples illustrating multiplex PCR CGE methods targeting GMO.

Multiplexing	Methods	Targets	References
Tetraplex	Taxon-specific Element-specific	Zein and LEC p35S and tNOS	Burrell *et al.* (2011)
Tetraplex	Taxon-specific Element-specific	SAD1 Cry1Ac, p35S, and tNOS	Basak *et al.* (2014)
Pentaplex	Taxon-specific Event-specific	ADH Bt11, GA21, MON810, and NK603	Nadal *et al.* (2006)
Hexaplex	Taxon-specific Event-specific	acp1 Bollgard, Bollgard II, RR, 3006-210-23, and 281-24-231	Nadal *et al.* (2009)
Hexaplex	Taxon-specific Event-specific	HMG DAS-59122-7, LY038, MON88017, MIR604, and 3272	Holck *et al.* (2010) and Holck and Pedersen (2011)
Octaplex	Event-specific	Bt11, Bt176, Huanong No. 1, GTS40-3-2, T25, MON88913, MON1445, and MIR604	Guo *et al.* (2011)
Octaplex	Taxon-specific Element-specific Event-specific	LEC and ssIIb pFMV and tNOS TC1507, MON531, NK603, and GA21	Guo *et al.* (2011)
Octaplex	Taxon-specific Element-specific Event-specific	SAD1 bar, chy, pAct, CP4-EPSPS, and Cry1Ab GT73 and OXY235	Guo *et al.* (2011)

Table 1.5 *continued*

Multiplexing	Methods	Targets	References
Nonaplex	Taxon-specific Event-specific	HMG T25, GA21, TC1507, MON863, MON810, NK603, Bt176, and Bt11	Heide *et al.* (2008b) and Heide *et al.* (2008a)

acp1: acyl carrier protein 1 gene from cotton; ADH: alcohol dehydrogenase I gene from maize; bar: phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; Chy: chymopapain gene from papaya; CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; HMG: major high-mobility group protein gene from maize; LEC: lectin gene from soybean; p35S: promoter of the 35 S cauliflower mosaic virus; pAct: promoter region of rice actin gene; pFMV: promoter of the figwort mosaic virus; SAD1: stearoyl-acyl carrier protein desaturase gene from cotton; ssIIb: starch synthase IIb gene from maize; tNOS: terminator of the nopaline synthase gene; Zein: Zein gene from maize.

Table 1.6: Representative examples illustrating multiplex PCR microarray methods targeting GMO.

Multiplexing	Techniques	Methods	Targets	References
Duplex	DualChip GMO	Element-specific	p35S and tNOS	Leimanis *et al.* (2006), Leimanis *et al.* (2008), and Hamels *et al.* (2009)
Duplex	DualChip GMO	Construct-specific Virus-specific	pNOS/nptII CaMV	Leimanis *et al.* (2006), Leimanis *et al.* (2008), and Hamels *et al.* (2009)
Triplex	DualChip GMO	Element-specific	pat, Cry1A(b), and CP4-EPSPS	Leimanis *et al.* (2006), Leimanis *et al.* (2008), and Hamels *et al.* (2009)
Triplex	NAIMA	Taxon-specific Element-specific	IVR p35S and tNOS	Morisset *et al.* (2008a)
Triplex	NAIMA	Taxon-specific Element-specific Event-specific	IVR p35S MON810	Morisset *et al.* (2008a)
Tetraplex	DualChip GMO	Plant-specific Taxon-specific	RBCL IVR, LEC, and CRU	Leimanis *et al.* (2006), Leimanis *et al.* (2008), and Hamels *et al.* (2009)
Octaplex	MQDA-PCR	Taxon-specific Element-specific Event-specific Other	HMG p35S and tNOS Bt176, Bt11, and MON810 IPC	Rudi *et al.* (2003)

Table 1.6 *continued*

Multiplexing	Techniques	Methods	Targets	References
Decaplex	PPLMD	Taxon-specific Element-specific Event-specific	SAD1, Zein, ACC, and LEC p35S, pFMV, and bar MON1445, Bt176, and GTS40-3-2	Prins *et al.* (2008)
Dodecaplex	MQDA-PCR	Taxon-specific Element-specific Event-specific Other	HMG p35S, tNOS, and Amp Bt176, Bt11, MON810, T25, GA21, CBH351, and DBT418 IPC	Rudi *et al.* (2003)

ACC: acetyl-CoA-carboxylase gene from colza; Amp: ampicillin resistance gene; bar: phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; CaMV: ORFIII from CaMV; CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; CRU: cruciferin gene from colza; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; HMG: major high-mobility group protein gene from maize; IPC: internal positive control; IVR: invertase gene from maize; LEC: lectin gene from soybean; nptII: neomycin phosphotransferase II gene; p35S: promoter of the 35 S cauliflower mosaic virus; pat: phosphinothricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*; pFMV: promoter of the figwort mosaic virus; pNOS: promoter of the nopaline synthase gene; RBCL: ribulose-1,5-biphosphate carboxylase oxygenase; SAD1: stearoyl-acyl carrier protein desaturase gene from cotton; tNOS: terminator of the nopaline synthase gene; Zein: Zein gene from maize.

Luminex technology Biotinylated targets amplified by single or multiplex PCR assays could be analyzed with the Luminex technology, potentially able to simultaneously detect up to 500 different targets in one sample using spectrally distinct sets of beads that are independently coupled to unique nucleic acid probes. After hybridization of biotinylated oligonucleotides to corresponding probe-bead complexes, the reader device individually analyzes each microsphere by flow cytometry in applying a laser excitation of 635 nm and 532 nm allowing, respectively, identifying the bead set and determining the presence or absence of the target (Figure 1.2) (Luminex Corporation 2015). This technology was firstly assessed in GMO detection by Fantozzi *et al.* (2008) (Table 1.7). In this study, the p35S and EPSPS elements, earlier individually amplified by PCR from the GTS-40-3-2 soybean event, were simultaneously detected (Fantozzi *et al.* 2008). Afterwards, the GM stacked LS28 × Cry1Ac rice and 281-24-236 × 3006-210-23 cotton events were identified on the Luminex platform using upstream, respectively, a pentaplex PCR or a hexaplex PCR (Table 1.7) (Choi *et al.* 2010; Choi 2011). This technology was also used to detect ten GM maize events through four sets of multiplex PCR assays (Table 1.7) (Han *et al.* 2013). Similarly, a liquid bead array approach allowing identifying thirteen GM maize was recently developed (Fu *et al.* 2015a).

Due to its potential high throughput, the Luminex technology seems to be a promising alternative in GMO detection. Moreover, the liquid bead array is considered as more sensitive and faster than the microarray system (Choi *et al.* 2010). Nevertheless, the drawback linked to the PCR complicates the setting of a unique multiplex assay targeting simultaneously all GM events. Furthermore, as only few studies using this technology in GMO detection have been reported to date, experiments have still to be carried out in order to provide effective and validated systems.

Digital PCR technology

To resolve difficulties observed during the relative quantification step in qPCR, especially when the copy numbers of GMO are low and/or PCR inhibitors are present, the digital PCR (dPCR) technology has been tested in GMO detection (Figure 1.2). Based on the binomial Poisson statistics, each partition of the fractionated sample is determined as positive (amplified target observed) or negative (no amplified target observed) by the dPCR technology allowing absolutely quantifying the number of nucleic acid targets from GMO present in any given sample. Two approaches of this end-point PCR system have up till now been used for this aim (Table 1.8). On the one hand, the chamber dPCR (cdPCR), partitioning the sample in several thousands of microfluidic chambers, was used to target GM maize MON810 event using a duplex PCR composed of the MON810 event-specific and maize taxon-specific methods. The detection

Table 1.7: Representative examples illustrating Luminex strategies targeting GMO.

Multiplexing	Methods	Targets	References
Simplex	Element-specific	p35S and CP4-EPSPS	Fantozzi *et al.* (2008)
Triplex	Taxon-specific Event-specific	Zein MIR604 and MON88017	Han *et al.* (2013)
Tetraplex	Event-specific	Bt176, MON810, NK603, and GA21	Han *et al.* (2013)
Tetraplex	Event-specific	Bt11, T25, MIR162, and MON89034	Han *et al.* (2013)
Pentaplex	Taxon-specific Element-specific	SPS Cry1Ac, tNOS, p35S, and LS28	Choi *et al.* (2010)
Hexaplex	Taxon-specific Element-specific Event-specific	SAD1 Cry1Ac, Cry1F, and pat 281-24-236 and 3006-210-23	Choi (2011)

CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; LS28: choline kinase; p35S: promoter of the 35 S cauliflower mosaic virus; SAD1: stearoyl-acyl carrier protein desaturase gene from cotton; SPS: sucrose phosphate synthase gene from rice; tNOS: terminator of the nopaline synthase gene; tORF23: open reading frame 23 terminator from *A. tumefaciens*; Zein: Zein gene from maize.

limits of this approach were also investigated (Bhat *et al.* 2009; Burns *et al.* 2010; Corbisier *et al.* 2010). Moreover, a strategy based on the cdPCR system was developed in order to cover a wide range of GMO by applying individually twenty-eight element-specific, thirty-six event-specific, and five taxon-specific methods (Table 1.8) (Brod *et al.* 2014). Afterwards, this strategy was applied with forty-eight markers, including seven transgenic elements-specific, fourteen event-specific, and five taxon-specific methods (Table 1.8) (Li *et al.* 2015). On the other hand, the droplet dPCR (ddPCR) approach, implying several thousands of droplets generated by a water-oil emulsion, was used in simplex or duplex PCR with the MON810 event-specific and maize taxon-specific methods (Morisset *et al.* 2013). Most recently, duplex assays, including one GMO-specific marker with one soybean, maize, or rice taxon-specific marker, were performed by using the ddPCR system to quantify twelve GM soybean, sixteen GM maize, and two GM rice events (Table 1.8) (Köppel *et al.* 2015; Köppel and Bucher 2015).

The dPCR technology could become a key tool in the field of GMO detection, mainly because an absolute, and not relative as in qPCR, quantification of the GM target is provided. The measurement does not require necessarily the use of reference material, solving issues related to the availability of an optimal reference material. Moreover, thanks to the partitioning of the sample, the PCR efficiency is less affected by the presence of inhibitors and allows reducing the uncertainty in the measurement, especially at low copy number, as observed with qPCR calibration curves generated by serial dilutions of the target. However, validated qPCR methods are not always simply transferable to the dPCR technology. Indeed, some optimization has to be carried out regarding, for instance, the design and the concentrations of primers and probes. In addition, given that maximum two different targets could be identified in one well, the low throughput power of the dPCR technology highlights its applicability more suitable at the identification/quantification level than at the screening step (Morisset *et al.* 2013; Strain *et al.* 2013; Köppel *et al.* 2015; Köppel and Bucher 2015).

Loop-mediated isothermal amplification

Due to its rapidity, specificity, sensitivity, and simplicity, the loop-mediated isothermal amplification (LAMP) method has been proposed to detect GMO (Figure 1.2). To this end, four primers specific to six distinct regions of the target are required, allowing, under isothermal condition, initiating the reaction and increasing the amplification speed by the formation of a loop structure. The amplification can be then directly visualized in the tube thanks to fluorescent dyes. Several LAMP markers were thus developed for this approach to target transgenic elements (Table 1.9) (Fukuta *et al.* 2004; Lee *et al.* 2009; Liu *et al.*

Table 1.8: Representative examples illustrating digital PCR strategies targeting GMO.

Multiplexing	Techniques	Methods	Targets	References
Simplex	cdPCR	Taxon-specific	HMG, LEC, GLU, and CRU	Brod *et al.* (2014)
		Element-specific	p35S, tNOS, Cry1Ab, Cry1F, bar, CP4-EPSPS, Cry3Bb, nptII, Cry1A.105, and Cry2Bb	
		Event-specific	MON531, MON88913, MON1445, MON15985 LLCOTTON25, GHB614, 3272, DAS-59122-7, Bt176, Bt11, GA21, MIR162, MIR604, MON810, MON863, MON88017, MON89034, NK603, T25, TC1507, Ms1, Topas19/2, OXY 235, Ms8, Rf3, GT73, T45, GTS40-3-2, A2704-12, MON89788, MON87701, DP-356043-5, A5547-127, BPS-CV-127, DP-305423-1, and TT51-1	
Simplex	cdPCR	Taxon-specific	ADH, CRU, PLD, LEC, and adhC	Li *et al.* (2015)
		Element-specific	p35S, pFMV, tNOS, Cry1Ab, bar, pat, and nptII	
		Event-specific	3272, Bt11, GA21, MON89034, MON810, MIR604, MON88017, TC1507, Bt176, GTS40-3-1, DP-305423-1, DP-356043-5, H7-1, and GT73	
Simplex	ddPCR	Taxon-specific	HMG	Morisset *et al.* (2013)
		Event-specific	MON810	
Duplex	cdPCR	Taxon-specific	HMG	Bhat *et al.* (2009), Burns *et al.* (2010), and Corbisier *et al.* (2010)
		Event-specific	MON810	

Table 1.8 *continued*

Multiplexing	Techniques	Methods	Targets	References
Duplex	ddPCR	Taxon-specific	LEC	Köppel *et al.* (2015) and Köppel and Bucher (2015)
		Event-specific	DP-356043-5, DP-305423-1, MON89788, GTS40-3-2, A5547-127, BPS-CV-127, A2704-12, MON87701, MON87708, MON87705, FG72, and MON87769	
Duplex	ddPCR	Taxon-specific	PLD	Köppel and Bucher (2015)
		Event-specific	LLRICE62 and KMD1	
Duplex	ddPCR	Taxon-specific	HMG	Köppel and Bucher (2015)
		Event-specific	Bt176, Bt11, MON810, NK603, Starllink, MON863, GA21, DAS-59122-7, MIR162, MIR604, 3272, T25, TC1507, MON88017, MON89034, and DAS-40278-9	
Duplex	ddPCR	Taxon-specific	HMG	Morisset *et al.* (2013)
		Event-specific	MON810	

ADH: alcohol dehydrogenase I gene from maize; adhC: alcohol dehydrogenase C gene from cotton; bar: phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; CRU: cruciferin gene from colza; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; GLU: glutamine synthetase gene from sugar beet; HMG: major high-mobility group protein gene from maize; LEC: lectin gene from soybean; nptII: neomycin phosphotransferase II gene; p35S: promoter of the 35 S cauliflower mosaic virus; pat: phosphinothricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*; pFMV: promoter of the figwort mosaic virus; phy: phytase gene from maize; PLD: phospholipase D gene from rice; pNOS: promoter of the nopaline synthase gene; tNOS: terminator of the nopaline synthase gene.

2009; Guan *et al.* 2010; Chen *et al.* 2011a; Chen *et al.* 2012a; Kiddle *et al.* 2012; Li *et al.* 2013c; Randhawa *et al.* 2013; Xu *et al.* 2013b; Zhang *et al.* 2013a; Cheng *et al.* 2014; Di *et al.* 2014; Huang *et al.* 2014; Li *et al.* 2014; Zahradnik *et al.* 2014; Wang *et al.* 2015a).
The LAMP strategy presents the advantage to tolerate several PCR inhibitors such as acidic polysaccharides (Zhang *et al.* 2013a). Its implementation does also not require any sophisticate devices. Indeed, the amplification could be carried out using a water bath or heating block (Xu *et al.* 2013b). Some of the developed LAMP methods have besides been successfully tested in the fields (Zhang *et al.* 2013a). Concerning the drawbacks, the design of four primers per target, which guarantee the high specificity and sensitivity of the LAMP, could be difficult. In addition, the identification of several GM targets using a multiplex assay is not applicable (Angers-Loustau *et al.* 2014).

DNA walking

In using PCR-based methods that required prior knowledge, the observed results are mostly generated in targeting elements derived from natural organisms. Therefore, they constitute merely an indirect proof of the presence of GMO in the tested food/feed matrices. In addition, when the observed signals do not correspond to known GMO, the presence of unknown GMO, containing at least one known element, could be only suspected. The only way to indubitably confirm the presence of GMO is provided by the characterization of sequences from the junctions between the transgenic cassette and the plant genome as well as the unnatural associations of transgenic elements.
To get this crucial information, several strategies of DNA walking, also called genome walking, have been reported (Figure 1.2 and Table 1.10). More precisely, this molecular technique allows identifying unknown nucleotide sequences adjacent to already known DNA regions in any given genome using specific primers to the known sequence combined to primers dictated by the DNA walking method used. Then, the final PCR products are usually sequenced by Sanger technology to be eventually analyzed with available databases (*e.g.*, NCBI and JRC GMO-Amplicons). Classically, three main categories of DNA walking are established, based on the characteristics of their first step (Leoni *et al.* 2011).
First, the restriction-based methods involve a digestion of the genomic DNA using appropriate restriction enzymes targeting sites close to sequences of interest, such as the junction between the known and unknown sequences. The obtained restriction fragments are then either self-circularized or ligated to DNA cassettes, named, respectively, inverted-PCR and cassette PCR methods (Leoni *et al.* 2011 and references therein). By this way, several sequences of transgene flanking regions and unnatural associations from transgenic *Arabidopsis thaliana*,

tobacco, shallot, potato, barley, grapefruit, tomato, banana, cotton (MON1445), colza (including GT73), soybean (GTS40-3-2 and MON89788), wheat (B73-6-1, B72-8-11, and B72-8-11b), rice (including TC-19, Bt Shanyou 63 (TT51-1), KeFeng-6, and KeFeng-8), and maize (CHB-351, Bt176, GA21, Bt11, MON88017, MON863 × NK603, MON863 × NK603 × MON810, T25, MON810, NK603, MON863, T25, DAS-59122-7, LY038, and 3272) were characterized (Table 1.10) (Rudenko *et al.* 1993; Knapp *et al.* 1994; Devic *et al.* 1997; Spertini *et al.* 1999; Zimmermann *et al.* 2000; Balzergue *et al.* 2001; Cottage *et al.* 2001; Windels *et al.* 2001; Zheng *et al.* 2001; Holck *et al.* 2002; Samson *et al.* 2002; Theuns *et al.* 2002; An *et al.* 2003; Rønning *et al.* 2003; Sallaud *et al.* 2003; Windels *et al.* 2003a; Windels *et al.* 2003b; Yuanxin *et al.* 2003; Sallaud *et al.* 2004; Salvo-Garrido *et al.* 2004; Collonnier *et al.* 2005; Côté *et al.* 2005; Taverniers *et al.* 2005; Rai 2006; Akritidis *et al.* 2008; Santos *et al.* 2009; Ji and Braam 2010; Raymond *et al.* 2010; Ruttink *et al.* 2010a; Cao *et al.* 2011; Cullen *et al.* 2011; Su *et al.* 2011; Taheri *et al.* 2012; Trinh *et al.* 2012a; Trinh *et al.* 2012b; Wang *et al.* 2012; Zhang *et al.* 2012; Zhang *et al.* 2013b; Majhi *et al.* 2014; Trinh *et al.* 2014; Zhang *et al.* 2015).

Second, the extension-based methods are defined by the extension of a sequence-specific primer. The resulting single-stranded DNA is subsequently ligated to either a DNA cassette or 3'-tailing (Leoni *et al.* 2011 and references therein). This strategy was successfully applied on GM maize (MON810), rice (LLRICE62), soybean (A2704-12), rapeseed (T45), and cotton (LLCOTTON25) events in order to characterize their transgenic cassettes and transgene flanking regions (Table 1.10) (Spalinskas *et al.* 2013a; Spalinskas *et al.* 2013b).

Third, the primer-based methods combine combinatorial (random and/or degenerate) primers to target-specific primers according to various PCR strategies (Leoni *et al.* 2011 and references therein). The transgenic *Arabidopsis thaliana*, tobacco, potato, barley, apple, banana, soybean, wheat (B73-6-1),rice (including KeFeng-6 and KMD1), and maize (including MON863 and MIR162) were thereby identified via the sequences of their transgene flanking regions and unnatural associations of elements (Table 1.10) (Myrick and Gelbart 2002; Côté *et al.* 2005; Tan *et al.* 2005; Yang *et al.* 2005b; Pan *et al.* 2006; Babekova *et al.* 2009; Santos *et al.* 2009; Cullen *et al.* 2011; Wang *et al.* 2011a; Wang *et al.* 2011b; Xu *et al.* 2013a; Yao *et al.* 2013; Bartlett *et al.* 2014; Liang *et al.* 2014; Ma *et al.* 2014).

However, the implementation of most of these DNA walking methods by the enforcement laboratories presents some difficulties such as an insufficient specificity, sensitivity, or yield. Moreover, some of them use laborious, complex, and lengthy techniques (*e.g.*, fingerprinting by capillary electrophoresis and genomic DNA library via restriction enzyme). Therefore, a DNA walking approach, corresponding better to the need of enforcement laboratories, has been developed and validated on unprocessed and processed food matrices

Table 1.9: Representative examples illustrating simplex LAMP strategies targeting GMO.

Methods	Targets	References
Taxon-specific	ADH	Kiddle *et al.* (2012)
	LEC	Guan *et al.* (2010) and Di *et al.* (2014)
	PLD	Chen *et al.* (2012a)
	IVR	Chen *et al.* (2011a)
Element-specific	p35S	Fukuta *et al.* (2004), Lee *et al.* (2009), Kiddle *et al.* (2012), Randhawa *et al.* (2013), Zhang *et al.* (2013a), Zahradnik *et al.* (2014), and Wang *et al.* (2015a)
	pFMV	Randhawa *et al.* (2013) and Wang *et al.* (2015a)
	aadA	Randhawa *et al.* (2013)
	uidA	Randhawa *et al.* (2013)
	nptII	Randhawa *et al.* (2013) and Wang *et al.* (2015a)
	Cry1Ab	Li *et al.* (2013c)
	tNOS	Lee *et al.* (2009), Kiddle *et al.* (2012), Zhang *et al.* (2013a), Di *et al.* (2014), and Wang *et al.* (2015a)
	pNOS	Lee *et al.* (2009)
	bar	Zhang *et al.* (2013a) and Wang *et al.* (2015a)
	pat	Wang *et al.* (2015a)
	Cry1Ac	Wang *et al.* (2015a)
	CP4-EPSPS	Wang *et al.* (2015a)
	Cry2A	Li *et al.* (2014)
	Cry3A	Li *et al.* (2014)
	phy	Huang *et al.* (2014)
Construct-specific	p35S/EPSPS	Lee *et al.* (2009)

Table 1.9 *continued*

Methods	Targets	References
	Ms8	Lee *et al.* (2009)
	Rf3	Lee *et al.* (2009)
	MON89788	Guan *et al.* (2010), Zhang *et al.* (2013a), and Di *et al.* (2014)
	GTS 40-3-2	Guan *et al.* (2010), Zhang *et al.* (2013a), and Di *et al.* (2014)
	DAS-59122-7	Chen *et al.* (2011a) and Zhang *et al.* (2013a)
	MON863	Chen *et al.* (2011a) and Zhang *et al.* (2013a)
	TC1507	Chen *et al.* (2011a) and Zhang *et al.* (2013a)
Event-specific	T25	Chen *et al.* (2011a) and Xu *et al.* (2013b)
	Bt11	Chen *et al.* (2011a)
	Bt176	Chen *et al.* (2011a)
	MON810	Chen *et al.* (2011a)
	B73-6-1	Cheng *et al.* (2014)
	KMD1	Chen *et al.* (2012a)
	Kefeng-6	Chen *et al.* (2012a)
	TT51-1	Chen *et al.* (2012a)

aadA: aminoglycoside 3'-adenylyltransferase; ACC: acetyl-CoAcarboxylase gene from colza; ADH: alcohol dehydrogenase I gene from maize; bar: phosphinothricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; Cry: gene encoding the *Bacillus thuringiensis* δ-endotoxin; IVR: invertase gene from maize; LEC: lectin gene from soybean; nptII: neomycin phosphotransferase II gene; p35S: promoter of the 35 S cauliflower mosaic virus; pat: phosphinothricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*; pFMV: promoter of the figwort mosaic virus; phy: phytase gene from maize; PLD: phospholipase D gene from rice; pNOS: promoter of the nopaline synthase gene; tNOS: terminator of the nopaline synthase gene; uidA: β-glucuronidase.

containing minute amounts of GM targets. As this DNA walking approach implies two seminested PCR rounds, the yield and the specificity of GM targets are increased, especially crucial in case of a low level presence of GMO. This approach, belonging to the PCR-based method category, has also the advantage to be fully integrated into the GMO routine analysis as the similar primers are used for the qPCR screening (detection of potential GMO presence) and the DNA walking (GMO identification). So, this simple and rapid approach could easily be applied by the enforcement laboratories, without any significant additional cost and equipment, to confirm signals previously obtained in qPCR (Table 1.10) (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b). Since DNA walking requires less prior knowledge about the sequence of interest than conventional PCR-based methods previously described, GMO with entirely or partially known sequences could be characterized. Therefore, in targeting key elements, such as p35S and tNOS that are highly frequent in GM crops, a broad range of GMO could be characterized (Yang *et al.* 2005b; Akritidis *et al.* 2008; Raymond *et al.* 2010; Ruttink *et al.* 2010a; Wang *et al.* 2011b; Spalinskas *et al.* 2013a; Fraiture *et al.* 2015b). In order to especially identify unauthorized GMO in European Union, a DNA walking approach using primers specific to the element t35S from the pCAMBIA vector, found in approximately 30% of transgenic plants, was developed (Fraiture *et al.* 2014; Fraiture *et al.* 2015a). However, the DNA walking strategy is not suitable to GMO containing only unknown elements.

Next generation sequencing technologies

Despite their higher throughput compared to qPCR, the multiplex strategies described above require the prior knowledge of at least a part of the GMO sequences. Once the information about these sequences is collected, the development of methods, each one targeting individually one sequence of interest, is carried out on a case-by-case basis. Then, the optimisation of unbiased multiplex assays presenting equal analytical performance compared to simplex assays remains laborious and intricate. Furthermore, the issues related to the detection of GMO containing no known sequences are still unsolved. Recently, NGS, allowing a massive parallel DNA sequencing, has been suggested to tackle these challenges.The NGS technology outperforms plainly the classical Sanger sequencing in terms of rapidity and throughput. Indeed, the powerful high throughput of NGS offers the possibility to sequence simultaneously many different samples, discriminable in using a wide range of barcodes (Buermans and Dunnen 2014; Liang *et al.* 2014; Willems *et al.* 2016). Two main strategies, sequencing samples that are earlier enriched with sequences of interest (targeted sequencing approach) or not (whole genome sequencing (WGS) approach), exist (Figure 1.2 and Table 1.11).

Targeted sequencing The targeted sequencing strategy is especially beneficial to target regions of interest from large and complex genomes, observed in most of plants. Even if a minimum of prior knowledge on sequences is needed to target the sequences of interest, it presents the advantage to use exclusively all the energy, in terms of time and cost, on the regions of interest. With this strategy, two substrategies could be used, involving the sequencing of either DNA library of PCR products (amplicon sequencing) or selected DNA fragments from a whole genome library (target enrichment sequencing) (Figure 1.2).
On the one hand, as the amplicon sequencing allows characterizing DNA fragments of interest previously enriched by PCR, this sequencing approach depends thus clearly on the PCR strategy adopted upstream as well as its inherent properties and performance. In order to detect GMO, Song *et al.* (2014) generated amplicons by PCR, using primers targeting maize endogen gene, Bt11 gene, Bt176 gene, soybean endogen gene, 35S/CTP4 construct, CP4-EPSPS element, p35S promoter, and tNOS terminator, from samples containing a low amount of GM targets (1% of Bt11 maize, 2% of Bt176 maize, 2% of GTS40-3-2 soybean, 1% of GTS40-3-2 soybean, 0.1% ofGTS40-3-2 soybean, or 0.01% of GTS40-3-2 soybean). Then, each kind of amplicons was individually sequenced using a variant of the 454 system called pyrosequencing on portable photodiode-based bioluminescence sequencer that is more sensitive, compact, and cost-efficient compared to the original 454 technology (Roche) (Table 1.11) (Wu *et al.* 2011; Song *et al.* 2014). This approach is relatively similar to the PCR screening with the additional value to provide, instead of positive or negative signals, the sequence of the amplified fragments, which is more reliable to prove the presence of GMO. Conversely to this approach, Liang *et al.* (2014) suggest an amplicon sequencing strategy allowing analyzing GMO for which the sequence information is only partially known. To this end, a DNA walking method (SiteFinding PCR), targeting the vip3Aa20 sequence, was coupled to NGS technologies, using the Illumina or Pacific Biosciences platforms, to characterize the sequences of the MIR162 maize event (Table 1.11). Even if the results were similar using the two different NGS platforms, the PacBio system shows the advantage to sequence DNA fragments with a size reaching up to 40 Kbp and to deal with DNA fragments presenting different sizes. Therefore, the PacBio system, in contrast to the Illumina technology, allows in many cases avoiding a *de novo* assembly step as the shearing of genomic DNA is not always required. Moreover, the use of NGS instead of the Sanger technology allows considerably increasing the throughput of DNA walking approaches. Indeed, in order to guarantee the entire representativeness of GMO present in a tested sample, all observed amplicons should be analyzed. However, the purification of the potential numerous amplicons excised from the electrophoresis gel and the subsequent Sanger sequencing could be cumbersome, especially in case of food/feed matrices containing several GMO sharing common targeted elements

Table 1.10: Representative examples illustrating DNA walking strategies targeting GMO.

DNA walking approaches	Characterized regions	Targets	References
	Restriction-based methods		
Inverse PCR	Transgene flanking regions	Bt11	Zimmermann *et al.* (2000) and Rønning *et al.* (2003)
Cassette PCR	Transgene flanking regions	GTS40-3-2	Windels *et al.* (2001)
	Transgene flanking regions	GT73	Taverniers *et al.* (2005)
	Transgene flanking regions	MON1445	Akritidis *et al.* (2008)
	Transgene flanking regions	TC-19	Majhi *et al.* (2014)
	Transgene flanking regions	TT51-1	Cao *et al.* (2011)
	Transgene flanking regions	KeFeng-6	Su *et al.* (2011)
	Transgene flanking regions	KeFeng-8	Wang *et al.* (2012)
	Transgene flanking regions	B73-6-1	Zhang *et al.* (2012)
	Transgene flanking regions	B72-8-11	Zhang *et al.* (2015)
	Transgene flanking regions	B72-8-11b	Zhang *et al.* (2013b)
	Transgene flanking regions	LY038	Trinh *et al.* (2012a)
	Transgene flanking regions	MON89788	Trinh *et al.* (2012a)
	Transgene flanking regions	3272	Trinh *et al.* (2012a)
	Transgene flanking regions and unnatural element associations	CHB-351	Windels *et al.* (2003a) and Raymond *et al.* (2010)
	Transgene flanking regions and unnatural element associations	Bt176	Taverniers *et al.* (2005) and Raymond *et al.* (2010)
	Transgene flanking regions and unnatural element associations	GA21	Taverniers *et al.* (2005) and Raymond *et al.* (2010)
	Transgene flanking regions and unnatural element associations	Bt11	Taverniers *et al.* (2005) and Raymond *et al.* (2010)

Table 1.10 *continued*

DNA walking approaches	Characterized regions	Targets	References
	Transgene flanking regions and unnatural element associations	T25	Collonnier *et al.* (2005) and Raymond *et al.* (2010)
	Transgene flanking regions and unnatural element associations	MON810	Holck *et al.* (2002) and Raymond *et al.* (2010)
	Transgene flanking regions and unnatural element associations	DAS-59122-7	Raymond *et al.* (2010) and Trinh *et al.* (2012a)
	Unnatural element associations	MON88017	Raymond *et al.* (2010)
	Unnatural element associations	MON863×NK603	Raymond *et al.* (2010)
	Unnatural element associations	MON863×NK603×MON810	Raymond *et al.* (2010)
	Unnatural element associations	NK603	Raymond *et al.* (2010)
	Unnatural element associations	MON863	Raymond *et al.* (2010)
Extension-based methods			
LT-RADE	Transgene flanking regions and unnatural element associations	MON810	Spalinskas *et al.* (2013a) and Spalinskas *et al.* (2013b)
	Transgene flanking regions and unnatural element associations	LLRICE62	Spalinskas *et al.* (2013a) and Spalinskas *et al.* (2013b)
	Transgene flanking regions and unnatural element associations	T45	Spalinskas *et al.* (2013a)
	Transgene flanking regions and unnatural element associations	A2704-12	Spalinskas *et al.* (2013a)
	Transgene flanking regions and unnatural element associations	LLCOTTON25	Spalinskas *et al.* (2013a)

Table 1.10 *continued*

DNA walking approaches	Characterized regions	Targets	References
	PCR-based methods		
TAIL-PCR	Transgene flanking regions	MON863	Yang *et al.* (2005b) and Pan *et al.* (2006)
	Transgene flanking regions	KeFeng-6	Wang *et al.* (2011b)
	Transgene flanking regions	B73-6-1	Xu *et al.* (2013a)
SiteFinding PCR	Transgene flanking regions	KMD1	Babekova *et al.* (2009)
	Unnatural element associations	MIR162	Liang *et al.* (2014)
APAgene GOLD Genome Walking Kit	Transgene flanking regions and unnatural element associations	Bt rice	Fraiture *et al.* (2014), Fraiture *et al.* (2015a), and Fraiture *et al.* (2015b)
	Transgene flanking regions and unnatural element associations	MON863	Fraiture *et al.* (2015b)

Table 1.11: Representative examples illustrating NGS strategies targeting GMO.

NGS strategies	NGS platforms	Targets	Target sizes	References
Targeted sequencing	HiSeq (Illumina)	vip3Aa2 from MIR162	150 bp to 2 Kbp	Liang *et al.* (2014)
	PacBio RS (Pacific Biosciences)	vip3Aa2 from MIR162	150 bp to 2 Kbp	Liang *et al.* (2014)
	454 system (Roche Applied Science)	ssIIb	157 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	Bt11 gene	324 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	Bt176 gene	206 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	LEC	118 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	p35S/CTP4	171 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	CP4-EPSPS	498 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	p35S	195 bp	Song *et al.* (2014)
	454 system (Roche Applied Science)	tNOS	180 bp	Song *et al.* (2014)

Table 1.11 *continued*

NGS strategies	NGS platforms	Targets	Target sizes	References
Whole genome sequencing	HiSeq (Illumina)	MON17903 soybean	1115 Mbp	Kovalic *et al.* (2012)
	HiSeq (Illumina)	MON87704 soybean	1115 Mbp	Kovalic *et al.* (2012)
	HiSeq (Illumina)	FP967 flax	373 Mbp	Young *et al.* (2015)
	HiSeq (Illumina)	LLRICE62 rice	385 Mbp	Wahler *et al.* (2013)
	HiSeq (Illumina)	TT51-1 rice	385 Mbp	Yang *et al.* (2013)
	HiSeq (Illumina)	T1c-19 rice	385 Mbp	Yang *et al.* (2013)
	HiSeq (Illumina)	Bt rice	385 Mbp	Willems *et al.* (2016)

CP4-EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain; CTP4: chloroplast transit peptide 4 from the *Arabidopsis thaliana* epsps gene; LEC: lectin gene from soybean; p35S: promoter of the 35 S cauliflower mosaic virus; ssIIb: starch synthase IIb gene from maize; tNOS: terminator of the nopaline synthase gene; VIP3A: vegetative insecticidal protein 3A.

(Liang *et al.* 2014; Fraiture *et al.* 2015b; Pacific Biosciences 2015).
On the other hand, the target enrichment sequencing approach involves the selection of sequences of interest from the whole genome DNA library. To capture them, appropriate hybridization methods could be used relying on magnetic beads or microarrays associated with specific probes. The efficiency of the hybridization step is thus crucial for this sequencing strategy. The DNA fragments containing entirely or partially the known regions could be then sequenced. However, even if this strategy has been applied to different plants, no study has to date been reported to our knowledge to detect GMO (Zhou and Holliday 2012; Clarke *et al.* 2013; DuBose *et al.* 2013; Dasgupta *et al.* 2015).
The analysis of preenriched DNA fragments of interest with NGS technology allows proving the presence of GMO in characterizing sequences entirely or partially known beforehand. However, given its relative high cost, expected to decrease over the time, and the prerequisite bioinformatics expertise, the targeted NGS strategy could not reasonably be currently applied routinely to all food/feed matrices by the enforcement laboratories (Buermans and Dunnen 2014; Liang *et al.* 2014; Willems *et al.* 2016).

Whole genome sequencing The WGS strategy allows in principle characterizing a sample without any prior knowledge (Figure 1.2). With this sequencing strategy, the entire DNA library, consisting of sheared genomic DNA ligated to adaptors, is sequenced. The generated reads are then treated with bioinformatics tools based on prior knowledge of tested GMO.
First, when no information about the transgenic cassette is available, the insert and its transgene flanking regions are identified by the analysis of all inferred contigs derived from reads that partially matched or unmatched with the endogenous plant-species reference genome (Yang *et al.* 2013). This WGS strategy was applied on the LLRICE62 event by using the available reference genome of *Oryza sativa* ssp. Japonica. As the results corresponded to the information from the developer dossier, the characterization of GMO with an unknown insert usingNGS was thus demonstrated (Table 1.11) (Wahler *et al.* 2013). Similarly, the T-DNA regions from the GM flax FP967 event and the transgenic rice TT51-1 and T1c-19 events were also characterized (Table 1.11) (Yang *et al.* 2013; Young *et al.* 2015). The success of this strategy is thus linked to the availability of good reference genomes for specific varieties and organisms. In case of no reference genome available, a strategy of *de novo* assembly, comparing all generated reads to find overlaps, has to be applied. However, this remains quite cumbersome with the large and complex plant genomes notably in terms of ploidy, repeated regions, and heterozygosity and with mixtures of different GMO (Kovalic *et al.* 2012; Schatz *et al.* 2012). To facilitate even so the *de novo* assembly, the strength of different NGS platforms can be associated. For instance, short reads from Illumina technology can

be aligned to long reads generated by the PacBio technology, constituting a substitute of reference genome (Au *et al.* 2012).
Second, with the condition that the sequence of at least one transgenic element is known, the insert is *de novo* assembled with reads that are matched and unmatched with a DNA transgene sequence library containing frequently used transgenic elements. This approach was tested on the transgenic rice TT51-1 and T1c-19 events (Table 1.11) (Yang *et al.* 2013).
Third, if the sequence of the insert is known, two kinds of bioinformatics analysis have been reported. On the one hand, the reads, corresponding not entirely to the reference genome, are mapped to the transgenic cassette sequence in order to determine the number of inserts and their transgene flanking regions. By this way, the GM rice TT51-1 and T1c-19 events and the GM soybean MON17903 and MON87704 events were characterized (Table 1.11) (Kovalic *et al.* 2012; Yang *et al.* 2013). On the other hand, Willems *et al.* (2016) have developed an analytical workflow, including three different approaches. The detection approach, consisting of comparing the reads to the reference sequence of the insert, allows detecting the presence of GMO in a given sample. To confirm the integration of the transgenic cassette and provide a rough localization of its flanking regions, the matched reads are then compared to the reference sequence of the host genome in the proof approach. By the simultaneous aligning of these selected reads to the host genome and the transgenic cassette, the identification approach allows determining precisely the localization of the transgenic cassette and the sequence of its flanking regions. This WGS strategy was initially assessed on pure transgenic GM rice (100% Bt rice). Conversely to all the other WGS strategies described above, food/feed matrices more likely to be encountered in GMO routine analysis, such as a GM/non-GM rice mixture (10% Bt rice) and a processed GM rice (100% Bt noodles), have also been tested (Table 1.11) (Willems *et al.* 2016). In this study, a statistical framework, predicting the probability to detect a sequence derived from a transgenic cassette and validated with experimental data originated from WGS, was also developed to estimate *in silico* the number of reads, derived from Illumina HiSeq device, required to characterize frequently encountered GMO. It was shown that samples composed of GMO at 100%, except for GM wheat owning a huge genome, could be wisely characterized at a standard price range. *A contrario*, the detection, and identification of GMO present at trace level are not reasonably achievable by WGS (Willems *et al.* 2016). Therefore, at the present time, only the previously described targeted sequencing approach can be applied on GM mixture containing GMO at trace level within reason.
The NGS technology is thus a promising alternative in the GMO detection field which offers the possibility to prove straightforward the presence of GMO in food/feed matrix via the characterization of their sequences. Moreover, the sequences obtained from unknown GMO will allow designing new PCR markers.

Nevertheless, the implementation of NGS in GMO routine analysis by the enforcement laboratories is still difficult due to its relatively high cost as well as the requirement of adequate computer infrastructures and qualified analysts in bioinformatics for dealing with the generated data (Buermans and Dunnen 2014; Liang *et al.* 2014; Willems *et al.* 2016).

1.3.1.3 Conclusion

In GMO routine analysis, qPCR remains the method of choice for the enforcement laboratories. However, as some technical hurdles could be encountered with this technology, alternative GMO detection methods have been developed to raise some of these challenges. In order to exploit at best the performance of all the above described strategies, their applicability could be considered according to the adopted strategy of GMO detection as well as the available information about the sequences of tested GMO (Figure 1.2). In case of fully characterized GMO, the methods based on conventional PCR are absolutely appropriate to rapidly detect individually GM targets low-prized (LAMP), to simultaneously detect several GM targets (CGE, microarray, and Luminex) or to precisely quantify the amount of GM targets without impact of inhibitors (dPCR). However, when tested matrices contain GMO for which only a part of their sequences is known, these strategies could generate unexplained signals for which the observed positive signals could not be related to known GM events. In targeting key DNA sequences, such as the elements p35S and tNOS that are frequently found in GM plants, the use of DNA walking or targeted sequencing by enrichment strategies allows indubitably confirming the presence of GMO via the sequences of transgenes flanking regions and unnatural associations of genetic elements. If no information is available, at this moment, only the WGS is conceivable to characterize this category of GMO.

Conflict of interests

The authors declare that they have no competing interests.

Acknowledgements

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contact UGMMONITOR (convention RF 11/6242).

1.3.2 Advances in the identification of genetically modified foods

The following subsection was accepted for publication as:
M.-A. Fraiture, S. Broeders, P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2016b). "Advances in the identification of genetically modified foods". *Advances in Food Authenticity Testing*. Ed. by G. Downey. 1st edition. Woodhead Publishing Series in Food Science, Technology and Nutrition. Cambridge, UK: Woodhead Publishing. Chap. 21. ISBN: 9780081002209. Accepted

Authors' contributions

M.-A. Fraiture collected all information and drafted the manuscript. S. Broeders, P. Herman, I. Taverniers, M. De Loose, D. Deforce and N. H. Roosens drafted the manuscript. All authors read and approved the final manuscript.

1.3.2.1 Introduction

In 2014, the planting of 181.5 million hectares of genetically modified organisms (GMOs) by 18 million farmers in 28 countries was reported (James 2014). To guarantee the traceability of the food/feed chain and the freedom of choice for the consumer, legislations concerning the detection and quantification of GMOs have been established by the relevant competent authorities worldwide (Kamle and Ali 2013). In this respect, several strategies, classified as indirect (protein-based methods) or direct (DNA-based methods), have been developed (Morisset *et al.* 2008b; Mazzara *et al.* 2013). DNA-based methods, especially the real-time Polymerase Chain Reaction (qPCR) technology, are generally favored because of their higher sensitivity and more exact quantification. The qPCR technique allows the analyst not only to screen for genetically modified (GM) elements but also to identify the integrated transgenic cassette as well as to quantify the GM event (Broeders *et al.* 2012b).
However, with the increasing amount of GMOs brought to the market, the necessity to identify GMOs present at trace levels (EC/619/2011 on Low Level Presence (LLP) of GMO in feed) and the growing possible presence of GMOs which are authorised in one country but not in another, the above mentioned methods are no longer sufficient (The European Commission 2011a). In order to strengthen the qPCR analysis in GMO routine analysis, a DNA walking strategy was recently developed to unequivocally prove the presence of GMO in food and feed samples via the identification and characterisation of sequences

of the transgene flanking regions and unnatural element combinations (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b).
Here, the use of qPCR for GMO identification as currently performed, is described and different theoretical outcomes typically encountered by the enforcement laboratories are reported. Additionally, alternative DNA-based approaches for the identification of GMOs which may be used in combination with qPCR are illustrated. Furthermore, we discuss the potential application, including advantages and disadvantages in routine GMO analysis, of several other technologies such as Loop-Mediated Isothermal Amplification (LAMP), PCR capillary gel electrophoresis (CGE), microarray, Luminex®, digital PCR (dPCR) and Next Generation Sequencing (NGS) (Fraiture *et al.* 2015c).

1.3.2.2 Processes used for identification of genetically modified foods

Classically, GMO routine analysis is performed using qPCR technology to detect, identify and quantify GM events (Figure 1.3) (Angers-Loustau *et al.* 2014). With the help of a specific primer pair, the sequence of interest is amplified and the corresponding fluorescence, emitted by using either the asymmetrical cyanine-dye binding to double-stranded DNA (SYBR®Green chemistry) or a fluorogenic probe specific to the targeted sequence (TaqMan® chemistry), is recorded in real-time (Navarro *et al.* 2015). GMO routine analysis is commonly composed of three successive steps (Broeders *et al.* 2012b). The potential presence of GMOs is first determined by qPCR screening. To this end, sequences recurrently found in GMOs (*e.g.* p35S (35S promoter from the *Cauliflower Mosaic Virus*) and tNOS (nopaline synthase terminator from *Agrobacterium tumefaciens*)) are employed. The additional use of taxon-specific targets as well as more discriminating markers targeting single transgenic regions (less frequent elements or specific constructs) permit a reduction in the number of subsequent analysis required (Broeders *et al.* 2012b; Angers-Loustau *et al.* 2014; Fraiture *et al.* 2014; Broeders *et al.* 2015). The outcome of this analysis, based on the positive and negative signals observed, is a list of all GM events potentially present in the sample being tested. Next, in order to identify the GM events truly present in the sample under investigation, the related event-specific methods, targeting the junction between the transgenic cassette and the plant genome, are applied (Broeders *et al.* 2012b). Finally, identified GM events are quantified using the corresponding event-specific and taxon-specific methods (Broeders *et al.* 2012b). In the European Union (EU), the event-specific qPCR methods used by enforcement laboratories have been validated at EU level in agreement with the established document "Minimum Performance Requirements for Analytical Methods of GMO Testing" and are reported in the Compendium of reference methods for GMO analysis (European Network of GMO Laboratories (ENGL) 2015; Joint Research Centre 2015a).

To facilitate the application of this workflow and to reduce the time-frame of the analysis, a 96-well, pre-spotted plate, including several taxon-specific, event-specific and construct-specific TaqMan® methods, was developed by the European Union Reference Laboratory for GM Food and Feed (EU-RL GMFF) allowing the simultaneous identification of thirty-nine GM events (Querci *et al.* 2010).

Besides the fact that the rising number of GMOs to be identified renders the laboratory work more time-consuming, analysis of the qPCR data is also intensified. This can be simplified by using analytical tools simultaneously integrating the results of several methods. For instance, CoSYPS (Combinatory SYBR®Green qPCR Screening) is a practical decision support system that allows the analyst to quickly determine which GMO is potentially present in a sample and should thus further be identified and eventually quantified. This system is based on the introduction of the experimentally obtained C_q and, in case of SYBR®Green chemistry, T_m values that are then compared to the corresponding data obtained during in-house validation (screening) and verification (event-specific methods) of the methods. Currently, CoSYPS includes data for the twenty SYBR®Green taxon-, element- and construct-specific screening methods developed by the Scientific Institute of Public Health in Belgium (WIV-ISP) as well as for all EU authorised GMO and LLP cases (Van den Bulcke *et al.* 2010; The European Commission 2011a; Broeders *et al.* 2012b). Alternatively, systems such as GMOseek and GMOfinder databases, allow, on the basis of name of the positive elements observed, giving rise to a theoretical list of the GM events potentially present in the sample. These results have, however, to be analysed with prudence since dissimilar sequences are sometimes named identically (Gerdes *et al.* 2012; Block *et al.* 2013). Recently, the JRC-GMO-Matrix has been developed to generate the list of GM events potentially present by integrating experimental positive and negative signals from EU-validated methods (Angers-Loustau *et al.* 2014; Joint Research Centre 2014).

Nevertheless, the approach described above still presents certain limitations. Firstly, taking into account that usually only one method per reaction is used, the laboratory work could become quite laborious and complex, especially due to the growing number of GMOs (Broeders *et al.* 2012a). Secondly, as the qPCR strategy is based on knowledge of the targeted sequences, the presence and quantity of only GMOs for which the elements of the transgenic cassette (screening) and the junction sequence (identification and quantification) are known can be assessed. In addition, the validated EU event-specific methods are only available for EU authorised GM events. This implicates that when the observed screening signals cannot be explained by the presence of EU authorised GM events, the presence of unknown/unauthorised GMO can only be suspected (Fraiture *et al.* 2014). Thirdly, Certified Reference Materials (CRM), mandatory

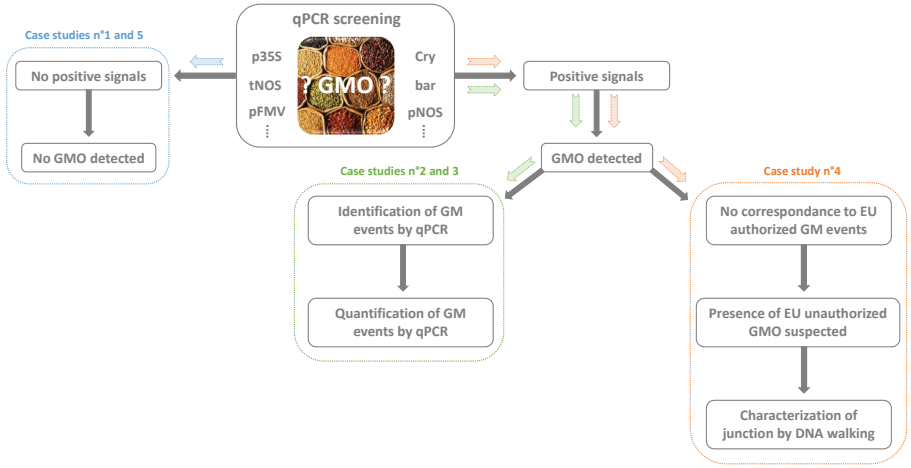


Figure 1.3: Traditional GMO routine analysis workflow. The case studies described in section 1.3.2.3 are indicated in blue (n°1 and 5), green (n°2 and 3) and orange (n°4).

for the quantification step, are also only available for the EU authorised GM events (Broeders *et al.* 2012b; Holst-Jensen *et al.* 2012; Fraiture *et al.* 2014). Finally, due to the sensitivity of qPCR to inhibitors which may negatively affect its efficiency (*e.g.*, polysaccharides, polyphenols, pectin, xylan or fat), the quantity of GMO may be underestimated or even remain undetected (Demeke and Jenkins 2010; Opel *et al.* 2010; Schrader *et al.* 2012). There is thus a need for several complementary and alternative methods to circumvent these drawbacks of singleplex qPCR methods and render GMO analysis more efficient and effective.

Multiplex qPCR strategy

To deal with the growing number of GMOs and the established one-by-one detection strategy, multiplex qPCR TaqMan® methods were developed which allow the detection of several targets in one reaction. Even if most of them combine only two or three markers, tetraplex, pentaplex and even hexaplex qPCR have also been elaborated (*e.g.* Waiblinger *et al.* (2008), Bahrdt *et al.* (2010), Dörries *et al.* (2010), Cottenet *et al.* (2013), Huber *et al.* (2013), Samson *et al.* (2013), Chaouachi *et al.* (2014), Köppel *et al.* (2014), Köppel *et al.* (2015), and Park *et al.* (2015)). However, the development of multiplex qPCR assays is not straightforward and several factors should be taken into account.

Additionally, the establishment of multiplex assays with similar performance (*e.g.* sensitivity and specificity) to singleplex qPCR assays could be complicated. Moreover, the throughput of this strategy is limited by the inherent properties of the qPCR technology which implies that fluorescence emitted by the dyes coupled to each marker should be sufficiently distinct, a factor that is also restrained by the current brand of qPCR instruments (Bahrdt *et al.* 2010; Joint Research Centre 2015a).

DNA walking

In the case that positive signals obtained in the qPCR screening step cannot be explained by EU authorised events, the presence of unknown/unauthorised GM events can be determined by the DNA walking technique (Fraiture *et al.* 2014 and references therein; Fraiture *et al.* 2015a and references therein; Fraiture *et al.* 2015b and references therein). Indeed, this technique allows characterisation of unknown nucleotide sequences adjacent to already known DNA regions in any given genome. Even though several DNA walking methods have already been suggested to characterise transgenic plants, most of them could not easily be implemented in GMO routine analysis due to their experimental complexity as well as insufficient specificity, sensitivity or yield. A rapid, simple and cost-effective DNA walking strategy was thus developed and validated on both processed and unprocessed food matrices containing minute amounts of GM targets (Fraiture *et al.* 2014 and references therein; Fraiture *et al.* 2015a and references therein; Fraiture *et al.* 2015b and references therein). Following qPCR screening, a PCR amplification is carried out using primers specific to the known sequence of the detected transgenic element(s) combined with degenerated random primers. With the aim of increasing the yield and specificity of the GM targets, two successive semi-nested PCR analyses are then performed. The purified final PCR products are subsequently sequenced using Sanger or Next-Generation-Sequencing (NGS) platforms and analysed using available databases (*e.g.* NCBI and JRC GMO-Amplicons) (Fraiture *et al.* 2014 and references therein; Joint Research Centre 2015b). The characterised transgene flanking regions and unnatural combinations of transgenic elements will prove the presence of GMO in the tested food/feed matrix beyond any doubt. Subsequently, these sequences can also be used to develop new event-specific qPCR methods. Moreover, as the same primers are used for the qPCR screening (potential detection of GMO) and the DNA walking (GMO identification), the present DNA walking system can be fully integrated into routine GMO analysis. Nowadays, in order to cover a broad range of GMOs, this DNA walking strategy has been developed to target the t35S sequence as found in the pCAMBIA vectors, frequently observed in EU unauthorised GMOs, as well as the p35S and tNOS sequences, commonly used in EU authorised and EU unauthorised

GMOs (Fraiture *et al.* 2014 and references therein; Fraiture *et al.* 2015a and references therein; Fraiture *et al.* 2015b and references therein).

1.3.2.3 Case studies

On the basis of results obtained in the qPCR screening step, subsequent analysis can present some variations (Figure 1.3). To illustrate this, theoretical case studies of typical outcomes observed in GMO routine analysis are described below.

Firstly, the simplest situation is encountered when only taxon-specific markers give a positive and specific amplification signal with none of the transgenic screening markers tested , designed to cover at least all EU authorised GM events and LLP cases, being detected (Case study n°1 in Figure 1.3). This means that no GM events containing the targeted sequences are present at a reliable detection level and no further analysis is thus undertaken. This case study highlights here the importance of the screening step, composed of markers targeting a broad spectrum of GMOs, to determine their presence or absence in any given sample.

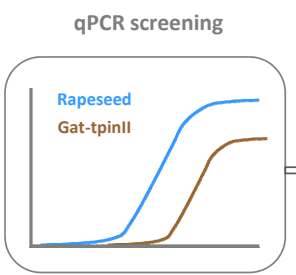
Secondly, matrices like raw and unprocessed materials are usually composed of only one plant species, indicated by the observation of only one positive taxon-specific marker. Similarly to case study n°1, screening methods that cover a broad range of GMOs are applied to the tested sample to determine the presence of GMO. When positive signals corresponding to transgenic elements are found, all the event-specific methods targeting EU authorised GM events and LLP cases potentially present for the identified species are tested in a subsequent identification step (Case study n°2 in Figure 1.3). However, depending on the identified species, this list may vary currently from one (*e.g.* rice species that includes only the LLRICE62 event) to eighteen (*e.g.* maize species) required identifications. Therefore, more discriminative screening methods are further used to decrease the number of reactions needed to be carried out downstream (Figure 1.4A). For instance, when the rapeseed species is detected, ten EU authorised GM events or LLP cases can be suspected. In the situation where only p35S and tNOS screening methods, known to target a broad spectrum of GMOs, are applied, seven GM events are still required to be identified. With the help of more discriminative screening methods, this list of GM rapeseed events can be reduced. If, among the panel of screening methods, a positive signal is only observed for the gat-tpinII method, the list of ten GM rapeseed events potentially present is reduced to only one (73496) (Figure 1.4A). The presence of this event will be then confirmed with the corresponding event-specific method in the identification step and subsequently quantified to determine the amount of this GM event present in the sample under test. This case study thus illustrates the usefulness of taxon-specific methods and the diverse panel of GM screening

methods, allowing coverage of a broad range of GM events as well as indication of a limited number of GM events to be analysed in further steps.
Thirdly, the benefit of using a panel of screening methods is even more apparent in the case of more complex matrices, such as feed, that may contain several GM events of different plant species. qPCR screening of this kind of sample will thus result in several detected taxon- and element-specific markers, providing a long list of GM events which are potentially present (Case study n°3 in Figure 1.3). In such samples, identification of numerous GMOs in different amounts is usual, from which several need to be quantified using the corresponding taxon-specific and event-specific methods. As an example, suppose that the qPCR screening results indicate that ingredients from the tested food/feed matrix belong to maize and cotton species; twenty-seven EU authorised GM events or LLP cases are then to be suspected as present (Figure 1.4B). According to the positive signals for the p35S and tNOS methods, twenty-four GM events must still be investigated during the identification step. However, using the panel of screening methods, this list could be reduced. Indeed, if negative signals are additionally observed with pFMV, pNOS, t35S, pat, bar, gat-tpinII and the pentaplex methods while the EPSPS, Cry1Ab/Ac and Cry3Bb methods are positive, ten GM events (GA21, MON 863, NK603, MON 88017, MIR604, MON 87460, 5307, MON 87427, MON 15985 and MON 531), instead of thirty-four, are finally designated for the identification step. Event-specific methods of these suspected ten GM events will thus be applied to determine which GM events are truly present. Next, the GM amount of identified GM events in the matrix tested will be quantified. In this case study, the importance of using a combination of screening markers targeting sequences from both elements present at high and low frequency in GMOs is demonstrated. Moreover, the use of a decision support system is crucial to simplify the analysis of the correspondence between all signals observed within the screening phase and the numerous EU authorised GM events and LLP cases which exist.
Fourthly, food or feed matrices can also contain unknown or unauthorised GM events (Case study n°4 in Figure 1.3). For instance, if the specific amplification signals from the qPCR screening cannot be related to any authorised GM event, the presence of unknown or unauthorised GMOs is suspected. To reveal the truth of this hypothesis, a DNA walking strategy can be applied starting from one of the detected transgenic screening elements. This technique permits further characterisation of the GM event through sequences of the transgene flanking regions and the unnatural combination of elements. Based on this sequence information, qPCR event-specific methods can be developed thereafter to allow not only identification but also, in combination with taxon-specific methods, quantification of these GMOs. To illustrate this case study, two examples are given. On the one hand, the presence of an unknown/unauthorised GMO can be detected using the t35S pCAMBIA screening method targeting a sequence

observed only in EU unauthorized GM events at a frequency of approximately 30% (Fraiture *et al.* 2014). A positive signal for this screening method automatically triggers application of the DNA walking strategy (Figure 1.5A). On the other hand, positive signals are observed for taxon-specific methods targeting rapeseed and cotton species as well as for the p35S method while all other screening methods tested show no amplification (Figure 1.5B). No correspondence can be established between the observed qPCR signals and all existing EU authorised GM events and LLP cases. Indeed, among the cotton and rapeseed species, nine GM events possess the p35S element in their transgenic cassettes. However, this marker is always present in combination with other elements that were not detected in the analysis. Therefore, the presence of unknown/unauthorised GMO is highly probable. The same supposition is also made when no GM events, potentially suspected in the tested sample based on the outcome of the screening, are identified using the corresponding event-specific methods. To confirm this hypothesis, the DNA walking strategy must be applied to characterise sequences surrounding the detected p35S element in order to get transgene flanking regions and/or unnatural combinations of transgenic elements. With the aim of detecting the entire population of GMOs present in the tested sample, all sequences generated by the DNA walking approach have to be analysed. However, this step could be quite time-consuming and laborious, especially in case of food/feed matrices composed of several GMOs harbouring the same transgenic element. Use of NGS technology in the sequencing step instead of the Sanger method could be highly helpful in this case (Fraiture *et al.* 2015b). However, when unknown GMOs, composed only of transgenic elements not targeted in the qPCR screening step, are present, the DNA walking strategy is not applicable (Case study n°5 in Figure 1.3). The absence of positive signals with the panel of screening methods cannot exclude the possible presence of unknown/unauthorised GMO in the sample under test. For these GMOs, detection still represents a real challenge.

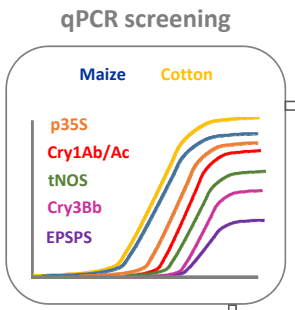
Although advances in GMO detection using additional techniques and strategies continue, there is still room for improvement to reach the ideal GMO analysis, able to deal in a time- and cost-effective way with the increasing number and diversity of GMOs, mixtures of several GMOs and the potential presence of unauthorised/unknown GMOs. Moreover, the problems of the current need for CRM and qPCR quantification drawbacks should be taken into account when developing new methodologies for routine analysis. To this end, alternative strategies are further proposed.

A



Species	GM events	p35S	tNOS	pFMV	pNOS	t35S	EPSPS	pat	bar	Cry1Ab/Ac	Gat-tpinII	Cry3Bb	Pentaplex
Rapeseed	MS8	-	+	-	-	-	-	-	+	-	-	-	-
	RF3	-	+	-	+	-	-	-	+	-	-	-	-
	GT73	-	-	+	-	-	+	-	-	-	-	-	-
	T45	+	-	-	-	+	-	+	-	-	-	-	-
	MON88302	-	-	+	-	-	+	-	-	-	-	-	-
	73496	-	-	-	-	-	-	-	-	-	+	-	-
	MS1	-	+	-	+	-	-	-	+	-	-	-	-
	RF1	-	+	-	+	-	-	-	+	-	-	-	-
	RF2	-	+	-	+	-	-	-	+	-	-	-	-
	Topas19/2	+	-	-	+	+	-	+	-	-	-	-	-

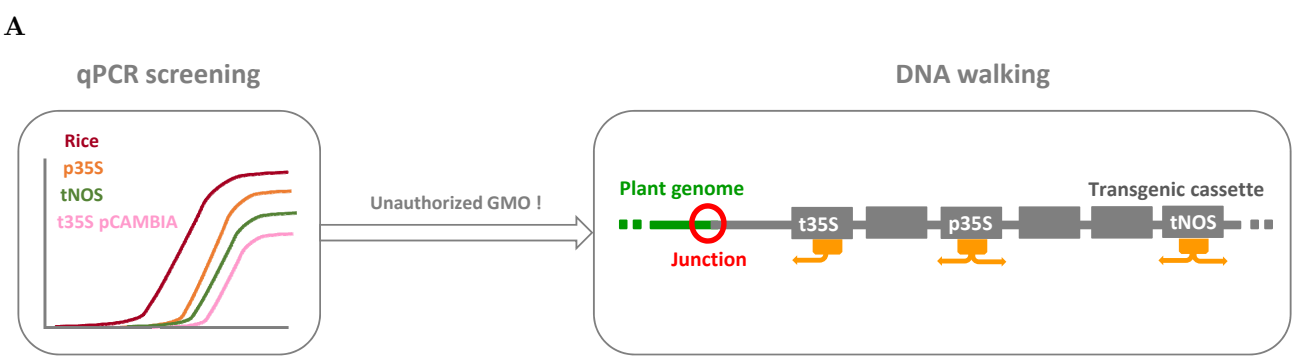
B



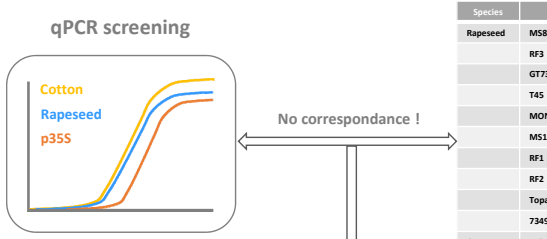
Species	GM events	p35S	tNOS	pFMV	pNOS	t35S	EPSPS	pat	bar	Cry1Ab/Ac	Gat-tpinII	Cry3Bb	Pentaplex
Maize	Bt11	+	+	-	-	-	-	+	-	+	-	-	-
	MON810	+	+	+	-	-	-	-	-	+	-	-	-
	T25	+	-	-	-	-	-	+	-	-	-	-	-
	DAS59122	+	-	-	-	-	-	+	-	-	-	-	-
	GA21	-	+	-	-	-	+	-	-	-	-	-	-
	MON863	+	+	-	-	-	-	-	-	-	-	+	-
	NK603	+	+	-	-	-	+	-	-	-	-	-	-
	MON88017	+	+	-	-	-	+	-	-	-	-	+	-
	MIR604	-	+	-	-	-	-	-	-	-	-	-	-
	MON87460	+	+	-	-	-	-	-	-	-	-	-	-
	5307	-	+	-	-	-	-	-	-	-	-	-	-
	MON87427	+	+	-	-	-	+	-	-	-	-	-	-
	TC1507	+	-	-	-	+	-	-	-	-	-	-	-
	MON89034	+	+	+	-	-	-	-	-	-	-	-	-
	MIR162	-	+	-	-	+	-	-	-	-	-	-	-
	3272	-	+	-	-	+	-	-	-	-	-	-	-
	Bt176	+	-	-	-	+	-	-	+	+	-	-	-
	DAS40278-9	-	-	-	-	-	-	-	-	-	-	-	+
Cotton	MON1445	+	+	+	-	-	+	-	-	-	-	-	-
	MON15985	+	+	-	-	-	-	-	-	+	-	-	-
	MON531	+	+	-	-	-	-	-	-	+	-	-	-
	LL25	+	+	-	-	-	-	-	+	-	-	-	-
	281-24-236 x 3006-210-23	-	-	-	-	-	-	+	-	+	-	-	-
	GHB614	-	-	-	-	-	-	-	-	-	-	-	+
	MON88913	+	-	+	-	-	+	-	-	-	-	-	-
	T304-40	+	+	-	-	-	-	-	+	+	-	-	-
	GHB119	+	+	-	-	+	-	-	+	-	-	-	-

Figure 1.4: Examples of results encountered with EU authorised GM events and LLP cases in routine GMO analysis: Case studies n°2 (A) and n°3 (B). For each matrix tested, positive signals observed in qPCR screening analysis for rapeseed species (light-blue), cotton species (yellow), maize species (dark-blue), gat-tpinII (brown), p35S (orange), tNOS (green), Cry1Ab/Ac

Figure 1.4 *(continued)*: (red), Cry3Bb (dark-pink) and EPSPS (purple) methods are represented by schematic amplification curves. In the tables, all positive (+) and negative (−) signals expected with the screening methods tested are indicated for each GM event. The current panel of screening methods includes p35S, tNOS, pFMV, pNOS, t35S, EPSPS, pat, bar, Cry1Ab/Ac, gat-tpinII, Cry3Bb and a pentaplex (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2013; Barbau-Piednoir *et al.* 2014; Broeders *et al.* 2015; in-house).



B



No correspondance !

Species	GM events	p35S	tNOS	pFMV	pNOS	t35S	EPSPS	pat	bar	Cry1Ab/Ac	Gat-tpinII	Cry3Bb	Pentaplex	t35S pCAMBIA
Rapeseed	MS8	-	+	-	-	-	-	-	+	-	-	-	-	-
	RF3	-	+	-	+	-	-	-	+	-	-	-	-	-
	GT73	-	-	+	-	-	+	-	-	-	-	-	-	-
	T45	+	-	-	-	+	-	+	-	-	-	-	-	-
	MON88302	-	-	+	-	-	+	-	-	-	-	-	-	-
	MS1	-	+	-	+	-	-	-	+	-	-	-	-	-
	RF1	-	+	-	+	-	-	-	+	-	-	-	-	-
	RF2	-	+	-	+	-	-	-	+	-	-	-	-	-
	Topas19/2	+	-	-	+	+	-	+	-	-	-	-	-	-
	73496	-	-	-	-	-	-	-	-	-	+	-	-	-
Cotton	MON15985	+	+	-	-	-	-	-	-	+	-	-	-	-
	MON1445	+	+	+	-	-	+	-	-	-	-	-	-	-
	MON531	+	+	-	-	-	-	-	-	+	-	-	-	-
	LL25	+	+	-	-	-	-	-	+	-	-	-	-	-
	281-24-236 x 3006-210-23	-	-	-	-	-	-	+	-	+	-	-	-	-
	GHB614	-	-	-	-	-	-	-	-	-	-	-	+	-
	MON88913	+	-	+	-	-	+	-	-	-	-	-	-	-
	T304-40	+	+	-	-	-	-	-	+	+	-	-	-	-
	GHB119	+	+	-	-	+	-	-	+	-	-	-	-	-

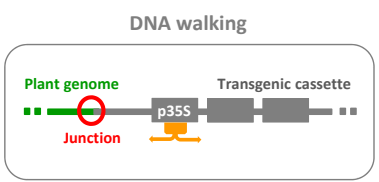


Figure 1.5: Examples of results encountered with EU unauthorised GM events in routine GMO analysis. For each matrix tested, positive signals observed in qPCR screening analysis for the rapeseed species (light-blue), cotton species (yellow), rice species (burgundy), p35S (orange), tNOS (green) and t35S pCAMBIA (light-pink) methods are represented by schematic amplification curves. In the table, all positive (+) and negative (−) signals expected with the screening methods tested are indicated for each GM event. The current panel of screening methods includes p35S, tNOS, pFMV, pNOS, t35S, EPSPS, pat, bar, Cry1Ab/Ac, gat-tpinII, Cry3Bb, a pentaplex and t35S pCAMBIA (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2013; Barbau-Piednoir *et al.* 2014; Fraiture *et al.* 2014; Broeders *et al.* 2015; in-house).

1.3.2.4 Future trends

Alternative multiplex strategies

To enable increased throughput of analysis, non-qPCR based multiplex assays have been investigated. While GMOs may be present at trace levels, the targets still need to be initially amplified by PCR to permit their analysis using CGE, microarray and Luminex® technologies. However, inherent properties linked to the necessary PCR step limit the multiplexing level in general to ten targets per PCR assay (Pla *et al.* 2012; Vega and Marina 2014).

With the CGE system, multiple GM targets, previously amplified with fluorescently-labelled primers, are easily discriminated even if the size range of the generated amplicons is similar. It also allows higher multiplexing than qPCR since tetraplex, pentaplex, hexaplex, octaplex and nonaplex PCR assays have already been developed to detect GMOs (Nadal *et al.* 2006; Heide *et al.* 2008a; Heide *et al.* 2008b; Nadal *et al.* 2009; Holck *et al.* 2010; Guo *et al.* 2011; Holck and Pedersen 2011; Basak *et al.* 2014). However, compared to the qPCR technology, the sensitivity of the CGE system is slightly lower (Milavec *et al.* 2014).

Regarding microarray technology, more than 250 000 targets, initially amplified by PCR using target-specific and/or universal primers, can be simultaneously detected, resulting in a much higher throughput than the current analyses. In the field of GMO detection, several multiplex PCR assays, from duplex to dodecaplex, have been established (Rudi *et al.* 2003; Leimanis *et al.* 2006; Leimanis *et al.* 2008; Morisset *et al.* 2008a; Prins *et al.* 2008; Hamels *et al.* 2009; Dobnik *et al.* 2010; von Götz 2010; Pla *et al.* 2012; Shao *et al.* 2014). Recently, with the help of constant improvements in the technology, one approach has allowed detection of unamplified GM DNA from three GM soybean and nine GM maize events at levels as low as or equal to 1% using a Comparative Genomic Hybridisation (CGH) microarray platform (Turkec *et al.* 2016). However, as is the case for the GCE approach, microarray technology never reaches the sensitivity obtained by qPCR.

Using Luminex® technology, up to 500 different targets can be identified. To this end, the target-specific amplicons are generated in a biotinylated form by PCR and subsequently hybridised to nucleic acid probes specific for the targets under investigation which are then independently coupled to spectrally distinct sets of beads. The fact that the beads are composed of different amounts of red and infrared color facilitates the high multiplexing property. Compared to the microarray system, this liquid bead array technology is potentially more sensitive and faster. Utilising singleplex, triplex, tetraplex, pentaplex or hexaplex PCR assays, several GMOs have already been detected by this technology (Fantozzi *et al.* 2008; Choi *et al.* 2010; Choi 2011; Han *et al.* 2013; Fu *et al.* 2015a;

Luminex Corporation 2015).

Digital PCR technology

Digital PCR (dPCR) technology, based on binomial Poisson statistics, allows absolute quantification, in contrast to the relative quantification with qPCR, of the number of nucleic acid targets present in the tested sample. Contrary to qPCR, this end-point PCR system is efficient even if the copy number of the target is low and/or PCR inhibitors are present. Moreover, no calibration curves are needed for the measurement, thus reducing drastically the need for a high amount of CRM. The fact that this technology is based on a stand-alone PCR would simplify its introduction for routine use especially if the current validated event-specific qPCR methods could be transferred. However, this process may require some optimisation. Nowadays, several GMOs have been quantified using the chamber dPCR (cdPCR) approach or the droplet dPCR (ddPCR) approach, in which samples are partitioned either into thousands of microfluidic chambers or thousands of droplets generated by a water-oil emulsion (*e.g.* Bhat *et al.* 2009; Burns *et al.* 2010; Corbisier *et al.* 2010; Morisset *et al.* 2013; Strain *et al.* 2013; Brod *et al.* 2014; Fu *et al.* 2015b; Köppel and Bucher 2015; Köppel *et al.* 2015; Li *et al.* 2015). To specifically identify the target, the throughput level of this dPCR technology is nonetheless limited to one singleplex or one duplex PCR assay per reaction. However, if the aim is to analyse the GMO content per type of ingredient, this throughput could be increased. As such, twelve EU authorised GMOs were recently targeted by combining one tetraplex dPCR assay with one decaplex dPCR assay (Dobnik *et al.* 2015).

Loop-mediated isothermal amplification

With the LAMP approach, sequences of interest are first amplified under isothermal conditions using four primers specific to six distinct regions of the target and then visualised with the help of color indicators, fluorescent dyes or precipitation processes. As no sophisticated equipment is necessary, this analysis can also be carried out on-field. In addition, a tolerance to PCR inhibitors has been observed with this technology (Morisset *et al.* 2008b; Zhang *et al.* 2013a). In GMO detection, several specific and sensitive element-specific, construct-specific, event-specific and taxon-specific methods have been successfully developed (*e.g.* Fukuta *et al.* 2004; Lee *et al.* 2009; Liu *et al.* 2009; Guan *et al.* 2010; Chen *et al.* 2011a; Chen *et al.* 2012a; Kiddle *et al.* 2012; Li *et al.* 2013c; Randhawa *et al.* 2013; Xu *et al.* 2013b; Zhang *et al.* 2013a; Cheng *et al.* 2014; Di *et al.* 2014; Huang *et al.* 2014; Li *et al.* 2014; Zahradnik *et al.* 2014; Wang *et al.* 2015a). Use of the LAMP system does however present some difficulties around the

design of four primers per target as well as the impossibility to individually identify targets within a multiplex PCR assay (Morisset *et al.* 2008b; Zhang *et al.* 2013a).

High-throughput sequencing technologies

Given the fact that characterisation of the GM junction sequences is an incontrovertible proof of GMO presence in a given sample, application of NGS technology in GMO detection has been investigated. Due to the possibility of barcoding individually different targets, this technology potentially offers a high-throughput. Among all NGS strategies, two main approaches are distinguished, namely targeted sequencing and whole-genome-sequencing (WGS) (Buermans and Dunnen 2014; Liang *et al.* 2014; Willems *et al.* 2016).

For the targeted sequencing strategy, sequences of interest are first selected to be then sequenced, thus requiring some prior knowledge.

On the one hand, sequences of interest can be enriched by PCR (amplicon sequencing) to compose a DNA library of final PCR products which are subsequently sequenced. Using a pyrosequencing device, this approach was used to characterise amplicons generated by GM-specific methods from samples presenting a low amount of GM soybean and GM maize targets (Wu *et al.* 2011; Song *et al.* 2014). Instead of C_q values with qPCR technology, the sequence of each targeted region is thus directly known. Moreover, the sequencing of amplicons from event-specific PCR methods using the Ion Torrent platform has allowed determination of the zygosity of transgenic maize lines, providing crucial information for the conversion between GM mass percentages and GM copy numbers (Fritsch *et al.* 2015). With the aim of characterising unknown regions, PCR products from a DNA walking approach targeting the vip3Aa20 element, notably found in the GM MIR162 maize, have also been sequenced on the Illumina and Pacific Biosciences platforms (Liang *et al.* 2014).

On the other hand, a DNA library can also be prepared from the whole genome followed by sequencing of selected sequences of interest (target enrichment sequencing). To this end, appropriate hybridisations to specific probes are carried out via magnetic beads or microarrays. However, to our knowledge, this approach has not yet been applied to GMO detection although successfully used in other fields (Zhou and Holliday 2012; Clarke *et al.* 2013; DuBose *et al.* 2013; Dasgupta *et al.* 2015).

Taking into account the expected improvements in terms of cost and bioinformatics tools, this targeted NGS strategy therefore represents a promising alternative for enforcement laboratories.

Regarding the WGS strategy, the DNA library contains the entire genome which is fully sequenced. The generated reads can then be treated with bioinformatics, eventually using knowledge available on the tested GMO.

For entirely unknown GMOs, the insert is characterised via all inferred contigs from reads that partially match or unmatch the reference genome of the plant species (Yang *et al.* 2013). In this way, LLRICE62, TT51-1, T1c-19 and FP967 events were identified (Wahler *et al.* 2013; Yang *et al.* 2013; Young *et al.* 2015). However, if no appropriate reference genome is available, a strategy of *de novo* assembly needs to be carried out. To this end, all generated reads are analysed to find overlaps between each other; this can be difficult due notably to the large size and complexity of plant genomes as well as the possible presence of several different GMOs in the tested sample (Kovalic *et al.* 2012; Schatz *et al.* 2012). In order to deal with some of these issues, long reads generated by the PacBio technology could constitute an alternative to the unavailability of reference genomes of the plant species concerned to which short reads from the Illumina technology can then be aligned (Au *et al.* 2012).
When the sequence of the transgenic insert is partially known, the generated reads can be compared to a DNA transgene sequence database containing transgenic elements commonly found in GMOs. The TT51-1 and T1c-19 events were, for instance, identified by this approach (Yang *et al.* 2013).
If the sequence of the insert is entirely known, two main bioinformatics analyses have been proposed. On the one hand, as tested with TT51-1, T1c-19, MON 17903 and MON 87704 events, the obtained reads which present only a partial similarity to the reference genome of the investigated plant species, are subsequently aligned to the sequence of the transgenic insert to locate and characterise the transgene flanking regions (Kovalic *et al.* 2012; Yang *et al.* 2013). On the other hand, the generated reads can also be analysed according to an analytical workflow composed of three steps. Firstly, the presence of GMO(s) is demonstrated if reads corresponding to the transgenic insert sequence are found. Next, these reads are compared to the plant-species genome to roughly locate the transgene flanking regions. Finally, characterisation of the transgenic insert with its precise localisation is determined by simultaneous aligning these selected reads to the plant reference genome and the sequence of the transgenic insert. This WGS strategy was successfully applied on pure GM rice, GM/non-GM rice mixtures and processed GM rice (Willems *et al.* 2016).
Although its implementation by enforcement laboratories is nowadays still difficult in terms of cost, staff training and bioinformatics analysis, the WGS strategy represents a promising support for routine GMO analysis, especially for totally unknown GMOs (Buermans and Dunnen 2014; Liang *et al.* 2014; Willems *et al.* 2016).

1.3.2.5 Sources of further information

See references Broeders *et al.* (2012b), Fraiture *et al.* (2015b), and Fraiture *et al.* (2015c).

Acknowledgements

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contact UGMMONITOR (convention RF 11/6242).

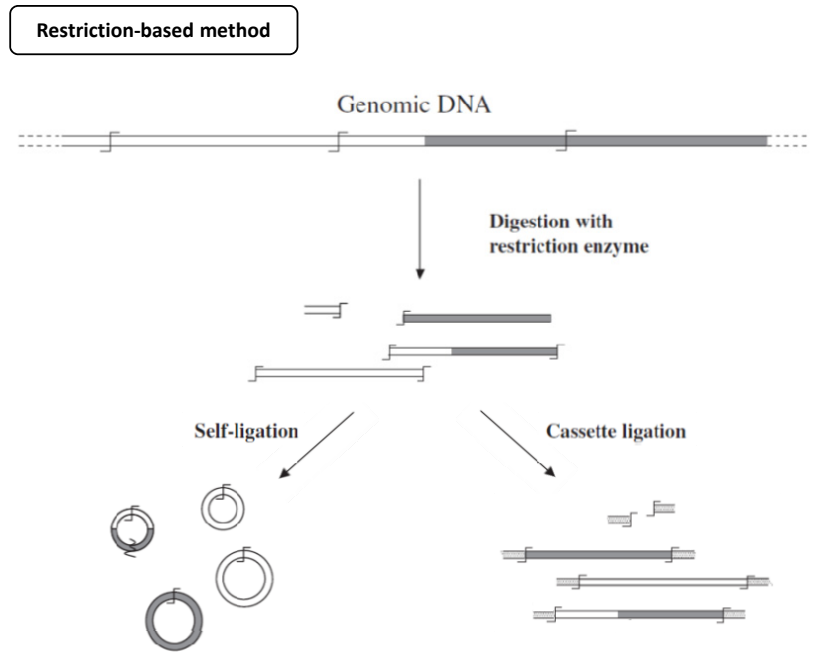
1.4 Technical details on used technologies

Given that the DNA walking and sequencing technologies were used in the present work, more technical details are provided in this section.

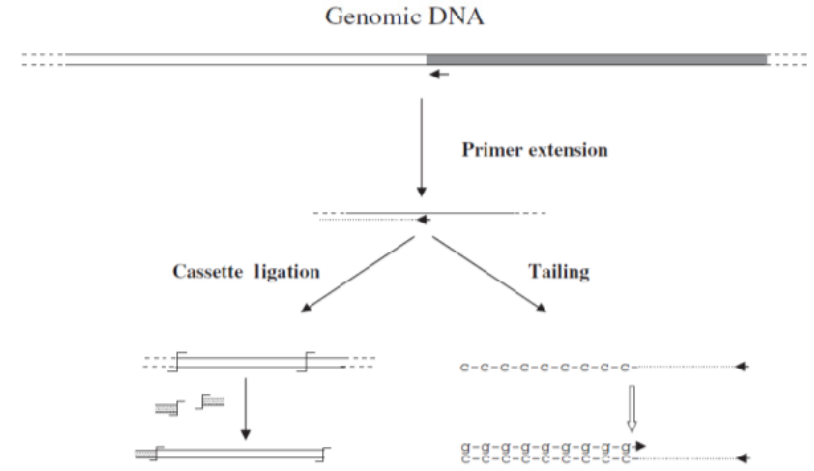
1.4.1 DNA walking

DNA walking, also called Genome walking, is a molecular technique allowing to identify unknown nucleotide sequences adjacent to already known DNA regions in any given genome by obtaining a final PCR amplification in which a specific primer to the known sequence is coupled with a primer dictated by the DNA walking strategy used. To characterize the final PCR amplicons, they could subsequently be sequenced.
Since the development of the first DNA walking strategies, several improvements have been suggested. Classically, all these methods are classified according to the three main following categories, based on the characteristics of their first step (Figure 1.6) (Leoni *et al.* 2011; Volpicella *et al.* 2012a; Volpicella *et al.* 2012b).
First, the restriction-based methods involve a digestion of the genomic DNA using appropriate restriction enzymes targeting sites close to the junction between the known and unknown sequences (Figure 1.6A). The obtained restriction fragments are then either self-circularized or ligated to DNA cassettes, named respectively inverted-PCR methods and cassette PCR methods (Leoni *et al.* 2011 and reference therein; Volpicella *et al.* 2012a; Volpicella *et al.* 2012b).
Second, the extension-based methods are defined by the extension of a sequence-specific primer (Figure 1.6B). The resulting single-strand DNA is subsequently either ligated to a DNA cassette or 3'-tailing (Leoni *et al.* 2011 and reference therein; Volpicella *et al.* 2012a; Volpicella *et al.* 2012b).
Third, the primer-based methods require the coupling of various combinatorial (random or degenerate) primers to known sequence specific primers according to various PCR strategies (Figure 1.6C) (Leoni *et al.* 2011 and reference therein; Volpicella *et al.* 2012a; Volpicella *et al.* 2012b).

A


Restriction-based method
Genomic DNA
Digestion with
restriction enzyme
Self-ligation
Cassette ligation

B


Extension-based method
Genomic DNA
Primer extension
Cassette ligation
Tailing

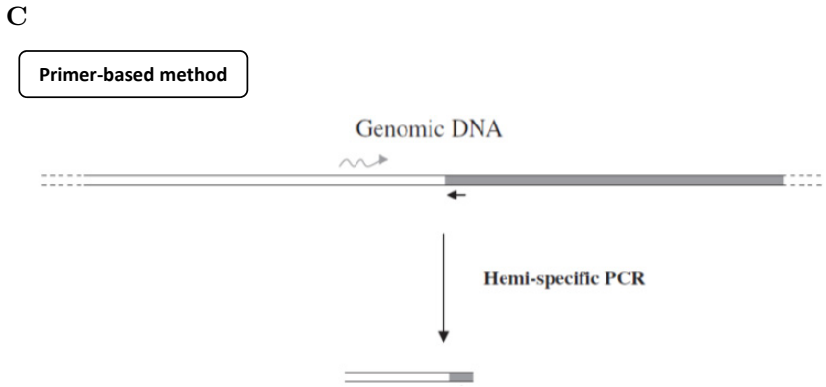


Figure 1.6: Main DNA walking strategies (adapted from Leoni *et al.* 2011).

1.4.2 Sequencing technologies

First generation sequencing Sanger method represents the first generation sequencing. It consists in cycle sequencing reactions including, in each cycle, the template denaturation, the target-specific primer annealing and the primer extension. The stochastic addition of a fluorescently labelled dideoxynucleotides (ddNTPs), unique for each kind of nucleotide, leads to the interruption of the primer extension step (Figure 1.7). Following to the high-resolution electrophoresis separation of the labelled products using a capillary-based polymer gel, the nucleotide identity of the sequence is determined by the specific emission spectra detected after laser excitation of the fluorescent labelling. This system is able to reach a read-length of around 1 Kbp with an accuracy of 99.999% (Sanger *et al.* 1977; Shendure and Ji 2008).

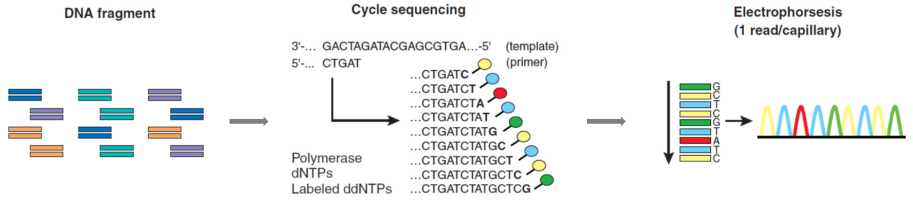


Figure 1.7: Schematic representation of the Sanger sequencing technology (adapted from Shendure and Ji 2008).

Second generation sequencing The second generation sequencing, also called NGS, is currently the most used sequencing technology. One of the leading

platform is the Illumina technology, which is developed on the "sequencing by synthesis" principle and is available since 2006. The DNA library preparation requires that sheared DNA, which a specific size range (*e.g.*, 100, 125, 150 or 300 bp) depending on the used device, is ligated to adaptors at both ends to be isothermally amplify by clonal bridge PCR using Bst polymerase. Indeed, each initial DNA template should be clonally pre-amplified as the sensitivity of the detection system is insufficient to detect the incorporation of only one nucleotide (Figure 1.8). To this end, forward and reverse adaptor-specific primers are immobilized, via a flexible linker at their 5'-end, on the surface of a solid glass substrate, implying that, for each initial DNA fragment, all generated amplicons (approximately 1000 clonal copies after 35 cycles) remain fixed near their original amplification point to form a clonal cluster physically distinguishable from the other clusters. Several millions of these clusters are produced within each lane on a single flow-cell. However, the cluster yield from short molecules is higher than from long molecules. To determine their sequence, adaptor-specific primers are hybridized on the single strand of linearized amplicons allowing the extension reaction with a modified polymerase and fluorescently labelled deoxynucleotides (dNTPs), specific for each kind of nucleotides. Due to a reversibly terminating moiety at the 3'-hydroxyl position of the nucleotides, only one labelled nucleotide is incorporated in each sequencing cycle (single-base extension). Following to the imaging acquisition, both fluorescent labels and terminating moiety from nucleotides are chemically cleaved before to start a new sequencing cycle. Several Illumina sequencing instruments have been commercialized and present different features. The accuracy of all these sequencers, decreasing toward the end, especially on the longer reads, is at least superior at 99.9% (Shendure and Ji 2008; Thudi *et al.* 2012; Buermans and Dunnen 2014; Anandhakumar *et al.* 2015; Illumina 2016).

Third generation sequencing Unlike to the first and second generation sequencing, the third generation sequencing, released by Pacific Biosciences in 2011, does not require any pre-amplification of the initial DNA template since this technology uses detection systems sufficiently sensitive to detect the extension of only one molecule. For the library preparation of this "sequencing by synthesis" technology, DNA fragments, which their size can reached up to 60 Kbp, are ligated at both ends by an identical Single-Molecule Real-Time (SMRT) loop adaptor presenting a hairpin structure to generate a circular library molecule (Figure 1.9). The active polymerization complex is formed by the annealing of the primers to the adaptor single-strand regions as well as the subsequent binding of DNA polymerase, a modified phi29 characterized by a low 3'→5' exonuclease activity, a high fidelity, no GC bias, an extreme processivity and a good strand displacement. Via the biotin-streptavidin affinity, the active polymerization complexes are then individually attached to the

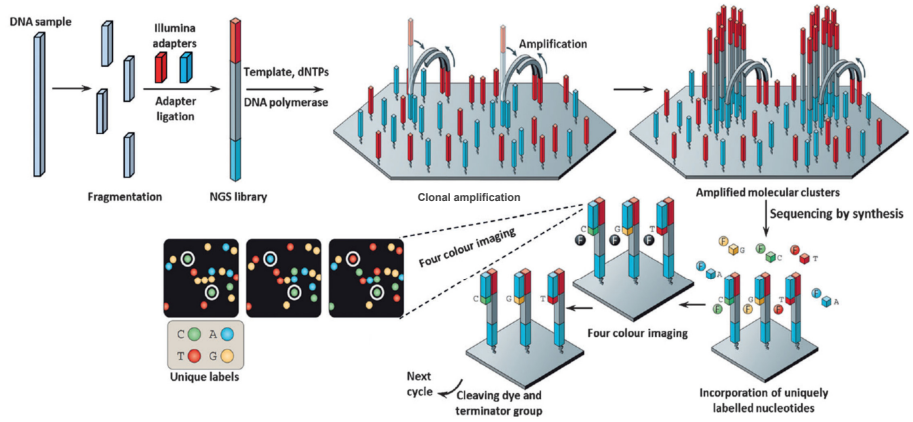


Figure 1.8: Schematic representation of the Illumina® sequencing technology (adapted from Anandhakumar *et al.* 2015).

bottom glass surface of one of the 150 000 zero-mode-waveguide (ZMW) wells of 50 nm-wide from one SMRT cell. During the sequencing, the extension of each fluorescently labelled nucleotide, distinct for each species, is recorded in real-time in each ZMW well. To this end, the light signal from the integrated fluorescent nucleotide, generated in all ZMW wells individually excited by green and red lasers through the glass surface, is detected by a confocal system. As only the bottom 30 nm of the ZMW wells is illuminated, only the incorporated nucleotides are excited while the free nucleotides above do not contribute to the observed signal. To pursue the molecule extension, the phosphate chain between the incorporate nucleotide and its fluorescent label is cleaved. According to the DNA fragment size, the sequence reading is carried out one (long molecules) or several (short molecules; Consensus Circular Sequencing (CCS)) times, which increases base accuracy. Following to the incorporation of the nucleotide, its terminal phospholinked fluorescent label is cleaved to allow the subsequent extension reaction. The current platform supplied by Pacific Biosciences (PacBio RSII) shows a consensus accuracy of at least 99.999% (Metzker 2010; Schadt *et al.* 2010; Buermans and Dunnen 2014; Rhoads and Au 2015; Pacific Biosciences 2016).

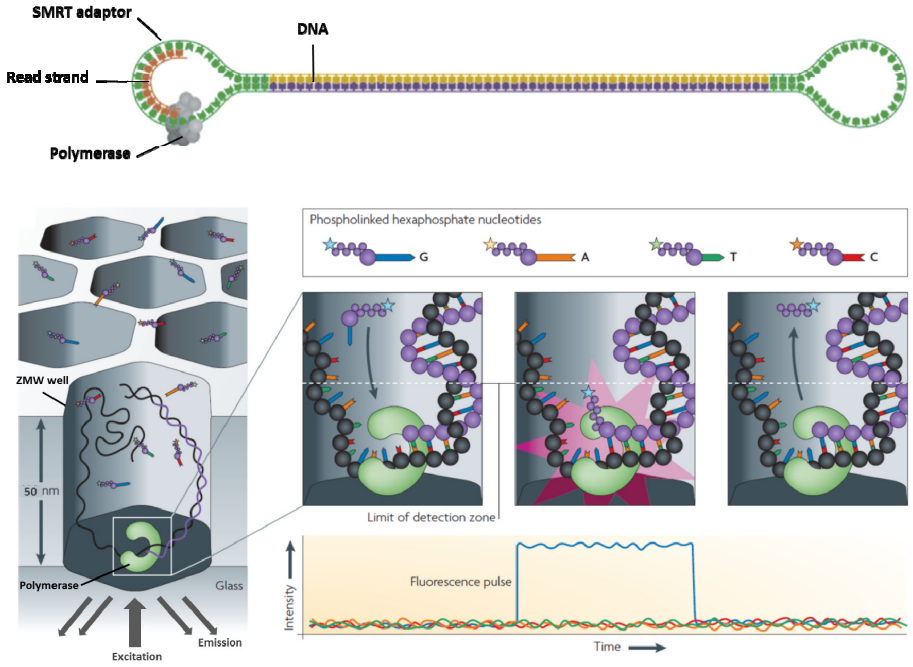


Figure 1.9: Schematic representation of the Pacific Biosciences® sequencing technology (adapted from Metzker 2010 and Rhoads and Au 2015).

Chapter 2

Objectives and outline

2.1 Rationale of the thesis

This PhD was carried out in collaboration with Prof. Dieter Deforce of the Laboratory of Pharmaceutical Biotechnology at the Ghent University (UGent), Dr. Nancy Roosens of the Platform Molecular Biology and Biotechnology (PBB) at the Scientific Institute of Public Health (WIV-ISP), Dr. Philippe Herman of the Biosafety and Biotechnology Unit (SBB) at WIV-ISP, Prof. Marc De Loose of the Technology and Food Sciences Unit (T&V) at the Institute for Agricultural and Fisheries Research (ILVO) and Dr. Isabel Taverniers of T&V at ILVO.
The research question of this PhD project, called UGMMONITOR, arises from the Belgium National Reference Laboratories (NRL) responsible for the enforcement of EU GMO legislations at the national level. Due to the ongoing and further expected expansion of GMO in terms of number, diversity and cultivated areas, the implementation of these legislations, including the control by the enforcement laboratories, is becoming even more and more complex, especially regarding EU unauthorized GMO. There is thus an urgent need, at national as well as international levels, to develop news tools for the enforcement laboratories to detect unauthorized GMO.
In order to cope with this challenge, this project aims to provide a rational and efficient GMO detection system, especially for unauthorized GMO. To this end, information regarding developed transgenic crops was collected, in a first time, to be able to develop new approaches, in a second time, that strengthen the current GMO detection strategy. To reach these objectives, the rice was chosen as a study case since it is one of the leading staple crops and also a model plant

in molecular biology.
It is important to notify that the present work is focussed on GM crops corresponding to Directive 2001/18/EC, excluding thus the GM animals and biotech events generated by new plant breeding techniques.

2.2 Outline of the thesis

This work is composed of three main parts. The outline of the thesis with the different chapters is presented in Figure 2.1. First, a general introduction about the current GMO detection system and the related needs as well as the objectives and outline of the thesis are respectively provided in the chapters 1 and 2. Second, the strategies developed to strengthen the current GMO detection system are detailed through six chapters. Chapter 3 is an overview of GM rice developed around the world allowing to highlight clue information to assess and improve the current GMO detection system. With the help of the information collected in chapter 3, a strategy, based on DNA walking anchored on elements commonly found in a broad spectrum of GMO, was developed and validated in chapters 4, 5 and 6. In chapter 7, this strategy was improved by using the NGS technology to increase its throughput and to simplify its workflow. In parallel, the potential to detect GMO with the NGS technology by WGS was investigated in chapter 8. Finally, according to all generated results and related discussions, the general conclusions of this PhD are given in chapter 9. In addition, the international context of this work, its relevance and its perspectives are presented in Chapter 10. It should be noted that the chapters 3 to 8 are composed of the full published/submitted peer-reviewed publications.

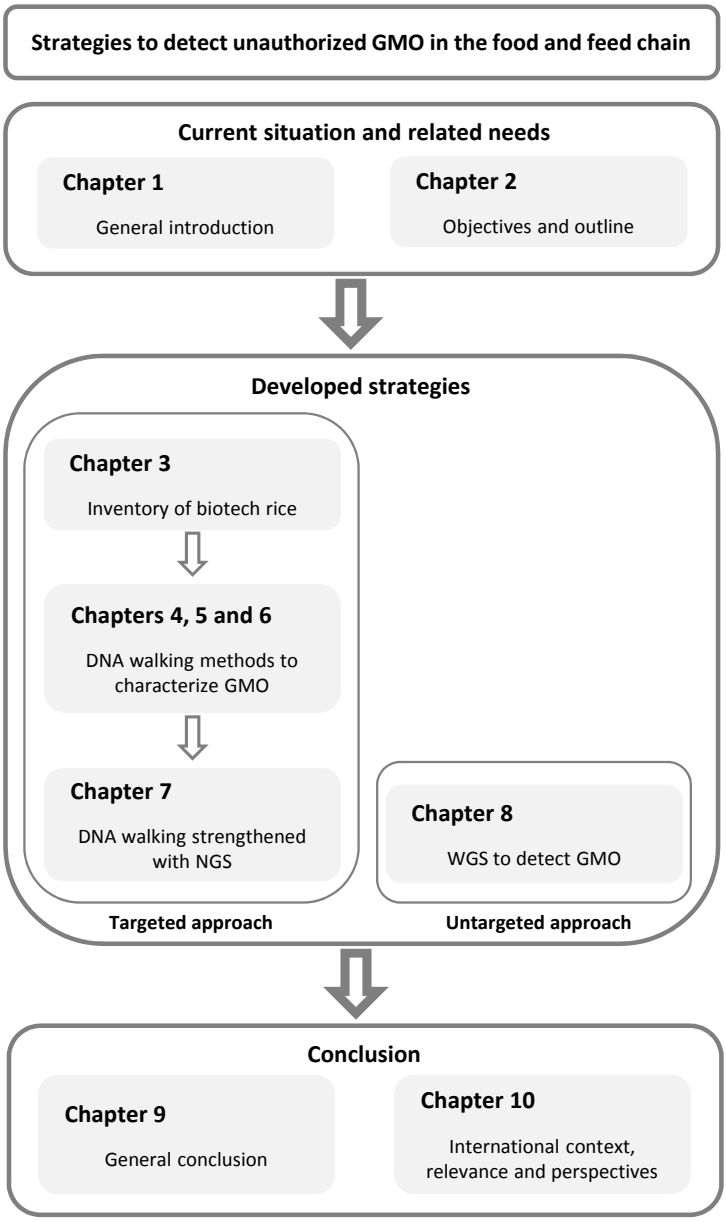


Figure 2.1: Schematic outline of the thesis.

Chapter 3

Biotech rice: current developments and future detection challenges in food and feed chain

In this chapter, an overview of GM rice developed around the world was realized to collect crucial information allowing to assess and strengthen the current GMO detection system. This research emphasised that the p35S and tNOS elements, commonly found in EU authorized GMO, are estimated to be present in around 93% of GM rice, as well as the t35S pCAMBIA element, absent from EU authorized GMO, is present in approximately 30% of GM rice. This last element could therefore be useful at this moment in terms of discrimination between the EU authorized and unauthorized GMO. In addition, as all these elements are derived from natural organism (CaMV for p35S and t35S pCAMBIA and *A. tumefaciens* for tNOS), the current qPCR GMO detection system used by the enforcement laboratories needs to be coupled to additional approaches (*e.g.* DNA walking and NGS) to be able to indubitably prove the presence of GMO via the characterization of the transgene flanking regions and the associations of elements that are typically found in transgenic constructs.

This chapter was previously published as:
M.-A. Fraiture, N. H. C. Roosens, I. Taverniers, M. De Loose, D. Deforce, and P. Herman (2016a). "Biotech rice: current developments and future detection challenges in food and feed chain". *Trends in Food Science & Technology* 52, pp. 66–79. DOI: 10.1016/j.tifs.2016.03.011

Authors' contributions

M.-A. Fraiture collected all information, analysed them and drafted the manuscript. P. Herman and N. H. Roosens helped to analyse data and drafted the manuscript. I. Taverniers, M. De Loose and D. Deforce drafted the manuscript. All authors read and approved the final manuscript.

Abstract

Background To improve agricultural practices and the food/feed security, plant breeding techniques were developed, including transgenesis commonly using *Agrobacterium tumefaciens* or biolistic technologies. To guarantee the traceability of GMO in food/feed chain and the consumer's freedom of choice, regulatory frameworks were established in many countries around the world, such as in Europe. Their implementations, including detection systems usually based on qPCR, are becoming complex and expensive regarding the number of analysis to perform. Moreover, the dispersion of publicly available information about developed GMO prevents to accurately estimate the efficiency of the standard detection system applied to unauthorized GMO.

Scope and approach To illustrate this problem, the case of rice, one of the leading staple crops, was investigated. An overview of worldwide developed biotech rice generated by transgenesis was thus conducted, based on 1067 peer-reviewed publications, and analysed regarding *inter alia* their expressed genes of interest and the corresponding traits, their transformation processes and the elements composing their transgenic cassettes. From this work, the power and weakness of the standard detection system, notably used by the European enforcement laboratories, are evaluated. To strengthen this system, especially with unauthorized GMO, additional strategies are suggested. Moreover, given the growing interest for biotech rice produced by new plant breeding techniques, related challenges for their detection are discussed.

Key findings and conclusions According to all collected information, suitable detection strategies, combining qPCR to additional technologies (*e.g.*, DNA walking and NGS), are proposed to cover most of inventoried biotech rice. The present approach, including the data centralization to subsequently suggest appropriated detection strategies, can be extended to biotech events from different species.

3.1 Introduction

To cope with the challenge of increasing the crop production, the evolution of current agricultural practices was envisaged via the support of conventional breeding methods by contemporary approaches. Therefore, various strategies have been used to develop biotech crops, such as transgenic crops, also named genetically modified (GM) crops or genetically engineered crops, opening new possibilities to reach the expected crop nutritional necessity and to ensure food security (Ahmad *et al.* 2012; He *et al.* 2014). As observed for several crops, genetic engineering had *inter alia* been applied to rice (*Oryza sativa*), currently one of the most important cereal crops that is cultivated in many countries through the world. For more than 3.5 billion people, rice is a staple food providing more than 20% of their daily calories intake, especially in developing countries. In addition, this crop is also intended to feed animals. In 2014, the annual production of paddy rice was estimated at 741.3 million tons (corresponding to 494.4 million tons of milled rice) where the majority was grown in Asia (674.4 million tons; 91%), mainly in China (208.1 million tons), India (155.5 million tons), Indonesia (70.6 million tons), Bangladesh (52.4 million tons) and Vietnam (44.9 million tons). The rest of paddy rice is harvested in Africa (27.6 million tons), South America (24.8 million tons), North and Central America (12.9 million tons), Europe (4.1 million tons) and Oceania (0.9 million tons) (Khush 2013; Fraiture *et al.* 2014; FAO 2015). Therefore, even if no biotech rice is nowadays cultivated worldwide at a commercial scale, its potential significance is clearly obvious in the near future (De Steur *et al.* 2014).

With the aim to guarantee the traceability on the markets as well as the freedom of choice to the consumers, several genetically modified organism (GMO) labelling systems have been established in several countries with a threshold varying from 0 to 5%. The labelling is either mandatory (*e.g.*, in Australia, Brazil, Chile, China, EU, India, Indonesia, Israel, Japan, Philippines, Russia, Saudi Arabia, South Korea, Taiwan and Thailand) or voluntary (*e.g.*, in Argentina, Canada and USA). On the European Union (EU) market, commercialised food and feed products containing at least 0.9% of EU authorized GMO have to be labelled to guarantee the freedom of choice of the consumers while the zero tolerance is applied on unauthorized GMO (Davison 2010; Kamle and Ali 2013). The implementation of these legislations is mainly carried out using real-time PCR (qPCR) technology, allowing to detect, identify and quantify GMO (Fraiture *et al.* 2015c). Three main steps are traditionally followed in GMO routine analysis. First, the presence of GMO is detected by screening. It includes simplex or multiplex methods targeting the most common transgenic elements, such as p35S (35S promoter from Cauliflower mosaic virus (CaMV)) and tNOS (nopaline synthase terminator from *Agrobacterium tumefaciens*). In

addition, some more discriminative markers are used to reduce the number of subsequent identification. In case of positive signals, the identity and the quantity of GMO are afterwards determined via event-specific methods. If the signals observed during the screening step do not correspond to any of the authorized GM events, the presence of unauthorized GMO is then suspected (Broeders *et al.* 2012a; Broeders *et al.* 2012b).

Given the ongoing and further expected expansion of GMO in terms of number, diversity and cultivated areas, the implementation of labelling legislations is becoming even more complex. Furthermore, the presence of some GM events could be prohibited or not according to the jurisdiction in reason of the asynchronous authorisations between many countries. Moreover, unlike the present commercialised GM crops which have been mainly developed by American and European companies, more and more GMO produced by national technology centres in developing countries are intended for local consumption. Consequently, these GM crops will probably not be submitted for EU approval. Therefore, the frequency of unauthorized GMO on the EU market is likely to significantly increase due to accidental contamination of non-transgenic raw material and processed food/feed matrices. In addition, unauthorized GMO concerns also GM crops that are currently unknown to the competent authorities (Stein and Rodríguez-Serezo 2009; Broeders *et al.* 2012a; Holst-Jensen *et al.* 2012; Parisi *et al.* 2016). The complexity of this problematic is particularly well illustrated by GM rice for which no events are nowadays authorized on the EU market. First, the problem of asynchronous approvals has been encountered with products originating from the USA. More precisely, the herbicide tolerant LLRICE601, in 2006, and LLRICE62, in 2007, both produced by Bayer CropScience and exclusively authorized in the USA, were identified in commercial rice matrices. Second, the insect resistant Bt Shanyou 63 and KeFeng-6 were found in food products originating from China in 2006 and 2010, respectively. These GM rice, produced by Asian research centres, were probably accidentally spread. Besides, the illegal propagation of seeds from field trials as well as their planting by Chinese farmers have been reported (Wang and Johnston 2007; Ruttink *et al.* 2010b; Wang *et al.* 2011b; Fraiture *et al.* 2014). Finally, the presence of unknown GM rice was also reported by the RAPID Alert System Database (The European Commission 2015a), allowing notably to notify the detection of unauthorized GMO on the EU market, such as in 2010 and 2011 in products imported from China, contaminated in all likelihood by accident (Figure 3.1). Furthermore, still according to the RAPID Alert System Database, practically 50% of analysed food/feed samples between January 2012 and May 2015 contained unauthorized GM rice, including Bt63 and GM Basmati rice, imported from Asia, mainly China. Due to the high level of EU unauthorized GM rice, the EU commission has notably decided to implement “Emergency measures regarding unauthorized genetically modified

rice in rice products originating from China and repealing Decision 2008/289/EC" (The European Commission 2011b).

The success of the qPCR strategy is directly linked to the availability of information on the targeted sequences, such as from transgenic elements (element-specific markers), association of elements (construct-specific markers) or from junctions between the transgenic cassettes and the plant genomes (event-specific markers). In addition, the quantification of identified GM events requires the availability of Certified Reference Materials (CRM) and taxon-specific methods (Broeders *et al.* 2012b; Holst-Jensen *et al.* 2012). Conversely to EU authorized events for which all these data are accessible in the Compendium of reference methods for GMO analysis, the identification of EU unauthorized events by qPCR strategy could be difficult (Fraiture *et al.* 2016b). Indeed, the lack of centralized information about unauthorized transgenic crops is problematic. For instance, for GM rice, only three herbicide tolerant (LLRICE601, LLRICE06 and LLRICE62), four insect resistant (Bt63, Huahui-1, Tarom molaii and GM rice 101096), one fungi resistant (GM rice 101097), six multiple biotic stress resistance (NIA-OS002-9, NIA-OS012-8, NIA-OS004-8, NIA-OS003-1, NIA-OS005-3 and NIA-OS006-4), two abiotic stress resistance (SeC and As-d) and two rice seed edible vaccines against Cedar pollen allergy (7Crp#10 and OsCr11) are currently reported in publicly available GMO databases. Those GM rice lines are mainly described according to the expressed genes of interest and the corresponding traits, the transformation methods used, the transgenic elements contained in the vectors and the related biosafety information (CERA 2012; Genetic Rights Foundation 2015; GMO Compass 2015; SCBD 2016; The European Commission 2016).

Therefore, in this review, after a brief description of the biotech rice history including the transformation technologies used, an overview of transgenic rice events developed through the world was built on the basis of more than 1000 available peer reviewed publications. All data were collected and analysed regarding notably the genes of interest expressed and their origins, the vectors and transgenic elements that composed them, the transformation technologies used and the status of the transgenic rice (laboratory development stage or field trial). In this way, the information on biotech rice centralized in this review can be used to complete publicly available databases as well as to develop and strengthen GMO detection strategies.

3.2 Biotech rice history

As efficient genetic engineering techniques and its genome sequence estimated at 430 Mb are available, rice represents currently a key crop model to develop biotech plants (Kathuria *et al.* 2007). Since most of the biotech rice are generated

using transgenesis techniques to date, this section concerns essentially transgenic rice, also called GM rice. However, due to the rise of biotech rice originated from new plant breeding techniques (NPBT), the sub-section 3.2.4 is entirely devoted to it. The rice crops generated by these promising alternatives are also named NPBT rice.

3.2.1 Transformation technologies

In 1988, the first transgenic rice plants were successfully developed by electroporation-mediated or polyethylene glycolmediated protoplast transformation methods (Figure 3.1). Based on electroporation and polyethylene glycol-mediated technologies, the fertility recovery of transgenic rice was then respectively reported in 1989 and 1990 (Hiei *et al.* 1997; Kathuria *et al.* 2007). In 1991, the biolistic transformation method, also named particle or microprojectile bombardment, was successfully used to generate transgenic rice plants. This direct DNA delivery system is often characterized by integration in multiple copies of transgenes into the recipient genome. Moreover, the transgenic rice plants present the risk of an unstable and aberrant expression of the gene of interest because of its likely fragmentation and rearrangement induced by this transformation technology (Hiei *et al.* 1997; Hoque *et al.* 2005; Kathuria *et al.* 2007).

Although *Agrobacterium tumefaciens*-mediated transformation is used since the 1980s, its application, initially limited to dicotyledonous plants, was slightly modified to be adapted on monocotyledonous plants such as rice (Hiei *et al.* 1997; Hoque *et al.* 2005). In 1990, the first transgenic rice calli were produced after application of the *Agrobacterium*-mediated transformation method. Afterwards, transgenic rice plants were regenerated from *Agrobacterium*-transformed calli (Hiei *et al.* 1997; Kathuria *et al.* 2007). To improve this strategy, several factors affecting the transformation efficiency have been studied. For instance, the activation of the T-DNA transfer process is promoted via the addition of phenolic compounds, such as acetosyringone, a potent inducer of virulence genes participating to the recognition of the host by *Agrobacterium*. Moreover, the bacterial strains and vectors used are decisive, particularly with recalcitrant rice genotypes such as some Indica cultivars. The choice of the competent rice tissue (age, cell type, cell cycle stage) has been also highlighted as a key parameter. In addition, the culture conditions are determining in the success of the transformation. Indeed, instead of kanamycin and G418 antibiotics, hygromycin B is preferably used as an antibiotic during the selection step of transformed rice as no natural resistance to hygromycin is present in rice. In addition, this antibiotic does not impact the transgenic rice regeneration and fertility (Hiei *et al.* 1997). Even if food/feed and environmental safety

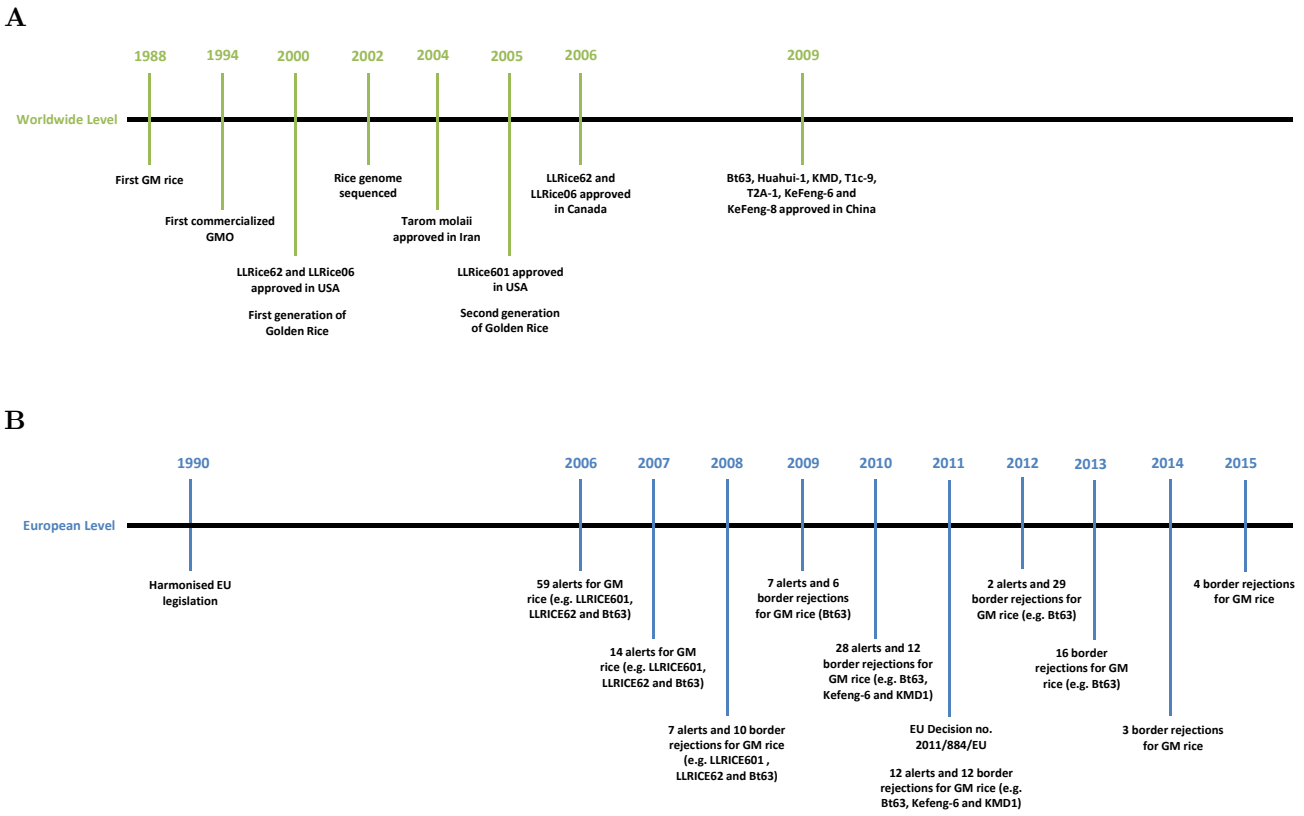


Figure 3.1: Timeline of transgenic rice history at the worldwide level (A) and of transgenic rice impact on the EU market (B).

concerns were raised in the past, to date only theoretical assumptions without evidence-based arguments have demonstrated any significant impact of antibiotic resistance marker genes related to the potential risk of horizontal gene transfers from transgenic plants to neighbouring bacteria and plants as well as a potential increase of antibiotic resistance among consumers. Anyway, partly due to regulatory requirements in the EU for phasing-out antibiotic resistance marker genes such as npt II in GMO, several strategies have been developed to remove microbial selection markers from transgenic plants (Hiei *et al.* 1997; Breyer *et al.* 2014). One alternative is to replace the microbial selection markers by vegetal selection markers that naturally confer an antibiotic resistance or by herbicide tolerance genes. In addition, other marker genes can be used instead of the classical microbial markers via three main strategies. First, the positive selection marker approach confers a metabolic or developmental advantage to the transformed cells without implying the death of untransformed cells. Second, just the opposite, the negative selection marker approach leads to a metabolic or developmental disadvantage in the transformed cells. Third, the reporter genes allow to visually select transgenic plants (Hiei *et al.* 1997; Breyer *et al.* 2014). In order to remove definitely the selection markers in the selected transgenic plants, several techniques, based on the site-specific recombination (*e.g.*, Cre/loxP and MAT), the intrachromosomal homologous recombination, the intra-genomic transpositional mechanism (*e.g.*, Ac/Ds transposable element) or the segregation via co-transformation, have been performed. This last approach is the simplest and widely used method that involved the introduction of two T-DNA regions, one with the gene of interest and one with the selection marker gene, allowing to subsequently generate transgenic plants presenting only one T-DNA by sexual crossing in successive generations (Breyer *et al.* 2014).

Unlike to direct DNA delivery transformation system, *Agrobacterium* strategy generates transgenic plants presenting a more predictable pattern of integration and a non-rearranged segment of DNA being inserted into the genome at a low copy number (Hoque *et al.* 2005). However, although the *Agrobacterium tumefaciens* strategy has generated progress in crop biotechnology, the less-known rhizosphere bacterium *Ensifer adhaerens* represents a potential alternative as it has been used to transform plants including Arabidopsis and potato. Besides, this approach has recently been applied to rice (Wendt *et al.* 2012; Zuniga-Soto *et al.* 2015).

3.2.2 Currently developed transgenic rice

Data collection As observed for other transgenic crops, the number and the diversity of transgenic rice is significantly increasing with time. Therefore, a large inventory of the currently developed transgenic rice has been carried out allowing

to provide crucial information to enforcement laboratories in order to detect and identify them. This kind of strategy was previously suggested notably regarding the detection of unauthorized GMO (Ruttink *et al.* 2010b). Based on the analysis of available online databases (Biosafety Clearing-House, Biosafety Scanner, CERA, GMO Compass, GMO register) as well as 1067 peer-reviewed articles published in 242 different journals, a list of transgenic ricewas drawn up. Initially, the scientific literature was reviewed using the Scopus database (Elsevier 2016) to cover the period from January 1991 to October 2015 with the coupled keywords "transgenic rice" (the search with the keywords "Genetically Modified Rice" or "GM Rice" provided far fewer data). In a second step, the relevant scientific literature was selected based on several keywords, including "herbicide", "insect", "resistance", "tolerance", "stress", "abiotic", "biotic", "bacteria", "virus", "fungi", "pathogen", "drought", "salt", "salinity", "cold", "disease", "deficiency", "heat", "metal", "chill", "oxidative", "iron", "water", "submergence", "yield", "grain", "biomass", "size", "fortification", "nutritional", "texture", "carotene", "vaccine", "allergen", "pharmaceutics", "health", "therapy", "inflammation", "immunogenicity", "microbe", "improvement" and "development". Furthermore, all pertinent peer-reviewed publications adjacent to the selected scientific literature were analysed (Additional file 1: Inventory of transgenic rice).
All the identified transgenic rice were recorded in an excel file according to their trait (one kind of trait per excel sheet), including herbicide tolerance, insect resistance, bacteria resistance, fungi resistance, virus resistance, multiple biotic stress resistance, abiotic stress resistance, abiotic and biotic stress resistance, grain yield improvement, nutritional quality improvement, pharmaceutical production and other innovations. Each transgenic rice was described in so far as possible in terms of rice variety, gene of interest expressed, donor organism of the gene of interest, vector used, transgenic elements used, transformation method used, knowledge level (KL) classification (ENGL *ad hoc* working group on "unauthorised GMOs" 2011), status of the transgenic rice (laboratory development stage or field trial), year of publication, developer country and references of the related peer-reviewed publications. Moreover, if necessary, more details on the expressed traits were added (Additional file 1).

Geographical distribution In order to have a worldwide picture of the developed transgenic rice, the geographical distribution of all inventoried developments described in peer-reviewed publications was investigated. Although some transgenic rice are developed in American (11.5%), European (8.9%), Oceanian (1.1%) and African (1%) laboratories, the majority of these research and development (R&D) studies is performed in Asia (77.5%), especially in China (47.8%) and Japan (20.2%) (Figure 3.2).

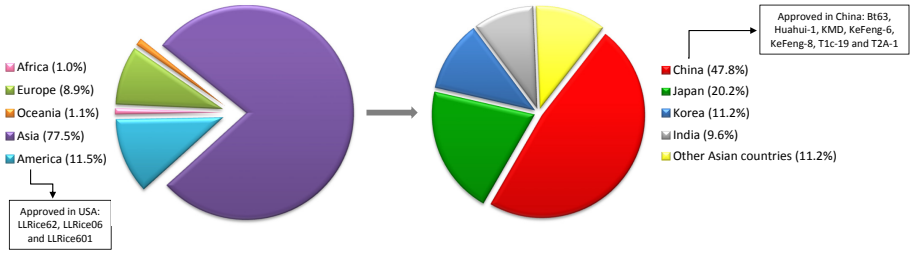


Figure 3.2: Contribution (%) of the five continents to the transgenic rice development in term of peer-reviewed publications, including a zoom on the Asian countries.

Developed traits Among all inventoried transgenic rice from all analysed articles and databases, most of the them have only been tested at the laboratory level (75.4%), but a significant amount (20.6%), approved or not for commercial cultivation, have already been subject to field trials. It should be noted that these data were not available for 4% of the listed transgenic rice. Concerning the traits identified in all inventoried transgenic rice, herbicide tolerance (6.1%) and insect resistance (20.3%), as found in the first transgenic rice lines approved for commercialisation, are still observed. Nevertheless, these traits are less developed in the course of time (Figure 3.3 and Additional file 2A-B). In fact, progressively, more and more transgenic rice present at the R&D step display a high diversity of new traits such as the resistance to biotic stress (17.2%), in targeting one specific category of phytopathogens (bacteria (5.8%), fungi (7.1%) or virus (4.4%)) as well as more than one category of phytopathogens (2.1%) (Figure 3.3). Moreover, the development of transgenic rice resisting to several abiotic stress (30.2%, *e.g.*, salinity, drought, cold and heavy metal) is increasing over time (Figure 3.3 and Additional file 2C). Some transgenic rice present also a resistance to both biotic and abiotic stresses (1.5%). In addition, some R&D centres focus their energy to improve directly the yield of grains (4.5%) (Fig. 3). The introduction of new traits in rice could also improve the consumer's quality of life in terms of nutrition and health. To this end, the modification of the rice grain nutritional composition (8.5%), including micronutrients (*e.g.*, iron, zinc, manganese), vitamin (*e.g.*, A and B9/folate) and essential amino acids for biofortification, is an auspicious strategy (Figure 3.3). This is actually the case with the folate biofortification which could prevent some birth defects. To reach the minimum daily level in folate, between 137 g and 281 g of biofortified rice have to be consumed (De Steur *et al.* 2014). A second well-known example is the Golden Rice Project, rewarded with the "Patents for Humanity Award" in 2015, with the aim to develop transgenic rice with β-carotene biofortified grain in order to struggle against the vitamin A

deficiency leading notably to blindness in children (Golden Rice Humanitarian Board 2015). Following to several improvements from the first generation of golden rice, the second generation of golden rice, transformed with a maize (*Zea mays*) phytoene synthase (psy) gene and a bacterial (*Erwinia uredorova*) phytoene desaturase (crt1) gene, yields a higher β-carotene concentration. The consumption of 77 g and 122 g of golden rice, by respectively young children and pregnant women, provides the minimum daily intakes in vitamin A (Paine *et al.* 2005; De Steur *et al.* 2014). After three seasons of field trials in the Philippines, the grains from the Golden Rice event R (GR2-R) presented the expected β-carotene level. However, as the average yield was inferior to the local varieties, new assays will be carried out with other Golden Rice versions such as GR2-E (James 2014). In addition, transgenic rice are developed in the pharmaceutical field (9.6%) to produce cytokines, vaccines, antibodies, albumin or other therapeutic proteins (Figure 3.3). Molecular farming of rice is envisaged to treat several critical health conditions such as allergy, autoimmune disorders, infectious diseases or even Alzheimer disease (Azegami *et al.* 2015). Among the 1673 transgenic rice listed in the current study, 413 of them are directly involved in fundamental research (*e.g.*, study of development, flowering, signalling) or in the implementation of new strategies such as optimisation of the transformation protocol and the production of marker-free transgenic rice.

In order to give rise to all these new properties in transgenic rice, a broad-spectrum of different gene families have been identified for each trait, including 16 gene families for herbicide tolerance, 23 gene families for insect resistance, 25 gene families for bacterial resistance, 41 gene families for fungi resistance, 13 gene families for virus resistance, 155 gene families for abiotic stress resistance, 29 gene families for grain yield improvement, 35 gene families for nutritional grain quality improvement and 62 gene families for drugs production (Table 3.1).

Transformation vectors Among all inventoried transgenic rice (1673 transgenic rice), approximately 200 different vectors were used for their transformation. The *Agrobacterium tumefaciens*-based method was predominantly used (75.7%; 1267 transgenic rice) compared to the biolistic one (13.5%) or other methodologies (2.6%). For 8.2% of the inventoried transgenic rice, the transformation method was not mentioned. Regarding more precisely the vectors used, most of inventoried transgenic rice (36.9%, 617 transgenic rice) were transformed using the family vector pCAMBIA. Based on all analysed articles published from 2001, the use of the pCAMBIA vector has increased over time to reach a plateau of around 30% from 2004 to 2013. Since 2014, the use of this vector has even increased as more than half of the described transgenic rice was transformed with a pCAMBIA cassette (Figure 3.4). It was already shown in 2007 that 30% of transgenic plants have been developed using vectors from the

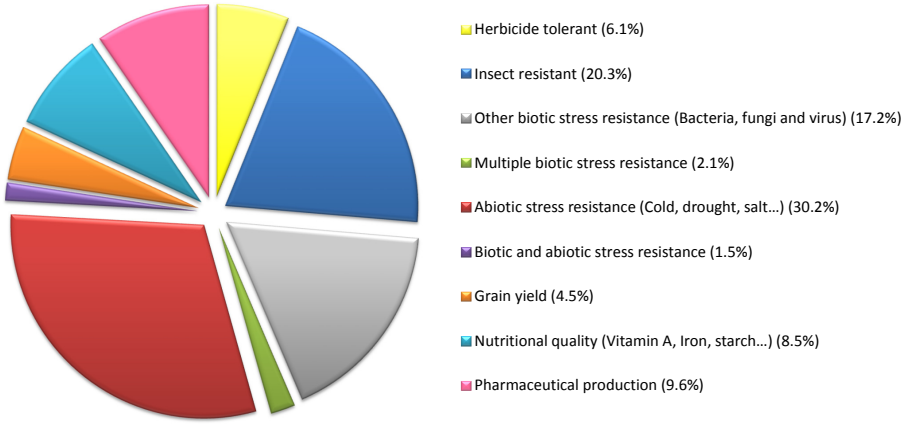


Figure 3.3: Observed traits (%) in the inventoried transgenic rice.

Table 3.1: Examples of gene families used to acquire new proprieties in transgenic rice.

Traits	Gene families	References
Herbicide tolerance	Acetolactate synthase (Als)	(Endo *et al.* 2012)
	Bialophos resistance (Bar)	(Christou *et al.* 1991)
	Protoporphyrinogen oxidase (PPO)	(Chun *et al.* 2013)
	5-enolphyruvylshikimate-3-phosphate synthase (EPSPS)	(Zhao *et al.* 2011)
	Glutathione S-transferase	(Hu *et al.* 2009)
	Cytochrome P450	(Ohkawa and Ohkawa 2002)
	atrazine chlorohydrolase	(Zhang *et al.* 2014b)
Insect resistance	Cry	(Breitler *et al.* 2004)
	GNA lectin	(Sudhakar *et al.* 1998)
	ASAL lectin	(Yarasi *et al.* 2008)
	Trypsin inhibitor	(Su *et al.* 2011; Yang *et al.* 2013)
	Potato proteinase inhibitor II (PINII)	(Bhutani *et al.* 2006)

Table 3.1 *continued*

Traits	Gene families	References
Bacteria resistance	Receptor kinase-like protein	(Afroz *et al.* 2012)
	WRKY transcription factor	(Liu *et al.* 2007)
	Thaumatin-like protein	(Shah *et al.* 2013)
	Ferredoxin-like amphipathic protein	(Tang *et al.* 2001)
Fungus resistance	Chitinase	(Sridevi *et al.* 2008)
	Pi-d2	(Chen *et al.* 2010)
	Defensin	(Kanzaki *et al.* 2002)
	AFP antifungal protein	(Coca *et al.* 2004)
	ACC synthase	(Seo *et al.* 2011)
	Cecropin	(Coca *et al.* 2006)
Virus resistance	Coat protein	(Sivamani *et al.* 1999)
	Spike protein	(Chaogang *et al.* 2003)
	Replicase	(Verma *et al.* 2012)
Abiotic stress resistance	Stress associated protein	(Ben Saad *et al.* 2012)
	D1-pyrroline-5-carboxylate synthetase	(Hien *et al.* 2003)
	Na^+/H^+ antiporter	(Ohta *et al.* 2002)
	Basic region/leucine zipper transcription factor	(Tang *et al.* 2012)
	Aquaporin	(Ayadi *et al.* 2014)
	Disulfide isomerase-like protein	(Chen *et al.* 2012b)
Grain yield	C4 pyruvate orthophosphate dikinase	(Fukayama *et al.* 2001)
	Cytosolic dehydroascorbate reductase	(Kim *et al.* 2013)
	Phytochome	(Garg *et al.* 2006)
Nutritional grain quality	Phytoene synthase (PSY)	(Ye *et al.* 2000)
	Carotene desaturase (crt)	(Ye *et al.* 2000)
	Ferritin	(Masuda *et al.* 2012)
	Waxy	(Terada *et al.* 2000)
	ScAcr3p	(Duan *et al.* 2012)
Pharmaceutical products	Major T-cell epitope	(Suzuki *et al.* 2012)
	Lactoferrin	(Humphrey *et al.* 2002)
	Lysozyme	(Humphrey *et al.* 2002)

Table 3.1 *continued*

Traits	Gene families	References
	Stilbene synthase	(Baek *et al.* 2014)
	Human serum albumin	(He *et al.* 2011)
	Major outer membrane protein	(Zhang *et al.* 2008)

pCAMBIA family (Fraiture *et al.* 2014). Although the pCAMBIA family vector has been mainly used, transgenic rice collected in the present bibliographical study were also transformed with other vectors such as pBI (3.7%), pIG (2.5%), pSB (2.1%), pANDA (2%), pCIB (1.8%), pBIG (1.7%), pGA (1.6%), pGPTV (1.1%) and pPZP (1%). For the rest of the inventoried transgenic rice, 30.5% of them presented a large variety of vectors only anecdotally used. It should also be mentioned that 15.1% of inventoried transgenic rice were transformed with vectors for which their name were not properly identifiable. However, for these transgenic rice with unnamed vectors, information about the elements composing their transgenic cassettes was available for 6% of them.

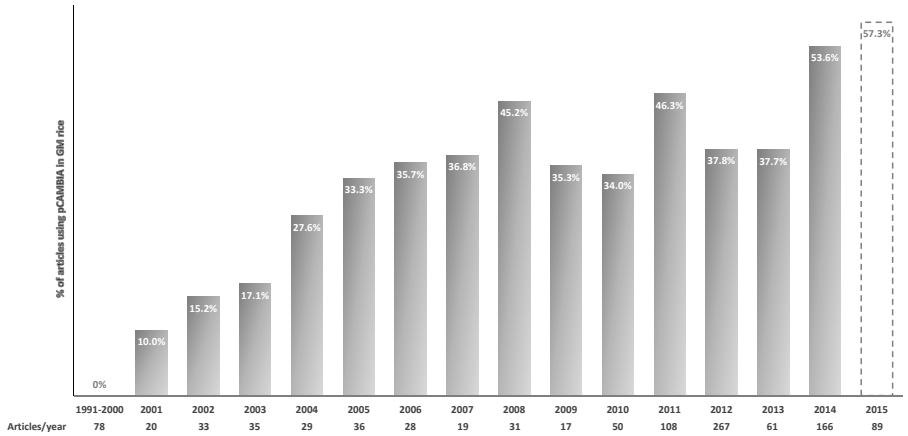


Figure 3.4: Percentage (%) of analysed peer-reviewed publications, by year, presenting the use of pCAMBIA vector to generate transgenic rice. The number of articles recorded per year is indicated below the corresponding period. Since the current year of 2015 is not entirely covered, the corresponding bar chart is built in dotted line.

Elements found in transgenic rice In order to collect more information about all inventoried transgenic rice (1673 transgenic rice), the elements composing

their transgenic cassettes were studied. Similarly to EU-authorized GMO (The European Commission 2016), the collected transgenic rice contain frequently the p35S promoter (69.8%, 1167 transgenic rice) and the tNOS terminator (62.6%, 1048 transgenic rice). Furthermore, some transgenic elements were often found in the collected transgenic rice, such as the UBI promoter (27.6%), the NOS promoter (8.4%), the ACT promoter (5.4%), the terminator t35S (38.5%; from pCAMBIA family vector or not), the PINII terminator (3.3%) and the GUS gene (18.4%). For the elements intended for the plant selection, the hygromycin B (54%) and kanamycin (9.5%) resistance genes (respectively *hph* and *npt II*) are mainly used. Among the 185 different transgenic elements identified, 175 different transgenic elements are only occasionally observed. It should also be mentioned that no transgenic element was identified for 11.1% of inventoried transgenic rice, including 9.1% of them transformed with an unnamed vector.

3.2.3 Commercialization of transgenic rice

Since the first approved GMO for commercialisation (Flavr-Savr™tomato) in 1994 in the United States, in 1995 in Canada and in Mexico and in 1997 in Japan, as well as the first significant transgenic crop cultivation surface reported in 1996, 181.5 million hectares of transgenic plants in 28 countries have been planted in 2014. This represents an increase of more than 100 fold from the 1.7 million planted hectares in 1996 (Figure 3.1) (CERA 2012; James 2014). As rice is one of the most important crops, many transgenic lines have been developed to improve agricultural productivity. In 1999, Liberty Link® (Bayer CropScience) rice varieties LLRICE06 and LLRICE62 were approved for release into the environment in the United States. These transgenic rice contains the bar gene from *Streptomyces hygroscopicus* in one complete copy (LLRICE62) or in several complete and partial copies (LLRICE06). This gene, encoding for phosphinotricin-N-acetyltransferase (PAT), confers herbicide tolerance by catalysing the conversion of the active form of glufosinate ammonium(L-phosphinotricin) in its inactive form. In 2000 and 2006, these two transgenic rice were notably then respectively approved for food/feed use in the USA and Canada (CERA 2012). Between 1998 and 2001, another herbicide tolerant Liberty Link® rice (LLRICE601), expressing also the bar gene, has been subjected to field trials in order to be approved for release into the environment in 2006 in the USA (Quirasco *et al.* 2008; CERA 2012). However, none of these Lyberty Links® rice lines have been planted in any country for commercialisation (Cao *et al.* 2011; De Steur *et al.* 2014).

In 2004, an insect resistant rice (Tarom molaii), containing the Bt gene Cry1Ab was approved for commercialisation in Iran (Figure 3.1). However, this authorisation was suspended by the National Biosafety Council of Iran in 2005

due to an inter-ministerial lack of consultation in order to assess the dossier (De Steur *et al.* 2014; Genetic Rights Foundation 2015).
In 2009, after several field trials in 2008 in collaboration with China National Rice Research Institute and Food and Environmental Safety Assessment, China's Ministry of Agriculture has delivered biosafety certificates for commercial production of two rice cultivar Minghui 63 lines (Bt Shanyou 63, also called Bt63 or TT51-1) and Huahui-1 in Hubei province. These insect resistant transgenic rice were previously tested in controlled field trials between 1999 and 2000, approved for environmental release between 2001 and 2002 and two preproduction field trials were performed between 2003 and 2004. These Bt rice contain the fused Cry1Ab/Cry1Ac genes, which confer insect resistance, under the control of the rice actin 1 gene promoter (pAct1) and NOS terminator (tNOS) (Lu 2010; Cao *et al.* 2011; Chen *et al.* 2011b). Compared to the WT variety, these Bt rice present a superior yield of 6-9% with a decreased use of pesticides of 80% (He *et al.* 2014). Concerning the potential commercialisation of transgenic rice, two insect resistant (Kemingdao 1 (KMD1; cv. Xiushui 11) and B827) and one bacterial resistant (Xa21) lines have already been submitted for regulatory assessment in China (Babekova *et al.* 2009; De Steur *et al.* 2014). Besides, in 2009, five insect resistant rice lines (KMD, T1c-19, T2A-1 and KeFeng variety (6 and 8 which both contain Cry1Ac and SCD genes)) have been approved by the National Biosafety committee of China, following preproduction tests, without any biosafety delivered certificate (Figure 3.1) (Chen *et al.* 2011b).

3.2.4 New plant breeding techniques applied to rice

In 1990, a harmonised EU legal framework regulating the deliberate release of GMO has been established (The Council Of The European Communities 1990). Even if this legislation has been revised and updated overtime, GMO are still defined as organisms "in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination" (The European Parliament and The Council Of The European Union 2001; The European Parliament and The Council Of The European Union 2003a). Given that this definition depends on process-based approaches used to produce GMO, the status of some NPBT, whether or not they are generating plants which fall into the scope of the EU GMO legislation, is still under discussion. These techniques, not technically achievable in the 90's, are *inter alia* the gene editing technology using zinc finger nucleases (ZFN) or oligonucleotide directed mutagenesis (ODM). More recently, other gene editing techniques were developed relying on meganucleases, transcription-activator like effector nucleases (TALEN) and the CRISPR/Cas system. In addition, cisgenesis, intragenesis, RNA-dependent DNA methylation (RdDM), grafting (non GM

scion on GM rootstock or the opposite), reverse breeding and agro-infiltration (true agro-infiltration, agro-inoculation or infection and floral dip) are considered as belonging to the NPBT. Even if the status of new crop varieties developed with NPBT is still undefined in the EU, decisions regarding some products obtained through NPBT have already been taken in Argentina, Australia, New Zealand and the USA (Lusser *et al.* 2011; Araki *et al.* 2014; Pauwels *et al.* 2014). In the coming years, the use of these NPBT could become widespread. Commercial breeders have already adopted some of these approaches towards the production of the first potential commercialised next-generation biotech crop. Research in this field is being conducted for crops such as herbicide tolerant oilseed rape and maize, herbicide tolerant and insect resistant cotton, fungal disease resistant potatoes, drought stress tolerant maize, scab disease resistant apple and reduced amylose content potatoes. In addition, more traits and/or crops developed by companies have still not been disclosed (Lusser *et al.* 2011; Wolt *et al.* 2016).

Via some of these NPBT, new rice varieties presenting an agronomic interest have also been developed. At our knowledge, these studies, carried out essentially in the USA (27.3%), China (36.4%) and Japan (36.4%), concern up to now only seven NPBT rice that present either an herbicide tolerance, a bacterial resistance, an abiotic stress resistance or a nutritional benefit. First, regarding the herbicide tolerance, the mutation of two amino acids (W548L, tryptophan to leucine, and S627I, serine to isoleucine) in the rice acetolactate synthase (ALS) gene generated by gene targeting via homologous recombination has allowed to provide a bispyribac-sodium (BS) herbicide hyper-tolerant rice line. Compared to BS tolerant plant from conventional breeding, the level of BS herbicide tolerance was superior. This rice line used as feed has also an additional nutritional value since these mutations have increased the branched-chain amino acid content that are not synthesized by animals (Endo *et al.* 2007; Endo *et al.* 2013). The tolerance to BS herbicide has also been acquired in rice via the biolistic introduction of chimeric RNA/DNA oligonucleotides to induce specific mutations (Pro171, Trp548 and Ser627) in the ALS gene (Okuzaki and Toriyama 2004). By combining the TALEN technology to the chimeric RNA/DNA oligonucleotides strategy, the development of transgenic rice with a tolerance to glyphosate herbicide has been attempted by the replacement of the base C317 by a base T inside the OsEPSPS gene (Wang *et al.* 2015b). Second, biotech rice with a resistance to bacterial blight stress was developed via the TALEN-mediated mutation of the natural binding site of the *Xanthomonas oryzae pv. Oryzea* TAL effector AvrXa7 or PthXo3 localised upstream of the Os11N3 (OsSWEET14) gene (Li *et al.* 2012). Third, the abiotic stress tolerance of rice was also investigated. Based on the CRISPR/Cas9 system, several genes were successfully targeted in rice, including OsDERF1, coding for the AP2 domain containing protein, that is implied in the drought stress resistance

(Zhang *et al.* 2014a). Finally, NPBT have allowed to improve the nutritional quality of rice. Indeed, the TALEN technology has been used to create fragrant rice from a rice variety devoid of this propriety following to a targeted knockout of OsBADH2 gene (Shan *et al.* 2013; Shan *et al.* 2015). Moreover, precise mutations were introduced by gene targeting based on homologous recombination in the OASA2 gene, coding for an α-subunit of anthranilate synthase that is involved in rice tryptophan biosynthesis. Compared to non-transformants, mature seeds from the obtained rice line present a higher accumulation of tryptophan without any phenotypic modifications. At the nutritional value level, these tryptophan fortified rice plants represent an interesting benefit in both human and livestock diets (Saika *et al.* 2011).

3.3 Detection methods targeting biotech rice and correlated challenges

In order to assess existing GMO detection strategies regarding all collected biotech rice, several DNA-based methods were investigated. By this way, the related benefits and difficulties in the detection of biotech rice originating from transgenesis or NPBT are discussed.

3.3.1 Transgenic rice detection

Knowledge level classification In 2011, the European Network of GMO Laboratories (ENGL) *ad hoc* working group "Unauthorized GMO" suggested a GMO classification, based on the knowledge level (KL) about the sequence of the insert. Given that the availability of this information is crucial for DNA-based GMO detection analysis, the inventoried transgenic rice for which the elements composing their transgenic cassettes are known, representing 1487 transgenic rice, were classified according to the four categories of this KL classification system, ranging from KL-1 to KL-4. For GMO from the KL-1 category, the DNA sequences from the inserts and the transgene flanking regions are known. After detection by element-specific and construct-specific markers in qPCR screening analysis, their identification is thus carried out using event-specific methods. Among all listed transgenic rice, only 0.2% (LLRICE62, LLRICE601 and LLRICE06) belong to this category. Concerning GMO from the KL-2 category, their transgenic cassettes are identical to GMO from the KL-1 category. However, their transgene flanking regions are unknown. Therefore, even if their presence could still be detected using element-specific and construct-specific methods, no event-specific method allows to identify

them. In this work, no collected transgenic rice was associated to the KL-2 category because their genetic constructs do not correspond to the ones found in fully characterized GMO from the KL-1 category. For the KL-3 category, GMO present a transgenic cassette where at least one known transgenic element is found in GMO from the KL-1 category. In addition, their transgene flanking regions are unknown. The majority of the listed transgenic rice (97.6%; 1451 transgenic rice) was classified in this KL-3 category. Within this KL category, the transgenic rice present frequently the p35S promoter and/or the tNOS terminator (93%, 1383 transgenic rice). More precisely, 22.5% (335 transgenic rice), 14.5% (216 transgenic rice) and 56% (832 transgenic rice) of the listed transgenic rice contain respectively only the p35S element, only the tNOS element or both of these elements. Therefore, the transgenic rice from the KL-3 category could be detected using element-specific methods targeting notably these common transgenic elements. However, similarly to GMO from the KL-2 category, no event-specific method makes their identification feasible. Among the KL-3 category, transgenic rice approved for commercialisation in some parts of the world, including Huahui-1, TT51-1, KeFeng-6, KeFeng-8, KMD1, T1c-19 and T2A-1, were notably reported. Finally, GMO assigned to the KL-4 category are only transformed with novel genetic elements and their transgene flanking regions are unknown, making their detection impossible with conventional qPCR approaches. However, only thirty-three transgenic rice (2.2%) corresponded to this criterion.

Detection strategies As mentioned before, classically, three steps are successfully applied in GMO routine analysis through qPCR, the gold standard technology.

First, the presence of GMO is detected in the screening phase using element-specific and construct-specific markers. Several screening markers have besides been successfully tested on transgenic rice approved for commercialisation (*e.g.* LLRice601, LLRice62, TT51-1, KeFeng-6 and KMD1) as well as other transgenic rice lines (*e.g.* Bt rice and KMD2) (Table 3.2) (Mäde *et al.* 2006; Akiyama *et al.* 2007; Quirasco *et al.* 2008; Gu *et al.* 2009; Kluga *et al.* 2013; Reiting *et al.* 2013; Fraiture *et al.* 2014). As most of the inventoried transgenic rice for which information about the elements composing their transgenic cassettes is available belongs to the KL-1, KL-2 and KL-3 groups (97.8%, 1454 transgenic rice), existing screening markers allow to target them (see sub-section 3.3.1) (Figure 3.5). Although the collected information regarding the transgenic elements is based on the name and not on the sequence (see sub-section 3.2.2), and consequently must be used with caution, this allows to estimate the efficiency of the current qPCR GMO detection system to target EU-unauthorized GMO. By this way, 93% of inventoried transgenic rice are covered in a first line through a minimum set of two screening markers targeting the highly frequent p35S and

tNOS elements (see sub-section 3.3.1). In order to increase the coverage, six additional screening markers are proposed to be applied after the initial qPCR screening using the p35S and tNOS markers. More precisely, given that many inventoried transgenic rice contained the t35S element (from the pCAMBIA family or not), the corresponding screening markers were also selected to be applied in a second line. In using two screening markers, which one is specific to the t35S element from the pCAMBIA family vector while the other one targets all t35S elements not originating from the pCAMBIA family vector, a coverage of 94.8% is reached. Among the rest of inventoried transgenic rice belonging to the KL-3 category without possessing at least one of these elements (3%), 1.7% of them contained the Cry1Ab and/or the Cry1Ac elements. Therefore, the use of the Cry1Ab/Ac screening marker in third line allows to cover 96.5% of inventoried transgenic rice. Concerning the remaining 1.3% of collected transgenic rice for which the detection is still possible, the pUBI (0.8%), pNOS (0.1%) or Bar (0.4%) elements were observed in their transgenic cassettes. With the help of the corresponding screening markers, the possible maximum coverage (97.8%) could thus be achieved. However, due to its natural presence in maize, the pUBI marker should be used with caution. Concerning the rest of the inventoried transgenic rice (2.2%), they remain uncovered by the set of screening markers since they belong to the KL-4 group.

Nevertheless, the maximum of coverage is only one aspect of the screening analysis. Indeed, the second one is the discriminative power, which allows to reduce the subsequent number of reaction to perform in the identification step. This step consists to identify GM events using the corresponding event-specific methods, targeting the transgene flanking regions, as well as differentiate EU-authorized and EU-unauthorized GMO. Therefore, even if the minimum set of two screening markers (p35S and tNOS) allows to cover a large spectrum of GMO, the six additional screening markers contribute to increase the discriminative power of the analysis.

However, this identification step is only applicable to GMO for which information about the insertion sites is known, such as for all transgenic rice approved for commercialisation (*e.g.* LLRice601, LLRice62, TT51-1, Huahui-1, KeFeng-6, KeFeng-8, KMD1, T1c-19 and T2A-1) or few other transgenic rice lines (*e.g.* Golden Rice 2) (Table 3.2) (Mazzara *et al.* 2006a; Mazzara *et al.* 2006b; Babekova *et al.* 2009; Köppel *et al.* 2010; Wu *et al.* 2010; Cao *et al.* 2011; Su *et al.* 2011; Wang *et al.* 2011b; Guertler *et al.* 2012; Wang *et al.* 2012; Li *et al.* 2013b; Qian 2013; Wang *et al.* 2014; Jacchia *et al.* 2015). For the majority of inventoried transgenic rice, classified in the KL-3 and KL-4 groups, their identification requires the development of new event-specific methods, which depend on the availability of their sequences. Consequently, only very few unauthorized GMO, originating mainly from the KL-2, KL-3 and KL-4 groups, could be identified using the current qPCR GMO detection system, especially with food/feed

matrices composed of several different GMO. Indeed, although the unauthorized GMO could be detected, their discrimination is difficult in using the signals obtained with the qPCR technology because same elements are found both in EU-authorized and EU-unauthorized GMO. In this context, the use of the t35S pCAMBIA screening marker will be helpful in order to target specifically some EU-authorized GMO (30%) since no EU-authorized GMO has nowadays been transformed with pCAMBIA constructs (Fraiture *et al.* 2014).

To overcome this challenge, the DNA walking technology is proposed as an additional tool allowing to characterize, in any given genome, unknown nucleotide sequence flanking from a short known DNA region, earlier detected through the qPCR screening analysis (Table 3.3). By this way, this approach allows to unequivocally demonstrate the presence of a GMO in food/feed matrix by the characterization of the transgene flanking regions and the unnatural associations of elements. A semi-routine strategy was developed and successfully applied to GM rice and maize by using methods specific to the p35S, tNOS and t35S pCAMBIA elements. Given that the cassette and the insertion sites for most of the inventoried transgenic rice are only partially known (KL-2 and KL-3 categories), the uncharacterized regions could thus be revealed by applying this DNA walking strategy (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b). However, as a minimum of knowledge is required, this strategy cannot deal with GMO containing exclusively unknown transgenic elements (KL-4 category). In that case, Next-Generation-Sequencing (NGS) offers a potential solution, especially in using the whole-genome-sequencing (WGS) approach (Table 3.3). To this end, a strategy of *de novo* assembly could be applied. Even if some difficulties could be encountered with large and complex plant genomes as well as with food/feed matrices composed of several different GMO, the analysis could be facilitated in combining the strength of different NGS platforms. For instance, reads from the Illumina technology could be aligned on substitutes of reference genomes generated by the PacBio technology (Fraiture *et al.* 2016b). However, even if it seems to offer promising solutions in the GMO detection field, the NGS technology is presently not easily implementable routinely by the enforcement laboratories and still requires a long-time frame to get results and qualified bioinformaticians dealing with NGS data. Among the available NGS data analysis tools, none are really dedicated specifically to GMO, except one developed by Yang *et al.* (2013). This last approach allows to map the reads, corresponding partially to the reference host genome, to the transgenic cassette sequence. In this way, the number of inserts and their transgene flanking regions could be determined (Yang *et al.* 2013). Regarding more universal bioinformatics tools, the CLC Genomics Workbench commercial software can be used. It presents the advantage to require only few bioinformatics background and provide easily interpretable output formats compared to other softwares like the Command-Line-Tools.

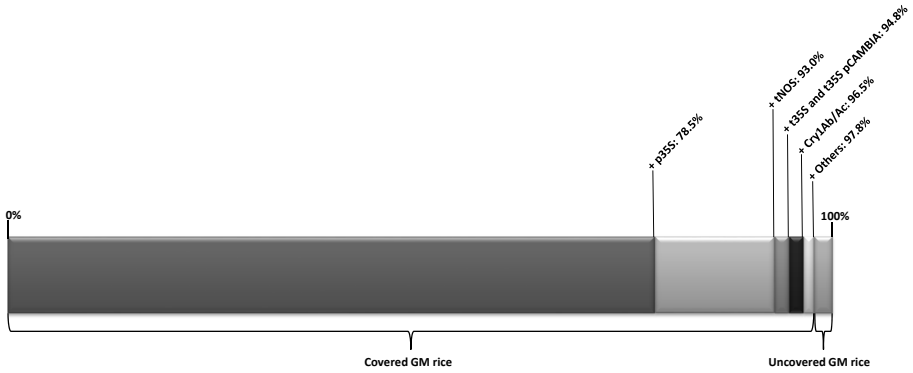


Figure 3.5: Covering (%) of inventoried transgenic rice, for which data regarding the elements composing their transgenic cassettes are known, through qPCR screening markers. In successively combining the p35S, tNOS, t35S, t35S pCAMBIA, Cry1Ab/Ac and other markers (pUBI, Bar and pNOS), most of these inventoried transgenic rice is targeted (97.8%). Uncovered transgenic rice belong to the KL-4 group (2.2%).

However, with this user-friendly tool, the range of possibilities is limited by the available pre-designed workflows (Willems *et al.* 2016). This kind of pre-designed workflows could also be created by qualified bioinformaticians in order to simplify the analysis for bioinformatics novices, with the advantage of being over time improved in-house if needed. In addition, even if the WGS approach could realistically be applied on a food/feed matrix exclusively composed of one kind of GM crop, such analysis still remains unreasonable, in term of cost, to detect GMO present at trace level. In the frame of the DECATHLON project, bioinformatics pipelines will be besides developed and assessed for appropriate implementations notably in GMO routine analysis(Decathlon Project 2016; Willems *et al.* 2016).

Besides the considerations upstream, alternatives approaches have also been considered to overcome some limits inherent to the qPCR technology, such as an insufficient throughput to deal with the increasing number of GMO, a relative quantification requiring certified materials that are essentially available for EU-authorized GMO and an incompatibility to carry out analysis directly onfield. These approaches allow notably to improve the speed, the transportability (*e.g.*, loop-mediated isothermal amplification (LAMP) and recombinase polymerase amplification (RPA)), the quantification accuracy (*e.g.*, digital PCR (dPCR)) or the throughput (*e.g.*, multi-labelling system based on electrochemical biosensor (MLSEB)) of the analysis (Table 3.3).

3.3.2 NPBT rice detection

Although only few applications have been nowadays reported for rice, the use of NPBT, including the most recent one like Crisp/Cas9, is expected to increase. Therefore, in order to assess the theoretical potential to detect these new plant varieties in a near future, the applicability of existing technologies has been carried out. Concerning the detection and the identification by qPCR of the plants generated from NPBT, several difficulties have already been highlighted by Lusser *et al.* (2011) in function of the technique used. First, as only small modifications are induced, with or without a repair template, the detection, but not the identification, of plants coming from the ZFN technology of respectively the ZFN-1 and ZFN-2 categories is feasible with the help of a minimum of available information. However, in case plants are modified by the ZFN-3 category, which induces large modification (*e.g.* gene), their detection and identification, are possible with *a priori* knowledge whereas, without any available information, only the detection step is conceivable. To attempt to prove these modifications from ZFN approaches, a strategy of full genome sequencing could be envisaged if no information is available. Second, for the mutations produced by the ODM method, only their detection could be performed based on prior information. But, similarly to ZFN approaches, the NGS technology could be applied when no information is available. Third, given that cisgenic plants and plants obtained through intragenesis present a large modification, PCR assays, using event-specific markers designed on the known sequence of the unnatural association of elements, can allow their detection and identification with a known case whereas sequencing strategies have to be investigated if no information is available. Fourth, the identification of RdDM plants is not applicable since the modifications generated could also occur naturally. Fifth, regarding reverse breeding, agro-infiltration and grafting, no discrimination is possible with plants coming from conventional breeding methods, preventing any detection and identification (Lusser *et al.* 2011).

It is thus clear that the precise mutations, as integrated in the reported NBPT rice (see sub-section 3.2.4), will be highly difficult to detect. Since each transgenic plant submitted for an approval on the EU market should be associated to a corresponding detection or identification method, it represents thus a real challenge for most of the plants coming from NPBT listed up to now, especially in absence of any prior information (Lusser *et al.* 2011). However, in case of the competent authorities agree finally that NPBT organisms do not fall under the GMO legislations, no related detection system will be mandatory.

3.4 Conclusion

Based on all the information collected from the inventoried biotech rice in the present study, the possibility to detect them have been evaluated, allowing to suggest suitable detection strategies.

On the one hand, the detection of transgenic rice could be considered in function of the KL classification. Indeed, the current qPCR strategy is still valid to detect GMO belonging to the KL-1, KL-2 or KL-3 categories since 97.8% of inventoried transgenic rice, for which information about the elements composing their transgenic cassettes are known, are covered by this way. To this end, a set of eight screening markers, including p35S and tNOS elements which are highly observed both in EU-authorized and EU-unauthorized GMO, was suggested. In addition, to increase the discrimination power of the analysis, the use of the t35S pCAMBIA marker allows to specifically target around 30% of EU-unauthorized GMO. Concerning their identification, the current qPCR strategy is however only applicable for GMO belonging to the KL-1 category. Therefore, for the majority of collected transgenic rice, classified in the KL-3 category, the observed screening qPCR positive signals, suggesting that GMO are present in the tested sample, need to be confirmed by alternatives approaches. Moreover, unlike the qPCR system, among the potentially detected GMO, some of the alternatives approaches could discriminated EU-authorized and EU-unauthorized GMO present within the same food/feed matrix. Indeed, in using DNA walking techniques, sequences from the transgene flanking regions and from unnatural associations of elements could be characterized in order to irrefutably prove the presence of GMO (Fraiture *et al.* 2015b). However, as a minimum of prior knowledge is required, this strategy is not convenient for the few transgenic rice belonging to the KL-4 category. In that case, only whole-genome-sequencing approaches seems relevant, even if some progresses are still required *inter alia* in terms of sensitivity, availability of reference genomes and ease of bioinformatics analysis.

On the other hand, for the rice produced by NPBT, regardless of the fact that no decision has yet been taken at the EU level regarding their potential GMO status, most of the introduced genetic modifications are too similar to those obtained with conventional breeding or natural processes, making their detection without prior knowledge challenging, or even technically impossible. The detection of some of these biotech plants could be envisaged in a near future using NGS approaches. Taking into account that the NGS technology and related analysis will be improved, the integrated modifications could be located via whole genome sequencing approaches. Moreover, for some of these biotech rice, the knowledge of their sequences, via developers or NGS analysis, could allow designing appropriated qPCR TaqMan® markers with probes specific to the integrated mutations.

Table 3.2: Representative examples of qPCR detection methods tested on transgenic rice approved for commercialisation. Those validated at the EU level are indicated by an asterisk.

Multiplexing	Method	Target	LLRICE601	LLRiCE62	TT51-1 (Bt63)	Huahui 1	KeFeng-6	KeFeng-8	KMD1	T1c-19	T2A-1	References
Simplex	Element-specific	p35S	X	X			X		X			Kluga *et al.* (2013)
Simplex	Element-specific	tNOS			X		X		X			Kluga *et al.* (2013)
Simplex	Element-specific	Cry1Ab/ Cry1Ac			X		X		X			Kluga *et al.* (2013)
Duplex	Element-specific	p35S	X	X			X		X			Kluga *et al.* (2013)
	Element-specific	tNOS			X		X		X			Kluga *et al.* (2013)
Simplex	Element-specific	p35S	X	X								Quirasco *et al.* (2008)
Simplex	Element-specific	t35S		X								Quirasco *et al.* (2008)
Simplex	Construct-specific	p35S-Bar	X	X								Quirasco *et al.* (2008)
Simplex	Construct-specific	Cry1Ab/ Cry1Ac-tNOS			X							Mäde *et al.* (2006)
Simplex	Construct-specific	p35S-hpt					X		X			Reiting *et al.* (2013)
Simplex	Construct-specific	hpt-t35S					X					Reiting *et al.* (2013)
Simplex	Construct-specific	hpt-tNOS							X			Reiting *et al.* (2013)
Duplex	Construct-specific	Cry1Ac-tNOS (two probes)					X					Akiyama *et al.* (2007)
Simplex	Event-specific	LLRICE601*	X									Mazzara *et al.* (2006a)
Simplex	Event-specific	LLRICE62*		X								Mazzara *et al.* (2006b)
Simplex	Event-specific	TT51-1			X							Wu *et al.* (2010), Cao *et al.* (2011), and Wang *et al.* (2014)

Table 3.2 *continued*

Multiplexing	Method	Target	LLRICE601	LLRiCE62	TT51-1 (Bt63)	Huahui 1	KeFeng-6	KeFeng-8	KMD1	T1c-19	T2A-1	References
Simplex	Event-specific	Huahui 1				X						Li *et al.* (2013a)
Simplex	Event-specific	KeFeng-6					X					Su *et al.* (2011), Guertler *et al.* (2012), Wang *et al.* (2011b), and Wang *et al.* (2014)
Simplex	Event-specific	KeFeng-8						X				Wang *et al.* (2012)
Simplex	Event-specific	KMD1							X			Babekova *et al.* (2009) and Wang *et al.* (2014)
Simplex	Event-specific	T1c-9								X		Qian (2013)
Simplex	Event-specific	T2A-1									X	Qian (2013)
Pentaplex	Taxon-specific	PLD	X	X	X							Köppel *et al.* (2010)
	Construct-specific	p35S-Bar*	X	X								Köppel *et al.* (2010)
	Event-specific	LLRICE601*	X									Köppel *et al.* (2010)
	Event-specific	LLRICE62*		X								Köppel *et al.* (2010)
	Event-specific	TT51-1			X							Köppel *et al.* (2010)

Table 3.3: Representative examples of alternative detection methods tested on transgenic rice.

Technology	Method	LLRiCE62	TT51-1 (Bt63)	KeFeng-6	KMD1	T1c-19	Other GM rice	References
dPCR	Construct-specific p35S-Bar	X						Köppel and Bucher (2015)
	Event-specific Bt63		X					Li *et al.* (2015)
	Event-specific KeFeng-6			X				Li *et al.* (2013b)
	Event-specific KMD1							Köppel and Bucher (2015)
LAMP	Element-specific Cry1Ab				X			Li *et al.* (2013c)
	Element-specific p35S			X		X		Zhang *et al.* (2012)
	Element-specific tNOS		X	X		X		Zhang *et al.* (2012)
	Element-specific bar					X		Zhang *et al.* (2012)
	Event-specific Bt63		X					Chen *et al.* (2012a)
	Event-specific KeFeng-6			X				Chen *et al.* (2012a)
	Event-specific KMD1				X			Chen *et al.* (2012a)
RPA	Element-specific p35S			X	X			Xu *et al.* (2014)
	Element-specific tNOS		X	X	X			Xu *et al.* (2014)
MLSEB	Specific probe GM rice						X	Huang *et al.* (2015)
DNA walking	Anchored to t35S pCAMBIA, p35S and tNOS						X	Fraiture *et al.* (2014), Fraiture *et al.* (2015a), and Fraiture *et al.* (2015b)
	Anchored to p35S and t35S	X						Spalinskas *et al.* (2013a) and Spalinskas *et al.* (2013b)
NGS	Whole genome sequencing using Illumina platform	X						Wahler *et al.* (2013)
	Whole genome sequencing using Illumina platform		X					Yang *et al.* (2013)
	Whole genome sequencing using Illumina platform						X	Willems *et al.* (2016)

Using biotech rice as a study case, the present approach could definitely be extended to biotech events belonging to other species. Indeed, similarly, publicly available information for non-rice biotech events developed worldwide, also usually scattered, could be centralized. By this way, clue information, regarding notably the integrated genetic elements, could be highlighted in order to subsequently establish an appropriated detection strategy. According to the collected data, the efficiency of the current qPCR GMO detection system used by the enforcement laboratories could be assessed. Moreover, if necessary, this detection system, especially for the unauthorized GMO, could be strengthened with additional methods, such as DNA walking and NGS.

Supporting information

Additional files 1 and 2 are available at http://dx.doi.org/10.1016/j.tifs.2016.03.011 (file formats: xlsx and docx).

Acknowledgements

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract UGMMONITOR (convention RF 11/6242).

Chapter 4

An innovative and integrated approach based on DNA walking to identify unauthorised GMOs

In this chapter, a discriminative qPCR screening method targeting the t35S pCAMBIA element was developed to detect EU unauthorized GMO. Then, a DNA walking strategy, anchored on the t35S pCAMBIA element and fully integrated to the GMO screening analysis in routine, was developed to be able to demonstrate the associations of elements that are typically found in transgenic constructs and the transgene flanking regions to prove the presence of GMO containing the t35S pCAMBIA element.

This chapter was previously published as:
M.-A. Fraiture, P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2014). "An innovative and integrated approach based on DNA walking to identify unauthorised GMOs". *Food Chemistry* 147, pp. 60–69. DOI: 10.1016/j.foodchem.2013.09.112

Authors' contributions

M.-A. Fraiture performed experiments, analysed data and drafted the manuscript. P. Herman, I. Taverniers, M. De Loose, D. Deforce and N. H. Roosens helped to design the study and drafted the manuscript. All authors read and approved the final manuscript.

Abstract

In the coming years, the frequency of unauthorised genetically modified organisms (GMOs) being present in the European food and feed chain will increase significantly. Therefore, we have developed a strategy to identify unauthorised GMOs containing a pCAMBIA family vector, frequently present in transgenic plants. This integrated approach is performed in two successive steps on Bt rice grains. First, the potential presence of unauthorised GMOs is assessed by the qPCR SYBR®Green technology targeting the terminator 35S pCAMBIA element. Second, its presence is confirmed via the characterisation of the junction between the transgenic cassette and the rice genome. To this end, a DNA walking strategy is applied using a first reverse primer followed by two semi-nested PCR rounds using primers that are each time nested to the previous reverse primer. This approach allows to rapidly identify the transgene flanking region and can easily be implemented by the enforcement laboratories.

4.1 Introduction

Rice (*Oryza sativa*), one of the most important crops in the world, is a staple food for more than three billion people. In addition, this cereal grain is also used in animal feed. The majority of rice is grown and consumed in Asia, particularly in China (Datta 2004; Kathuria *et al.* 2007; James 2009; Chen *et al.* 2011b). On the European (EU) market, most of the rice is currently imported from Asia (Stein and Rodríguez-Serezo 2009). In order to provide food to the growing worldwide population (approximately eight billion in 2020), rice production should increase significantly (25–40%). To this end, genetically modified (GM) rices are developed to ensure sufficient rice production in spite of the lack of arable land. According to the scientific literature on GM rice, the research in laboratories mainly target improving biotic (insect, virus, fungi, bacteria) and abiotic (drought, salinity, cold) tolerances (Datta 2004; High *et al.* 2004; Kathuria *et al.* 2007; Chen *et al.* 2011b; Ahmad *et al.* 2012; Yu *et al.* 2012). The development of GM rice is highly supported by the Chinese government (Chen *et al.* 2011b; Xia *et al.* 2011). Since 2009, two insect resistant GM rices (Huahui-1 and Bt Shanyou 63) are cultivated on a large scale for commercialisation in China. In addition, other insect resistant (Tarom molaii) and herbicide tolerant (CL121, CL141, CFX51, IMINTA-1, IMINTA-4, PWC16, LLRICE62, LLRICE06 and LLRICE601) GM rices are nowadays commercialised worldwide (Chen *et al.* 2011b; Tan *et al.* 2011; Xia *et al.* 2011; Wang *et al.* 2012; Genetic Rights Foundation 2015; CERA 2012). On the EU market, the introduction and the control of genetically modified organisms (GMOs) in the food and feed chain are submitted to the EU legislation in order to guarantee the freedom of choice to the consumer (The European Parliament and The Council Of The European Union 2003a; The European Parliament and The Council Of The European Union 2003b). However, the continuous enforcement of this legislation is complex for several reasons. First of all, the number (around 30 to 120 GMOs) and the diversity (2 to 15 genes) of commercialised GMOs will increase significantly in the 5 coming years (Stein and Rodríguez-Serezo 2009). Moreover, numerous vectors will be used, such as the pBin19, pBI121, pPZP and pCAMBIA families (Komori *et al.* 2007). Second, in addition to genes conferring insect resistance or herbicide tolerance, a larger range of traits will be developed (*e.g.*, abiotic stress tolerance, disease resistance and nutritional allegations). Third, the present commercialised GM crops are principally developed by American and EU companies which have a major interest in being authorised to commercialise their products on the EU market. Nevertheless, in 2015, more and more GMOs intended for local consumption will be developed by Asian technological centres. These GM crops are unlikely to be submitted for EU approval. Thus, it is very likely that the frequency of unauthorised GMOs on the EU market will significantly

increase by their accidental (or adventitious) presence in raw materials and processed food or feed (Stein and Rodríguez-Serezo 2009). In addition, the possible escape of GMOs from field-trials or during development stages could also be another source of unauthorised GMOs (Zapiola *et al.* 2008; Holst-Jensen *et al.* 2012). Indeed, although the presence of GM rice is to date not authorised on the EU market, the herbicide tolerant LLRice601, in 2004, and the insect resistant Bt Shanyou 63, in 2006, and KeFeng-6, in 2010, have been detected in food products imported from China (Stein and Rodríguez-Serezo 2009; The European Commission 2011b; Wang *et al.* 2011b). In 2012, more than 50% of the GMOs detected in food/feed matrices, reported in the RAPID Alert System Database, concerned unauthorised GM rices imported from Asia, mainly China (The European Commission 2015b). To address the increasing number of alerts, the EU commission decided to implement "Emergency measures regarding unauthorised genetically modified rice in rice products originating from China and repealing Decision 2008/289/EC" (The European Commission 2011b).

To ensure an efficient GMO detection in food and feed products on the EU market, several screening methods have been developed, mainly based on the most common recombinant elements in GM crops like transcription-regulating sequences. These elements are p35S (cauliflower mosaic virus (CaMV) 35S promoter) and tNOS (*Agrobacterium tumefaciens* nopaline synthase terminator) (Barbau-Piednoir *et al.* 2010). The majority of these methods have been performed with the SYBR®Green and TaqMan® technologies (Barbau-Piednoir *et al.* 2010; Reiting *et al.* 2010; Mbongolo Mbella *et al.* 2011; Broeders *et al.* 2012a; Barbau-Piednoir *et al.* 2013; Kluga *et al.* 2013). However, the detection of elements derived from natural organisms (viruses and bacteria) can be misinterpreted. One of the most common examples is a p35S positive signal which could also mean the identification of the host CaMV in *Brassica* species (Broeders *et al.* 2012a; Broeders *et al.* 2012b). Therefore, additional markers have been developed to discriminate the presence of the transgenic crop or the natural organism, such as CRT (targeting the transcriptase gene of CaMV virus) used for routine analysis in-house and CaMV (targeting the ORFIII of CaMV virus) (Chaouachi *et al.* 2008; Broeders *et al.* 2012a; Broeders *et al.* 2012b; Broeders *et al.* 2013).

However, the strategy described above is merely an indirect proof of the potential GMO presence in food matrix. Direct proof can only be supplied by the characterisation of the junction between the transgenic integrated cassette and the plant genome. To get this crucial information, DNA walking methods have been used to identify this unknown nucleotide sequence flanking already known DNA regions in any given genome (Leoni *et al.* 2011; Volpicella *et al.* 2012b). Classically, three classes of strategies exist: (a) restriction-based methods, involving a preliminary restriction digestion of the genomic DNA (Triglia *et al.* 1988; Shyamala and Ames 1989; Jones and Winistorfer 1992;

Theuns *et al.* 2002; Leoni *et al.* 2011); (b) extension-based methods, defined by the extension of a sequence-specific primer and subsequent tailing of the resulting single-strand DNA molecule (Mueller and Wold 1989; Min and Powell 1998; Hermann *et al.* 2000; Leoni *et al.* 2011); and (c) primer-based methods, coupling various combinatorial (random or degenerate) primers to sequence-specific primers (Parker *et al.* 1991; Leoni *et al.* 2011). Up to now, some studies have been published about the junction characterisation of transgenic plants such as thale cress (*Arabidopsis thaliana*) (Windels *et al.* 2003b; Ruttink *et al.* 2010a), potato (*Solanum tuberosum*) (Côté *et al.* 2005; Cullen *et al.* 2011), rice (*O. sativa*) (KeFeng-6, KeFeng-8, LLRICE62, Bt Shanyou 63 (TT51-1)) (Cao *et al.* 2011; Su *et al.* 2011; Wang *et al.* 2011b; Wang *et al.* 2012; Spalinskas *et al.* 2013b), maize (*Zea mays*) (MON810, MON863, MON88017, NK603, LY038, DAS59122-7, T25, 3272, Bt11, BT176, CHB351, GA21) (Holck *et al.* 2002; Rønning *et al.* 2003; Windels *et al.* 2003b; Collonnier *et al.* 2005; Taverniers *et al.* 2005; Yang *et al.* 2005b; Raymond *et al.* 2010; Trinh *et al.* 2012a; Spalinskas *et al.* 2013b), cotton (*Gossypium hirsutum*) (MON1445) (Akritidis *et al.* 2008), canola (*Brassica napus*) (GT73) (Taverniers *et al.* 2005) and soybean (*Glycine max*) (MON89788, GT40-3-2) (Windels *et al.* 2001; Raymond *et al.* 2010; Trinh *et al.* 2012a). However, most of the methods in these studies cannot easily be used in routine analysis by the enforcement laboratories: techniques are laborious and complex (finger-printing by capillary electrophoresis, genomic DNA library via (unpredictable) restriction enzyme) with regard to a method exclusively based upon PCR, require a lengthy procedure with generally multiple steps to get results, or present a lack of specificity, yield or data concerning the compatibility with a low amount of target.

The aim of the present study is to supply an integrated approach to identify unauthorised GMOs: A first real-time PCR screening allows the detection of the terminator 35S (t35S) of the pCAMBIA family vectors to indicate the potential presence of unauthorised GMOs in food matrices (Figure 4.3). Then, an appropriate DNA walking method, anchored on the sequence used for the screening followed by two semi-nested PCRs to identify the junction, confirms the GMO presence.

4.2 Materials and methods

4.2.1 Plant material

Grains of transgenic Bt rice (*O. sativa* L. *Japonica cv Ariete*) and its wild-type (WT) were used in this study to develop the methodology (Breitler *et al.* 2004). This transgenic rice was transformed by *A. tumefaciens* with the

binary vector pCAMBIA1300, which contains the synthetic Cry1B gene from *Bacillus thuringiensis* conferring insect resistance. The Certified Reference Materials (CRM) in the form of seeds powders or genomic DNA (gDNA) were obtained from the American Oil Chemists' Society and the Institute for Reference Materials and Measurements and were used to test the specificity (AOCS, Urbana, USA 2015; IRMM, Geel, Belgium 2015). These materials were characterised as previously described (Broeders *et al.* 2013). The list of all plant material is shown in Table 4.1.

4.2.2 DNA extraction, concentration and purity

Bt rice grains were ground to obtain a homogeneous powder. DNA was extracted using a CTAB-based procedure (ISO 21571 2005) in combination with the Genomic-tip20/G (QIAGEN, Hilden, Germany). This DNA extraction method, adapted from the EU-RL GMFF validated method, is composed of four main successive steps: (1) Extraction of proteins, polysaccharides and organic components, (2) Precipitation of DNA in presence of C-hexadecyl-Trimethyl-Ammonium-Bromide (CTAB), (3) Purification of DNA using a tip20 column and (4) Precipitation of DNA with isopropanol (ISO 21571 2005; European Union Reference Laboratory for GM Food and Feed and Feed 2006b). DNA concentration was measured by spectrophotometry using the Nanodrop® 2000 (ThermoFisher, DE, USA) device and DNA purity was evaluated by the A260/A280 and A260/A230 ratios. DNA extraction, concentration and purity of CRMs were carried out as previously described (Broeders *et al.* 2013).

4.2.3 Development and assessment of oligonucleotide primers

The oligonucleotide primers were designed to target the t35S sequence of the pCAMBIA vector (Figure 4.1). To get universal oligonucleotide primers detecting all pCAMBIA vectors, all t35S pCAMBIA sequences were compared via the software "ClustalW2". The oligonucleotide primers were thus designed manually on the conserved region of all pCAMBIA vectors. To be convenient for the DNA walking approach, these oligonucleotide primers were chosen at the nearest extremity of the walking direction. Note that the t35S pCAMBIA element is the starting position and the walking direction is defined on the rice genome through the left border of the transgenic cassette (Cambia, Canberra, Australia 2015; EMBL-EBI 2015). Via a different combination, the same oligonucleotide primers were usable for qPCR assays. The oligonucleotide primers and the obtained amplicon sequences are indicated in Table 4.2 and Figure 4.1. The specificity of oligonucleotide primers was initially evaluated

Table 4.1: Specificity of t35S pCAMBIA marker tested on plant materials observed *in silico* and *in vitro* by qPCR SYBR®Green assay. The positive and the negative signals obtained are indicated by + and −, respectively.

Species	Plant materials	GMO% (m/m)	*In silico*	*In vitro*
Maize (*Zea mays*)	WT[a]	0	−	−
	MON 810[a,c]	1	−	−
	MON 89034[a]	1	−	−
	MON 863[a,c]	1	−	−
	MON 88017[a,c]	1	−	−
	Bt176[a,c]	1	−	−
	3272[a,c]	10	−	−
	DAS59122[a,c]	9.87	−	−
	TC1507[a,c]	1	−	−
	T25[a,c]	100	−	−
	Bt11[a,c]	1	−	−
	GA21[a,c]	1	−	−
	NK603[a,c]	1	−	−
	MIR604[a,c]	1	−	−
	MIR162[a,c]	1	−	−
	98140[a]	1	−	−
Soybean (*Glycine max*)	WT[a]	0	−	−
	GTS40-3-2[a,c]	10	−	−
	A2704-12[a,c]	100	−	−
	A5547-127[a,c]	100	−	−
	MON 89788[a]	1	−	−
	305423[a]	1	−	−
	356043[a]	1	−	−
	MON 87701[a]	1	−	−
Oilseed rape (*Brassica napus*)	WT[a]	0	−	−
	RT73[a]	1	−	−
	MS8[a,c]	100	−	−
	Rf3[a,c]	100	−	−
	T45[a,c]	100	−	−
	Ms1[a,c]	1	−	−
	Rf1[a,c]	1	−	−
	Rf2[a,c]	1	−	−
	Topas 19/2[a,c]	1	−	−
Cotton (*Gossypium hirsutum*)	WT[a]	0	−	−
	MON 1445[a,c]	1	−	−
	MON 15985[a,c]	100	−	−
	MON 531[a,c]	100	−	−

Table 4.1 *continued*

Species	Plant materials	GMO% (m/m)	*In silico*	*In vitro*
	LL25[a,c]	100	–	–
	GHB614[a]	100	–	–
	GHB119[a]	10	–	–
	281-24-236×3006-210-23[a]	1	–	–
	MON 88913[a]	1	–	–
Potato (*Solanum tuberosum*)	WT[a]	0	–	–
	EH92-527-1[a,c]	1	–	–
Sugar beet (*Beta vulgaris*)	WT[a]	0	–	–
	H7-1[a]	100	–	–
Rice (*Oryza sativa*)	WT[a]	0	–	–
	LLRICE62[a,c]	1	–	–
	Bt rice	100	+	+
Plasmid	Sybricon t35S pCAMBIA[b]	/	+	+

[a] CRM (Certified Reference Materials).
[b] Plasmid pUC18 containing t35SpCAMBIA amplicon.
[c] GMOs containing the p35S and/or tNOS elements (Barbau-Piednoir *et al.* 2010).

in silico using the program "wprimersearch" from the software "wEMBOSS", which mimics PCR amplification (Barbau-Piednoir *et al.* 2013; EMBnet 2015) (Table 4.1).

4.2.4 qPCR SYBR®Green technology

As previously described, for all qPCR assays, a standard 25 μl reaction volume was applied containing 1 × SYBR®Green PCR Mastermix (Diagenode, Liège, Belgium), 250 nM of each primer and 5 μl of DNA (10 ng/ μl). The qPCR cycling program consisted of a single cycle of DNA polymerase activation for 10 min at 95°C followed by 40 amplification cycles of 15 s at 95°C (denaturing step) and 1 min at 60°C (annealing–extension step). The program for melting curve analysis was performed by gradually increasing the temperature from 60 to 95°C in 20 min (±0.6°/20 s) (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2013). All runs were performed on an iQ™5 real-time PCR detection system (BioRad, Hemel Hempstead, UK) or an ABI 7300 qPCR system (Applied Biosystems, CA, USA) for the specificity assessment and the rest of the analysis, respectively. Concerning the qPCR method acceptance parameters, evaluation of specificity,

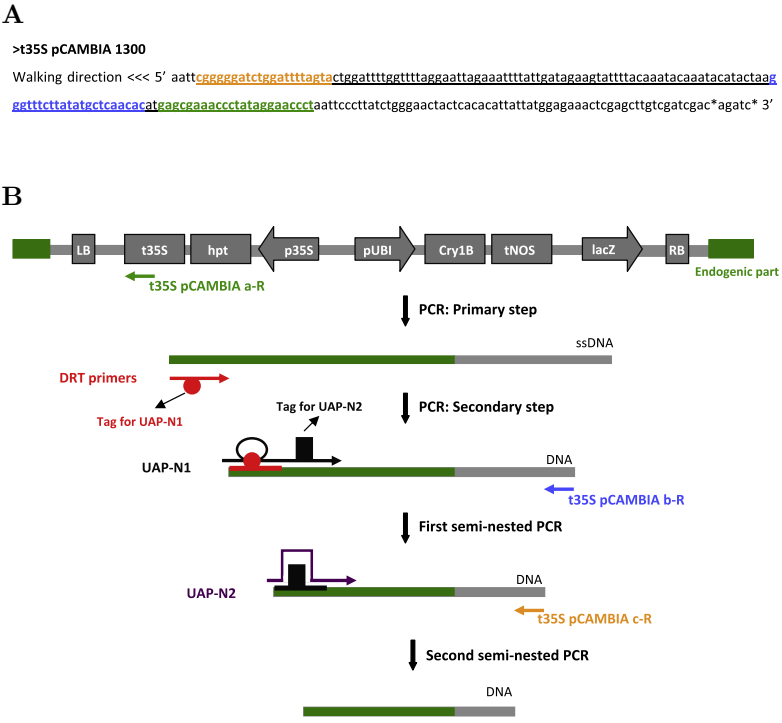


Figure 4.1: DNA walking strategy.
(A) Designed primer position of t35S pCAMBIA a-R (green), t35S pCAMBIA b-R (blue) and t35S pCAMBIA c-R (orange) to target t35S pCAMBIA sequence (highlighted) from t35S pCAMBIA 1300 sequence. This sequence is identical for all pCAMBIA vectors (1200, 1201, 1281Z, 1291Z, 1301, 1302, 1303, 1304, 1380, 1381Xa, 1381Xb, 1381Xc, 1381Z, 1390, 1391, 1391Xa, 1391Xb, 1391Xc and 1391Z except for 2200, 2201, 2300, 2301 which are lightly shorter at the 3' end*). (B) Principle of DNA walking approach based on a double semi-nested PCR. Three reactions are carried out to amplify the targeted sequence. First, a two-step PCR taking place in one tube is carried out. In the first step, single strand DNA (ssDNA) fragments are produced by a single primer extension reaction using t35S pCAMBIA a-R primer (green). This reaction is repeated four times in four individual tubes. In the second step, four different DRT primers (A–D) (red) are immediately added individually to the four reaction tubes. Second, the first semi-nested PCR is obtained in combining t35S pCAMBIA b-R primer (blue) and a long universal tagging primer (UAP-N1) (black). Finally, the second semi-nested PCR is carried out by the combination of t35S pCAMBIA c-R primer (orange) and a short universal tagging primer (UAP-N2) (purple) in order to increase the yield of the specific product.

Table 4.2: Oligonucleotide sequences used for qPCR assays, DNA walking approach and PCR amplifications. In the purpose of an integrated approach, the same t35S pCAMBIA oligonucleotides are used for the qPCR as well as for the DNA walking. PLD marker (Phopholipase D gene from rice (*Oryza sativa*)) is used as the rice endogene control.

Method	Oligonucleotide name	Oligonucleotide sequence	Product size (bp)	Reference
SYBR®Green qPCR	p35S F p35S R	AAAGCAAGTGGATTGATGTGATA GGGTCTTGCGAAGGATAGTG	75	Barbau-Piednoir *et al.* (2010)
SYBR®Green qPCR	tNOS F tNOS R	GATTAGAGTCCCGCAATTATACATTTAA TTATCCTAGKTTGCGCGCTATATT	69	Barbau-Piednoir *et al.* (2010)
SYBR®Green qPCR	t35S F t35S R	Data not shown Data not shown	107	In-house, Broeders *et al.* (2012a)
SYBR®Green qPCR	PLD F PLD F	GCTTAGGGAACAGGGAAGTAAAGTT CTTAGCATAGTCTGTGCCATCCA	80	Mbongolo Mbella *et al.* (2011)
SYBR®Green qPCR	t35S pCAMBIA c-F t35S pCAMBIA a-R	CGGGGGATCTGGATTTTAGTA AGGGTTCCTATAGGGTTTCGCTC	137	This study
DNA walking	t35S pCAMBIA c-R t35S pCAMBIA b-R t35S pCAMBIA a-R	TACTAAAATCCAGATCCCCCG GTGTTGAGCATATAAGAAACCC AGGGTTCCTATAGGGTTTCGCTC	/	This study
PCR	Rice chromosome II-F pCAMBIA-R	CGAAAAGAAGATGGCAGGAT CTGTCGATCGACAAGCTCGAGT	490	This study
PCR	Rice chromosome III-F pCAMBIA-R	TTTCTTTCGCTTCTGCAGGT CTGTCGATCGACAAGCTCGAGT	515	This study

Table 4.3: Analysis of Bt rice identity *in silico* and *in vitro* by qPCR SYBR®Green assay. The positive and the negative obtained signals are indicated by + and −, respectively.

	p35S		tNOS		t35S		t35S pCAMBIA		PLD	
	In silico	*In vitro*	*In silico*	*In vitro*	*In silico*	*In vitro*	*In silico*	*In vitro*	*In silico*	*In vitro*
WT Bt rice	−	−	−	−	−	−	−	−	+	+
Bt rice	+	+	+	+	−	−	+	+	+	+

sensitivity and inter-run repeatability was carried out as previously described (Broeders *et al.* 2013). In brief, the specificity of the t35S pCAMBIA c-F and the t35S pCAMBIA a-R primers was tested on several WTs, GMOs and LLPs (Low Level Presence) by qPCR SYBR®Green method using C_t and T_m values as criteria (Tables 4.1 and 4.2) (The European Commission 2011a). Sensitivity and repeatability were determined for t35S pCAMBIA primers on Bt rice using the qPCR SYBR®Green method on serial dilutions going from 2000 to 0.1 haploid genome equivalents (HGEs) (Tables 4.2 and 4.4). From these serial dilutions, the PCR efficiency and linearity (R^2) were estimated.

4.2.5 Cloning, sequencing and plasmid registration

The t35S pCAMBIA amplicon was cloned into a pUC18 plasmid (INVITROGEN, CA, USA) to obtain the t35S pCAMBIA Sybricon as previously described (Sambrook and Russell 2001; Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2013). Briefly, the t35S pCAMBIA amplicon was first subcloned into the pCR®2.1-TOPO® Vector using the TOPO TA Cloning® Kit (INVITROGEN, CA, USA) according to the manufacturers' instructions. After EcoRI restriction, the correct amplicon was then cloned into the vector pUC18 (INVITROGEN, CA, USA). The plasmid was sequenced via a Genetic Sequencer 3130XL using the Big Dye Terminator Kit v3.1 (Applied Biosystems, CA, USA) and was tested using the qPCR reaction conditions and the specific primers as indicated in point 4.2.4. The t35S pCAMBIA Sybricon plasmid was registered under "Safe Deposit" at the "Belgian Culture Collection for Micro-organisms" in the "Plasmid and DNA Library Collection" (BCCM/LMBP, Gent, Belgium; BCCM number: LMBP 8352). Authenticity was assessed by the BCCM/LMBP prior to acceptance and certification (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2013).

4.2.6 DNA walking approach

DNA walking and double semi-nested PCR The assay was performed using 100 ng of 100% Bt rice DNA (Figure 4.1). Degenerated random tagging (DRT) and Universal tagging primers (UAP-N1 and N2) were provided by APAgene™ GOLD Genome Walking Kit (BIO S&T, Montréal, Canada). Recombinant *Taq* DNA Polymerase (10342; INVITROGEN, CA, USA) was used to synthesise DNA. The three gene-specific primers for t35S pCAMBIA were designed as described above (Section 4.2.3). The t35S pCAMBIA a-R primer was used to perform the DNA walking and then the t35S pCAMBIA b-R and the t35S pCAMBIA c-R primers were applied in the first and the second semi-nested PCR rounds, respectively. PCR mixes and conditions were carried out according

Table 4.4: LOD_6 (Limit Of Detection with 6 repeats) of t35S pCAMBIA screening marker for 100% Bt rice in SYBR®Green qPCR assay obtained at 5 HGEs (Haploid Genome Equivalent) (in bold). The inter-run repeatability is determined as the relative standard deviation of the test results (RSD_r %).

Dilution	HGE/well	Final DNA concentration (ng/μl)	Final DNA concentration/well (ng)	Dilution factor	RSD_r (%)
Stock		4	/		
S1	2000	0.2	1	20	1.5
S2	1000	0.1	0.5	2	1.7
S3	400	0.04	0.2	2.5	1.1
S4	100	0.01	0.05	4	0.9
S5	50	0.005	0.025	2	1.8
S6	20	0.002	0.01	2.5	0.9
S7	10	0.001	0.005	2	2.0
S8	**5**	**0.0005**	**0.0025**	**2**	**2.2**
S9	2	0.0002	0.001	2.5	
S10	1	0.0001	0.0005	2	
S11	0.1	0.00001	0.00005	10	

to the manufacturers' instructions. The final PCR product was separated by electrophoresis on a 1% agarose gel (INVITROGEN, CA, USA) (100 V, 400 mA, 60 min). The amplicons were retrieved by excising the specific band from the gel and were purified using the QIAEX® Agarose Gel Extraction Kit (QIAGEN, Hilden, Germany).

Cloning and sequencing Two sequencing strategies have been used. On the one hand, the purified amplicons were directly sequenced using the t35S pCAMBIA c-R primer to get information on the sequences including the junction between the transgenic integrated cassette and the plant genome (direct sequencing). On the other hand, each purified amplicon was cloned into the pCR®2.1-TOPO® Vector using the TOPO TA Cloning® Kit (INVITROGEN, CA, USA) according to the manufacturers' instructions. A PCR was carried out on colonies using PCR™2.1-TOPO® and t35S pCAMBIA c-R primers and analysed by electrophoresis on a 1% agarose gel (INVITROGEN, CA, USA) (100 V, 400 mA, 60 min). The colonies possessing a fragment of the correct size were further cultured. The plasmids were extracted, using the QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany) according to manufacturers' manual, to be sequenced (classic sequencing). All sequencing reactions were performed on a Genetic Sequencer 3130XL using the Big Dye Terminator Kit v3.1 (Applied Biosystems, CA, USA) (Sambrook and Russell 2001; Broeders *et al.* 2013). The obtained sequences were aligned via the software "ClustalW2" and then analysed using the software "Nucleotide BLAST NCBI" (EMBL-EBI 2015; NCBI 2015).

Verification of the transgene flanking regions by PCR amplification The transgene flanking regions identified by DNA walking were verified by PCR amplification. The PCR was carried out using a reverse primer designed in downstream to the t35S pCAMBIA a-R primer on the pCAMBIA construct and a forward primer designed on the rice chromosome II or III (Table 4.2). These oligonucleotide primers were initially evaluated *in silico* using the program "wprimersearch" from the software "wEMBOSS" (EMBnet 2015). A standard 25 μl reaction volume is applied containing 0.625 U of DreamTaq™ DNA Polymerase (Fermentas, CA, USA), 1 × DreamTaq™ Buffer (Fermentas, CA, USA), 0.2 mM of dNTPs, 250 nM of each primer and 5 μl of DNA (10 ng/μl). The PCR program consisted of a single cycle of 10 min at 95°C (initial denaturation) followed by 35 amplification cycles of 30 s at 95°C (denaturation), 30 s at 60°C (annealing) and 1 min at 72°C (extension) and finishing by a single cycle of 10 min at 72°C (final extension). The run was performed on an iQ™5 real-time PCR detection system (BioRad, Hemel Hempstead, UK). The PCR products were analysed by electrophoresis on a 1% agarose gel (INVITROGEN,

CA, USA) (100 V, 400 mA, 60 min). The PCR products were purified using USB® ExoSAP-IT® PCR Product Cleanup (Affymetrix, CA, USA) according to the manufacturers' instructions. All sequencing reactions were performed on a Genetic Sequencer 3130XL using the Big Dye Terminator Kit v3.1 (Applied Biosystems, CA, USA) (Sambrook and Russell 2001; Broeders *et al.* 2013). The obtained sequences were analysed using the software "Nucleotide BLAST NCBI" (EMBL-EBI 2015; NCBI 2015).

4.3 Results and discussion

4.3.1 Selection of the t35S pCAMBIA screening marker and development of a qPCR SYBR®Green assay

Considering the high diversity of genetic elements integrated in GM rices, our attention was focused on rice transformation vectors. Because of its presence in 30% of transgenic plants and, more particularly, in 65 and 53 peer reviewed publications on GM rices in 2011 and 2012, respectively, the pCAMBIA family vector was considered as a strategic target to detect a large spectrum of unauthorised GMOs (Kathuria *et al.* 2007; Komori *et al.* 2007; Ahmad *et al.* 2012; Yu *et al.* 2012; Elsevier 2015). The t35S pCAMBIA screening marker was thus developed to identify unauthorised GMOs containing a pCAMBIA family cassette. The t35S pCAMBIA-specific SYBR®Green screening method, generating an amplicon of 137 bp, was performed for integration in to the CoSYPS (Combinatory SYBR®Green qPCR Screening) for GMO detection, composed of 18 SYBR®Green methods (RBCL, LEC, ADH, CRU, PLD, SAD1, GLU3, p35S, tNOS, pFMV, pNOS, t35S, Cry1Ab/Ac, Cry3Bb, pat, bar, epsps and CRT), which is able to run in a single 96-well plate (Vaïtilingom *et al.* 1999; Yang *et al.* 2005a; European Union Reference Laboratory for GM Food and Feed and Feed 2006a; Barbau-Piednoir *et al.* 2010; Van den Bulcke *et al.* 2010; Mbongolo Mbella *et al.* 2011; Barbau-Piednoir *et al.* 2012; Broeders *et al.* 2012a; Broeders *et al.* 2012b; Broeders *et al.* 2013).

Analysis of Bt rice *in silico* and *in vitro* The general structure of pCAMBIA vector is composed notably of p35S, tNOS and t35S elements (Cambia, Canberra, Australia 2015). Bioinformatics studies have shown that the common methods, including those described in the compendium of reference methods for GMO analysis, allow detection of p35S and tNOS (Data not shown, Table 4.3) (Höhne *et al.* 2002; Corbisier *et al.* 2005; Fernandez *et al.* 2005; Reiting *et al.* 2007; Waiblinger *et al.* 2008; Barbau-Piednoir *et al.* 2010; Joint Research Centre 2011).

Concerning the t35S element of the pCAMBIA family vectors, its sequence was slightly different at the 5' end compared to the authorised GMOs and LLPs events containing a t35S element (A2704-12, A5547-127, Bt11, Bt176, DAS59122, GHB119, LLRICE62, T25, TC1507 and Topas-19-2). Therefore, this element was not detected by the t35S SYBR®Green detection method developed previously in-house (Broeders *et al.* 2012a; personal communication). All these bioinformatics data were confirmed *in vitro* by qPCR SYBR®Green assay (Table 4.3).

Development of the t35S pCAMBIA screening marker In order to discriminate unauthorised GMOs containing pCAMBIA family vectors, the t35S pCAMBIA screening marker was developed. To this end, the sequence of the t35S pCAMBIA element was analysed. The majority of the pCAMBIA vectors (1200, 1201, 1281Z, 1291Z, 1300, 1301, 1302, 1303, 1304, 1380, 1381Xa, 1381Xb, 1381Xc, 1381Z, 1390, 1391, 1391Xa, 1391Xb, 1391Xc, 1391Z, 2200, 2300, 2301), except 0380 and 0390, possessed the t35S element. Its sequence was practically identical (slightly shorter by 5 bp at the 3' end for 2200, 2201, 2300 and 2301). The t35S pCAMBIA a-R and t35S pCAMBIA c-F primers were designed manually in the conserved region of the pCAMBIA family vector to discriminate exclusively this element (Tables 4.1 and 4.2). The specificity of this marker was tested initially *in silico* with the software wEMBOSS.

qPCR SYBR®Green assay To develop the t35S pCAMBIA marker, the specificity of t35S pCAMBIA c-F and a-R primers was tested *in vitro* on all authorised GMOs and LLPs events by the qPCR SYBR®Green assay (Tables 4.1 and 4.2). As expected, only the Bt rice, containing a pCAMBIA cassette, was detected after 40 cycles with a C_t value at 22.70 and a T_m value at 73°C, indicating that the screening marker was specific. All the other WT, GMO and LLP materials tested did not give a signal after 40 cycles. Then, the sensitivity of this marker was determined via the limit of detection with 6 repeats (LOD_6). The LOD_6 is defined as the amplicon copy number that affords a positive PCR result (expressed as C_t-value) upon six-fold measurement of the target sequence in the same DNA sample (Table 4.4). To this end, DNA from 100% Bt rice was diluted to 4 ng/μl and 4 independent dilution series were prepared (in nuclease-free water) starting from this concentration. The dilution series (from 1 to 0.00005 ng/μl of DNA) were prepared prior to setting up each of the qPCR runs. For each assay, a range from 2000 to 0.1 HGEs was tested in a qPCR SYBR®Green assay. The HGE content of the DNA extracts was calculated according to the size of the rice genome (0.5 pg) (Arumuganathan and Earle 1991). The LOD_6 was obtained at 5 HGEs (corresponding to 0.0025% of unauthorised GMOs) with a mean C_t value of 35.28 Ct and a mean T_m

value of 72.52°C. Because the LOD_6 was observed below a HGE of 20, the t35S pCAMBIA screening marker was considered as being sensitive. The experiment was also evaluated as repeatable according to the RSD_r percentage (relative standard deviation, RSD, of the test results). In addition, the PCR efficiency (80%) and the linearity ($R^2 = 0.9954$) were assessed as acceptable.

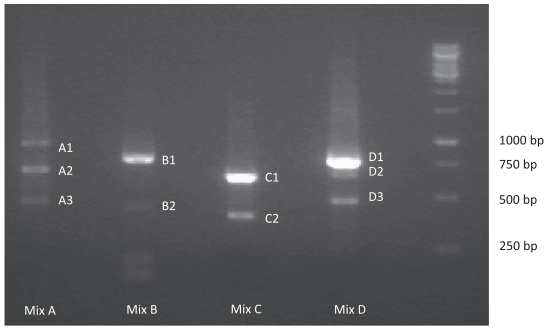
4.3.2 DNA walking approach

Following a positive signal observed in qPCR SYBR®Green assay, the second step was to characterise the junction between the transgenic cassette and the plant genome to confirm the presence of pCAMBIA unauthorised GMOs in food/feed matrices (Figure 4.3). Therefore, a DNA walking approach has been developed.

***In silico* study** In order to supply an integrated approach, an additional oligonucleotide primer, named t35S pCAMBIA b-R, was designed manually, on the conserved region of the t35S pCAMBIA sequence, localised between the t35S pCAMBIA a-R and t35S pCAMBIA c-R primers previously used for the qPCR SYBR®Green assay (Figure 4.1 and Table 4.2). The specificity of this primer was confirmed *in silico* via the software wEMBOSS (data not shown).

Characterisation of the junction The amplicons resulting from the double semi-nested PCR were visualised on a 1% agarose gel. For each kind of DRT primers mixes (A–D), amplicons were observed with an approximate size of 300 bp up to 1000 bp (Figure 4.2A). The identity of the amplicons was confirmed by direct sequencing of the purified PCR products. The sequencing of the plasmids containing these amplicons allowed identifying the t35S pCAMBIA c-R and UAP-N1/UAP-N2 primers and determining the exact size of the amplicons (408–944 bp) (Figure 4.2). All analysed amplicons presented a sequence including the junction between the pCAMBIA vector and the rice genome. Two transgenic insertions have been detected. For the majority of the amplicons (A2, A3, B1, C2, D1, D2 and D3), the pCAMBIA cassette was integrated on a genomic sequence (OSJNBb0111B07) from the chromosome III of *O. sativa* japonica Group coding for a putative uncharacterised protein. For the three other amplicons (A1, B2 and C1), the transgene flanking region was localised on a genomic sequence (OSJNBa0016G10) from the chromosome II of *O. sativa* japonica Group coding for a putative uncharacterised protein. These transgene flanking regions present a shorter left ends compared to the pCAMBIA cassette situated on the chromosome III. This variability of length could be explained by the fact that a left end integrates less precisely than a

A



B

>Chromosome II

GGAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGAACACGCGTCGTTTACCTTCGGGGCGTCATCCTCGACGACTTCGACGACG
GCGACGCCAAAGTTCCTTCGCCTAGTCGTCAAACGCCAAGGCCAAAACTTACACGCCAAGAAGAGCTAGCAGCAGATCATTAAACAAAA
CGGCACACCTCACCGACAATCCAATACTCCTACGAAATACCGATGCGAGGATTATACTAGATCCCACCGGGCCCACACGTCAGCGACCCT
CGTGCGACCCCAAAGGCGAGGAGCGGAGGCCGCCGCCCTCCGGCGGTGTGCTCTCCTCGGCCTTGCGCGAAAAGAAGATGGCAGGATT
AAAATACTCTAGGATATTCCGGAGCCCCCTTATCCAGTTCTCGTCACTGGCACGTGGGCCCCACCGTGGGCCCCACGGGAGGAAAAGGA
TGGGGGACCGCACGAGTTTTTACACTCTCACCCGTAACATTACACGTAATAAATTTCACATATACCCCTGTGGGGGTTTTTACACGCAGCC
CCCTCTACGGCTCCCCCGCGAGCGTGGCCCGCGCCGCTGACACGTGGGCCCCCAGCCGACTTAATGGCCCAACAACGCCGAATTAATTCG
GGGGATCTGGATTTTAGTA

>Chromosome III

GGAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGAACACGCGTCGTTTACCTCCGGGGACCCCCTGACATCAAAACAAAAGCAG
CACCAATTTAAATTTTCCCTATTTCAGCCAAGAAATGAAACTGCAATGTTGCAATGCAATGCCATTTGGTATGTGCAGTGTGTGCAGGAG
GAGGAGACACGAGCGGCAGCTATAGCCGGAGATTTAACCTTGATGAAGGTGAGTCACTTACTACAATGCAGTACTTGCTGCTGCGGACG
TAGCTGCTGTGGATGACCTGCACGAACACCGGGTTCCCGGGCTGAACCGGCACGGGGACGCGCGAGATCTCCTCCCTGTCTTCCTCCCTC
AGGTTGTACTCGGCCACCCGTCCGTGCCTGCTGAAGTAGTACTCCCCGTAGCGCCCGCCTTCCGCGGCCGCGACAACATCTTCATCTTCCC
CATCGTTTTCTTTCGCTTCTGCAGGTGGCAGACACGCATATAAGTTACAATGCCGATTTGGGAGCTTTTAGCTGCTGCATATTCTACCACAT
TGTTACCCAGATTCTCCTAATATGTAGTTGTATAATCCAGAAAATAAACTAGAACCAGAAGCTAGAAAACCTAGCTTCTCCAAATTCTTAG
AAGCTGACTTCCAAACAGCGGATTCTCATCTTAAGCTCACCAAACAGGGCCTATTGTGGTGTAAACAAATTGACGCTTAGACAACTTAAT
AACACATTGCGGACGTCTTTAATGTACTGAATTAACGCCGAATTAATTCGGGGGATCTGGATTTTAGTA

Figure 4.2: Characterisation of the junction between the integrated transgenic pCAMBIA cassette and the rice genome.
(A) Visualisation of the amplicons obtained with the different DRT mixes (A–D) on a 1% agarose gel. (B) Amplicon sequences presenting the junction between the pCAMBIA 1300 vector (underlined) and the rice genome identified on the chromosome II and the chromosome III, respectively. The t35S pCAMBIA c-R (in bold) and the UAP-N1/UAP-N2 primers are dotted-underlined. These sequences were obtained by classic sequencing of the plasmids.

right end (Gheysen *et al.* 1991; Krizkova and Hrouda 1998). To confirm the two chromosomal insertions, a PCR amplification using primers annealing to the pCAMBIA construct and the rice chromosome II or III was carried out (Table 4.2). The presence of PCR amplification as well as the sequencing of

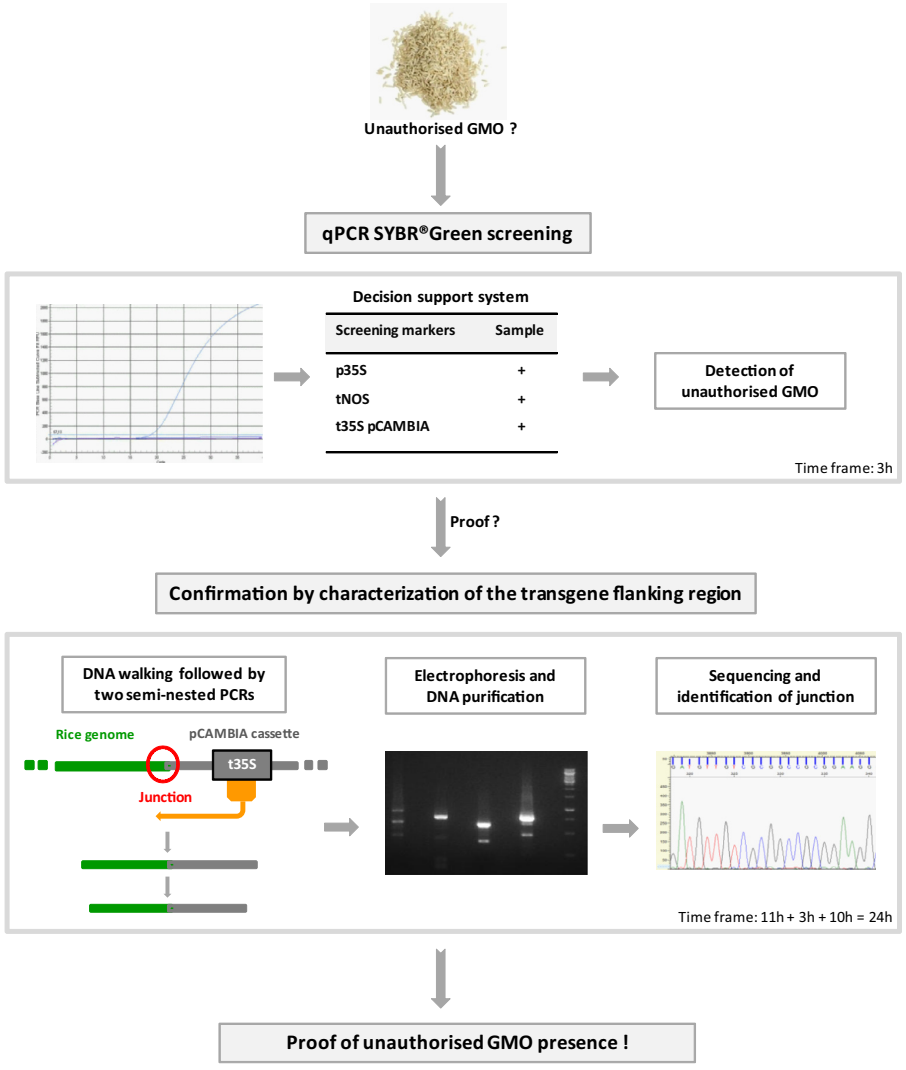


Figure 4.3: Workflow of the present integrated approach.

these amplicons allowed verifying properly the transgene flanking regions (data not shown).
As all the obtained amplicons provided the transgene flanking regions, these results demonstrated the high efficiency, specificity and reliability of the present integrated approach. In addition, it's interesting to note that the two different transgene flanking regions were not identically detected in function of the DRT primers used. Indeed, all the types of DRT primers allowed identifying the junction on the chromosome III while only the A, B and C DRT primers have detected the junction on the chromosome II. This system using four different DRT primers thus presents the advantage to increase the likelihood to detect unauthorised GMOs, independently of the tested matrices.
The proposed strategy is based on the presence of known transgenic elements. Consequently, the success of this integrated approach is limited to the knowledge level of transgenic elements making up unauthorised GMOs. Therefore, in spite of the good performance of this method, the strategy is not appropriate to detect GMOs constituted of only unknown elements. To this end, other technologies are more suitable such as "Next Generation Sequencing" methods. However, this last technique is at the present time not easily implementable in GMO routine analysis due to its high cost and its long time frame for data processing.

4.4 Conclusion

Considering the numerous unauthorised GM rices detected in food/feed matrices on the EU market listed in 2012, as well as their expected increase in the coming years, this study supplies to the enforcement laboratories a strategy to ensure the unauthorised GMO detection in the food and feed chain in semi-routine analysis (Stein and Rodríguez-Serezo 2009; The European Commission 2015b). The proposed integrated approach is composed of two main steps (Figure 4.3). On the one hand, the potential presence of unauthorised GMOs, containing a pCAMBIA family vector, in food/feed matrices is detected via the qPCR SYBR®Green technology. The key choice to target the pCAMBIA family vector, via its element t35S, will allow detection of a large spectrum of unauthorised GMOs. The t35S pCAMBIA marker was developed to be specific, sensitive, efficient, repeatable and to be integrated into the CoSYPS. On the other hand, once this marker is indicated as positive for a given food/feed matrix, the potential presence of unauthorised GMOs, containing a pCAMBIA vector, is demonstrated by the characterisation of the junction between the integrated cassette and the plant genome using a DNA walking method starting from the t35S pCAMBIA a-R primer. This method is then followed by two semi-nested PCR rounds using the t35S pCAMBIA b-R and t35S pCAMBIA c-R primers, respectively. With regard to the previous articles describing methods

characterising the junction sequences of GMOs, the present DNA walking approach possesses the substantial advantage to be easily implementable in semi-routine use thanks to the simplicity of a method exclusively based on PCR. In addition, its short time frame to get the results (less than three days including DNA walking method, DNA purification and direct sequencing) and its relatively low cost clearly represent a crucial benefit for the enforcement laboratories.

Acknowledgements

This research is funded by the Federal Public Service Health, Food Chain Safety and Environment (convention RF 11/6242) through the Project UGMMONITOR. Sequencing is performed at the Platform Biotechnology and Molecular Biology at the Scientific Institute of Public Health. The authors would like also to thank Emmanuel Guiderdoni (from Biotrop and Crop Protection Programmes, Cirad-Amis, France) for his kindness to provide rice grains, Nina Papazova and Sylvia Broeders for their precious help, Maud Delvoye for her technical assistance and Inge Huyghe for the critical review of the manuscript.

Chapter 5

Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products

The approach developed in the previous chapter was here tested on GM rice present in various percentages in typical rice products in order to assess its feasibility in GMO routine analysis.

This chapter was previously published as:
M.-A. Fraiture, P. Herman, I. Taverniers, M. De Loose, F. Van Nieuwerburgh, D. Deforce, and N. H. Roosens (2015a). "Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products". *Food chemistry* 173, pp. 1259–1265. DOI: 10.1016/j.foodchem.2014.09.148

Authors' contributions

M.-A. Fraiture performed experiments, analysed data and drafted the manuscript. P. Herman, I. Taverniers, M. De Loose, D. Deforce and N. H. Roosens helped to design the study and drafted the manuscript. All authors read and approved the final manuscript.

Abstract

To identify unauthorised GMOs in food and feed matrices, an integrated approach has recently been developed targeting pCAMBIA family vectors, highly present in transgenic plants. Their presence is first assessed by qPCR screening and is subsequently confirmed by characterising the transgene flanking regions, using DNA walking. Here, the DNA walking performance has been thoroughly tested for the first time, regarding the targeted DNA quality and quantity. Several assays, on model food matrices mimicking common rice products, have allowed to determine the limit of detection as well as the potential effects of food mixture and processing. This detection system allows the identification of transgenic insertions as low as 10 HGEs and was not affected by the presence of untargeted DNA. Moreover, despite the clear impact of food processing on DNA quality, this method was able to cope with degraded DNA. Given its specificity, sensitivity, reliability, applicability and practicability, the proposed approach is a key detection tool, easily implementable in enforcement laboratories.

5.1 Introduction

In the European (EU) market, to preserve freedom of choice for the consumer and to protect the food and feed chain, authorisations for genetically modified (GM) food or feed products are subjected to the EU legislation (The European Parliament and The Council Of The European Union 2001; The European Parliament and The Council Of The European Union 2003a; The European Parliament and The Council Of The European Union 2003b). The commercialisation as well as the detection of genetically modified organisms (GMOs) make up an integrated part of these EU regulations. In this context, several methods to detect GMOs in food and feed matrices have been developed. These methods, based essentially on real-time PCR technologies, target the most common elements present in GMOs (Broeders *et al.* 2012a). Screening methods, usually used as the first step in GMO analysis, allow the detection of GMOs in a given food/feed sample. In case of positive responses for some GM targets, it allows to narrow down the number of EU-authorised GM events to be identified using event-specific methods in a second step (Broeders *et al.* 2012b; Joint Research Centre 2015a). The screening step can also indirectly indicate the potential presence of EU-unauthorised GM events. Indeed, if the transgenic elements identified during the screening step cannot be explained by EU-authorised events, the presence of unauthorised GMOs can be suspected (The European Commission 2011a; Broeders *et al.* 2012a). However, screening methods are only able to provide an indirect proof of GMO presence. In addition, targeted screening elements often originate from natural organisms, such as the Cauliflower Mosaic Virus (CaMV). As a consequence, an "unexplained" transgenic element identified in the screening (*e.g.* p35S) might also find its origin in the presence of the corresponding donor organism (*e.g.* CaMV) in a food/feed sample (Broeders *et al.* 2012a; Broeders *et al.* 2012b). Therefore, the presence of GMO in food and feed matrices can only be confirmed by the identification of the junction between the transgenic integrated cassette and the plant genome, which represents the unique signature of a GMO.

In order to prove indubitably the presence of an unauthorised GMO, different DNA walking methods have been carried out on transgenic plants (Zhang *et al.* 2012; Spalinskas *et al.* 2013b; Fraiture *et al.* 2014 and references therein). However, most of these DNA walking strategies are not easily implementable in GMO routine analysis by the enforcement laboratories for several reasons. First, some techniques are laborious, complex or insufficiently specific. Second, although some of them are based on PCR, which is simple, mastered and frequently used by the enforcement laboratories, DNA walking approaches are not integrated into the screening strategy. In addition, their ability to detect low amounts of target is expected to be weak (Spalinskas *et al.* 2013b). To design a strategy corresponding to the need of the enforcement laboratories, we

have recently developed an integrated approach to rapidly detect and identify unauthorised GMOs in food and feed matrices in two main steps (Fraiture *et al.* 2014). First, a qPCR SYBR®Green screening allows to determine the potential presence of unauthorised GMOs in a given sample. Second, their presence is confirmed by DNA walking. This last method is based on PCR, which is applied using a first reverse primer followed by two semi-nested PCR rounds using primers that are each time nested to the previous reverse primer. This approach is integrated since the same primer sequences are used for both the unauthorised GMO detection (screening method) and its characterisation (DNA walking method). Moreover, the use of two semi-nested PCR rounds allows us to increase the yield and the specificity of the method, especially in the case of a low level presence of GMOs.

The detection of a target using DNA-based methods implies limitations related to the nature of the tested food/feed sample. Two different categories of limitations can affect this detection system (Ballari and Martin 2013; Fernandes *et al.* 2013). On the one hand, the detection of a weak concentration of the target in food and feed matrices, which is usually the case for unauthorised GMOs, requires a method sufficiently sensitive (ENGL *ad hoc* working group on "unauthorised GMOs" 2011; Broeders *et al.* 2012a). On the other hand, the performance of the detection method could also be affected by the state of the sample. Indeed, food processing, defined as any food manipulation step (physical, chemical or mechanical) from the raw material to the final product, is known to induce DNA damage. This process could thus reduce the size of the obtained amplicons due to the fragmentation of high molecular weight DNA strands (Lipp *et al.* 2001; Peano *et al.* 2004; Gryson 2010; Ruttink *et al.* 2010a; Arun *et al.* 2013; Ballari and Martin 2013; Fernandes *et al.* 2013). To our knowledge, few of these potential impacts (quantity and quality of the target) on DNA walking methods have been examined to date (Nielsen *et al.* 2008).

Previously, we have developed an integrated approach, based on DNA walking, to identify unauthorised GMOs (Fraiture *et al.* 2014). In the extension of this study, the sensitivity of the method has been assessed using different amounts of Bt rice. In addition, the method has been tested on model food matrices mimicking rice food mixtures (rice and maize) and processed rice food (rice flour and rice noodles). In this way, the present study will provide the enforcement laboratories with crucial information concerning the applicability, the practicability and the dynamic range of the proposed method (ISO 24276 2006). To our knowledge, the performance criteria measured in the present study, applied to unauthorised GMOs, have never been described for DNA walking methods published so far.

5.2 Materials and methods

5.2.1 Reagents and devices

Genomic-tip20/G was purchased from QIAGEN (Hilden, Germany). SYBR®Green PCR Mastermix was provided by Diagenode (Liège, Belgium). APAgene™ GOLD Genome Walking Kit was obtained from BIO S&T (Montréal, Canada). Agarose was bought from INVITROGEN (CA, USA). Wizard® SV Gel and PCR Clean-Up System and pGEM®-T Easy Vector Systems were purchased from PROMEGA (WI, USA). Big Dye Terminator Kit v3.1 was obtained from Applied Biosystems (CA, USA). Nanodrop® 2000 (Thermo-Fisher, DE, USA), iQ™5 real-time PCR detection system (BioRad, Hemel Hempstead, UK) and Genetic Sequencer 3130XL were used.

5.2.2 Plant materials and sample preparation

As previously described, transgenic Bt rice grains (*Oryza sativa* L. *Japonica cv Ariete*), transformed with the binary vector pCAMBIA1300 harbouring the synthetic Cry1B gene to confer an insect resistance, and its wild-type (WT) were used in this study to develop and optimise the methodology (Breitler *et al.* 2004; Fraiture *et al.* 2014).

From these rice grains, in-house rice flour and noodles were prepared by mixing WT rice grains and Bt rice grains to obtain rice products containing 0%, 0.1% or 1% of Bt rice (w/w) (Figure 5.2). First, to prepare the rice flour, one volume of rice grain samples, containing 0%, 0.1% or 1% of Bt rice, mixed with two volumes of Milli-Q water were incubated at room temperature (RT) overnight. The mixtures were then ground, filtered and dried at RT. Second, to prepare rice noodles, the previously described rice flours were mixed with warm water to obtain a homogenous dough allowing us to form noodles (Quynh 2010). The fresh noodles were then dried at 30°C with ventilation for 30 min in order to mimic the traditional sun-drying method (Hsieh and Luh 1991). As control, unprocessed WT grain samples, containing 0%, 0.1% or 1% of Bt rice grains compared to WT rice grains (w/w), were also prepared.

The Certified Reference Materials (CRM) of the WT maize (non GM MON810 maize counterpart (ERM-BF413ak)), in the form of seed powders, were obtained from the Institute for Reference Materials and Measurements (IRMM, Geel, Belgium 2015). This material was characterised as previously described (Broeders *et al.* 2013).

5.2.3 DNA extraction, concentration and purity

DNA, from homogeneous powder of ground rice grains, rice flour and rice noodles, was extracted using a CTAB-based procedure (ISO 21571), in combination with the Genomic-tip20/G, which was adapted from the EU-RL GMFF (European Union Reference Laboratory for GM Food and Feed) validated method (ISO 21571 2005; European Union Reference Laboratory for GM Food and Feed and Feed 2006b). DNA concentration was measured by spectrophotometry using the Nanodrop® 2000 device and DNA purity was evaluated as falling within the acceptance criteria according to the A260/A280 (∼1.8) and A260/A230 (∼2.0–2.2) ratios. DNA extraction, concentration and purity of WT maize were carried out as previously described (Broeders *et al.* 2013).

5.2.4 qPCR SYBR®Green assay

All qPCR assays were carried out in a standard 25 μl reaction volume containing 1 × SYBR®Green PCR Mastermix, 250 nM of each primer and 5 μl of DNA. The qPCR cycling programme consisted of a single cycle of DNA polymerase activation for 10 min at 95°C, followed by 40 amplification cycles of 15 s at 95°C (denaturing step) and 1 min at 60°C (annealing-extension step). Melting curve analysis was performed by gradually increasing the temperature from 60 to 95°C over 20 min (±0.6°/20 s) (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2013). All runs were performed on an iQ™5 real-time PCR detection system. For each assay, a "No Template Control" (NTC) was included.

As a decision support system to identify pCAMBIA unauthorised GM rice, the *p35S* (Promoter of the 35S cauliflower mosaic virus; Forward: AAAGCAAGTGGATTGATGTGATA; Reverse: GGGTCTTGCGAAGGATAGTG), *tNOS* (Terminator of the nopaline synthase gene; Forward: GATTAGAGTCCCGCAATTATACATTTAA; Reverse: TTATCCTAGKTTGCGCGCTATATT), *pld* (Phospholipase D gene from rice; Forward: GCTTAGGGAACAGGGAAGTAAAGTT; Reverse: CTTAGCATAGTCTGTGCCATCCA) and *t35S* pCAMBIA (Forward: CGGGGGATCTGGATTTTAGTA; Reverse: AGGGTTCCTATAGGGTTTCGCTC) markers were used on 100 ng of DNA (Barbau-Piednoir *et al.* 2010; Mbongolo Mbella *et al.* 2011; Fraiture *et al.* 2014).

In addition, the potential presence of inhibitors in the analysed rice (grains, flours and noodles) and maize materials was evaluated via an inhibition test, based on the difference of C_t values (δC_t) as a function of the DNA concentration (Broeders *et al.* 2012b). To this end, all rice and maize materials were analysed at two different DNA concentrations (100 and 10 ng) by qPCR, using the *pld* and *adh* (Alcohol dehydrogenase I gene from maize; Forward: TCTCTTCCTCCTTTAGAGCTACCACTA; Reverse:

AATCGATCCAAAGCGAGATGA) markers, respectively (Mbongolo Mbella *et al.* 2011; Broeders *et al.* 2012a). The inhibition was excluded if the calculated $\delta\delta C_t$ value, based on the difference between the observed δC_t value and the theoretical δC_t value (corresponding to 3.3 for a dilution 10), was equal to or less than 0.5 (Broeders *et al.* 2012b).
Moreover, qPCR assays, using the *pld* and *t35S* pCAMBIA markers, were carried out on the samples analysed by DNA walking. Only the DNA from WT maize was tested at 100 ng instead of 520 ng, due to the presence of inhibitors (data not shown).

5.2.5 DNA walking approach

General As previously described, the present DNA walking strategy was performed using a first reverse target-specific primer (*t35S* pCAMBIA a-R: AGGGTTCCTATAGGGTTTCGCTC) and a degenerated random tagging primer (DRT). Two semi-nested PCR rounds were then applied using target-specific primers (*t35S* pCAMBIA b-R: GTGTTGAGCATATAAGAAACCC; *t35S* pCAMBIA c-R: TACTAAAATCCAGATCCCCCG), that are each time nested to the previous reverse target-specific primer, combined with universal tagging primers (UAP-N1 and UAP-N2) (Fraiture *et al.* 2014). PCR mixes and conditions were carried out according to the manufacturers' instructions of APAgene™ GOLD Genome Walking Kit. The final PCR product was separated by electrophoresis on a 1% agarose gel (100 V, 400 mA, 60 min).

Assessment of the method sensitivity, the impact of food mixture and food processing First, to determine the limit of detection (LOD), the DNA walking approach was carried out on 100 ng of rice DNA templates composed of 100% of Bt rice (200 000 haploid genome equivalents (HGEs) of Bt rice) or WT rice containing 1% (2000 HGEs of Bt rice), 0.1% (200 HGEs of Bt rice), 0.05% (100 HGEs of Bt rice), 0.025% (50 HGEs of Bt rice), 0.001% (20 HGEs of Bt rice), 0.005% (10 HGEs of Bt rice) or 0% of Bt rice. In addition, as the sensitivity of the DNA walking method may be limited by the visualisation of obtained amplicons on agarose gel, a cloning strategy of direct ligation was carried out on final PCR products from the four different DRT primer mixes at a Bt rice concentration of 0.025% (50 HGEs). Second, the potential effect of background DNA from food mixtures was tested by mixing WT maize DNA with 200 000 HGEs, 2000 HGEs, 200 HGEs or 0 HGE of Bt rice DNA. Finally, to evaluate the potential impact of food processing, the DNA walking method was applied to 100 ng of DNA from "home-made" WT rice flour and noodles containing 1%, 0.1% or 0% of Bt rice. The WT rice grain samples containing 0%, 0.1% or 1% of Bt rice were used as a control. The HGE contents of the DNA extracts

were calculated according to the size of the rice genome (0.5 pg) and the maize genome (2.6 pg) (Arumuganathan and Earle 1991). A NTC was included in the assay.

Cloning and sequencing Two different sequencing methods were applied. On the one hand, the amplicons were excised from the gel and were purified using the Wizard® SV Gel and PCR Clean-Up System. The purified amplicons were directly sequenced using the *t35S* pCAMBIA c-R primer or cloned into the pGEM®-T Easy Vector Systems, according to the manufacturers' instructions, in order to be then sequenced. On the other hand, the final PCR products from the DNA template 0.025% (50 HGEs) and 0% (WT rice) of Bt rice were cloned into the pGEM®-T Easy Vector Systems. A PCR was carried out on colonies using pGEM®-T Easy Vector (T7: TAATACGACTCACTATAGGG; SP6: ATTTAGGTGACACTATAGAAT) combined with rice primers (Rice chromosome II: CGAAAAGAAGATGGCAGGAT; Rice chromosome III: TTTCTTTCGCTTCTGCAGGT) and was analysed by electrophoresis on a 1% agarose gel (100 V, 400 mA, 60 min). The colonies presenting a fragment of the correct size were then sequenced. All sequencing reactions were performed on a Genetic Sequencer 3130XL using the Big Dye Terminator Kit v3.1. The sequences were aligned via the software "ClustalW2" and analysed using the software "Nucleotide BLAST NCBI" (EMBL-EBI 2015; NCBI 2015).

5.3 Results and discussion

5.3.1 General

To assess the potential impact of the quality and quantity of targeted DNA on the DNA-based method, the developed DNA walking strategy identifying pCAMBIA-transformed unauthorised GMOs was thoroughly assessed. To this end, three different assays have been established in order to evaluate the limit of detection (LOD) of this method as well as the effect of food mixtures and food processing. All tested samples were first analysed by qPCR using a decision support system to confirm the identity of the materials (Fraiture *et al.* 2014). This system, based on three screening markers (p35S, tNOS and t35S pCAMBIA), allows to effectively confirm the presence of the pCAMBIA target in the sample. In addition, an inhibition test was applied on the rice and maize materials using the endogenous *PLD* and *ADH* markers respectively. Based on the obtained C_t values, the calculated $\delta\delta C_t$ values of rice grains (0.23), rice flour (0.19), rice noodles (0.48) and maize (0.24) were less than 0.5, indicating that none of these samples at 100 ng of DNA was subject to inhibition.

5.3.2 Sensitivity assessment

In order to determine the LOD of the DNA walking approach, DNA samples containing 100% to 0% of Bt rice were prepared and analysed (Table 5.1).
As the first step of GMO analysis in enforcement laboratories, all these DNA samples were submitted to a qPCR SYBR®Green analysis using the *PLD* and *t35S* pCAMBIA screening markers in order to identify potential unauthorised GMOs. As expected, all samples containing Bt rice (100% to 0.005%) presented a positive signal, inversely proportional to the DNA target amount, for the two markers. Moreover, the WT rice (0%) showed only a positive signal of the same amplitude for the *PLD* marker and no signal was observed for the NTC with these screening markers (Table 5.1; data not shown).
As a second step to confirm the presence of unauthorised GMOs, the DNA walking approach was applied to these samples using the four different DRT primers (A–D) (Figure 5.1; Table 5.1). The first analysis of the 100% Bt rice (200 000 HGEs) had shown the presence of two different insertions of the pCAMBIA vector in the rice genome. Indeed, as previously mentioned, the transgenic cassette was integrated in the *Oryza sativa* japonica genome in chromosome III (OSJNBb0111B07) and chromosome II (OSJNBa0016G10) (Fraiture *et al.* 2014). In the present study, the sensitivity of this method was assessed using samples containing 1% (2000 HGEs) to 0.005% (10 HGEs) of Bt rice in order to determine its LOD. Up to 100 HGEs of Bt rice (0.05% Bt rice), the transgene flanking regions were detected on both chromosomes II and III via DRT A and D primers, respectively. At this Bt rice concentration, the size of the remaining amplicons was approximately 950 and 850 bp corresponding to the junctions on chromosomes II and III, respectively. With a lower amount of Bt rice, only the transgene flanking region on the chromosome III was identified as low as 10 HGEs of Bt rice (0.005% Bt rice) using DRT D primer mix. Among the tested materials, DRT B and C primers presented a weaker sensitivity compared to DRT A and D primers. Indeed, the transgene flanking regions on chromosomes II and III were detected as low as 200 HGEs of Bt rice (0.1% Bt rice) using DRT C and B primers, respectively. The decrease of target concentration thus seems to affect the detection power of the DNA walking approach. This performance drop was related to the affinity of the DRT primers mixes used and not to the size of the amplicons.
As the sensitivity of the method may be limited by the visualisation of PCR amplicons on agarose gel, a cloning strategy of PCR products by direct ligation was applied on a Bt rice amount of 0.025% (50 HGEs) where no amplicon was obtained with the different DRT primers, except for the D mix. However, a maximum of 3.3% of tested colonies from the A, B and C mixes (1 over 30) allowed to detect the transgene flanking regions on the chromosome II (DRT A primer) and chromosome III (DRT B primer) while 80% of tested colonies

from the DRT D mix presented the transgenic insertion on the chromosome III. This strategy did not thus allow to improve the sensitivity of the present method since the majority of the cloned amplicons corresponded to an aspecific amplification when no PCR products were visible on agarose gel.
These results demonstrate a high sensitivity of the present DNA walking method (as low as 100 HGEs and 10 HGEs for the junction on chromosomes II and III, respectively) that is crucial for detecting traces of unauthorised GMOs in food matrices. In addition, the dynamic range of the proposed method depends essentially on the DRT primer mixes used and not on the size of the obtained amplicons. For instance, only mix A was still able to identify the pCAMBIA insertion on the chromosome II at 100 HGEs of Bt rice while the transgenic junction on the chromosome III was detected only by the mix D from 100 to 10 HGEs of Bt rice (Table 5.1). Moreover, this last DRT mix cannot provide the junction on the chromosome II at any concentration of the target, as previously observed by Fraiture *et al.* (2014). These data highlight the importance of using the four different DRT primers, as the sensitivity of the method depends mainly on the affinity of the primers, in order to increase the probability of detecting the two different transgene flanking regions, especially in the presence of low amounts of target. Concerning the specificity of the method, no aspecific amplification was observed, except for the WT rice (0% Bt rice) combined with the DRT D primer mix (Figure 5.1). The obtained amplicon corresponds to an *Oryza sativa* Japonica genomic sequence of chromosome X (OSJNBa0050N08.1) which codes for a putative retro-element protein. It was generated by the adventitious presence of the UAP-N2 integrated tag near a rice sequence which is its reverse complement in order to allow this amplification using the UAP-N2 primer (data not shown). In addition, as expected, the NTC presented no amplification, suggesting that no reaction between the used primers was generated, independently of the DRT primers used.

5.3.3 Study of potential food mixture impact on DNA walking method

In order to evaluate the potential impact of a food mixture on the DNA walking approach, DNA samples composed of 200 000 HGEs of WT maize mixed with 200 000 HGEs, 2000 HGEs, 200 HGEs or 0 HGE of Bt rice were prepared. The WT maize was chosen as DNA background given its frequent presence in rice-based products, such as vermicelli. In addition, the qPCR SYBR®Green analysis using the *PLD* and *t35S* pCAMBIA screening markers, was applied (Table 5.2). The obtained signals were inversely proportional to the DNA target amount present in the tested matrices. As expected, the results were comparable in the presence and absence of maize (Table 5.1).

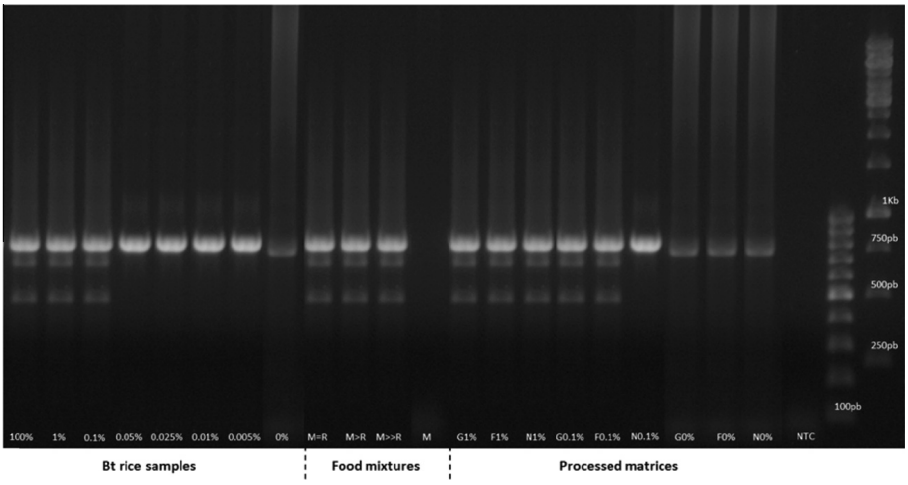


Figure 5.1: Visualised amplicons on 1% agarose gel from all the tested matrices using the DRT D primer mix: (i) Bt rice (200 000 HGEs; 100%) and WT rice containing 2000 (1%), 200 (0.1%), 100 (0.05%), 50 (0.025%), 20 (0.01%), 10 (0.005%) or 0 HGEs (0%) of Bt rice; (ii) WT maize containing 200 000 (M = R), 2000 (M > R), 200 (M $\gg$ R) or 0 HGEs (M) of Bt rice; (iii) WT rice grains (G), flour (F) and noodles (N) containing 1%, 0.1% or 0% of Bt rice. The "No Template Control" is symbolised by NTC.

Table 5.1: Sensitivity analysis of the DNA walking method using the four different mixes of DRT primers (A–D). For each tested sample, the Bt rice amount in WT rice is indicated in percentage (100% to 0%) with its corresponding Bt rice HGEs. The obtaining of these samples was based on the HGEs. Following the qPCR analysis using the *PLD* and *t35S* pCAMBIA screening markers, the observed C_t values with the standard deviations are indicated for each sample. The detection of transgene flanking regions on the chromosomes II (X2) or III (X3) is symbolised by + (3/3), (+) (1–2/3) or −(0/3). For each result, the experiment was carried out in triplicate. The approximate size of amplicons, only obtained three times, is indicated between brackets in base-pair under the corresponding signal.

		100%	1%	0.1%	0.05%	0.025%	0.01%	0.005%	0%
		200 000 HGEs	2000 HGEs	200 HGEs	100 HGEs	50 HGEs	20 HGEs	10 HGEs	0 HGEs
PLD		17.2 ± 0.4	17.3 ± 0.4	17.4 ± 0.4	17.9 ± 0.7	17.4 ± 0.3	17.3 ± 0.4	17.3 ± 0.3	17.3 ± 0.3
t35S pCAMBIA		21.8 ± 0.3	27.6 ± 0.5	31.9 ± 0.1	33.9 ± 0.2	34.9 ± 0.3	36.0 ± 0.6	37.1 ± 0.1	0.0 ± 0.0
Mix A	X2	+ *(950)*	+ *(950)*	+ *(950)*	+ *(950)*	−	−	−	−
	X3	+ *(700; 500)*	+ *(500)*	+ *(700)*	−	−	−	−	−
Mix B	X2	+ *(450)*	(+)	−	−	−	−	−	−
	X3	+ *(800)*	+ *(800)*	+ *(800)*	−	−	−	−	−
Mix C	X2	+ *(650)*	+ *(650)*	(+)	−	−	−	−	−
	X3	+ *(450)*	−	−	−	−	−	−	−
Mix D	X2	−	−	−	−	−	−	−	−
	X3	+ *(850; 800; 500)*	+ *(850; 800; 500)*	+ *(850; 800; 500)*	+ *(850)*	+ *(850)*	+ *(850)*	+ *(850)*	−

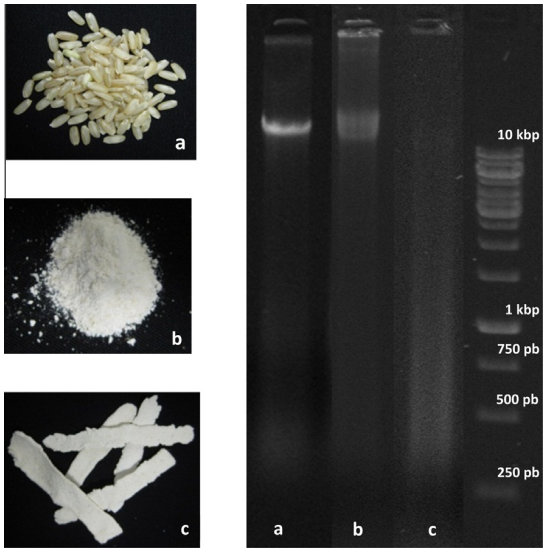


Figure 5.2: Genomic DNA (100 ng) extracted from rice grains (a), flour (b) and noodles (c).

A DNA walking assay was carried out on these samples using the four different DRT primers (A–D) (Figure 5.1; Table 5.2). For all the samples containing Bt rice, the transgene flanking regions on the chromosome II and III were detected similarly to samples composed exclusively of rice (Table 5.1). However, only the pCAMBIA insertion on the chromosome II was not identified at 2000 HGEs (1%) of the target using the DRT B mix. Furthermore, no aspecific amplification was generated.
Moreover, the sizes of the obtained amplicons were similar to the corresponding samples without maize (Figure 5.1; Tables 5.1 and 5.2).
These results thus indicate the high specificity and reliability of the developed DNA walking approach applied to a food matrix, independently of the addition of another ingredient.

5.3.4 Study of potential food processing impact on DNA walking strategy

As food processing is well-known to cause DNA damage, the DNA walking approach was carried out on rice matrices processed in-house (Ballari and Martin 2013). In order to gradually test the potential effect of food processing,

rice flour and rice noodles were generated. These products are composed of WT rice and 1%, 0.1% or 0% of Bt rice. As a control, unprocessed WT rice samples (rice grain), containing 1%, 0.1% or 0% of Bt rice, were also prepared. To evaluate the impact of food processing on DNA quality, extracted DNAs were observed on 1% agarose gel (Figure 5.2). Although the DNA yields were similar, DNA degradation followed the level of food processing. Indeed, rice flour and rice noodles presented a DNA slightly and strongly degraded compared to the unprocessed materials, respectively.

First, these processed food products were analysed by qPCR SYBR®Green using the *PLD* and *t35S* pCAMBIA markers (Table 5.3). As expected, the Bt rice samples presented a positive signal for the two screening markers (*PLD* and *t35s* pCAMBIA), inversely proportional to the amount of target. A difference of C_t was observed between the unprocessed and processed samples, suggesting an impact of food processing.

Second, the DNA walking approach was evaluated on processed rice products (Figure 5.1; Table 5.3). Compared to the unprocessed materials, the detection power of this system had decreased according to the level of DNA damage caused by food processing. However, the intensity of this effect differed in function of the DRT mix used. Indeed, no amplicon was generated on processed food with the DRT B primer while the mix D was able to detect the junction localised on the chromosome III similarly to unprocessed materials. Concerning the obtained amplicons, their sizes were in the same range as those of unprocessed materials (Figure 5.1; Table 5.1). Nonetheless, the DNA degradation had implied a disappearance of some of them, such as in mix D tested on the 0.1% Bt rice noodles sample. As a whole, based on the combination of results from the four different DRT primer mixes (A–D), the DNA walking approach was able to confirm the presence of the target at its lowest tested concentration (0.1%), in both rice flour and noodles, identifying the transgene flanking regions on chromosomes II and III. Regarding the specificity, no additional aspecific amplification was observed.

The high sensitivity and specificity of the proposed DNA walking approach was confirmed by all these results. This method presents the important advantage to be able to cope with processed food that is essential for all analysis applied to food matrices. It should be noted, however, that, although the DNA from rice noodles was strongly degraded by the food processing, the impact of higher temperatures was not investigated in this study (Figure 5.2).

In addition, the sensitivity of the method is clearly linked to the affinity of DRT primers used. Indeed, a primer with a poor affinity for the targeted sequence, such as mix B, presents an obvious difficulty to detect the transgene flanking regions in a given sample submitted to food processing. Therefore, the importance of using the four different DRT primers is highlighted in order to maximise the detection power of the DNA walking method, independently of

Table 5.2: Performance analysis of the DNA walking method tested on food mixtures (rice/maize) using four different mixes of DRT primers (A–D). For each sample, the amount of Bt rice and WT maize is based and indicated in HGEs (200 000, 2000, 200 or 0 HGEs). The indicated percentage corresponds to the Bt rice HGEs tested as in point 5.3.2. Following the qPCR analysis using the *PLD* and *t35S* pCAMBIA screening markers, the observed C_t values, with the standard deviations are indicated for each sample. The detection of transgene flanking regions on the chromosomes II (X2) or III (X3) is symbolised by + (3/3), (+) (1–2/3) or −(0/3). For each result, the experiment was carried out in triplicate. The approximate size of amplicons, only obtained three times, is indicated in base-pair under the corresponding signal.

WT Maize		200 000 HGEs	200 000 HGEs	200 000 HGEs	200 000 HGEs
Bt Rice		200 000 HGEs (100%)	2000 HGEs (1%)	200 HGEs (0.1%)	0 HGE
PLD		17.5 ± 0.1	25.1 ± 0.1	29.8 ± 0.7	0.0 ± 0.0
t35S pCAMBIA		21.2 ± 0.2	28.3 ± 0.5	31.9 ± 0.2	0.0 ± 0.0
Mix A	X2	+	+	+	−
		(950)	*(950)*	*(950)*	
	X3	+	+	+	−
		(700; 500)	*(700)*	*(500)*	
Mix B	X2	+	−	−	−
		(450)			
	X3	+	+	+	−
		(800)	*(800)*	*(800)*	
Mix C	X2	+	+	(+)	−
		(650)	*(650)*		
	X3	+	−	−	−
		(450)			
Mix D	X2	−	−	−	−
	X3	+	+	+	−
		(850; 800; 500)	*(850; 800; 500)*	*(850; 800; 500)*	

Table 5.3: Performance analysis of the DNA walking method tested on unprocessed (grains) and processed rice food (flour and noodles), using four different DRT primers (A–D). For each tested sample, the amount of Bt rice in WT rice, indicated in percentage (1%, 0.1% or 0%), is based on the rice grain weight. Following the qPCR analysis using the *PLD* and *t35S* pCAMBIA screening markers, the observed Ct values with the standard deviations are indicated for each sample. The detection of transgene flanking regions on the chromosomes II (X2) or III (X3) is symbolised by + (3/3), (+) (1–2/3) or −(0/3). For each result, the experiment was carried out in triplicate. The approximate size of amplicons, only obtained three times, is indicated in base-pair under the corresponding signal.

		Grains 1%	Flour 1%	Noodles 1%	Grains 0.1%	Flour 0.1%	Noodles 0.1%	Grains 0%	Flour 0%	Noodles 0%
PLD		17.6 ± 0.1	18.4 ± 0.5	19.9 ± 0.8	17.2 ± 0.2	18.6 ± 0.4	18.7 ± 0.6	17.3 ± 0.3	18.7 ± 0.6	19.6 ± 0.7
t35S		27.2 ± 0.2	29.3 ± 0.3	30.1 ± 0.5	30.3 ± 0.2	31.6 ± 0.2	31.9 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
pCAMBIA										
Mix A	X2	+ *(950)*	+ *(950)*	+ *(950)*	+ *(950)*	+ *(950)*	(+)	−	−	−
	X3	+ *(500)*	+ *(500)*	−	+ *(500)*	+ *(500)*	−	−	−	−
Mix B	X2	(+)	−	−	−	−	−	−	−	−
	X3	+ *(800)*	−	−	+ *(800)*	−	−	−	−	−
Mix C	X2	+ *(650)*	+ *(650)*	(+)	(+)	(+)	(+)	−	−	−
	X3	−	−	−	−	−	−	−	−	−
Mix D	X2	−	−	−	−	−	−	−	−	−
	X3	+ *(850; 800; 500)*	+ *(850; 800; 500)*	+ *(850; 800; 500)*	+ *(850; 800; 500)*	+ *(850; 800; 500)*	+ *(850)*	−	−	−

the processing state of the tested matrix.

5.4 Conclusion

An integrated PCR-based DNA walking approach has recently been developed to identify unauthorised GMOs including a pCAMBIA family cassette that is frequently present in transgenic plants (Fraiture *et al.* 2014). For the first time, the analytical performance of this method has been here thoroughly assessed, in terms of sensitivity as well as applicability to a range of model food samples mimicking common rice-based mixtures and processed products.
The results obtained in this study suggest the good specificity, sensitivity, reliability, practicability and applicability of the developed DNA walking strategy on food mixtures, processed food matrices and low amounts of target, especially interesting in the case of unauthorised GMOs present at trace level.
At the moment, other new high-throughput technologies to identify unauthorised GMOs are emerging, such as Next-Generation Sequencing (NGS). Recently, this last technology has successfully characterised the transgene flanking region of GM soybean and rice (LLRice62, TT51-1 and T1c-19) (Kovalic *et al.* 2012; Wahler *et al.* 2013; Yang *et al.* 2013). However, these approaches were tested only on unprocessed materials composed of 100% targeted GMOs and the detection of low amounts of GMO mixture and processed food, although crucial, remains challenging and extremely expensive. At the present time, NGS technology is not easily implementable routinely in the enforcement laboratories and still requires a long time-frame to get results, high-cost and qualified bioinformaticians for dealing with NGS data. Therefore, the proposed DNA walking strategy is currently a key molecular tool to easily prove, without significant additional cost and equipment, the presence of unauthorised GMOs in any given food/feed matrix.

Acknowledgements

This research is funded by the Federal Public Service Health, Food Chain Safety and Environment (convention RF 11/6242) through the Project UGMMONITOR. Sequencing is performed at the Platform Biotechnology and Molecular Biology at the Scientific Institute of Public Health. The authors would also like to thank Emmanuel Guiderdoni (CIRAD, UMR AGAP, Biological Systems department, Montpellier, France) for his kindness in providing rice grains, Sylvia Broeders for her critical view of the problems of GMO detection in routine analysis and Chuong Dai Do Thi for her rice culinary advices.

Chapter 6

Integrated DNA walking system to characterize a broad spectrum of GMOs in food/feed matrices

To target a broader spectrum of GMO, the integrated approach, developed and validated in the chapters 4 and 5 respectively, was strengthened with two bidirectional DNA walking methods anchored on the p35S and tNOS elements, which are frequently found in GMO.

This chapter was previously published as:
M.-A. Fraiture, P. Herman, L. Lefèvre, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2015b). "Integrated DNA walking system to characterize a broad spectrum of GMOs in food/feed matrices". *BMC Biotechnology* 15, 76. DOI: 10.1186/s12896-015-0191-3

Authors' contributions

M.-A. Fraiture performed experiments, analysed data and drafted the manuscript. L. Lefèvre participated to the DNA walking experiment applied on Bt and WT rice samples. P. Herman, I. Taverniers, M. De Loose, D. Deforce and N. H. Roosens helped to design the study and drafted the manuscript. All authors read and approved the final manuscript.

Abstract

Background In order to provide a system fully integrated with qPCR screening, usually used in GMO routine analysis, as well as being able to detect, characterize and identify a broad spectrum of GMOs in food/feed matrices, two bidirectional DNA walking methods targeting p35S or tNOS, the most common transgenic elements found in GM crops, were developed. These newly developed DNA walking methods are completing the previously implemented DNA walking method targeting the t35S pCAMBIA element.

Methods Food/feed matrices containing transgenic crops (Bt rice or MON863 maize) were analysed using the integrated DNA walking system.

Results First, the newly developed DNA walking methods, anchored on the sequences used for the p35S or tNOS qPCR screening, were tested on Bt rice that contains these two transgenic elements. Second, the methods were assessed on a maize sample containing a low amount of the GM MON863 event, representing a more complex matrix in terms of genome size and sensitivity. Finally, to illustrate its applicability in GMO routine analysis by enforcement laboratories, the entire workflow of the integrated strategy, including qPCR screening to detect the potential presence of GMOs and the subsequent DNA walking methods to characterize and identify the detected GMOs, was applied on a GeMMA Scheme Proficiency Test matrix. Via the characterization of the transgene flanking region between the transgenic cassette and the plant genome as well as of a part of the transgenic cassette, the presence of GMOs was properly confirmed or infirmed in all tested samples.

Conclusion Due to their simple procedure and their short time-frame to get results, the developed DNA walking methods proposed here can be easily implemented in GMO routine analysis by the enforcement laboratories. In providing crucial information about the transgene flanking regions and/or the

transgenic cassettes, this DNA walking strategy is a key molecular tool to prove the presence of GMOs in any given food/feed matrix.

6.1 Background

In 2014, 181.5 million hectares of genetically modified organisms (GMOs) have been planted in 28 countries (James 2014). On the European Union (EU) market, the commercialization of GMOs in the food/feed chain is subject to the EU legislation (The European Parliament and The Council Of The European Union 2001; The European Parliament and The Council Of The European Union 2003a; The European Parliament and The Council Of The European Union 2003b), which is becoming more and more complex to implement due to the increasing number and diversity of GMOs (Stein and Rodríguez-Serezo 2009; James 2014). The majority of EU-authorized GMOs (78.6%) harbours the transgenic p35S element (Cauliflower mosaic virus (CaMV) 35S promoter), the transgenic tNOS element (*Agrobacterium tumefaciens* nopaline synthase terminator) or both of them, with an occurrence respectively reported of 60.7, 53.6 and 35.7% (Barbau-Piednoir *et al.* 2010; Broeders *et al.* 2012a; CERA 2012; GMO Compass 2015).

To ensure the correct enforcement of the EU legislation, several GMO detection methods have been developed, mainly based on SYBR®Green and TaqMan® real-time PCR technologies. Usually, a screening is first performed with qPCR methods targeting the most common transgenic elements present in genetically modified (GM) crops (*e.g.* p35S and tNOS). These strategies, covering a broad spectrum of GMOs, allow to indicate the potential presence of GMOs in tested samples (Barbau-Piednoir *et al.* 2010; Reiting *et al.* 2010; Mbongolo Mbella *et al.* 2011; Broeders *et al.* 2012a; Broeders *et al.* 2012b; Kluga *et al.* 2013). In case of positive responses, EU-authorized GMOs are subsequently identified and quantified using EU event-specific methods. If some observed positive screening elements, like p35S and tNOS, are not explained by these event-specific methods, the presence of EU-unauthorized GMOs can be indirectly suspected (Broeders *et al.* 2012a). However, as most of the targeted elements originate from natural organisms (*e.g.* p35S from CaMV and tNOS from *Agrobacterium tumefaciens*), the confirmation of their presence can be irrefutably provided only by the characterization of the transgene flanking regions between the plant genome and the integrated cassette (Broeders *et al.* 2012a; Broeders *et al.* 2012b; Fraiture *et al.* 2014). To this end, DNA walking strategies have notably been proposed in order to get this crucial information allowing to identify GM crops (Babekova *et al.* 2009; Xu *et al.* 2013a; Cao *et al.* 2014; Fraiture *et al.* 2014 and references therein; Liang *et al.* 2014; Majhi *et al.* 2014; Fraiture *et al.* 2015a and references therein; Zhang *et al.* 2015). However, these methods are not usually used in GMO routine analysis because they are not easily implementable by the enforcement laboratories. Recently, an integrated DNA walking strategy, better corresponding to the need of the enforcement laboratories, was developed to rapidly detect and identify EU-unauthorized GMOs, without significant

additional cost and equipment (Fraiture *et al.* 2014). This method targets the t35S element from the pCAMBIA vector, which is frequent (30%) in transgenic plants and is absent in EU-authorized GMOs. This DNA walking approach, based on PCR, has the advantage to be fully integrated into the initial qPCR analysis as the same primers are used for the qPCR screening (detection) and the DNA walking (identification) (Komori *et al.* 2007; Fraiture *et al.* 2014). In addition, this approach was assessed as highly sensitive and able to deal with rice based mixtures and processed products, which is essential in GMO routine analysis (Fraiture *et al.* 2015a).
Here, the concept of this integrated PCR-based DNA walking strategy has been adapted to also target p35S and tNOS, the most common transgenic elements found in GMOs, in order to characterize a broader spectrum of GMOs as well as to strengthen the initial DNA walking system targeting t35S from pCAMBIA. For each element, two DNA walking directions, starting from a position anchored on the sequences used for the p35S or tNOS SYBR®Green qPCR screening, have been established (Barbau-Piednoir *et al.* 2010). First, the p35S and tNOS bidirectional DNA walking methods were developed on Bt rice, as previously used for the t35S pCAMBIA method. Second, these DNA walking methods were assessed using the certified reference material (CRM) of the GM maize MON863 (9,85%), which represents a more complex matrix due to its large genome and its low target content. Finally, in order to illustrate its applicability in routine analysis, a GeMMA Scheme Proficiency Test food matrix was submitted to the entire integrated strategy, including the qPCR screening using the p35S, tNOS and t35S pCAMBIA markers to detect the presence of GMOs and, then, the DNA walking methods, corresponding to the qPCR positive responses, allowing to characterize them.

6.2 Results and discussion

In order to characterize a broad spectrum of both EU-authorized and -unauthorized GMOs, two novel DNA walking methods, based on the p35S and tNOS transgenic elements, have been developed. These methods were designed similarly to the t35S pCAMBIA DNA walking method targeting only EU-unauthorized GMOs (Fraiture *et al.* 2014). In the interest to provide an integrated approach, for each DNA walking method, the same primers allow the detection of the potential presence of GMOs containing the targeted elements (qPCR screening) as well as their characterization and identification insofar as possible (DNA walking).

6.2.1 *In silico* study

Since the DNA walking approach is integrated into the screening step, the SYBR®Green primers published by Barbau-Piednoir *et al.* (2010) were used to target the p35S and tNOS elements.
As three primers are required by the DNA walking method for each targeted element, an additional primer (b) intermediate to the screening primers (a and c) was designed (Table 6.1). The specificity of these primers was successfully assessed *in silico*, against all EU-authorized GMOs, LLPs (Low Level Presence) and corresponding WTs (Wild-Type), using the software wEMBOSS (data not shown) (EMBnet 2015). Moreover, for each of the targets, two walking directions were established (p35S-F, p35S-R, tNOS-F and tNOS-R) in order to extend the GMO coverage of the integrated DNA walking strategy.

6.2.2 Development of the DNA walking methods

Assessment of p35S DNA walking methods

For the p35S approach, several amplicons were observed from 100% Bt rice, corresponding to 200 000 HGEs (Haploid Genome Equivalent), for the four different degenerated random tagging (DRT) primers (A-D), including 18 amplicons for the p35S-F DNA walking method (amplicons n° 1 to 18) and 12 amplicons for the p35S-R DNA walking method (amplicons n° 24 to 35) (Figure 6.1A). The size range of these amplicons was approximately from 100 bp to 1 Kbp and from 250 bp to 2 Kbp for the p35S-F and p35S-R DNA walking method, respectively. All these amplicons were consecutively analysed by sequencing to evaluate the specificity of the methods (Additional File 1).
All these characterized sequences corresponded specifically to the position of the p35S element in the transgenic cassette (Figure 6.1B) (Breitler *et al.* 2004). As expected, these sequences present the continuity of the p35S element [GenBank:AF234296] for the p35S-F DNA walking method (amplicons n° 1 to 18) and the p35S promoter [GenBank:AF234296] regulating the hygromycin resistance gene (hpt) [GenBank:AAF65337] for the p35S-R DNA walking method (amplicons n° 24 to 35) (Figure 6.1b and Additional File 1).
For the WT rice sample, few amplicons (amplicons n° 19 to 23 for the p35S-F method and amplicons n° 36 to 42 for the p35S-R method) were observed (Figure 6.1A) and identified as corresponding to the rice genome (Additional File 2). They are probably due to the use of DRT primers which can potentially generate a background of aspecific products, especially in absence or in low amounts of targeted sequences (Leoni *et al.* 2011).

Assessment of tNOS DNA walking methods

The use of the tNOS DNA walking approach with the two walking directions using the four different DRT primers (A-D) on the 100% Bt rice sample produces several amplicons, including 17 amplicons for the tNOS-F DNA walking method (amplicons n° 43 to 59) and 22 amplicons for the tNOS-R DNA walking method (amplicons n° 66 to 87) (Figure 6.1A). The tNOS-F and tNOS-R DNA walking methods gave respectively amplicons with a size range of approximately 100 bp to 1.5 Kbp and 200 bp to 2 Kbp. To assess the specificity of the methods, all these PCR products were examined by sequencing (Additional File 1).
On the one hand, as expected, regarding the tNOS element localisation in the transgenic cassette, 100% of the analysed amplicons coming from the tNOS-F DNA walking method have allowed to characterize the transgene flanking regions between the rice genome and the right border of the integrated pCAMBIA cassette via the amplicon sequences containing both the tNOS element and the rice genome (Figure 6.1b and Additional File 1). None of the obtained amplicons presented an unexpected sequence. As the Bt rice presents two transgenic insertions, two types of transgene flanking regions were characterized: one localised between the transgenic cassette [GenBank:AY836546.1] and a genomic sequence from chromosome II of *Oryza sativa* japonica Group [GenBank:OSJNBa0016G10] identified using the amplicons generated by the DRT C primers (amplicons n° 51 to 56) and one situated between the pCAMBIA cassette [GenBank:AY836546.1] and a genomic sequence from chromosome III of *Oryza sativa* japonica Group [GenBank:OSJNBb0111B07] identified using the amplification coming from the DRT A, B and D primers (amplicons n° 43 to 50 and n° 57 to 59) (Fraiture *et al.* 2014; Fraiture *et al.* 2015a). These results yet clearly demonstrate the importance to use four different DRT primer mixes. Indeed, the difference in affinity of these DRT primers allows increasing the likelihood to successfully characterize all targets (Fraiture *et al.* 2014; Fraiture *et al.* 2015a). In addition, the right border of the pCAMBIA cassette on chromosome II was shorter of two base-pairs compared to the one on chromosome III (Additional File 1). These two transgene flanking regions were also properly confirmed by sequencing of PCR products obtained in using primers annealing to the pCAMBIA cassette and chromosome II or III (Table 6.1 and Additional File 3).
On the other hand, as expected, all PCR products generated from the tNOS-R DNA walking method allow to characterize the continuity of the tNOS element (amplicons n° 71, 82 and 87) as well as, for the longer ones, the flanking region between the tNOS element [GenBank:HQ593861.1] and the Cry1B gene [GenBank:KC414884.1] conferring an insect resistance (amplicons n° 66 to 70, n° 72 to 81 and n° 83 to 86) (Figure 6.1b and Additional File 1). 100% of the analysed amplicons corresponded to the expected sequences.

Table 6.1: Oligonucleotide primers used for the real-time PCR assays, the DNA walking approaches and the PCR confirmation of the transgenic junctions.

Methods	Oligonucleotide names	Oligonucleotide sequences	Product sizes (bp)	References
SYBR®Green qPCR	p35S F	AAAGCAAGTGGATTGATGTGATA	75	Barbau-Piednoir *et al.* (2010)
	p35S R	GGGTCTTGCGAAGGATAGTG		Barbau-Piednoir *et al.* (2010)
SYBR®Green qPCR	tNOS F	GATTAGAGTCCCGCAATTATACATTTAA	69	Barbau-Piednoir *et al.* (2010)
	tNOS R	TTATCCTAGKTTGCGCGCTATATTT		Barbau-Piednoir *et al.* (2010)
SYBR®Green qPCR	t35S pCAMBIA c-F	CGGGGGATCTGGATTTTAGTA	137	Fraiture *et al.* (2014)
	t35S pCAMBIA a-R	AGGGTTCCTATAGGGTTTCGCTC		Fraiture *et al.* (2014)
DNA Walking	p35S-F a (p35S R)	GGGTCTTGCGAAGGATAGTG		Barbau-Piednoir *et al.* (2010)
	p35S-F b	TGTGCGTCATCCCTTACGTCAGT	/	This study
	p35S-F c	TATCACATCAATCCACTTGCTTT		Barbau-Piednoir *et al.* (2010)
DNA Walking	p35S-R a (p35S F)	AAAGCAAGTGGATTGATGTGATA		Barbau-Piednoir *et al.* (2010)
	p35S-R b	ACTGACGTAAGGGATGACGCACA	/	This study
	p35S-R c	CACTATCCTTCGCAAGACCC		Barbau-Piednoir *et al.* (2010)
DNA Walking	tNOS-F a (tNOS F)	GATTAGAGTCCCGCAATTATACATTTAA		Barbau-Piednoir *et al.* (2010)
	tNOS-F b	TTAATACGCGATAGAAAACAAAAT	/	This study
	tNOS-F c	AAATATAGCGCGCAAMCTAGGATAA		Barbau-Piednoir *et al.* (2010)

Table 6.1 *continued*

Methods	Oligonucleotide names	Oligonucleotide sequences	Product sizes (bp)	References
DNA Walking	tNOS-R a (tNOS R)	TTATCCTAGKTTGCGCGCTATATTT		Barbau-Piednoir *et al.* (2010)
	tNOS-R b	ATTTTGTTTTCTATCGCGTATTAA	/	This study
	tNOS-R c	TTAAATGTATAATTGCGGGACTCTAATC		Barbau-Piednoir *et al.* (2010)
PCR junction	Rice chromosome II	CCCCTAATTTCTCACAGGCC	848	This study
	tNOS-F c	AAATATAGCGCGCAAMCTAGGATAA		Barbau-Piednoir *et al.* (2010)
PCR junction	Rice chromosome III	AGGTACTCAAGCCTTTTCCAGC	1105	This study
	tNOS-F c	AAATATAGCGCGCAAMCTAGGATAA		Barbau-Piednoir *et al.* (2010)

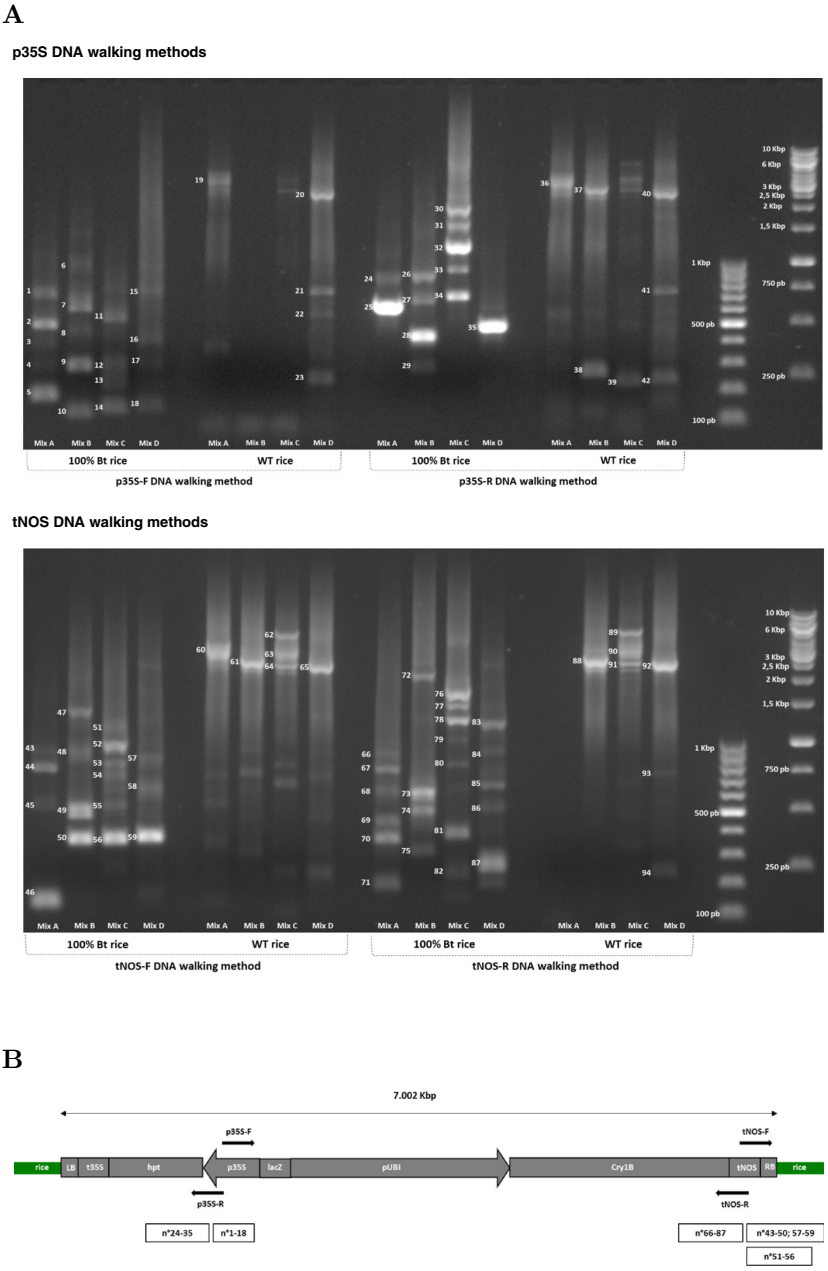


Figure 6.1: Development of the bidirectional p35S and tNOS DNA walking methods on 100% Bt rice.

Figure 6.1 *(continued)*: (A) Visualisation of the obtained amplicons, numeroted from 1 to 94, using the p35S and tNOS DNA walking methods applied on 100 ng of 100% Bt rice and WT rice. For each method, four different DRT primer mixes (A-D) have been used. (B) For each DNA walking method, a schematic representation of the potential start position and direction, applied on the transgenic cassette of the Bt rice, is illustrated by the black arrows. Below the transgenic cassette, the sequence covering of the obtained amplicons from the 100% Bt rice is schematically represented by rectangles. The corresponding amplicon numbering is indicated in the Figure 6.1a.
LB: left border; t35S: CaMV 35S terminator; hpt: hygromycin phosphotransferase gene; p35S: CaMV 35S promoter; lacZ: LacZ alpha fragment; pUBI: maize ubiquitin promoter; Cry1B: synthetic Cry1B gene; tNOS: *Agrobacterium tumefaciens* nopaline synthase terminator; RB: right border; rice: rice genome [Schema adapted from Breitler *et al.* (2004)].

Similarly to the p35S DNA walking methods, the bidirectional tNOS approach presents uniquely specific amplifications further to the analysis of the Bt rice sample while few aspecific amplicons (Figure 6.1B), corresponding to the rice genome, were generated from the WT rice material (amplicons n° 60 to 65 for the tNOS-F method and amplicons n° 88 to 94 for the tNOS-R method) (Additional File 2).

6.2.3 Practical application of the DNA walking methods

Analysis of GM maize

To test the developed p35S and tNOS bidirectional DNA walking methods on a more complex food matrix than rice in term of genome size and target amount, GM maize MON863 9.85% (ERM-BF416c), corresponding to 3 788 HGEs, was selected as it possesses both the p35S and tNOS elements in its transgenic cassette (Barbau-Piednoir *et al.* 2010; Joint Research Centre 2014).
First, the presence of these elements in the tested CRM sample was confirmed by SYBR®Green qPCR screening (Additional File 4). Then, several amplicons were generated by each DNA walking method with a size ranging from approximately 200 bp to 4 Kbp (Figure 6.2A). In order to obtain the most informative sequences, the amplicon with the highest size for each DRT primer mix in each applied DNA walking method was selected to be sequenced (Additional File 5).
Most of the selected amplicons from the p35S-F DNA walking method present the 5' transgene flanking region between the maize genome [GenBank:DQ490951.2] and the p35S promoter [GenBank:KJ608136.1] from the transgenic cassette of

MON863, as previously published (Figure 6.2b and Additional file 5) (Yang *et al.* 2005b; Zhu *et al.* 2008). This transgene flanking region, confirming the presence of GM maize MON863, is also targeted by the EU event-specific qPCR method to identify and quantify this GMO (Joint Research Centre – European Commission 2005). Only one tested amplicon (n° 3) showed an aspecific sequence corresponding to the WT maize genome [GenBank:AC196084; *Zea mays* BAC clone CH201-52A17 from chromosome 5] (Figure 6.2a and Additional File 2). A possible explanation is that the tested sample contains primarily WT maize material and only a relative low amount of the target.

For the p35S-R DNA walking method, two different types of sequences were observed due to the presence of two p35S promoters in the transgenic cassette of GM MON863 maize (Figure 6.2B) (Yang *et al.* 2005b; Pan *et al.* 2006). On the one hand, the continuity of the p35S promoter [GenBank:KJ608136.1; *Zea mays* transgenic line MON863 promoter region] regulated the neomycin phosphotransferase gene (nptII) from *A. tumefaciens* [GenBank:AAF65400.1] which confers a resistance to kanamycin (Figure 6.2b and Additional File 5). On the other hand, a part of the p35S promoter [GenBank:JX139718.1], followed by the 5' upstream sequence of the Wheat major chlorophyll a/b binding protein gene (wtCAB) [GenBank:X05823.1] and by the Rice Actin intron (rAct) [GenBank:EU155408.1; X63830.1], regulated the synthetic Cry3Bb1 gene [GenBank:CS409981.1; GX181970.1], providing insect resistance (Figure 6.2b and Additional File 5). The information acquired from these both types of sequences allows a better characterization of the transgenic cassette.

Using the tNOS-F DNA walking method, all analysed amplicons showed a part of the tNOS terminator from *A. tumefaciens* [GenBank:JN400387.1] followed by a modified p35S promoter, including in upstream of four repeats of a short activating sequence (21 bp), referred to as 4-AS1 promoter [GenBank:JX139718.1] (Yang *et al.* 2005b; Pan *et al.* 2006; Zhu *et al.* 2008). This modified p35S promoter was then followed by the 5' upstream sequence of the Wheat major chlorophyll a/b binding protein gene (wtCAB) [GenBank:X05823.1] and the Rice Actin intron (rAct) [GenBank:EU155408.1; X63830.1] to regulate the Cry3Bb1 gene [GenBank:CS409981.1; GX181970.1] (Figure 6.2b and Additional File 5).

The tNOS-R DNA walking method provided the sequence of the nptII selection marker [GenBank:AAF65400.1] followed by the tNOS terminator [GenBank:JN400387.1] (Figure 6.2b and Additional File 5). An aspecific sequence corresponding to the WT maize genome [GenBank:AC196084], identical to the amplicon n° 3 from the p35S-F DNA walking method, was observed for the amplicons n° 14 and 15 (Additional File 2).

Most of the tested amplicons (81.25%) derived from all the DNA walking methods presented a sequence corresponding to the GM targets. Based on these data, the presence of GM MON863 maize in the tested sample was clearly

identified by isolation and sequencing of its junction between the maize genome and the transgenic cassette. In addition, this strategy allows to reconstruct 2.727 Kbp of the integrated transgenic cassette, going from the left border to a part of the gene Cry3Bb1, in agreement with the published information (Figure 6.2b and Additional File 5) (Yang *et al.* 2005b; Pan *et al.* 2006; Zhu *et al.* 2008). These results also highlight that the proposed DNA walking strategy is able to identify GMOs from different plant species.

Analysis of the food matrix

In order to illustrate its applicability in GMO routine analysis by the enforcement laboratories, the entire workflow of the integrated system was applied on a food matrix (GeM SU34-A) from a GeMMA Scheme Proficiency Test containing 1.2% of GM maize MON863 event, corresponding to 461 HGEs.

First, similarly to the GMO routine analysis, the GeMMA food matrix was submitted to the SYBR®Green qPCR screening using the p35S, tNOS and t35S pCAMBIA screening markers allowing to detect the potential presence of GMOs (Table 6.1). As expected, a positive signal was observed for the p35S and tNOS screening markers while the t35S pCAMBIA screening marker gave a negative signal (Additional File 4), suggesting the potential presence of GMOs in the tested food matrix.

Second, based on the positive signals obtained from the screening qPCR analysis, the bidirectional p35S and tNOS DNA walking approaches were selected to be applied on the sample. In doing so, the potential presence of GMOs will be confirmed by the characterization of their sequences.

All applied DNA walking methods were able to produce amplicons in a size range from approximately 200 bp to 1.5 Kbp (Additional File 6). In order to follow an efficient workflow suitable for GMO routine analysis, only one amplicon, chose for its large size as well as for its ease to be selected on an electrophoresis gel, was sequenced for each DNA walking method (Additional File 6).

With all these DNA walking methods, 100% of the analysed amplicons presented sequences specific to the GM target. Indeed, when using p35S-F DNA walking, the transgene flanking region between the maize genome [GenBank:DQ490951.2] and the p35S promoter from the transgenic cassette of MON863 [GenBank:KJ608136.1] was identified, proving the presence of this GMO in the tested sample (Yang *et al.* 2005b; Zhu *et al.* 2008). The p35S-R DNA walking method presented the continuity of the p35S promoter [GenBank:KJ608136.1] regulating the nptII selection marker [GenBank:AAF65400.1]. From the tNOS-F DNA walking method, a part of the tNOS terminator [GenBank:JN400387.1], followed respectively by the 4AS-1 promoter [GenBank:JX139718.1] and the Wheat major chlorophyll a/b binding

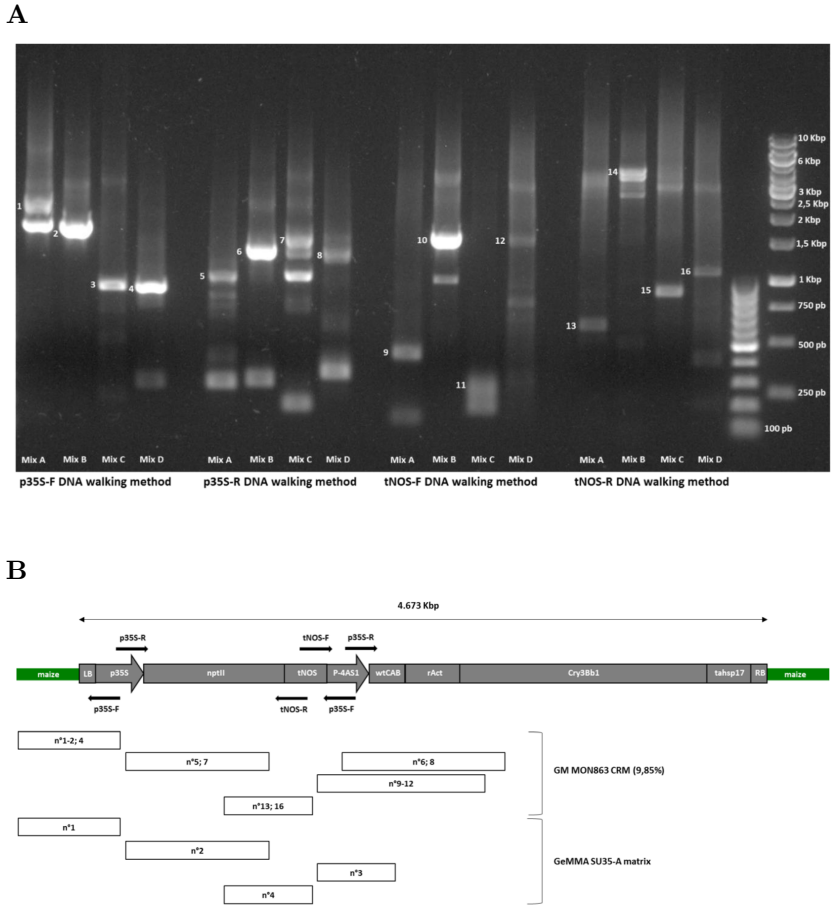


Figure 6.2: Application of the bidirectional p35S and tNOS DNA walking methods on GM maize matrices.
(A) Visualisation of the obtained amplicons using the p35S and tNOS DNA walking methods applied on 100 ng of the GM MON863 maize CRM (9.85%). For each method, four different DRT primer mixes (A-D) have been used. The analyzed amplicons are indicated by a numerotation going from 1 to 16. (B) For each DNA walking method, a schematic representation of the potential start position and direction, applied on the transgenic cassette of the GM maize MON863, is illustrated by the black arrows. Below the transgenic cassette, the sequence covering of the selected amplicons from the GM MON863 maize CRM (9.85%) and the GeMMA proficiency test food matrix (GeMMA SU35-A) is schematically represented by rectangles. The corresponding amplicon numbering is indicated in the Figure 6.2A and Additional File 6.

Figure 6.2 *(continued)*: LB: left border; p35S: CaMV 35S promoter; nptII: neomycin phosphotransferase II gene; tNOS: *Agrobacterium tumefaciens* nopaline synthase terminator; p4-AS1: modified CaMV 35S promoter; wtCAB: wheat major chlorophyll a/b binding protein gene; rAct: rice actin intron; Cry3Bb1: synthetic Cry3Bb1 gene; tahsp17: wheat heat shock protein terminator; RB: right border; maize: maize genome [Schema adapted from Yang *et al.* (2005b) and Pan *et al.* (2006)].

protein gene (wtCAB) [GenBank:X05823.1], was detected. Via the tNOS-R DNA walking method, the sequence of the nptII gene [GenBank:AAF65400.1] regulated by the tNOS terminator [GenBank:JN400387.1] was observed (Figure 6.2b and Additional File 5 and 6).
All these sequences indubitably prove the presence of the GM MON863 maize event in the GeMMA food matrix sample though the identification of its junction between the maize genome and the transgenic cassette as well as the partial reconstruction of its transgenic cassette, in agreement with the published information (Yang *et al.* 2005b; Pan *et al.* 2006; Zhu *et al.* 2008).
Similarly to the t35S pCAMBIA DNA walking method, the good specificity of the newly developed DNA walking methods (p35S-F, p35S-R, tNOS-F, tNOS-R) was illustrated in this study since almost all of the sequences from the analysed amplicons generated from the Bt rice (100%), MON863-9.85% (81.25%) and Gemma proficiency test (100%) matrices corresponded to the GM targets (Fraiture *et al.* 2014; Fraiture *et al.* 2015a). The success of this strategy is mainly due to the specificity of the target-specific primers, allowing to initially amplify the targets by PCR and, then, to enrich them by two successive semi-nested PCRs (Table 6.1).

6.3 Conclusion

In order to provide an integrated system able to detect, characterize and identify a broad spectrum of both EU-authorized and -unauthorized GMOs in food/feed matrices, two bidirectional DNA walking methods targeting p35S or tNOS, the most common transgenic elements, were developed to be anchored on the sequences used for the p35S or tNOS qPCR SYBR®Green screening described by Barbau-Piednoir *et al.* (2010). These DNA walking methods also allow to strengthen the previously published t35S pCAMBIA DNA walking method in order to currently target around 75% of the GM crops (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; personal communication).
First, the p35S and tNOS bidirectional DNA walking methods were developed and assessed for their specificity using 100% Bt rice. These methods were

evaluated as highly specific since no aspecific amplifications were generated in presence of the target. Second, the developed DNA walking methods were tested on a more complex maize food matrix, in term of genome size, containing approximately 10% of the GM maize MON863 event. Finally, the entire workflow of the integrated system, including the detection of the potential presence of GMOs by qPCR screening with the p35S, tNOS and t35S pCAMBIA markers and, subsequently, the confirmation of their presence using the DNA walking methods corresponding to the previously obtained qPCR responses, was applied on a GeMMA Scheme Proficiency Test matrix, containing 1.2% of the GM maize MON863 event, to illustrate its applicability in GMO routine analysis by the enforcement laboratories. For all tested matrices, the p35S and the tNOS bidirectional DNA walking methods were successfully applied as the GMO presence was proven via the characterization of the junction between the transgenic cassette and the plant genome as well as of a part of the transgenic cassette.

In addition to its clear benefit in GMO detection, this integrated system has the advantage to present a simple procedure and a short time-frame to get the results. However, in order to analyse even more easily the PCR products derived from the DNA walking methods, some adaptations in the entire DNA walking workflow could be done regarding the purification of the generated amplicons excised from the electrophoresis gel and the subsequent sequencing using Sanger technology. Indeed, even if the initial DNA walking workflow remains simple, in case of matrices containing several GMOs, harbouring the same targeted element, the purification of the potential numerous amplicons excised from the electrophoresis gel and the subsequent Sanger sequencing could be cumbersome. This situation could be for instance encountered with matrices presenting a low amount of EU-unauthorized GMOs mixed with EU-authorized GMOs harbouring the elements p35S and/or tNOS, very frequently observed in GM crops. In this scenario, the obtained amplicons will present different sequences, representing potentially one GMO per observed DNA fragment. Therefore, the simplified workflow, consisting in selecting the largest size amplicons to obtain the most informative sequences, does not guarantee the entire representativeness of GMOs present in the tested sample. Consequently, it's preferable to analyse all amplicons observed on the electrophoresis gel and to eventually them using Sanger technology, which may be a quite laborious work. In the future, this difficulty could be circumvented in replacing the step related to the purification of the amplicons excised from the electrophoresis gel and the subsequent Sanger sequencing by a high-throughput Next-Generation-Sequencing approach, as suggested by Liang *et al.* (2014).

6.4 Methods

6.4.1 Plant material

Grains of an insect resistant transgenic Bt rice (*Oryza sativa* L. *Japonica cv Ariete*), transformed by *Agrobacterium tumefaciens* with the binary vector pCAMBIA 1300 containing the synthetic Cry1B gene from *Bacillus thuringiensis*, and its corresponding wild-type (WT) were used in this study (Breitler *et al.* 2004). The CRM of the GM maize MON863 9.85% (ERM-BF416c) in the form of seed powder was obtained from the Institute for Reference Materials and Measurements (IRMM, Geel, Belgium). The food matrix (GeM SU34-A), coming from a GeMMA Scheme Proficiency Test, is a maize flour, tumble blended for 50 h, containing 1.2% (w/w) of 100% GM maize MON863.

6.4.2 DNA extraction, concentration and purity

Using a CTAB-based procedure (ISO 21571) in combination with the Genomic-tip20/G (QIAGEN, Hilden, Germany), DNA was extracted from a homogenous powder of rice grain obtained by manual grinding. Adapted from the EU-RL GMFF validated method, this DNA extraction method was carried out by four main successive steps: extraction of proteins, polysaccharides and organic components, precipitation of DNA in the presence of C-hexadecyl-Trimethyl-Ammonium-Bromide (CTAB), purification of DNA using a tip20 column and precipitation of DNA with isopropanol (ISO 21571 2005; European Union Reference Laboratory for GM Food and Feed and Feed 2006b). DNA concentration was measured by spectrophotometry using the Nanodrop® 2000 (ThermoFisher, DE, USA) device and the DNA purity was evaluated using the A260/A280 and A260/A230 ratios. DNA extraction, concentration and purity of the CRM and the food matrix (GeM SU34-A) were carried out as previously described (Broeders *et al.* 2013).

6.4.3 qPCR SYBR®Green technology

All qPCR assays were performed as described in Barbau-Piednoir *et al.* (2010) and Fraiture *et al.* (2014) using the primers indicated in Table 6.1. More precisely, a standard 25 μl reaction volume was applied containing 1 × SYBR®Green PCR Mastermix (Diagenode, Liège, Belgium), 250 nM of each primer and 5 μl of DNA (10 ng/μl). The qPCR cycling program consisted of a single cycle of DNA polymerase activation for 10 min at 95°C followed by 40 amplification cycles of 15 s at 95°C (denaturing step) and 1 min at 60°C (annealing-extension

step). The program for melting curve analysis was performed by gradually increasing the temperature from 60 to 95°C in 20 min (±0.6°/20 s). All runs were performed on an iQ™5 real-time PCR detection system (BioRad, Hemel Hempstead, UK). For each assay, a "No Template Control" (NTC) was included.

6.4.4 DNA walking approach

Development and assessment of oligonucleotide primers

Two DNA walking approaches have been developed to target the p35S or tNOS elements. For each method, three target-specific primers are required to carry out first the DNA walking (a) and then the first (b) and the second (c) semi-nested PCR rounds. To provide an integrated approach, the design of the target-specific primers a and c is based on the sequences from the SYBR®Green real-time PCR screening markers p35S or tNOS published by Barbau-Piednoir *et al.* (2010). An intermediate primer, corresponding to the target-specific primer b, was additionally designed. From each targeted transgenic element, two walking directions, called forward (F) and reverse (R) methods, have been performed (Figure 6.1b and Table 6.1). Using the program "wprimersearch" from the software "wEMBOSS", that mimics PCR amplification, the specificity of oligonucleotide primers was initially assessed *in silico* (EMBnet 2015).

DNA walking strategy

DNA walking and double semi-nested PCR reactions The DNA walking strategy previously described by Fraiture *et al.* (2014) was adapted in this study to target the transgenic element p35S or tNOS. Similarly, a first reverse target-specific primer (a) and one kind of the degenerated random tagging primer (DRT) mix (A-D) were applied, in a first step, followed by two semi-nested PCR rounds using target-specific primers (b and c), that are each time nested to the previous reverse target-specific primer, combined to universal tagging primers (UAP-N1 and UAP-N2) (Fraiture *et al.* 2014). All these methods were applied on 100 ng of DNA from 100% of Bt rice and its corresponding WT as well as on 100 ng of DNA from the food matrix (GeM SPU34-A) and its corresponding CRM (GM maize MON863 9.85%). Moreover, a NTC was included for each assay. PCR mixes and conditions were carried out according to the manufacturer's instructions (APAgene™ GOLD Genome Walking Kit from BIO S&T, Montréal, Canada). The final PCR products were analysed by electrophoresis on a 1% agarose gel (INVITROGEN, CA, USA) (100 V, 400 mA, 60 min) in view to further analysis allowing to identify the generated sequences.

Analysis and workflow In order to assess the specificity of the developed p35S and tNOS bidirectional methods, all the visualized amplicons produced from the 100% Bt rice and WT rice were excised from agarose gel and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, WI, USA) to then be sequenced and identified.
Next, to test the developed methods on a maize matrix, the CRM of maize MON863 (9,85%) was used and a workflow convenient for the GMO routine analysis was followed. For each DNA walking method, only the longest and easily selectable amplicon observed for each DRT primer mix was excised from agarose gel and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, WI, USA) to be sequenced.
To test the applicability of the entire integrated system, a simplified workflow was used for the food matrix (GeM SPU34-A). Only the longest and easily selectable amplicon observed for each DNA walking method was excised from agarose gel and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, WI, USA) to be sequenced.

6.4.5 Cloning and sequencing

Three different sequencing approaches were used to obtain the sequence of the selected amplicons. First, a direct sequencing was applied using the corresponding target-specific c primer or the UAP-N2 primer. Second, in case of an unsatisfying size or quality of the obtained sequences, two other sequencing approaches were carried out. On the one hand, a cloning strategy was performed. The amplicons were cloned into the pGEM®-T Easy Vector Systems (PROMEGA, WI, USA), according to the manufacturer's instructions. A PCR was carried out on colonies using pGEM®-T Easy Vector primers (T7: TAATACGACTCACTATAGGG; SP6: ATTTAGGTGACACTATAGAAT) and was analyzed by electrophoresis on a 1% agarose gel (INVITROGEN, CA, USA) (100 V, 400 mA, 60 min). The colonies presenting a fragment of the correct size were then sequenced (Sambrook and Russell 2001). On the other hand, an "enrichment" strategy, based on a PCR amplification using the corresponding target-specific c primer and the modified UAP-N2 primer coupled to the T7 sequence (UAP-N2_T7: TTTAATACGACTCACTATAGGGGGAAGCAGTGGTATCAACG), was used. To this end, a standard 25 μl reaction volume was applied containing 0.625 U of DreamTaq™ DNA Polymerase (Fermentas, CA, USA), 1 × DreamTaq™ Buffer (Fermentas, CA, USA), 0.2 mM of dNTPs, 250 nM of each primer and 5 μl of the purified amplicon. The PCR program consisted of a single cycle of 3 min at 95℃ (initial denaturation) followed by 45 amplification cycles of 30 s at 95℃ (denaturation), 30 s at 50℃ (annealing) and 4 min at 72℃ (extension)

and finishing by a single cycle of 10 min at 72℃ (final extension). The run was performed on an iQ™5 real-time PCR detection system (BioRad, Hemel Hempstead, UK). The PCR products were analysed by electrophoresis on a 1% agarose gel (100 V, 400 mA, 60 min; INVITROGEN, CA, USA) and purified using USB® ExoSAP-IT® PCR Product Cleanup (Affymetrix, CA, USA), according to the manufacturer's instructions, to be then sequenced via the T7 primer.
All sequencing reactions were performed on a Genetic Sequencer 3130XL using the Big Dye Terminator Kit v3.1 (Applied Biosystems, CA, USA). The sequences were aligned and analysed using "ClustalW2" software and "Nucleotide BLAST NCBI" software, respectively (EMBL-EBI 2015; NCBI 2015).

6.4.6 Verification of the transgene flanking regions by PCR amplification

The two different transgene flanking regions between the right border of the pCAMBIA cassette and the rice genome identified by the tNOS-F DNA walking method were verified by PCR amplification using the tNOS-F c primer combined to a primer designed on the rice chromosome II or III (Table 6.1). These oligonucleotide primers were initially evaluated *in silico* using the program "wprimersearch" from the software "wEMBOSS" (EMBnet 2015). A standard 25 μl reaction volume was applied containing 0.625 U of DreamTaq™ DNA Polymerase (Fermentas, CA, USA), 1 × DreamTaq™ Buffer (Fermentas, CA, USA), 0.2 mM of dNTPs, 250 nM of each primer and 5 μl of Bt rice DNA (5 ng/ μl). The PCR program consisted of a single cycle of 3 min at 95℃ (initial denaturation) followed by 35 amplification cycles of 30 s at 95℃ (denaturation), 30 s at 55℃ or 60℃ respectively for the rice chromosome III or II (annealing) and 1 min at 72℃ (extension) and finishing by a single cycle of 10 min at 72℃ (final extension). The run was performed on an iQ™5 real-time PCR detection system (BioRad, Hemel Hempstead, UK). The PCR products were analysed by electrophoresis on a 1% agarose gel (100 V, 400 mA, 60 min; INVITROGEN, CA, USA) and purified using USB® ExoSAP-IT® PCR Product Cleanup (Affymetrix, CA, USA), according to the manufacturers' instructions, in order to be sequenced. All sequencing reactions were performed on a Genetic Sequencer 3130XL using the Big Dye Terminator Kit v3.1 (Applied Biosystems, CA, USA). The identity of the obtained sequences was analysed by comparing to the software "Nucleotide BLAST NCBI" (NCBI 2015).
Concerning the verification of the 5' transgene flanking region of MON863, the sequence obtained from the DNA walking strategy was compared to the available data published by Zhu *et al.* (2008) using the "ClustalW2" software. The resulting alignment is provided in the Additional File 7.

Supporting information

Additional files 1–7 are available at http://bmcbiotechnol.biomedcentral.com/articles/10.1186/s12896-015-0191-3 (file formats: docx and pdf).

Conflict of interests

The authors declare that they have no competing interests.

Acknowledgements

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract UGMMONITOR (convention RF 11/6242). Sequencing is performed at the Platform Biotechnology and Molecular Biology at the Scientific Institute of Public Health. The authors would like also to thank Emmanuel Guiderdoni (CIRAD, UMR AGAP, Biological Systems department, Montpellier, France) for his kindness to provide rice grains, Sylvia Broeders and Nina Papazova for their critical view on the problematic of GMO detection in routine analysis and Sigrid De Keersmaecker for her highly precious advices.

Chapter 7

An integrated strategy combining DNA walking and NGS to detect GMO

The approach previously described in the chapters 4, 5 and 6 could become quite laborious with samples containing multiple GMO harbouring the same targeted elements. Therefore, this DNA walking strategy was here improved with the help of NGS in order to set up its throughput as well as to simplify its workflow to be more easily integrated in GMO routine analysis. This improved strategy may greatly facilitate the discrimination between the EU unauthorized GMO from the EU authorized GMO.

This chapter was previously submitted as:
M.-A. Fraiture, P. Herman, N. Papazova, M. De Loose, D. Deforce, T. Ruttink, and N. H. Roosens. "An integrated strategy combining DNA walking and NGS to detect GMO". *Food Chemistry*.

Authors' contributions

M.-A. Fraiture performed experiments, analysed data and drafted the manuscript. N. Papazova carried out the CoSYPS analysis of the Kuwaiti sample. P. Herman, M. De Loose, D. Deforce, T. Ruttink and N. H. Roosens helped to design the study and drafted the manuscript. All authors read and approved the final manuscript.

Abstract

Recently, we developed a DNA walking system for the detection and characterization of a broad spectrum of GMOs in routine analysis of food/feed matrices. Here, we present a novel version with improved throughput and sensitivity, by coupling the DNA walking system to Pacific Bioscience® Next-Generation-Sequencing technology. The performance of the novel strategy was thoroughly assessed via several assays. First, we tested detection and identification capability on unprocessed materials with a high or a low GMO content. Second, the potential impacts of food processing were investigated using rice noodle samples. Third, GMO mixtures and a real-life sample were analyzed to illustrate the applicability of the proposed strategy in GMO routine analysis. In all tested samples, the presence of multiple GMOs was unambiguously proven by the characterization of transgene flanking regions and the combinations of elements that are typical for transgene constructs.

7.1 Introduction

The traceability of food and feed products on the market and the right of consumers to know the exact ingredients have led to the establishment of legislations concerning the introduction and the control of genetically modified organisms (GMO) in the food and feed chain. In enforcement laboratories worldwide, the presence of GMO in food and feed matrices is routinely monitored through qPCR analyses. More precisely, the presence of GMO is initially assessed by qPCR screening using a panel of methods that target a broad range of common GMO elements and can discriminate between certain genetically modified (GM) events. Based on the positive and negative signals observed by these screening methods, a list of potential GM events present in the tested matrix is drawn up and the corresponding event-specific methods are then used to confirm their presence (Broeders *et al.* 2012b). However, this system is not specifically designed to identify unknown GMOs. Indeed, in the situation where no correspondence is established between the set of positively confirmed known GMOs and the signals observed during screening, the presence of unknown GMO can be inferred, but remains to be proven by alternative methods. This is, in part, because several targeted screening elements are derived from natural organisms (*e.g.*, p35S from Cauliflower Mosaic Virus (CaMV) and tNOS from *Agrobacterium*) that may naturally be present in the tested sample. In addition, most of these screening elements are commonly found in European (EU) authorized but also in EU unauthorized GMOs, obscuring their independent detection. Indeed, the explanation of observed qPCR screening signals by positive observations of known EU authorized GMOs does not prove the absence of EU unauthorized GMOs *per se* (Ruttink *et al.* 2010a; Broeders *et al.* 2012a; Broeders *et al.* 2012b; Holst-Jensen *et al.* 2012).

To solve this issue, an integrated DNA walking strategy has recently been developed to strengthen the current qPCR system regarding the detection of EU unauthorized GMOs (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b). After detection of a positive signal during the routine qPCR screening for the p35S, tNOS or t35S pCAMBIA elements that are frequently found in GMOs, we propose to perform this DNA walking method to amplify their respective unknown flanking regions. Full-length sequencing of the generated amplicons allows identifying the transgene flanking regions and/or the combinations of elements that are typical for transgene constructs. This DNA walking strategy therefore allows to unambiguously confirm the presence of GMO and to discriminate EU authorized and unauthorized GMOs (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b).

We previously demonstrated the simplicity and time-efficiency of this DNA walking strategy with samples containing only few GMOs. However, the workflow of this DNA walking strategy had to be improved to make the analysis

more efficient for samples containing multiple GMOs. More precisely, gel purification of individual DNA walking amplicons is required prior to sequencing using the Sanger technology. From a practical point of view, these steps could become quite laborious with food and feed matrices that contain multiple GMOs harboring the same targeted elements. When we take into account that all amplicons should be analyzed to guarantee that all GMO events present in the tested sample are amplified, detected and sequenced, the purification and sequencing steps of the potentially numerous amplicons generated represent a long and tedious work. With the aim to adapt this workflow into a versatile high-throughput method, the steps after the DNA walking PCRs were replaced in this study by a high-throughput Next Generation Sequencing (NGS) approach, as recently suggested (Liang *et al.* 2014; Fraiture *et al.* 2015b). To this end, several DNA walking PCRs were performed in parallel to be then, per sample, pooled and indexed with a unique barcode during the NGS library preparation. All indexed sequencing libraries were subsequently combined and sequenced together in one run on a PacBio® instrument. We assessed the capacity to characterize transgene flanking regions and/or combinations of elements that are typical for transgene constructs of the improved DNA walking strategy on various food/feed matrices: (i) a pure unprocessed GM material (rice grain) to test feasibility, (ii) raw materials (rice grain) containing 1%, 0.1% or 0.01% of GM rice to evaluate sensitivity, (iii) processed matrices (rice noodles) to estimate the potential impact of reduced DNA integrity due to food processing, (iv) complex mixtures with multiple GM rice, maize and/or soybean events, (v) a real life sample originating from the Kuwaiti market. We demonstrate that the proposed DNA walking strategy, fully integrated to the current GMO routine analysis, allows to efficiently identify known and unknown GMOs with the p35S, tNOS and t35S pCAMBIA elements in typical food/feed matrices. In addition, this strategy could easily by implemented by the enforcement laboratories since it takes the advantage of readily available high-throughput NGS sequencing technology to deliver high resolution DNA sequencing data, while data analysis does not require an exceptionally high level of bioinformatics expertise.

7.2 Materials and methods

7.2.1 Plant materials and sample preparation

Rice grains from transgenic Bt rice (*Oryza sativa L. Japonica cv Ariete*), transformed by *Agrobacterium tumefaciens* with the binary vector pCAMBIA 1300 containing the synthetic Cry1B gene from *Bacillus thuringiensis*, and its corresponding wild-type (WT) line were used in this study (Breitler *et al.* 2004).

Samples composed of 1%, 0.1% and 0.01% of Bt rice were then prepared by mixing DNA from WT and Bt rice in given proportions.
Using this rice grains, rice noodle samples composed of 100% of Bt rice (Bt noodles 100%) or WT rice containing 1% (Bt noodles 1%) of Bt rice (w/w) were also prepared as previously described (Fraiture *et al.* 2015a).
In addition, samples containing DNA from different GMO were prepared. Mixture-1 was composed of 2000 Haploid Genome Equivalents (HGE) of Bt rice and 2000 HGE of GM maize MON863. Mixture-2 contained 2000 HGE of Bt rice, 2000 HGE of GM maize MON863 and 2000 HGE of GM soybean GTS-40-3-2. The same GM events, each one at 20 HGE, were used to produce Mixture-3.
Certified Reference Material (CRM) of 9.8% GM maize MON863 (ERM-BF416c) and 10% GM soybean GTS40-3-2 (ERM-BF410dk) were obtained from the Institute for Reference Materials and Measurements (IRMM, Geel, Belgium).
A food matrix with positive detection of p35S and tNOS markers was obtained from Kuwaiti market (Kuwaiti matrix).

7.2.2 DNA extraction, concentration and purity

Rice grain was manually ground to a fine homogenous powder and used for DNA extraction using a CTAB-based procedure (ISO 21571) in combination with the Genomic-tip20/G (QIAGEN) procedure adapted from the EU-RL GMFF validated method (ISO 21571 2005; European Union Reference Laboratory for GM Food and Feed and Feed 2006b). This DNA extraction method comprises four successive steps: extraction of proteins, polysaccharides and organic components; precipitation of DNA in the presence of C-hexadecyl-Trimethyl-Ammonium-Bromide (CTAB); purification of DNA using a tip20 column; and precipitation of DNA with isopropanol (ISO 21571 2005; European Union Reference Laboratory for GM Food and Feed and Feed 2006b). DNA concentration was measured by spectrophotometry using the Nanodrop® 2000 (ThermoFisher) and DNA purity was evaluated using the A260/A280 and A260/A230 ratios. DNA extraction, concentration and purity of CRM and food matrix were carried out as previously described (Broeders *et al.* 2013).

7.2.3 qPCR analysis

The qPCR screening assays performed on the Bt rice 100%, Bt rice 1%, Bt rice 0.1%, Bt rice 0.01%, Bt noodles 100%, Bt noodles 1%, Mixture-1, Mixture-2 and Mixture-3 samples using the primers for p35S, tNOS, t35S pCAMBIA, PLD (Phospholipase D gene from rice), ADH (Alcohol dehydrogenase I gene from maize) and LEC (Lectin gene of soybean) markers (Tables A.1 and

A.2 in appendix B) (Vaïtilingom *et al.* 1999; Barbau-Piednoir *et al.* 2010; Mbongolo Mbella *et al.* 2011; Fraiture *et al.* 2014). PCR reactions contained 1 × SYBR®Green PCR Mastermix (Diagenode), and 250 nM of each primer in 25 μl. The tested DNA amount was the same as the one used for the DNA walking analysis (see sub-section 7.2.4). The qPCR program consisted of initial DNA polymerase activation for 10 min at 95°C followed by 40 amplification cycles of 15 sec at 95°C (denaturing step) and 1 min at 60°C (annealing-extension step). Melting curve analyses were performed by gradually increasing the temperature from 60 to 95°C in 20 min (±0.6°/20 sec).

The qPCR screening applied on 25 ng of DNA from the Kuwaiti matrix was carried out using p35S, tNOS, t35S pCAMBIA, PLD, ADH, LEC, CRU (Cruciferin gene from oilseed rape), pFMV (Promoter of the figworth mosaic virus), Cry3Bb (Gene encoding the *Bacillus thuringiensis* δ-endotoxin 3Bb), PAT (Phosphinotricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*), BAR (Phosphinotricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*) and CP4/EPSPS (5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain CP4) markers (Table A.1) (Mbongolo Mbella *et al.* 2011; Barbau-Piednoir *et al.* 2012; Broeders *et al.* 2013; Broeders *et al.* 2015). Based on the positive and negative signals obtained from the qPCR screening analysis, a list of potentially detected GM events was made using the CoSYPS (Combinatory SYBR Green qPCR Screening) decision support system (Figure 7.3) (Van den Bulcke *et al.* 2010). The presence of these GM events was then tested by a subsequent identification step using the Taqman qPCR event-specific methods for the 3272 maize, 98140 maize, Bt11 maize, DAS 40278-9 maize, DAS59122 maize, GA21 maize, MIR 604 maize, MIR 162 maize, MON810 maize, MON87427 maize, MON87460 maize, NK603 maize, T25 maize and TC1507 maize events, as recommended (Table A.1) (Joint Research Centre 2015a). All runs were performed on an iQTM5 real-time PCR detection system (BioRad) or an ABI 7300 qPCR system (Applied Biosystems, CA, USA). For each assay, a "No Template Control" (NTC) was included.

7.2.4 DNA walking

The DNA walking approach was carried out as previously described by Fraiture *et al.* (2014), Fraiture *et al.* (2015a), and Fraiture *et al.* (2015b) (Table A.1). First, a target-specific primer (a) and one kind of the degenerated random tagging primer (DRT) mixes (A-D) were applied. In a second and third semi-nested PCR, target-specific primers (b and c) are combined, respectively, with universal tagging primers (UAP-N1 and UAP-N2). Using the p35S-F, p35S-R, tNOS-F, tNOS-R and t35S pCAMBIA DNA walking methods, DNA from rice

samples composed of 100% of Bt rice (200 000 HGE of Bt rice) or WT rice containing 1% (2000 HGE of Bt rice; Bt rice 1%), 0.1% (200 HGE of Bt rice; Bt rice 0.1%) and 0.001% (20 HGE of Bt rice; Bt rice 0.01%) of Bt rice were analyzed. Similarly, Bt noodles 100% (200 000 HGE of Bt rice), Bt noodles 1% (2000 HGE of Bt rice), Mixture-1, Mixture-2 and Mixture-3 were also tested. In addition, the p35S-F, p35S-R, tNOS-F and tNOS-R DNA walking methods were applied on 100 ng of DNA from the Kuwaiti matrix.

7.2.5 Library preparation and sequencing

All PCR products generated by the DNA walking methods were purified using the QIAquick PCR Purification Kit (QIAGEN), according to the manufacturer's instructions. These samples were then dried with a Speedvac DN120 (ThermoSavant) prior to shipment to the sequencing provider (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA) for PacBio® library preparation and PacBio sequencing (Cold Spring Harbor Laboratory 2016). The quality of the samples was assessed using a 2100 Bioanalyzer (Agilent). Libraries were prepared with the PacBio® 2Kb Template Prep Kit (Pacific Biosciences), according to the manufacturer's instructions. All DNA walking PCR amplicons were pooled per sample prior to library preparation. During library preparation, a unique barcoded SMRTbell adapter was coupled to the PCR products generated for each of the ten samples (Bt rice 100%, Bt rice 1%, Bt rice 0.1%, Bt rice 0.01%, Bt noodles 100%, Bt noodles 1%, Mixture-1, Mixture-2, Mixture-3 and Kuwaiti matrix). After purification with the AMPure PB Beads Kit (Pacific Biosciences), SMRTbell templates were bound to primers and polymerases using the DNA/Polymerase Binding Kit (Pacific Biosciences). The MagBead Kit (Pacific Biosciences) was used to pool all SMRTbell templates into one SMRT® Cell on a PacBio RS II System (Pacific Biosciences) and sequenced using the DNA Sequencing Reagent Kit (Pacific Biosciences).

7.2.6 Sequencing data analysis

All generated sequences were demultiplexed by CSHL, yielding sequences between 11 and 7030 bp. For each set of sequences per sample, clusters of sequences with a high similarity (90% of identity) were created using the "CD-HIT Suite: Biological Sequence Clustering and Comparison" web server (Huang *et al.* 2010). From each cluster, the longest sequence was selected to further characterization.

For the Bt rice 100%, Bt rice 1%, Bt rice 0.1%, Bt rice 0.01%, Bt noodles

100% and Bt noodles 1% samples, these sequences were analyzed via public databases (National Center for Biotechnology Information (NCBI) and JRC GMO-Amplicons) using the tool nBLAST.
For the mixtures and real-life samples, these sequences were first compared by BLAST to a private database from the Belgian Scientific Institute of Public Health (WIV-ISP). This database contains the sequences from all EU authorized GMO and low level presence (LLP) cases. The sequences presenting no similarity to the private database were then compared by nBLAST to the public databases (NCBI and JRC GMO-Amplicons).
Among all generated sequences (284212), 3.1% were not further analyzed because they presented no significant similarities to the private and public databases.
All final sequences analysis (BLAST and alignment) was supported by the use of the CLC bio Genomics Workbench v 8.0 software.

7.3 Results and discussion

7.3.1 Performance assessment of the strategy on unprocessed matrices

The strategy coupling the DNA walking to the PacBio sequencing technology was first tested on a pure unprocessed GM material consisting of 100 ng (200 000 HGE) of Bt rice (Bt rice 100%), which contains the p35S, tNOS and t35S elements (Table A.2 in Appendix B). Using the DNA walking methods targeting these transgenic elements, combinations of elements that are typical for transgene constructs as well as transgene flanking regions were identified. All analyzed sequences were specific to the Bt rice event (Table A.3 in Appendix B). Indeed, the left and right transgene flanking regions, located on a genomic sequence from chromosome II of *O. sativa* [KT184679; AF234296; OSJNBa0016G10] and a genomic sequence from chromosome III of *O. sativa* [KT184678; AF234296; OSJNBb0111B07], were observed (Figure 7.1). Regarding the combinations of elements that are typical for transgene constructs, the obtained sequences revealed the pUBI promoter and tNOS terminator regions flanking the insect resistance Cry1B gene, the junction between the p35S and pUBI elements, and the p35S promoter and the t35S terminator flanking the hygromycin resistance (hpt) gene [KT184680; KT184677; KT184676; AF234296; KC414884.1; S94464.1] (Figure 7.1). Based on the DNA sequence of these elements, which do not normally occur in this configuration in nature, the presence of a GMO in the tested sample was proven and most of its construct was characterized (Figure 7.1). The identity of this GMO event was also revealed since the observed genome insertion sites corresponded to a Bt rice event that was previously characterized.

This is supporting the need for event-specific identification to discriminate EU authorized GMO from EU unauthorized GMOs. In addition, the results from the Bt rice 100% sample obtained with the novel DNA walking strategy were comparable to the ones previously obtained using the DNA walking strategy combined with gel purification and Sanger sequencing. Nonetheless, the sequenced amplicons presented here cover a larger region of the transgene insert due to the sequencing technology used (Fraiture *et al.* 2014; Fraiture *et al.* 2015b; Willems *et al.* 2016).

In GMO routine analysis, most of the encountered samples contain GMOs at trace level. Therefore, the sensitivity of the strategy was investigated through 100 ng of unprocessed WT rice samples containing a low percentage of unprocessed GM material, such as 2000 HGE (Bt rice 1%), 200 HGE (Bt rice 0.1%) or 20 HGE (Bt rice 0.01%) (Figure 7.1). First, with the Bt rice 1% sample, the left and right transgene flanking regions on chromosome II of *O. sativa* and chromosome III of *O. sativa* were detected. The transgenic cassette was also characterized through the exact configuration of the p35S, t35S, hpt, pUBI, Cry1B and tNOS elements (Figure 7.1). Second, in the case of the Bt rice 0.1% sample, three of the four transgene flanking regions were identified and the t35S, p35S, hpt, Cry1B and tNOS elements were detected inside the transgenic cassette (Figure 7.1). Third, the analysis of the Bt rice 0.01% sample allowed to cover only the two left transgene flanking regions and the left part of the transgenic cassette composed of the p35S, hpt and t35S elements. The analysis of the same sample by the previously published DNA walking strategy using gel purification and Sanger sequencing was not able to detect the left transgene flanking region on the rice chromosome II, as observed here (Figure 7.1) (Fraiture *et al.* 2015a; Fraiture *et al.* 2015b). This indicates that the improved DNA walking strategy has a high sensitivity.

Among a total of 79 492 sequences obtained for the Bt rice 1%, 0.1% and 0.01% samples, 10 133 (12.7%) sequences corresponded however exclusively to the rice genome [AP014957.1; AP014960.1; AP014961.1; AP014962.1; AP014963.1; AP014964.1; AP014965.1; AP014966.1; AP014967.1; AP014968.1; AP012527.1; KM088016.1] (Table A.3). The sequencing of off-target fragments on the host genome is more often observed in samples with a low percentage of GMO targets. Nevertheless, even in that cases, at least 80-90% of all sequenced data are derived from the targets, generating clear and unambiguously identifiable signals of the presence of GMOs (Table A.3). This can be explained because the use of degenerated random primers could amplify aspecific fragments, despite using three nested PCRs that increase the specificity. Alternatively, this could also be due to the fact that the final PCR products are here directly sequenced without prior gel purification of amplicons with specific fragment lengths (Fraiture *et al.* 2015a).

According to all these results, the proposed DNA walking strategy is able to

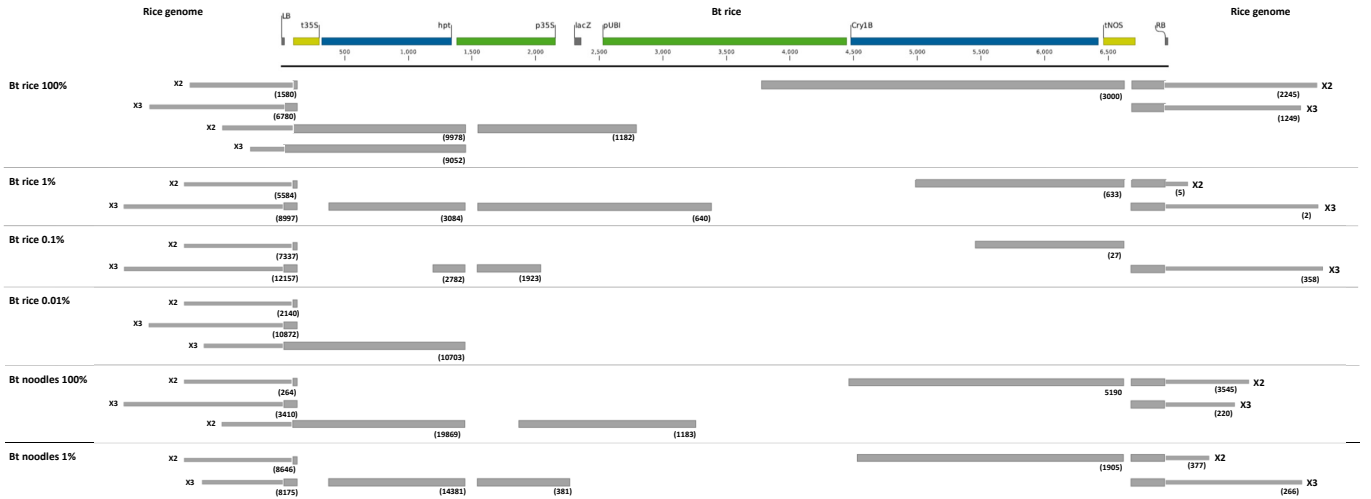


Figure 7.1: Characterization of Bt rice by the DNA walking/NGS strategy for the Bt rice 100%, Bt rice 1%, Bt rice 0.1%, Bt rice 0.01%, Bt noodles 100% and Bt noodles 1% samples. Below the transgenic cassette of Bt rice, the longest generated amplicons targeting the transgene flanking regions and constructs for each sample are schematically represented by grey rectangles. The number of sequences supporting each rectangle is indicated below in parentheses. The region corresponding to chromosome II and III from the rice genome are respectively indicated by X2 and X3. LB: left border; t35S: CaMV 35S terminator; hpt: hygromycin phosphotransferase gene; p35S: CaMV 35S promoter; lacZ: LacZ alpha fragment; pUBI: maize ubiquitin promoter; Cry1B: synthetic Cry1B gene; tNOS: *Agrobacterium tumefaciens* nopaline synthase terminator; RB: right border. The scale is indicated in bp.

detect GMOs, even at trace level. Indeed, the p35S-F, p35S-R, tNOS-F, tNOS-R and t35S pCAMBIA DNA walking methods presented a limit of detection in the range between 20 and 200 HGE, which is similar to the range (10 to 385 HGE) observed with the previously published DNA walking strategy using gel purification and Sanger sequencing technology (Fraiture *et al.* 2015a; Fraiture *et al.* 2015b).

7.3.2 Performance assessment of the strategy on processed matrices

Given that food processing affects DNA integrity, hampering the amplification of long intact GMO fragments, its potential impact on the proposed strategy was investigated using Bt noodles 100% and Bt noodles 1%, respectively prepared from rice grains with high (100%) or low (1%) percentage of GMO (Fraiture *et al.* 2015a).

All analyzed sequences of the Bt noodles 100% sample were specific to the target, similar to that observed for the unprocessed Bt rice 100% sample (Figure 7.1 and additional file 3). Moreover, all transgene flanking regions as well as a large part of the configuration of elements in the transgenic cassettes were identified. The regions characterized for the Bt noodles 1% sample were comparable to the ones from the unprocessed Bt rice 1% sample (Figure 7.1).

Taken together, these results indicate that the proposed strategy is able to cope with processed food matrices, similar to the previously published DNA walking strategy using the Sanger sequencing technology (Fraiture *et al.* 2015a). Indeed, even if the amount of long DNA fragments was strongly decreased by food processing, some amplicons of large sizes, reaching up to 6184 bp, were generated.

7.3.3 Performance assessment of the strategy on GMO mixtures

In GMO routine analyses, the tested samples generally contain multiple GMO ingredients. In that case, with the current qPCR GMO detection and decision system, the presence of EU unauthorized GMO could be concealed by the identification of one or more EU authorized GMOs. Indeed, if these can explain the signals observed in the qPCR screening analysis, the sample is considered as conform to regulations and no further analysis is required, even if the screening elements may be derived from additional unauthorized GMOs. Given that the proposed strategy has the potential to detect unauthorized GMOs in samples that also contain authorized GMOs, three samples mimicking

food/feed mixtures were prepared and analyzed to validate the strategy. The EU authorized MON 863 maize and GTS-40-3-2 soybean events, both containing the p35S and tNOS elements, were mixed with the EU unauthorized Bt rice event, containing the p35S, tNOS and t35S elements, to simulate scenario's with concealed unauthorized GMOs (Figure 7.2 and Table A.2). These three Mixture samples were first submitted to a qPCR analysis. The observation of positive signals for the p35S, tNOS and t35S pCAMBIA elements, as expected, led to the application of the corresponding DNA walking methods on the three Mixture samples (Figure 7.2 and Table A.2).

In the Mixture-1, eleven different classes of fragments were revealed. Through comparison to the private database, five of these fragments (n°1-5) were clearly derived from the MON863 maize event (Figure 7.2 and Figure A.1 in Appendix B). As six other fragments (n°6-11) showed no similarity to the private database, they were compared to the public databases. This revealed that they were derived from a Bt rice event [AF234296; OSJNBa0016G10; KT184678; OSJNBb0111B07; KT184680; KT184677; KT184676; AF234296; KC414884.1; S94464.1] (Figure 7.2 and Figure A.1) (Fraiture *et al.* 2014; Fraiture *et al.* 2015b).

Among the fourteen different classes of fragments detected in Mixture-2, five (n°1-5) and three (n°12-14) corresponded respectively to the MON863 maize and GTS40-3-2 soybean events, according to the private database (Figure 7.2 and Figure A.1). Comparison of the six other fragments (n°6-11) to public databases further confirmed the presence of a Bt rice event [AF234296; OSJNBa0016G10; KT184678; OSJNBb0111B07; KT184680; KT184677; KT184676; AF234296; KC414884.1; S94464.1] (Figure 7.2 and Figure A.1) (Fraiture *et al.* 2014; Fraiture *et al.* 2015b).

For the Mixture-3, eight different classes of fragments were observed. Comparison to the private database revealed that five (n°1-5) and one (n°13) fragments respectively showed the presence of the MON863 maize and GTS40-3-2 soybean events (Figure 7.2 and Figure A.1). Comparison of the fragments n°7 and n°11 to the public databases identified a Bt rice event [AF234296; KT184678; OSJNBb0111B07] (Figure 7.2 and Figure A.1).

As previously observed with other matrices containing low percentages of targets, few generated sequences from the mixtures corresponded entirely to the rice genome sequence [AP014957.1; AP014960.1; AP014961.1; AP014962.1; AP014963.1; AP014964.1; AP014965.1; AP014966.1; AP014968.1; AP012527.1], the maize genome [AC196084.4; DQ490951.2] and the soybean genome [AC235405.1; JX463295.1; X02623.1; XM_014761742.1].

A crucial advantage of the proposed strategy is its capacity to characterize and discriminate each GMO ingredient, based on the entire set of GMO specific sequences amplified from the tested sample. With the current qPCR GMO detection and decision system, only the p35S and tNOS screening markers could

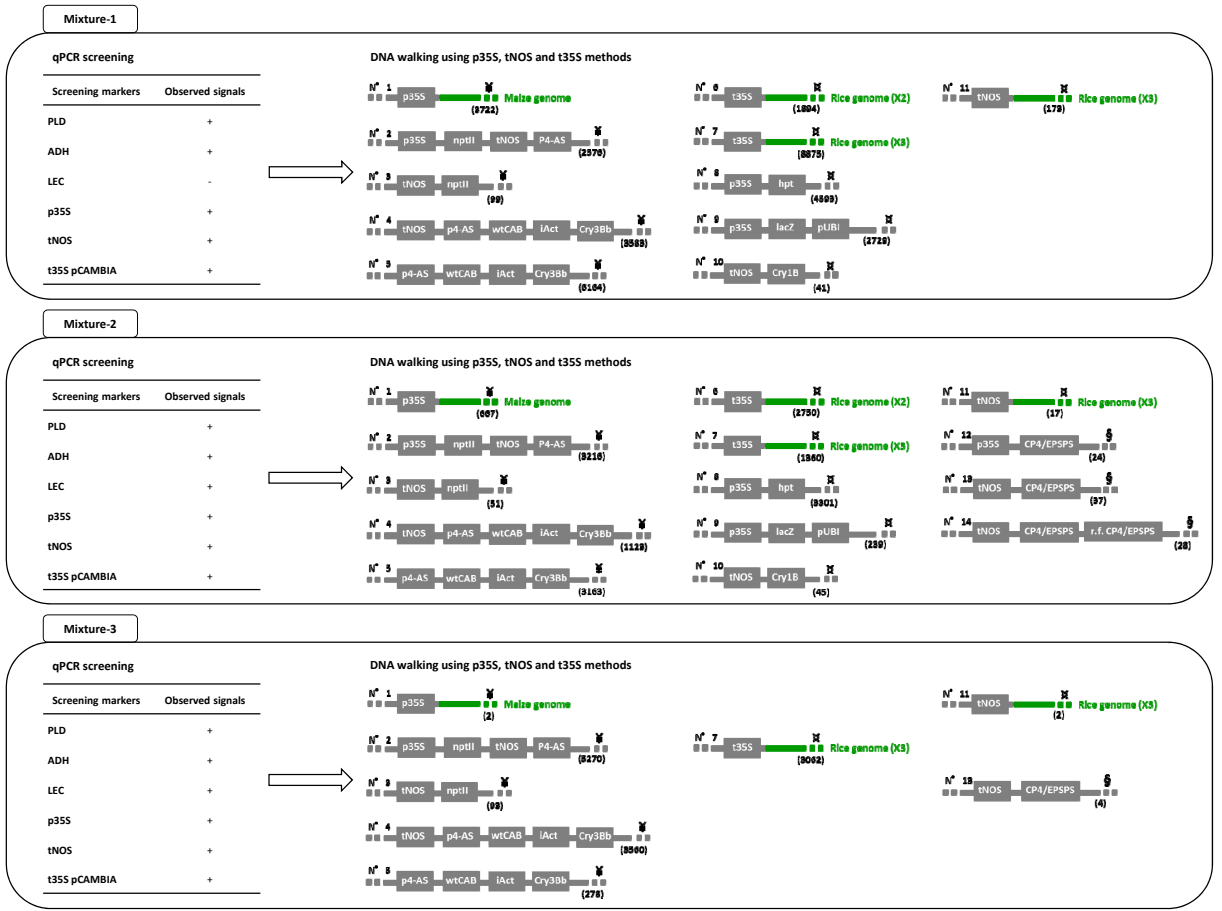


Figure 7.2: Schematic representation of the mixture sample analysis, including the qPCR and DNA walking assays.

Figure 7.2 *(continued)*: The positive and negative qPCR signals are represented respectively by + and −. The number of sequences supporting each different fragment characterized by the DNA walking analysis is indicated below in parentheses. The region corresponding to chromosome II and III from the rice genome are respectively indicated by X2 and X3. The fragments corresponding to MON863 maize, Bt rice and GTS40-3-2 soybean are respectively indicated by ¥, ¤and §. PLD: Phopholipase D gene from rice; ADH: Alcohol dehydrogenase I gene from maize; LEC: Lectin gene of soybean; t35S: CaMV 35S terminator; hpt: hygromycin phosphotransferase gene; p35S: CaMV 35S promoter; lacZ: LacZ alpha fragment; pUBI: maize ubiquitin promoter; Cry1B: synthetic Cry1B gene; tNOS: *Agrobacterium tumefaciens* nopaline synthase terminator; nptII: neomycin phosphotransferase II gene; p4-AS1: modified CaMV 35S promoter; wtCAB: Wheat major chlorophyll a/b binding protein gene; iAct: Rice Actin intron; Cry3Bb: synthetic Cry3Bb gene; CP4/EPSPS: *Agrobacterium tumefaciens* (strain CP4) 5-enolpyruvylshikimate-3-phosphate synthase; r.f. CP4/EPSPS repeated fragment of CP4/EPSPS.

be explained by the identification of the MON863 maize in the three Mixtures and the GT40-3-2 soybean in Mixture-2 and Mixture-3 (Figure 7.2). The t35S pCAMBIA marker observed in the three Mixtures indicates the presence of one or more EU unauthorized GMOs because this element is absent from the EU authorized GMO and LLP cases. However, there are no qPCR methods to prove the presence of event-specific elements of unauthorized GMOs because these are not made available to enforcement laboratories. Here, we demonstrate that all sequences generated with the DNA walking methods that do not correspond to the MON863 maize in the Mixture-1, 2 and 3 or to the GT40-3-2 soybean in the Mixture-2 and 3, directly prove the presence an EU unauthorized GMO and allow its identification (Bt rice).

7.3.4 Real-life sample analysis

Finally, to illustrate the application of the entire workflow of the proposed strategy, a real life sample (Kuwaiti matrix) was tested and compared to the current GMO routine analyses.

First, the sample was submitted to routine qPCR screening analysis to detect the potential presence of GMOs (Figure 7.3). According to the taxon-specific markers, the tested sample contained maize and traces of soybean ingredients. Moreover, the p35S, tNOS, PAT and CP4/EPSPS markers were positive. Based on all observed positive and negative signals from the qPCR screening analysis, a list of fourteen potentially present EU authorized GM events and LLP cases (3272 maize, 98140 maize, Bt11 maize, DAS 40278-9 maize, DAS59122 maize, GA21 maize, MIR 604 maize, MIR 162 maize, MON810 maize, MON87427 maize, MON87460 maize, NK603 maize, T25 maize and TC1507 maize) was established using the CoSYPS decision support system (Figure 7.3) (Van den Bulcke *et al.* 2010). Next, the event-specific methods corresponding to these GMO were tested, leading to the identification of the MON810, NK603 and TC1507 events. Some traces from the Bt11 event were also observed. According to enforcement regulations, this would conclude the analysis because all detected screening elements can be explained by the combined presence of the MON810, NK603 and TC1507 events.

Second, because the p35S and tNOS screening markers were positive in the screening analysis, the corresponding DNA walking methods (p35S-F, p35S-R, tNOS-F and tNOS-R) were applied on the sample. Eight different classes of sequences were observed, numbered from n°1 to 8, and corresponded to unique configurations of elements (Figure 7.3 and Figure A.2 in Appendix B). Even if these sequences allowed at this stage to prove the presence of GMO in the tested sample, all of them were then compared to confidential dossiers of EU authorized GMO and LLP cases in order to identify the GM events involved. This revealed

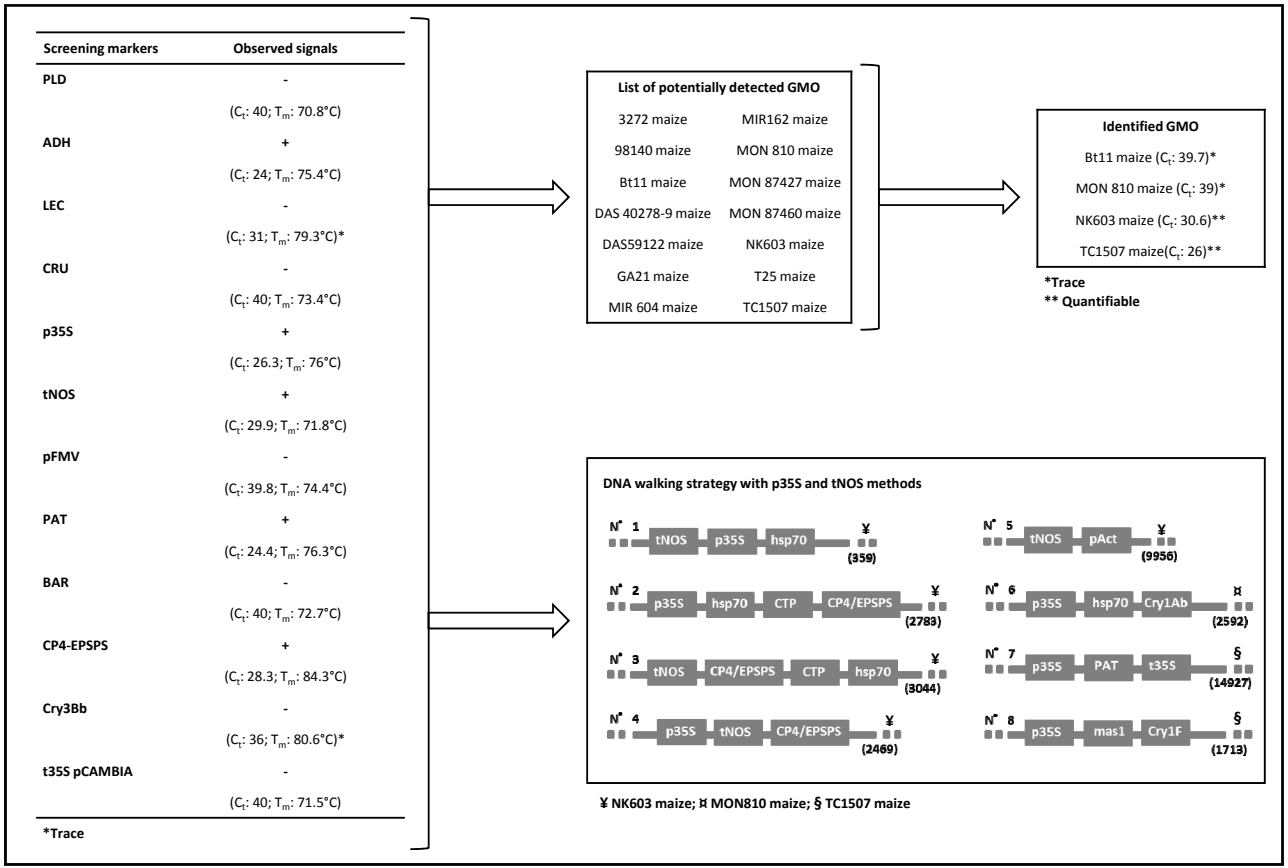


Figure 7.3: Schematic representation of the Kuwaiti matrix analysis, including the qPCR and DNA walking assays.

Figure 7.3 *(continued)*: The positive and negative qPCR signals are represented respectively by + and −. The number of sequences supporting each different fragment characterized by the DNA walking analysis is indicated below in parentheses. PLD: Phopholipase D gene from rice; ADH: Alcohol dehydrogenase I gene from maize; LEC: Lectin gene of soybean; CRU: Cruciferin gene from oilseed rape; pFMV: Promoter of the figworth mosaic virus; Cry3Bb: Gene encoding the *Bacillus thuringiensis* δ-endotoxin 3Bb; BAR: Phosphinotricin-N-acetyltransferase gene from *Streptomyces hygroscopicus*; CP4/EPSPS: *Agrobacterium tumefaciens* strain CP4 5-enolpyruvylshikimate-3-phosphate synthase gene; Cry1Ab: synthetic Cry1Ab gene; Cry1F: synthetic Cry1F gene; CTP: *Arabidopsis thaliana* chloroplast transit peptide; hsp70: maize hsp70 intron; mas1: mannopine synthase region; p35S: CaMV 35S promoter; pAct: rice Actin promoter; PAT: phosphinotricin acetyltransferase gene; t35S: CaMV 35S terminator; tNOS: *Agrobacterium tumefaciens* nopaline synthase terminator.

that the fragments n°1 to 5 were derived from parts of the transgenic cassette from the maize NK603 event. The fragment n°6 was clearly derived from parts of the transgenic cassette from the maize MON810 event. The fragments n°7 and n°8 matched with parts of the transgenic cassette from the maize TC1507 event. These results were consistent with the qPCR screening and identification analyses. Even if it was not the case with the tested sample, the use of the DNA walking strategy could also allow to reveal the presence of additional GMOs, which would otherwise remain concealed with the current qPCR GMO detection and decision system. Indeed, if no similarity was established between the identified GM events and other sequences, the presence and identity of EU unauthorized GMOs could have been highlighted and proven.

7.4 Conclusion

We demonstrate that the proposed strategy was able to detect and identify known and unknown GMOs in matrices frequently encountered in GMO routine analysis by the enforcement laboratories. A DNA walking strategy starting from screening elements and coupled to long-range NGS sequencing is used for the characterization of transgene flanking regions and combinations of elements that are typical for transgene constructs. This strategy resolves the observation of unexplained signals in routine qPCR screening analysis as well as reveals the presence of unknown GMOs that could otherwise be undetected or concealed by the current qPCR GMO detection and decision system.
By using the NGS technology from Pacific Biosciences® instead of Sanger sequencing, the proposed strategy was here improved in terms of easiness of practical application, throughput and sensitivity. Several PCR products from different food/feed matrices could be pooled together for sequencing, thus increasing the throughput of the strategy. In this study, 196 PCR products from 10 different samples were sequenced together in one run to obtain information for 6 events, and this is not the limit as hundreds or thousands of duplicated sequences were obtained per class of sequences (Figures 7.1, 7.2 and 7.3). In principle any NGS sequencing platform can be used because adaptor ligation kits for amplicons are available for all current NGS platforms. We chose the PacBio system for two main reasons. On the one hand, as the DNA walking strategy can produce amplicons with a size range varying approximately from 250 bp to 6 Kbp, the capacity of the PacBio technology to deal with heterogenic library sizes is a clear advantage. On the other hand, the PacBio technology currently provides the longest read-length (up to 60 Kbp), thus sequencing each amplicon in its entirety. This property avoids any shearing of the amplicons during library preparation as well as de novo assembly to reconstruct the amplicons during downstream data analysis. In case of food/feed matrix composed of

several unknown GMO, this last point could be nearly impossible to carry out without collapsing and merging common sequence elements (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b; Pacific Biosciences 2016).
In contrast to the proposed targeted strategy, the Whole-Genome-Sequencing (WGS) approach to identify unauthorized GMOs theoretically needs no *a priori* knowledge (Willems *et al.* 2016). However, this untargeted approach is currently not feasible for samples containing low levels of GMOs, which are the majority of the samples analyzed by the enforcement laboratories. This issue could probably be overcome with the ongoing increases in sequence yield, the further innovations in target enrichment by sequence capture or amplification, the novel long-range sequencing technologies, as well as in combining the specific abilities from several NGS platforms, such as in aligning short reads from Illumina to long reads from PacBio (Au *et al.* 2012).
Therefore, the proposed strategy represents at the present time a rational option to detect and characterize GMOs, in particular unauthorized GMOs. In addition, with the aim to provide an overview of GMOs present on the market and to reduce the cost of the analysis, several samples could be analyzed simultaneously in one experiment with the proposed strategy.

Conflict of interests

The authors declare no conflict of interest.

Acknowledgements

This research that yielded these results, was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contact UGMMONITOR (convention RF 11/6242). The authors thank Emmanuel Guiderdoni (CIRAD, UMR AGAP, Biological Systems department, Montpellier, France) for his kindness to provide rice grains, Sylvia Broeders for her critical view on the issues of GMO detection in routine analysis and Maud Delvoye for technical support.

Chapter 8

Statistical framework for detection of genetically modified organisms based on next generation sequencing

Even if the targeted approach previously developed in the chapters 4 to 7 was successful, a minimum of prior knowledge is required. To overcome this issue, an untargeted approach was also investigated. To this end, a WGS strategy was thus assessed in this chapter to detect GMO and a statistical framework was developed to predict the feasibility of this WGS strategy on samples frequently encountered in routine analysis.

This chapter was previously published as:
S. Willems, M.-A. Fraiture, D. Deforce, S. C. J. De Keersmaecker, M. De Loose, T. Ruttink, P. Herman, F. Van Nieuwerburgh, and N. Roosens (2016). "Statistical framework for detection of genetically modified organisms based on next generation sequencing". *Food Chemistry* 192, pp. 788–798. DOI: 10.1016/j.foodchem.2015.07.074

Authors' contributions

S. Willems analysed data and drafted the manuscript. M.-A. Fraiture contributed to design the experiment, helped to analyse data, prepared DNA samples and drafted the manuscript. T. Ruttink, F. Van Nieuwerburgh and N. H. Roosens helped to design the study, analysed the data and drafted the manuscript. D. Deforce, S. De Keersmaecker, P. Herman and M. De Loose helped to design the study and drafted the manuscript. All authors read and approved the final manuscript. S. Willems, M.-A. Fraiture and D. Deforce have contributed equally as first author. F. Van Nieuwerburgh and N. H. Roosens have contributed equally as last author.

Abstract

Because the number and diversity of genetically modified (GM) crops has significantly increased, their analysis based on real-time PCR (qPCR) methods is becoming increasingly complex and laborious. While several pioneers already investigated Next Generation Sequencing (NGS) as an alternative to qPCR, its practical use has not been assessed for routine analysis. In this study a statistical framework was developed to predict the number of NGS reads needed to detect transgene sequences, to prove their integration into the host genome and to identify the specific transgene event in a sample with known composition. This framework was validated by applying it to experimental data from food matrices composed of pure GM rice, processed GM rice (noodles) or a 10% GM/non-GM rice mixture, revealing some influential factors. Finally, feasibility of NGS for routine analysis of GM crops was investigated by applying the framework to samples commonly encountered in routine analysis of GM crops.

8.1 Introduction

In recent years, the number and diversity of genetically modified (GM) crops on the market have drastically increased (James 2013). Legislations related to GMO (genetically modified organism) commercialisation differ from country to country, but it is internationally agreed that GMOs can only be commercialised after thorough safety assessments. To this end, GMO developers have to perform molecular characterisation of each novel GMO subjected to authorisation. This molecular characterisation includes the determination of the inserted DNA sequence via the evaluation of the number of inserts using Southern blot analysis and Polymerase Chain Reaction (PCR). Furthermore, Sanger sequencing of the junction of the transgene insert and the host genome is used to determine its precise location as well as the detection of possible presence of the backbone sequence of the transformation vector. This approach is relatively time-consuming and requires customised experiments, carefully designed for each event (Kovalic *et al.* 2012).

The DNA sequence data of the insert junctions is also used for the development and validation of the event-specific detection method, required for subsequent GMO monitoring in food and feed products by EU enforcement laboratories (The European Parliament and The Council Of The European Union 2003a; The European Parliament and The Council Of The European Union 2003b). These laboratories use quantitative real-time PCR (qPCR) to screen for the presence of commonly used DNA elements in GMOs and then, using event-specific methods provided by the GMO developers, to identify a GMO (Broeders *et al.* 2012a). To increase the efficiency of GMO detection, qPCR methods are being used that run on a 96-well plate with multiplex qPCR for simultaneous detection. Moreover, Decision Support Systems have been developed to deal with the complexity of multiple PCR signals (Brodmann *et al.* 2002; Foti *et al.* 2006; Waiblinger *et al.* 2008; Bahrdt *et al.* 2010; Dörries *et al.* 2010; Huber *et al.* 2013; Köppel *et al.* 2014; Morisset *et al.* 2014). If the presence of unauthorised GMOs (UGMs) is suspected, additional analyses, like DNA walking, are performed to identify the junction between the host genome and the transgene sequence to identify or better characterise the UGM (Ruttink *et al.* 2010a; Fraiture *et al.* 2014). Although this methodology has been optimised for use by enforcement laboratories, the DNA walking method can be laborious in the case of a complex mixture.

While GMO analysis has benefitted from multiplexing PCR methods, limitations like a maximum of six targets per qPCR experiment (Bahrdt *et al.* 2010) and unbiased primer design with equal analytical performance for a multiplex assay compared to simplex assays remain. Furthermore, the qPCR strategy *per se* implies the prior knowledge of at least part of the sequence of the transgene integrated in the host genome as well as the subsequent development of an

efficient assay targeting this sequence. Collecting these sequences and designing the corresponding method for each new sequence target case by case remains challenging today, especially for unknown GMOs. This poses a major problem as GMOs remain undetectable when no method targeting the transgene element has been used. Recently, Next Generation Sequencing (NGS) has been proposed to tackle these challenges.

NGS, allowing massive parallel DNA fragment sequencing, was of great importance to sequence several complete plant genomes and is being used in the sequencing of many more plant genomes (Michael and Jackson 2013). As a consequence, the use of NGS has been proposed to provide an informative and cost-effective alternative to the current Southern blot-based method for molecular characterisation of plant GMOs. One of these alternatives assumes the availability of a reference genome of the GM crop and the sequence of the inserted transgene cassette. Based on this information, Kovalic *et al.* (2012) used NGS to characterise the junctions on both sides of a specific transgene cassette. Other approaches have been developed to exploit the potential of NGS for GMO detection and analysis when a reference genome of the GM crop is available, but only partial or no prior knowledge of the sequence of the transgene insert is available (Wahler *et al.* 2013; Yang *et al.* 2013). Liang *et al.* (2014) have dealt with GMOs by developing a targeted strategy combining a chromosome walking method, based on SiteFinding-PCR, and NGS technology. In this study, a part of the cassette is known and targeted (partial *a priori* knowledge). The NGS technology is not used for full characterisation of the GM crop but rather as a high-throughput sequencing technology that is more time-efficient than Sanger sequencing to individually sequence DNA fragments. These pioneer studies in the context of NGS-based GMO detection showed the applicability of NGS to circumvent the limitations posed by the qPCR strategy and Sanger sequencing. The major benefit of NGS is its independence of *a priori* knowledge of the transgene sequence. Because NGS is a relatively new technique applied to GMO detection, the infrastructure and expertise amongst scientists of enforcements laboratories, mainly molecular biologists, is often not present. A key component for short term implementation of NGS is therefore the development of bioinformatics capacity by enforcement laboratories. This includes the availability of computing infrastructure, the development or implementation of adequate software and the development of expertise in order to manage, analyse and gain new information from NGS data. A second challenge is related to the nature of the DNA that needs to be analysed by NGS during GMO analysis in routine; including the large size of plant genomes, lack of good reference genomes for specific varieties or organisms due to large intraspecific genome variability in plants, DNA samples with traces of GMO material and degraded DNA due to food processing. While some of these issues have already been tackled, *i.e.* large intraspecific variability can be

circumvented by an initial alignment against the transgenic cassette (Yang *et al.* 2013), the applicability of NGS for routine analysis has not been previously investigated.

To accommodate NGS within routine GMO detection, a first priority is capturing transgene information with NGS. The focus on a specific sequence (transgene insert) within a given genome, as opposed to reconstructing the entire genome sequence, means that statistical methods for the estimation of sequencing depth versus coverage of whole genomes, like the Lander–Waterman theory (Sims *et al.* 2014), are not applicable. Therefore, a novel conceptual statistical framework is developed in this article to draw a better picture of the present feasibility of NGS technology for routine GMO analysis. This statistical framework was validated by NGS data from a GM rice (Bt rice), with known transgene insert and flanking regions, and is based on three approaches: (1) detecting potential transgene inserts, (2) proving their integration in the host genome, (3) identifying the specific junctions. All these approaches start with an alignment against an *a priori* known insert and only the aligned reads are subsequently investigated to avoid large intraspecific variability in plants. To assess the potential applicability of NGS on different types of food matrices, 100% Bt rice grains, 10% Bt rice grains mixed with 90% non-GM rice grains and 100% Bt rice noodles were analysed. To evaluate the robustness of these three approaches, they were implemented on two different data analysis platforms: an easy-to-use commercial software platform, the "CLC Genomics Workbench", allowing potential use of NGS by "bioinformatics novices', and a "Command-Line" platform allowing greater control of the workflow and parameters, but demanding a higher level of expertise in bioinformatics. This newly developed statistical framework allows to determine the probability that a given GMO can be detected when its presence in a sample is known. Based on this probability, an estimate of the number of reads necessary to be able to detect a transgene cassette, to prove integration in the host genome and to identify several common GMO events and mixtures can be calculated.

8.2 Materials and methods

8.2.1 Statistical framework

Three approaches, addressing different levels of complexity in the analysis of GMOs, are used to analyse shotgun sequencing libraries, sequenced as paired-end reads from a sample that consists of a single GMO. The "detection approach" was used to detect the presence of a transgene cassette, referred to as the insert. The "proof approach" allows to provide the evidence that the insert is effectively

integrated in the non-GM genome, referred to as the host genome, and gives a crude localisation of the insert in the host genome. The "identification approach" delivers the precise identification and localisation of the junctions between the host genome and the insert (Figure 8.1).

Calculation of probabilities to successfully detect a sequence aligned to a transgene For each approach, the probability to successfully detect a theoretical read in an NGS sample of a known GMO, $P(+|\text{GMO})$, was calculated. False positives were not considered and as a result the probability of an unknown sample containing a GMO when testing positive $P(\text{GMO}|+)$ was not determined. For a GMO, the length of the GM genome is the sum of the length of the non-GM genome (H) and the length of the insert (I). A partial insertion is defined as an insert with a large part of the insert deleted. In this case the length of the partial insertion is considered as the length of the insert (I). After sequencing of the GMO, this gives a total of different mates (T_s), with an average read length for each mate (R), equal to $H + I - R + 1$ or a total of different theoretical paired-end reads (T_p), with an average paired-end distance (D), mates included, equal to $H + I - D + 1$.
To be able to detect the presence of a known insert, only sequences that fall completely in the inserted region can be detected using a global alignment. As a consequence, partial insertions that are smaller than the read length ($I < R$) are impossible to detect with this method. If $I \geqslant R$, there are $I - R + 1$ different theoretical mates that possibly align.
To be able to prove that the insert is integrated in the host genome as well as to give a rough location, a theoretical paired-end read needs one mate globally aligned to the host genome and the other mate globally aligned to the insert. Similarly to the detection of an insert, it is only possible to find such kind of sequences when $I \geqslant R$. If the paired-end distance of a theoretical paired-end read (D) is large enough ($D \geqslant I + R$), a mate globally aligned to the insert will always have a mate globally aligned to the host genome. Otherwise, a theoretical paired-end read, with each mate globally aligned to either the insert or the host genome, will span a junction only if the junction is not located on either of the mates. The length of this sequence is equal to $D - 2 \cdot R$, so there are $D - 2 \cdot R + 1$ theoretical paired-end reads for each different junction.
To be able to identify these junctions, a sequence needs to locally align its 5' and 3' tail to respectively the host genome and the insert with a minimum overlap of nucleotides (M) for each tail or vice versa. This is impossible in cases with a very small partial insertion ($M > I$). If the read length R is large compared to the insert ($I + M < R$), a theoretical mate that locally aligns to the insert with overlap O is then guaranteed to have at least M basepairs overlap with the host genome reference. In this case, there are exactly $I - M + 1$ different theoretical mates that locally align to the insert with O basepairs

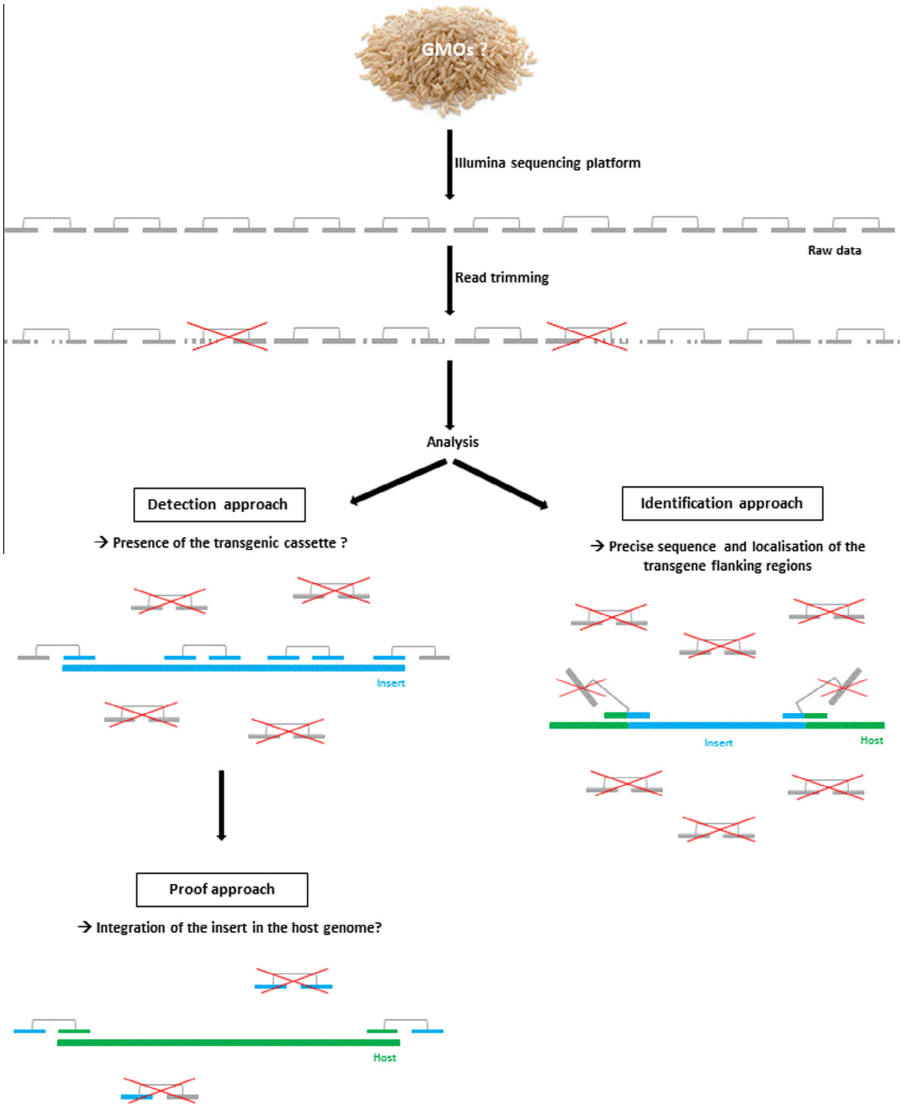


Figure 8.1: GMO analysis workflow based on NGS. From a given matrix, extracted DNA is used for shotgun library construction and sequenced on an Illumina platform to obtain millions of raw paired-end reads. These are first trimmed based on sequencing quality scores and then filtered so only paired-end reads remain with each mate having a length of 30 bp or larger. To determine the presence of GMOs, the filtered reads are then analysed using three different

Figure 8.1 *(continued)*: approaches. On the one hand, the detection approach selects all paired-end reads with one mate globally aligned to the reference sequence of the insert, revealing its presence in the tested sample. The corresponding mates of the detected reads are subsequently analysed in the proof approach to confirm the integration of the transgenic insert in the host genome by globally aligning these mates to the reference sequence of the host genome. This approach also allows a rough localisation of the transgene flanking regions. On the other hand, all filtered reads are analysed with the identification approach to determine the exact localisation and sequence of the flanking regions by locally aligning them to the host genome and transgenic insert simultaneously.

overlap. Finally, if $I \geqslant R - M$, sequence covers a junction only if the junction is covered by the $R - 2 \cdot M + 1$ bp in the middle of the sequence. For each junction there are thus $R - 2 \cdot M + 1$ different theoretical mates.
The ratio of targeted theoretical reads over the total of theoretical reads is now given by the following formulae:

(1) p_1: The ratio of theoretical mates globally aligned to the insert, as needed for the detection approach.
   (a) $p_1 = \frac{I-R+1}{T_\mathrm{s}}$ if $I \geqslant R$
   (b) $p_1 = 0$ if $I < R$ (small partial inserts)

(2) p_2: The ratio of theoretical paired-end reads covering a single junction with one mate globally aligned to the host genome and the other to the insert, as needed for the proof approach.
   (a) $p_2 = \frac{D-2\cdot R+1}{T_\mathrm{p}}$ if $I > D - R$
   (b) $p_2 = \frac{I-R+1}{T_\mathrm{p}}$ if $I \leqslant D - R$ and $I \geqslant R$
   (c) $p_2 = 0$ if $I < R$ (small partial inserts)

(3) p_3: The ratio of theoretical mates covering a single junction with one mate locally aligned to the host genome and the other locally aligned to the insert, as needed for the identification approach.
   (a) $p_3 = \frac{R-2\cdot M+1}{T_\mathrm{s}}$ if $I > R - M$
   (b) $p_3 = \frac{I-M+1}{T_\mathrm{s}}$ if $I \leqslant R - M$ and $I \geqslant M$ (small partial inserts)
   (c) $p_3 = 0$ if $I < M$ (very small partial inserts)

Given these ratios of theoretical paired-end reads or mates, it is straightforward to calculate the probability $P(-|\text{GMO})$ that no targeted reads are found after sequencing N paired-end reads, all originating from a GMO:

(1) P_1 (no insert detected, while reads originate from a GMO) $= (1 - p_1)^{2N}$, if N paired-end reads are considered as $2N$ independent mates.

(2) P_2 (no proof of integration of insert detected, while reads originate from a GMO) $= (1 - 2p_2)^N$, assuming either one of the two junctions suffices as a proof of this insertion.

(3) P_3 (no identification of junctions possible, while reads originate from a GMO) $= (1 - 2p_3)^{2N}$, if N paired-end reads are considered as $2N$ independent mates and either one of the two junctions suffices for identification.

Conversely, the probability to detect at least one read in a pure sample extracted from a GMO is equal to $P(+|\text{GMO}) = 1 - P(-|\text{GMO})$.

Estimation of the number of paired-end reads needed to have a probability P of finding at least one targeted read For a GMO, the number N of paired-end reads that are needed to have a probability P of finding at least one targeted read can easily be retrieved by rewriting the formulae in the previous paragraph and is given by the next formulae where p_i is defined as before:

(1) $N_1 = \frac{\ln(1-P)}{2\ln(1-p_1)}$ paired-end reads are needed for the detection approach.

(2) $N_2 = \frac{\ln(1-P)}{\ln(1-2p_2)}$ paired-end reads are needed for the proof approach.

(3) $N_3 = \frac{\ln(1-P)}{2\ln(1-2p_3)}$ paired-end reads are needed for the identification approach.

while the above formula are completely general and can be used for any pure sample of a GMO, they can be greatly simplified for most common cases. In general the host genome length H is large compared to the insert length I, which in turn is large compared to the paired-end distance D and read length R. As a result the probabilities p_1, p_2 and p_3 will be small, so $\ln(1 - p_\text{i}) \approx -p_\text{i}$. Furthermore, the total number of reads T_s and T_p can be simplified to H, and the constant 1 in the numerator can be omitted for all probabilities p_i. In summary, this yields the following simple approximations for the number of paired-end reads N needed to have a probability P of finding at least one targeted read:

(1) $N_1 \approx \frac{H}{2\cdot(R-1)} \cdot \ln(1-P)$

(2) $N_2 \approx \frac{H}{2\cdot(2\cdot R-D)} \cdot \ln(1-P)$

(3) $N_3 \approx \frac{H}{4\cdot(2\cdot M-R)} \cdot \ln(1-P)$

The following parameters are thus of importance to make an estimate:

- *A priori* known or estimated
  - o I: length of the insert reference.
  - o H: length of the host genome reference.
- Definable by user
  - o R: sequenced read length (average).
  - o D: sequenced paired-end distance (average), including mates, thus larger than twice the read length.
  - o M: minimum overlap length between each tail of a mate and the host genome /insert reference, thus smaller than halve a read length. Software often has default parameters for M, dependent on read length R.
  - o P: probability to find at least one targeted read.
- Calculated result
  - o N: number of quality filtered paired-end reads needed to have a probability P of finding at least one targeted read.

Modifications for more complex cases It is possible to adjust the presented formulae to different scenarios that better reflect food and feed matrices complexity. These matrices usually contain only traces of a GMO or might contain a mixture of different ingredients. In such cases the ratio of targeted reads should be multiplied by the DNA ratio r of the GMO over the rest of the sample. For instance a mixture of 10% GM rice (genome size of 400 Mbp) and 90% non-GM maize (genome size of 2300 Mbp) has a DNA ratio of $\frac{10\cdot 400}{90\cdot 2300+10\cdot 400} \approx 0.019$ for the GM rice. Since in the simplified version the number of needed paired-end reads is linearly dependent on the ratio of targeted reads, it follows that the linear dependence of the DNA ratio is not only valid for ratios of targeted reads, but also for the number of paired-end reads needed to be able to detect at least a single read for each approach. By calculating the probability of detecting exactly $x-1, x-2, \ldots, 0$ reads, it is also possible to calculate the probability of detecting at least x reads instead of detecting at least one read.

8.2.2 Generation of NGS data from different food matrices

Three DNA samples were generated from transgenic Bt rice (see Supplementary Text S1 for a description of DNA extraction and the inserted cassette): (1) Bt rice grains (named 100% Bt rice sample), (2) Bt rice grains processed into noodles as described in Fraiture *et al.* (2015a) (named 100% Bt noodles sample), (3) mixture of 10% Bt rice DNA with 90% of the corresponding non-GM rice DNA (named 10% Bt rice sample).

Library preparation and sequencing Two Illumina shotgun sequencing libraries were generated, one from 5 μg of the 100% Bt rice sample and the other from 5 μg of the 10% Bt rice sample. DNA was fragmented to 300–400 bp using Covaris S2 sonication and an indexed sequencing library was prepared using an Illumina TruSeq DNA Sample Preparation kit. The two resulting libraries were sequenced simultaneously on a single Rapid Run flow cell with 2 lanes, one per library, with an Illumina HiSeq 2000 sequencer, generating 2 $\times$ 100 bp paired-end reads for each sequenced fragment. After base calling using the Illumina CASAVA version 1.8 software, raw sequences were obtained.
The 100% Bt noodles sample was sequenced several months later, using updated protocols and techniques. In this case, an Illumina shotgun sequence library was generated from 1 μg of the 100% Bt noodles sample. DNA was fragmented to $\pm$400 bp using Covaris S2 sonication and a sequencing library was made using the NEBNext Ultra DNA Library Prep Kit with 8 enrichment PCR cycles. Size selection was performed on the resulting library using an Invitrogen 2% E-gel, selecting fragments between 400 and 600 bp. The library was sequenced on half of a Rapid Run flow cell lane on an Illumina HiSeq 1500 sequencer, generating 2 $\times$ 100 bp paired-end reads for each sequenced fragment. Base calling and primary quality assessments were performed using Illumina's Basespace genomics cloud computing environment.

8.2.3 Implementation of the framework

Two different platforms were used to analyse the NGS data: (1) freely available programs such as BWA (Li and Durbin 2009) and Bowtie2 (Langmead and Salzberg 2012) combined with Python and Perl scripts were used on a computer running Linux Ubuntu 14, referred to in this manuscript as the Command-Line-Tools, (2) the commercial software package CLC Genomics Workbench 7 (CLC bio 2015) running on Windows 7 Enterprise, referred to in this manuscript as the CLC Genomics Workbench.
For the host genome reference the sequence of *Oryza sativa* was used, more specifically the MSU6 build of *O. sativa* of length 374 332 026 bp available

via Illumina's IGenomes (Illumina 2015a), which includes pseudomolecules representing the mitochondria, plastids and Syngenta sequences. The reference of the inserted pCAMBIA cassette was obtained from Breitler *et al.* (2004) as a personal communication and consists of 7002 bp.

Command-Line-Tools Using a custom Perl script, the sequenced paired-end reads were trimmed when the average quality in a sliding window of 10 bp fell below Q20 and were filtered for sequences shorter than 30 bp after trimming (Del Fabbro *et al.* 2013). Only paired-end reads were retained.
For the detection approach, the mates (each paired-end read consists of two mates) of all quality filtered paired-end reads were considered as single-ended and were aligned end-to-end (global alignment) to the insert using BWA with default parameters (BWA manual version 0.7.7-r441). Results were converted to SAM format (Li *et al.* 2009) and only aligned mates were selected.
For the proof approach, corresponding mates of those previously aligned in the detection approach were retrieved with a custom python script. These mates were then aligned to the host reference genome using default BWA parameters, similarly to module 1 described by Yang *et al.* (2013). Results were then converted to SAM format and unaligned mates were discarded.
For the identification approach, the mates of all quality filtered paired-end reads were considered as single-end and were partially aligned (local alignment) to the insert using Bowtie2. A length of 20 bp for the part initially aligned before elongation starts (seed), located at the beginning or end of a sequence with a maximum of one mismatch, was used instead of default Bowtie2 parameters, as found in the Bowtie2 manual version 2.2.1. Only mates that aligned were selected from the resulting SAM file. Mates with a CIGAR string (Li *et al.* 2009) matching a global alignment were discarded and the remaining mates were aligned against the host genome reference with the same parameters. Only aligned mates were selected from the resulting SAM file, again discarding mates with a global alignment. The resulting mates were divided in groups corresponding to different junctions, similar to the study published by Kovalic *et al.* (2012).

CLC Genomics Workbench Similar to the Command-Line-Tools, a separate stand-alone analysis was done with the CLC Genomics Workbench.
All sequenced paired-end reads were trimmed with the NGS Core Tool "Trim Sequences" with an ambiguous trim length of 2, quality limit of 0.05 and minimum length of 30. Only paired-end reads were retained.
The quality filtered paired-end reads were globally aligned to both the insert and the host genome simultaneously using the NGS Core Tool "Map Reads to Reference" with similarity fraction 0.8, length fraction 1.0 and default

parameters, as found in the CLC Genomics Workbench 7 Manual. Only paired-end reads with at least one mate aligned to the insert were selected for downstream analysis.
To verify that this insert was effectively integrated in the host genome, the option "Find Broken Pair Mates" was used on the selected paired-end reads to retrieve mates that did not align to the insert but to the host genome instead. All quality filtered paired-end reads that were not globally aligned to the insert nor the host genome, were selected to identify the junction sequences. These paired-end reads were locally aligned against the insert with the NGS Core Tool "Map Reads to Reference" with similarity fraction 0.8 and length fraction 0.3. Aligned paired-end reads were selected and realigned against the host genome with the same parameters.

8.3 Results and discussion

8.3.1 Statistical framework

In Section 8.2.1, a statistical framework was developed to investigate the use of NGS in routine analysis of GMOs. The formulae of this framework predict the number of NGS reads needed to have a probability P to detect transgene sequences, to prove their integration into the host genome or to identify the specific transgene event in a sample with known composition based on a number of parameters. To verify if the developed statistical formulae are good predictors, they were implemented using Command-Line-Tools and compared with the experimental results from the 100% Bt rice, 10% Bt rice and 100% Bt noodles samples (Section 8.3.2). We identify and discuss several influential factors that have an impact on the formulae of the statistical framework.

Validation based on experimental results The statistical framework takes several parameters as input. To estimate some of these values, a global alignment against the host genome reference was carried out with all N quality filtered paired-end reads, using BWA with default parameters.
The *a priori* parameters used for all samples were $I = 7002$ bp and $H = 374\,332\,026$ bp. Experimentally, two different insertions were previously identified in Bt rice, one of length $I_2 = 6868$ on chromosome II and one of length $I_3 = 6936$ on chromosome III. Furthermore, literature suggests that the length H of the host genome of *O. sativa japonica* is actually 385 Mbp (Kawahara *et al.* 2013) instead of 374 Mbp, the length of the used reference. The ratio r of the genome reference length over the actual genome length of 0.97 was used to correct the ratio of targeted reads.

For the 100% Bt rice sample, the user definable parameters were $R = 100$, $N = 28$ and $D = 350$ (average). However, the experimental values R and D were respectively approximated at 86.4 (average) and 208.35 (average). Although the average read length R is usually a good approximation, it should be noted that the average paired-end distance has a large spread and is skewed (data not shown). To calculate the probability P of detecting no single reads, the number of quality filtered reads 91 371 164 is used as N, instead of the number of raw reads. Similar parameters were applicable for the 10% Bt rice and 100% Bt noodles samples (Supplemental Table S1).
In the 100% Bt rice sample, 26.6% of the reads originate from the mitochondria and plastids according to the global alignment against the host genome, while the statistical framework was developed for pure genomic DNA. The ratio p_{i} for each approach was thus corrected by a ratio r of 0.34 to correct for the abundance of mitochondrial DNA. For the 10% Bt rice and 100% Bt noodles samples, the percentage of reads that aligned to the mitochondria/plastids was respectively 18.2% and 11.3%.
The ratios of targeted reads p_{i} were used to estimate the number of paired-end reads that are to be expected ($E = p_{\mathrm{i}} \cdot N$) after sequencing N quality filtered paired-end reads. Since two identical insertions were present, the ratio of targeted reads for the detection approach was equal to the sum of the ratio for each individual approach. The probabilities for the proof and identification approach were calculated separately for each insertion, since the insertions are independent. The number of expected reads (Table 8.1) was compared to the number of experimental reads (Table 8.2) and were found to be in agreement. The largest deviations concern the 10% Bt rice sample where absolute values are low (< 5) and these results were disregarded due to the low statistical significance of few reads.

Identification of influential factors It should be noted that several assumptions and simplifications have been made to develop the formulae. First, all theoretical reads are assumed to be perfect and to not contain any errors. Although this assumption is not true in reality, it affects both targeted and untargeted reads. It can thus be assumed that the ratio of targeted reads over total reads (p) is mostly unaffected by this property, even though the number of quality filtered reads will be reduced when errors are present. Similarities between the insert and the host genome add an extra level of complexity to the analysis. Currently, the host genome and insert of a GMO are often of a different species or even of a different kingdom, *i.e.* Plantae and Bacteria, and are genetically different. However, in a near future, many new GMOs are expected to be developed with *cis*-genic inserts and there might be cases where this property has a major influence on the analysis (Espinoza *et al.* 2013; Holme *et al.* 2013). Another important assumption is that all reads are equally likely to

Table 8.1: Theoretical formulae of the statistical framework applied to the 100% Bt rice, 10% Bt rice and 100% Bt noodles samples for all three approaches. For the detection approach, reads from both inserts were analysed simultaneously, as they cannot be identified separately. For the proof and identification approach the reads of the inserts were investigated independently. All used parameters are shown in Supplemental Table S1. Experimentally detected true positive reads from the Command-Line-Tools are shown in brackets.

		Detection approach	Proof approach		Identification approach	
			Chromosome II	Chromosome III	Chromosome II	Chromosome III
100% Bt rice	Ratio of theoretical targeted reads over theoretical possible reads (millionfold)	25.891	0.070	0.070	0.060	0.060
	Probability P to detect at least one read	1.00	1.00	1.00	1.00	1.00
	Expected reads (truly detected reads)	4 731 (3 186)	13 (23)	13 (17)	22 (12)	22 (22)
10% Bt rice	Ratio of theoretical targeted reads over theoretical possible reads (millionfold)	2.921	0.008	0.008	0.007	0.007
	Probability P to detect at least one read	1.00	0.68	0.68	0.85	0.85
	Expected reads (truly detected reads)	406 (284)	1 (2)	1 (0)	2 (0)	2 (2)
Bt Noodles	Ratio of theoretical targeted reads over theoretical possible reads (millionfold)	31.573	0.082	0.082	0.100	0.100
	Probability P to detect at least one read	1.00	1.00	1.00	1.00	1.00
	Expected reads (truly detected reads)	4 265 (5 485)	11 (30)	11 (11)	27 (22)	27 (25)

Table 8.2: Overview of the number of detected reads per approach for both the Command-Line-Tools and the CLC Genomics Workbench applied to the three samples; 100% Bt rice, 10% Bt rice and 100% Bt noodles. The detection approach was designed to detect an insert by finding reads that align to the used insert reference. The proof approach was designed to prove integration of the insert in the host genome by finding mates of detected reads in the detection approach that align to the host genome reference. The identification approach identified junctions between the host genome reference and the insert by locally aligning reads to both the insert and the host genome. In brackets true/false positives are shown.

Sample name		100% Bt rice	10% Bt rice	Bt noodles
Total paired-end reads		123 574 914	93 206 312	69 931 700
Command-Line-Tools	Quality filtered paired-end reads	91 371 164	69 464 211	67 539 855
	Detection approach	3 186	284	5 485
	Proof approach	51 (40/11)	2 (2/0)	98 (41/57)
	Identification approach	49 (34/15)	9 (2/7)	77 (47/30)
CLC Genomics Workbench	Quality filtered paired-end reads	107 455 990	81 491 366	68 981 939
	Detection approach	3 876	339	5 691
	Proof approach	88 (74/14)	6 (4/2)	134 (55/79)
	Identification approach	952 (20/932)	538 (1/537)	514 (24/490)

be sequenced. However, it has been shown that regions with a high GC content are underrepresented in *i.e.* Illumina sequencing and it was found that some regions cannot be sequenced at all with Illumina (Rieber *et al.* 2013). With enough prior knowledge about these issues though, the calculated probabilities can be adjusted accordingly.

Aside from these assumptions and simplifications, there are some limitations in defining all parameters, although they have a major influence on the analysis. First, a proper reference sequence of both the insert and the host genome is required. In reality this is not always the case, as shown in this study where the reference genome was only 374 Mbp as opposed to the literature suggesting it should be 385 Mbp (Kawahara *et al.* 2013), implying a different sequence. Since only reads with at least one mate aligned against the insert are used for downstream analysis, differences in the reference genome sequence will only be a limitation if they are near an insert site. Therefore, older draft genomes or other cultivars, as used in this study, are expected to have little influence. However, the plant genome size may vary greatly within a species or between cultivars (Ohri 1998; Greilhuber 2005), having a large effect on the statistical formulae. Furthermore, the formulae were developed for a pure sample. This study showed that even for a pure sample a significant part of the sequence data is not derived from the GMOs chromosomal DNA, but from the mitochondrial genome instead. This is not surprising for rice with a single diploid nuclear genome of almost 400 Mbp and a mitochondrial genome of almost 500 Kbp with a copy number of potentially over a 100 per cell (Bendich and Gauriloff 1984). The main difficulty is that it is not easy to estimate the relative amount of mitochondrial DNA a priori. Not only do different species have a different mitochondrial DNA size, but even within a single organism the number of mitochondria per cell is variable between tissues or organs (Mackenzie and McIntosh 1999; Tian *et al.* 2006). In addition, it can be difficult to determine/control experimental parameters properly. For instance, the number of high quality paired-end reads N is difficult to know beforehand (Kircher *et al.* 2011). It is highly dependent on the quality of the raw unfiltered reads. These reads are produced by a whole sequencing process where different batches of reagents are used, errors in detection of fluorescence are possible and cluster density is variable. Most experimental procedures for library preparation generate a range of DNA fragment sizes. This uncertainty can greatly be reduced by size selecting fragments of the library on a gel. In addition, despite variation in library insert sizes, the paired-end distance D after sequencing on an Illumina HiSeq instrument is typically in the range of 100–300 bp, due to the competitive efficiency of small fragments during the cluster formation by bridge-PCR in Illumina instruments.

8.3.2 Experimental results

Results of the 100% Bt rice sample Two different platforms, the CLC Genomics Workbench and a combination of Command-Line-Tools, were used to implement the statistical framework. After quality filtering, the detection, proof and identification approach (Figure 8.1) were applied to the 100% Bt rice sample on both platforms. Multiple mates globally aligned to the insert on both platforms by using the detection approach. By using the proof approach, their corresponding mates aligned in a small range on chromosome II in the region 22 981 000–22 982 000 and on chromosome III in the region 23 613 000–23 614 000, indicating two independent insert sites (Figure 8.2). By using the identification approach, both platforms identified four junctions; (1) at position 22 981 764 of chromosome II and at position 94 of the insert; (2) at position 22 981 674 of chromosome II and at position 6962 of the insert; (3) at position 23 613 353 of chromosome III and position 22 of the insert; (4) at position 23 613 341 of chromosome III and position 6958 of the insert (Figure 8.3). False positives, due to PCR artefacts, chimeric reads or genomic similarities between the insert sequence and the host genome, were filtered out by inspection of the alignments and their quality and mapping scores. These results are summarised in Table 8.2.

Effect of different samples The 10% Bt rice and 100% Bt noodles samples were analysed in a similar way as the 100% Bt rice sample (Table 8.2).
Degraded DNA in the Bt noodles sample did not impair the construction of the shotgun sequencing library, because the fragment size of the degraded DNA was larger than the selected fragment size of 300–400 bp for sonication (Supplemental Figure S1).
A factor 10–20 more reads, aligned to the insert, were detected in the 100% Bt rice sample compared to the 10% Bt rice sample in all the described approaches and platforms. Since the number of quality filtered reads is 1.3 times higher for the 100% Bt rice sample than for the 10% Bt rice sample, the results agreed with an expected factor 13.

Effect of different approaches The detection approach was used to detect the presence of the insert in the sample and provided a minimum of 284 hits for all the samples (Table 8.2). Analysis of the read mapping (*i.e.* Figure 8.2 for the 100% Bt rice sample) at nucleotide resolution showed few mismatches in their global alignments, suggesting no or few false positive hits. This number of properly aligned mates highlights the power and significance of the detection approach.
The proof approach was used to prove the integration of the insert within the

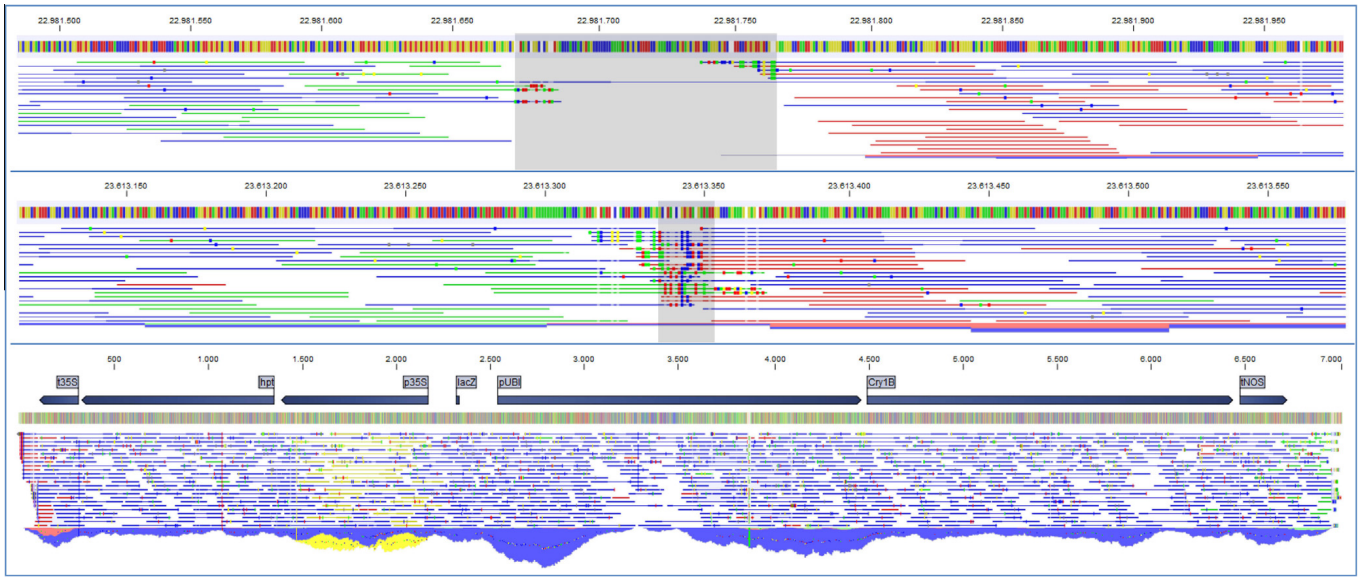


Figure 8.2: Global alignment of 100% Bt rice reads using CLC Genomics Workbench. Region 22 981 500–22 981 950 of chromosome II of the host genome (top), region 23 613 150–23 613 550 of chromosome III of the host genome (centre) and the complete insert of length 7002 (bottom), including globally aligned reads using the CLC Genomics Workbench. Reference nucleotides are shown as vertical bars with the four different bases in different colours on top of each image. Below this reference all reads from the 100% Bt rice sample that are globally aligned to this region are shown. Reads with corresponding mates are indicated in blue with a thin line connecting them. Green and red coloured reads do not have their respectively reverse and forward mate pairs aligned in this region. Yellow coloured reads indicate ambiguous reads with multiple possible alignments, in this case corresponding to a repeated region in the promoter p35S on the insert. Mismatches are shown on each read. A clear deletion is present on chromosome II, while a smaller one is

Figure 8.2 *(continued)*: detected on chromosome III, indicated by a grey shaded box, this part is replaced by the insert. When the end of a read originates from the insert, but is aligned to chromosome II or chromosome III, multiple mismatches can be detected, *i.e.* around position 22 981 750 of chromosome II. A single read seems to span the complete insert on chromosome II, although this is unlikely and it is more plausible that a minor contamination with the non-GM type occurred. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

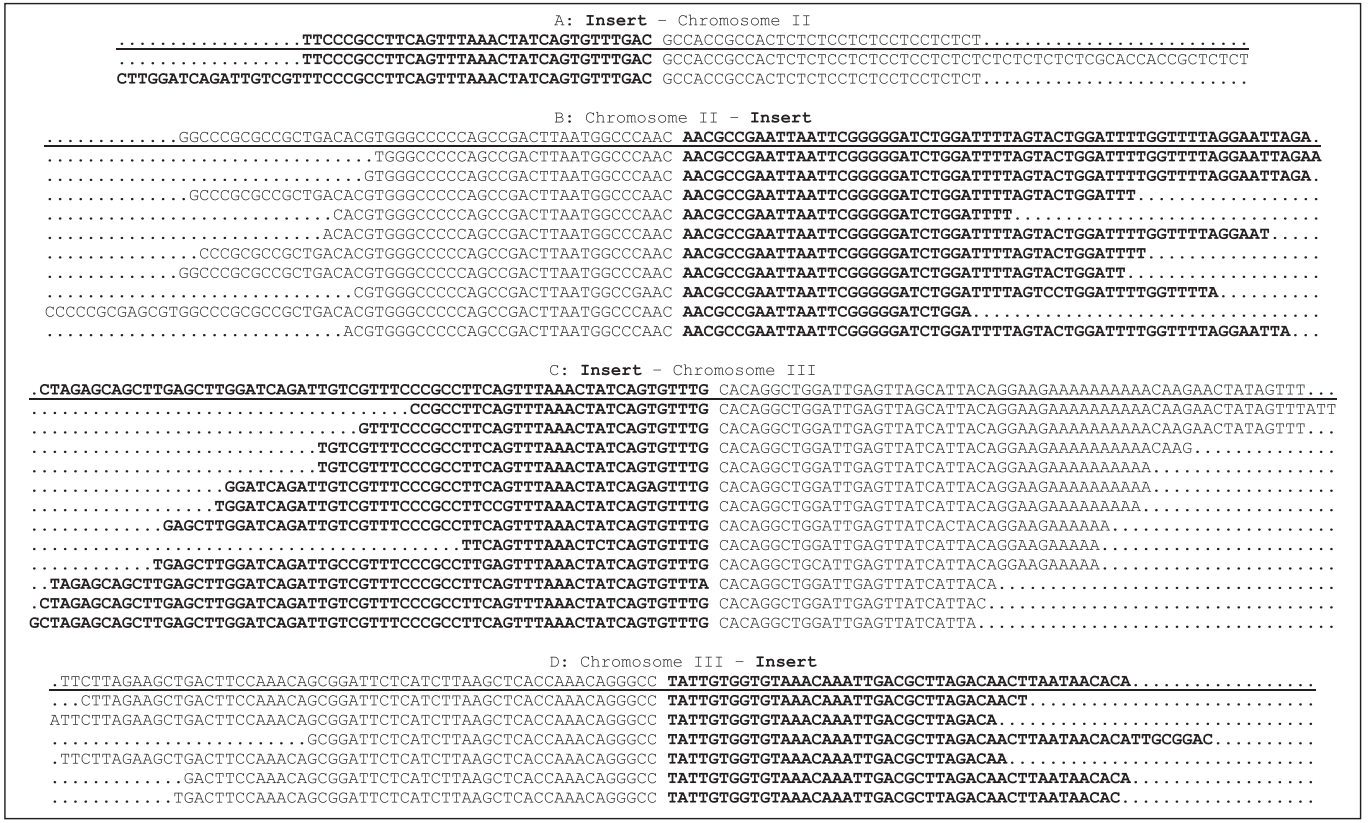


Figure 8.3: Single-end reads covering the junctions for the 100% Bt rice sample, detected with the identification

Figure 8.3 *(continued)*: approach using the Command-Line-Tools. The consensus sequence is underlined. The transition between host genome and insert is indicated by a gap. The part of each read belonging to the insert is indicated in bold.
(A) Junction with transition on insert position 6962 and chromosome II position 22 981 674. (B) Junction with transition on chromosome II position 22 981 764 and insert position 94. (C) Junction with transition on insert position 6958 and chromosome III position 23 613 341. (D) Junction with transition on chromosome III position 23 613 353 and insert position 22.

host genome. In the design of this approach only a subset of the quality filtered paired-end reads, those with aligned mates in the detection approach, was used. For all samples, multiple true positive hits were found, although some false positive hits were observed as well (Table 8.2). For the 100% Bt rice and 100% Bt noodles samples, the proof approach has provided strong evidence of insert integration into the host genome, while the result for the 10% Bt rice sample is of low significance with a minimum of two detected mates.
Mates covering junctions were detected using the identification approach. A sufficient number of true positive hits were detected to make a strong identification for the 100% Bt rice and 100% Bt noodles samples. The identification of each specific junction, compared to the detection of either the left or right junction necessary for identification, proved to be less reliable, as suggested by the presence of only two identifiable mates for the junction on position 22 981 674 of chromosome II and position 6962 of the insert of the 100% Bt rice sample with the Command-Line-Tools. A moderate and high level of false positives was respectively found for the Command-Line-Tools and for the CLC Genomics Workbench. Since the overlap was relatively small (28 bp) and some mismatches were allowed, the presence of false positive hits was expected. Inspection of false positive hits relied on the fact that the 5' tail of a mate should align to the insert, while the other tail should align to the host or vice versa in regions that were deemed interesting by the proof approach. The consensus sequence for all junctions was in perfect agreement (100% identity) with the DNA sequences originating from the DNA walking technique of Fraiture *et al.* (2014) (personal communication).

Effect of different platforms The CLC Genomics Workbench provides intuitive implementation and easily interpretable output formats such as figures and graphs, at the cost of full control of all parameters. Due to this limited control, the results are prone to false positives which are not straightforward to avoid and can be hard to identify graphically. An example is the high number of false positives in the identification approach, due to the lack of a parameter that specifies a seed location for the alignment. The CLC Genomics Workbench thus sacrifices some robustness for user-friendliness.
The Command-Line-Tools rely on textual/tabular information, extendable with other tools for visualisation that were not investigated in this article. Different software tools are available, each with their own benefits and drawbacks (Ruffalo *et al.* 2011), but in this article only BWA and Bowtie2 were used in combination with custom Python and Perl scripts. While some false positives are inherent to sequencing technology, textual/tabular representation provides easy identification of false positives since they are often single occurrences with low alignment/mapping qualities, as opposed to true positive hits where multiple hits were found per region. Remaining false positives can often be filtered with

the right software tools or custom scripts in subsequent steps. An analysis on the Command-Line-Tools is thus less affected by false positives than the CLC genomics Workbench, but knowledge of several tools and/or programming languages is essential.

8.3.3 Feasibility of using NGS data for GMO detection

All three approaches used to detect, to prove and to identify GMO events provided the same cassettes, junction sequences and number of insertions as those described in previous studies (Fraiture *et al.* 2014; personal communication). Command-Line-Tools and commercial software for bioinformatics analysis were able to come to the same results for the used samples. These samples; 100% Bt rice, 10% Bt rice and 100% Bt noodles, were of limited complexity. There are reference sequences available for both the insert and the host genome and the flanking regions of the two insertions are known.
To explore the feasibility of NGS for routine analysis using the statistical framework, it was applied to some theoretical samples containing common GMOs as shown in Table 8.3. For instance, to identify at least one paired-end read that aligns to a 7 Kbp transgene cassette in the rice genome of 384 Mbp with probability of 0.95, about 7 million paired-end reads need to be generated. Larger genomes, like wheat, will need 300 million paired-end reads to achieve the same result. Based on this information, it can be concluded that pure samples consisting of 100% GMO can, at the time of writing, reasonably be characterised with a single lane on an Illumina Hiseq2500 Rapid Run, yielding roughly 300 million paired-end reads per experiment (Illumina 2015b), at a standard price range (Genohub 2015). The required number of sequencing runs and associated costs increase when samples with only trace amounts of 1% GMO or less are investigated. For instance, for a wheat genome sample with trace amounts of 0.01% GMO, more than 30 billion paired-end reads, equal to a hundred Rapid Run lanes, are necessary to be able to only detect the insert with a probability of 0.95. Even this amount of data does not yield a high probability of detecting reads proving host genome integration or identifying the event.

8.4 Conclusion

The laborious analysis of an increasing number of GMOs using qPCR technology and the ineffectiveness in detecting “unknown and new GMOs” creates a need for alternatives to the current qPCR technology. In this context, NGS, allowing ‘detection-by-sequencing’ of GMOs in food and feed matrices, was proposed since

it circumvents the need to design specific primers to amplify target sequences for each specific GM event. Although some previous studies have shown the successful use of NGS to detect and identify GMOs, only limited information is available about the feasibility/applicability of NGS for routine GMO analysis, hampering its implementation in enforcement laboratories. In the present study a statistical framework was developed that offers preliminary, yet practical information that needs to be considered before NGS becomes routine use in GMO analysis.

Three approaches are considered in the framework: the "detection approach" to detect transgene sequences, the "proof approach" to prove integration of transgenes into the host genome and the "identification approach" to identify the specific transgene event. For each approach, formulae were developed to calculate the probability P of detecting at least one read in an NGS experiment with N reads or vice versa the number of reads N needed for a probability P to detect at least one targeted read. This framework was validated by using experimental data from a 100% pure Bt rice grains sample, a 10% Bt rice grains mixed with 90% non-GM rice grains sample and a noodles containing 100% pure processed Bt rice sample. Robust experimental results were obtained, regardless of implementation of the framework, on both the CLC Genomics Workbench and by using Command-Line-Tools. The experimental results of all three samples agreed with the theoretical results of the "detection approach", "proof approach" and "identification approach".

There are several assumptions and drawbacks of the approaches in the statistical framework. While whole genome complexities are avoided in the analysis by aligning reads to the transgene reference before aligning them to the host genome reference, the reference sequences of the transgenic cassette and host genome are always required *a priori*. Furthermore, the statistical framework was developed to calculate the probability to detect a GMO in a sample with known composition $P(+|\text{GMO})$, and is not fit to calculate the probability that a GMO is truly present when a sample with unknown composition is analysed $P(\text{GMO}|+)$. To achieve the latter, the statistical framework must be further developed to calculate the probability of false positives when no GMO is present in a sample.

Finally, the framework was applied to a range of different samples commonly encountered in routine analysis. It was shown that it is theoretically possible to use NGS to detect and identify samples of 100% GM crops. However, diluted samples and mixtures require large NGS experiments, with billions to trillions of reads and their associated costs, to yield a high probability of finding targeted reads for each approach.

It is concluded that the developed statistical framework can be used to estimate the number of NGS reads needed to detect a GMO in a given sample, and to help decide whether it will be useful to perform a NGS experiment. When the

Table 8.3: Common GMO samples and the number of reads N (in millions) needed to find at least one targeted read with a certainty of $P = 0.95$ for each approach as proposed in the statistical framework. Read length $R = 100$, insert length of GMO cassette $I = 7000$ (unstacked and homozygous), overlap length $M = 30$ and paired-end distance $D = 300$ are assumed for each sample. All samples are assumed to be pure genomic DNA. In case of mixtures the DNA ratio r of the GMO can be calculated as explained in Section 8.2.1. Experiments requiring more data than currently obtained with a single lane on an Illumina Rapid Run (300 million paired-end reads) are indicated in italics.

Species	Genome size (H) in Mbp	Number of reads reads (in millions) needed for the		
		Detection approach $N = \frac{(\ln 1-P)}{2\cdot\ln\left(1-r\cdot\frac{1-R+1}{H-R+1}\right)}$	Proof approach $N = \frac{(\ln 1-P)}{\ln\left(1-2\cdot r\cdot\frac{D-2\cdot R+1}{H-D+1}\right)}$	Identification approach $N = \frac{(\ln 1-P)}{2\cdot\ln\left(1-2\cdot r\cdot\frac{R-2\cdot M+1}{H-R+1}\right)}$
100% GM Rice (*Oryza sativa*)	385 (diploid)	0.08	5.71	7.03
100% GM Sugar beet (*Beta vulgaris*)	758 (diploid)	0.16	11.24	13.85
100% GM Soybean (*Glycine max*)	1115 (diploid)	0.24	16.54	20.37
100% GM Oilseed rape (*Brassica napus*)	1235 (tetraploid)	0.27	18.32	22.56
100% GM Cotton (*Gossypium hirsutum*)	2250 (tetraploid)	0.49	33.37	41.10
100% GM Maize (*Zea mays*)	2300 (diploid)	0.50	34.11	42.01
100% GM Wheat (*Triticum aestivum*)	17000 (hexaploid)	3.69	*252.12*	*310.53*
1% GM Rice + 99% WT Rice	385 (diploid)	8.36	*570.97*	*703.28*
0.01% GM Rice + 99.99% WT Rice	385 (diploid)	*835.65*	*57 096.88*	*70 327.91*
1% GM Wheat + 99% WT Wheat	17000 (hexaploid)	*368.99*	*25 211.61*	*31 053.32*
0.01% GM Wheat + 99.99% WT Wheat	17000 (hexaploid)	*36 898.63*	*2 521 083.57*	*3 105 081.42*
50% GM Rice + 50% WT Maize	385 (diploid) + 2300 (diploid)	0.58	39.82	49.05
0.1% GM Soy + 99.9% WT Oilseed	1115 (diploid) + 1235 (tetraploid)	*536.07*	*36 627.44*	*45 114.57*

composition of a sample is unknown, the framework can still be used to estimate how many NGS reads are needed to form a hypothesis about the presence of a specific GMO, but no significance testing can be done and any results of an NGS experiment need to be confirmed by targeted molecular analysis in an independent analysis afterwards.

Supporting information

Supplementary Text S1, Supplemental Table S1 and Supplemental Figure S1 are available at http://www.sciencedirect.com/science/article/pii/S0308814615011073 (file format: docx).

Conflict of interests

The authors declare no conflicts of interest.

Acknowledgements

The research that yielded these results, was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the Project UGMMONITOR (convention RF 11/6242) and an Ylieff Grant. The authors also thank Emmanuel Guiderdoni (CIRAD, UMR AGAP, Biological Systems department, Montpellier, France) for his kindness to provide rice grains.

Chapter 9

General conclusion

The aim of this PhD was to collect first crucial information on GM crops to develop then strategies allowing to detect and prove the presence of GMO, especially EU unauthorized, in food and feed chain in order to strengthen the current GMO detection system used by the enforcement laboratories. To this end, two approaches were suggested, including a targeted approach and an untargeted approach.

On the one hand, as a minimum of prior knowledge is required to develop an efficient and appropriate targeted approach, information about EU unauthorized GMO was collected. To this end, an inventory of GM rice, as a study case, developed around the world was initially carried out (Chapter 3). Based on the collected information, the current qPCR screening analysis was assessed as being able to cover 97.8% of the GM rice, via mainly the p35S and tNOS elements that are frequently found in both EU authorized and unauthorized GMO. This work allowed also to highlight some issues with the current qPCR GMO detection system. Indeed, given that most of the targeted elements, including p35S and tNOS, are originating from natural organisms, their detection cannot indubitably prove the presence of GMO in the tested sample. Moreover, these elements are useless in term of discrimination since they are present in both EU authorized and unauthorized GMO. To cope with this last point, this work allowed to identify one candidate marker for the qPCR screening step. The t35S pCAMBIA element was indeed considered as a key discriminative target since this element is absent from the EU authorized GMO and present in around 30% of EU unauthorized GMO (Chapter 3). Therefore, in this thesis, a new qPCR screening marker targeting the t35S pCAMBIA element was developed allowing to increase the discriminative power of the qPCR

screening analysis in order to detect EU unauthorized GMO (Chapter 4). Next, a DNA walking strategy, anchored on the t35S pCAMBIA element, was developed and validated to indubitably prove the presence of GMO that contain this element via the characterization of the sequences from their transgene flanking regions and from the associations of elements typically found transgenic constructs (Chapters 4 and 5). To complement this DNA walking strategy and to cover a broader spectrum of GMO, both EU authorized and unauthorized, DNA walking methods anchored on the p35S and tNOS elements were also developed and validated (Chapter 6). Furthermore, with the help of the NGS technology, the workflow of this DNA walking strategy was simplified to be more easily implemented by the enforcement laboratories. By this way, each GM event present in the tested sample, even at trace level in a mixture, could be individually identified (Chapter 7). To date, this DNA walking strategy, anchored on the t35S pCAMBIA, p35S and tNOS elements, was successfully applied on several typical food/feed matrices encountered in GMO routine analysis, suggesting its conceivable implementation by the enforcement laboratories.

On the other hand, an untargeted approach was developed and assessed to detect and prove the presence of GMO in food/feed matrices in using the WGS strategy, which requires theoretically no prior knowledge. Based on this work, a statistical framework was then provided in order to predict the probability to detect and characterize GMO in food/feed matrices frequently encountered in GMO routine analysis (Chapter 8). This approach was however assessed as not implementable at the present time in GMO routine analysis. This is due to the fact that this approach was not able to deal with samples containing GMO from plant species with complex genome (*e.g.* wheat), GMO in low amounts and GMO mixtures.

Through this PhD, useful information as well as key strategies were provided to the enforcement laboratories in order to improve their current GMO detection system.

Chapter 10

Broader international context, relevance and future perspectives

As previously described in chapter 1, the detection of GMO, to guarantee the traceability in the food and feed chain as well as the freedom of choice for consumers, is a worldwide problematic, even though quite extreme in EU.

This PhD was included in the context of the UGMMONITOR research project (convention RF 11/6242) of the contractual research regarding the food safety and plant and animal health policy. This project was financed by the Federal Public Service Health, Food Chain Safety and Environment (FPS Health) that allocates grants for scientific research supporting the food safety and the plant and animal health policy. Besides the scientific relevance, the project was also assessed by the FPS members in terms of opportunity, consistency and its potential impact on the economy and public health. Concretely, the following criteria were taken into account: the current priorities of the federal authority, the needs for the sector, the value and applicability of the expected results, the solution-orientation approach of the research, the added value in regard to ongoing or existing research projects/areas, the potential contribution to the policy decision making, the seriousness of the problematic and the budgetary, societal and ethical impacts.

The thematic of this project is thus relevant *per se*. In particular, in this PhD thesis, several impacts regarding the EU enforcement laboratories can be highlighted. Indeed, the qPCR screening analysis is classically composed of

markers that target sequences found in both EU authorized and EU unauthorized GMO (Figure 10.1A). To increase the discriminative power of this analysis, the t35S pCAMBIA qPCR screening method, detecting specifically the presence of EU unauthorized GMO (∼30%), was thus developed (Figure 10.1B). This qPCR method has been presented at the EU level and selected by the European Network of GMO laboratories (ENGL) members to be added in the Compendium of reference methods for GMO analysis (Joint Research Centre 2011), becoming thus a reference method for the EU official enforcement laboratories. This method is currently being validated at the EU level, via at least twelve independent laboratories, to verify its performance. In addition, this qPCR method is already implemented at the national level by the NRL-GMO from the Scientific Institute of Public Health in Belgium (WIV-ISP), in agreement with the competent authorities (Federal Agency for the Safety of the Food Chain (FASFC)). Furthermore, the integrated DNA walking strategy was developed to confirm the presence EU unauthorized GMO, only suspected or even concealed with the current qPCR analysis. This DNA walking strategy was up to now developed to anchor on the t35S pCAMBIA sequence, frequently observed in EU unauthorized GMO, as well as the p35S and tNOS sequences, commonly found in EU authorized and EU unauthorized GMO (Figure 10.1B). By this way, all GMO possessing these elements could be characterized (Fraiture *et al.* 2014; Fraiture *et al.* 2015a; Fraiture *et al.* 2015b). At the national level, the DNA walking method anchored on the t35S pCAMBIA element is already implemented by the NRL-GMO from the Scientific Institute of Public Health in Belgium (WIV-ISP) and may be used in agreement with the competent authorities (FASFC) to prove the presence of EU unauthorized GMO if a positive signal is observed for the t35S pCAMBIA marker during the qPCR screening analysis. All these improvements offered by this PhD thesis regarding the current tools and strategies used by the enforcement laboratories are clearly illustrated in Figure 10.1.

In addition, the DNA walking strategy, integrated into the qPCR analysis, as well as the Whole-Genome-Sequencing approach were presented at the EU level to respectively the 19th (19-20/06/2013) and 22th (1-2/12/2014) ENGL plenary meeting, arousing a great interest. The peer-reviewed publications related to these studies are also available on the private area of the ENGL website, that is accessible to the official enforcement laboratories. In addition, projects proposing to deeply investigate the use of NGS (targeted and untargeted approaches) were and will be submitted for funding by FPS.

In a near future, the improved DNA walking strategy could be applied in order to reach an ideal situation applicable at the present state (Figure 10.1C). More precisely, the potential detection of GMO via the qPCR screening could be automatically followed by a DNA walking analysis, regardless of the relationship

between the observed screening signals and the EU authorized GMO. Additional DNA walking methods targeting newly identified key transgenic elements could also be over time developed to strengthen this DNA walking strategy. Moreover, new event-specific qPCR methods could also be developed based on the sequences from EU unauthorized GMO characterized using the DNA walking strategy or sequencing approaches.

With the view to provide an ideal GMO detection system, the WGS technology represents a promising alternative in the distant future (Figure 10.1D). Due to its potential high-throughput, this technology will allow to analyse a large amount of samples, earlier individually barcoded, in one sequencing run in order to decrease the cost of the analysis. Nonetheless, several challenges should be circumvented to make possible the successful implementation of this WGS strategy in GMO routine analysis by the enforcement laboratories. First, the WGS approach is, at the present time, not able to reasonably deal with the detection and identification of GMO present at trace level, such as in GMO mixtures frequently encountered in GMO routine analysis (Willems *et al.* 2016). Second, even if this technology could theoretically be performed without any prior knowledge, the availability of good reference genomes is primordial. Otherwise, a strategy of de novo assembly could be envisaged, taking into account that its application is difficult with large and complex plant genomes as well as with mixtures composed of several different GMO. Another option consists to generate a substitute of reference genome, through the long reads from PacBio technology, where short reads from Illumina technology will be aligned. Last, the implementation of the WGS strategy needs the development of a simple, rapid and standard protocol for the library preparation step and a user-friendly workflow, requiring a minimum of bioinformatics background, for the analysis of the generated data. Therefore, several technological progress are still needed for the implementation in GMO routine analysis of this ideal strategy. In collecting several samples at some points, instead of only a few samples over time, as currently done by the competent authorities, this ideal strategy could thus allow to analyse several samples in only one experiment to perform a widespread monitoring of GMO, authorized or not, that are present on the market.

Even if this PhD thesis was essentially focused on GM crops falling within the scope of the EU GMO legislation, the scientific community, including the developers and the enforcement laboratories, is also becoming more and more interested in the detection and characterization of biotech organisms derived from Genome Editing (GE) techniques, for which the status regarding the EU GMO legislation is still undefined (Lusser *et al.* 2011; Lusser *et al.* 2012). In this context, the feasibility of the suggested targeted (DNA walking strategy) and untargeted (WGS) approaches is therefore also shortly discussed.

According to the GE techniques used, two main kinds of modifications are generated in the targeted organisms. In the first category, a large modification could be introduced in the organisms, such as in an herbicide tolerant maize where a sequence of interest (PAT gene) was integrated at a specific maize genome location using the ZFN-3 technique. In that case, the targeted approach is applicable only if a minimum of knowledge is available. Otherwise, the use of the untargeted approach is needed to compare the genome from the biotech organism to the reference genome. In the second category, a small modification could be produced using, for instance, the CRISPR technique. To confirm the presence of the generated biotech organisms, only the untargeted approach could theoretically be envisaged. However, given that the modifications induced by this category of techniques are highly similar to conventional breeding or natural processes, the detection and characterization of the resulting biotech organisms remains clearly challenging or even impossible (Shukla *et al.* 2009; Lusser *et al.* 2011; Lusser *et al.* 2012; Wolt *et al.* 2016).

Nowadays, no decision has still been taken at the EU level regarding the status of the organisms derived from GE techniques. However, in contrast to the biotech organisms from the first category, the ones belonging to the second category will probably not be considered by the EU legislation as GMO (Devos *et al.* 2014; Hartung and Schiemann 2014; Andersen *et al.* 2015; Kanchiswamy *et al.* 2015). Due to the expected high impact on the developers, for which the production of non-GMO products is less expensive and has a better public acceptance compared to GMO products, as well as on the enforcement laboratories, for which the detection methods have potentially to be adapted to target also the GE organisms, this hot topic is also discussed outside the EU and some decisions have besides already been taken. For instance, a common white button mushroom (*Agaricus bisporus*) modified with the CRISPR technique to resist to browning was considered as non-GMO product by the US competent authority, allowing consequently the cultivation and commercialisation in the USA without any further specific oversights (Lusser *et al.* 2011; Waltz 2016; Wolt *et al.* 2016).

The increasing use of the GE techniques to produce improved organisms concerns both plant and animal species. In this latter case, the creation of biotech animals has considerably be simplified and accelerated with the GE techniques that are cheaper, faster, easier, more precise, more successful and more reliable than the conventional breeding techniques. Although the first GM animal (AquAdvantage salmon presenting a faster growth than its non-GM counterpart) approved for human consumption in November 2015 in the USA is considered as a GMO due to its classical transgenic cassette integrated into the host genome, the next coming biotech animals in the pipeline for the food chain are expected to mainly be produced via the GE techniques, such as the hornless cows and the pigs resistant to the porcine reproductive and respiratory syndrome virus

(Fahrenkrug *et al.* 2010; Butler *et al.* 2015; Laible *et al.* 2015; Lievens *et al.* 2015; Seruggia and Montoliu 2015; U.S. Food and Drug Administration (FDA) 2015; Hatada and Horii 2016; Strauss and Sax 2016; Whitworth *et al.* 2016).

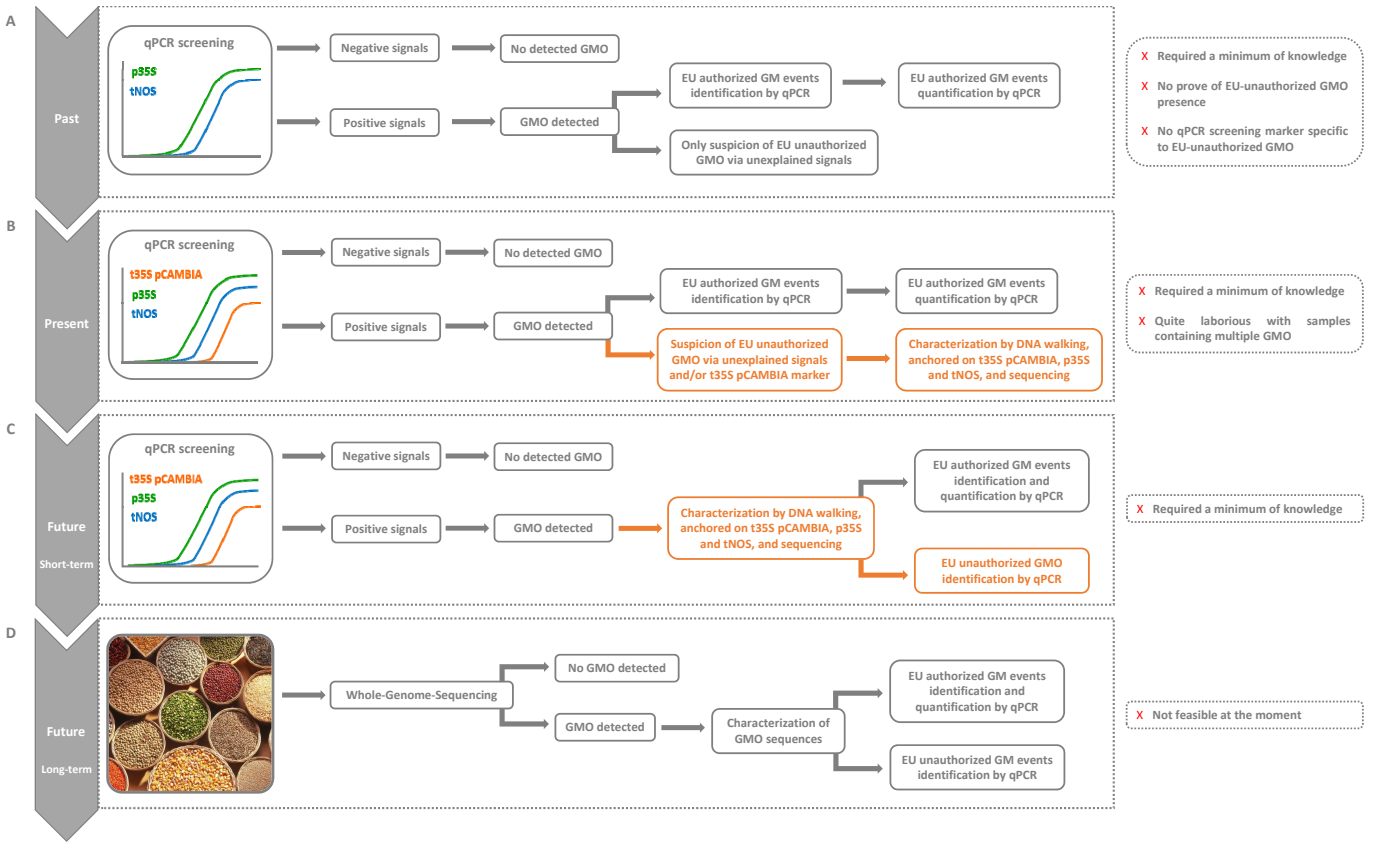


Figure 10.1: Schematic representation of the GMO detection system currently used by the enforcement laboratories

Figure 10.1 *(continued)*: (past; A), improved by the strategies developed in this PhD thesis (present; B), in the ideal situation feasable at the present state (future: short term; C) or in the ideal situation unenforceable at the present state (future: long-term; D). The direct outcome of this PhD thesis is indicated in orange. The drawbacks of each approah are indicated on the left side by X.

Appendix A

Supporting information for chapter 7

A.1 Additional tables

Table A.1: Oligonucleotide sequences used for qPCR and DNA walking assays. PLD: Phopholipase D gene from rice; ADH: Alcohol dehydrogenase I gene from maize; LEC: Lectin gene of soybean; CRU: Cruciferin gene from oilseed rape; pFMV: Promoter of the figworth mosaic virus; Cry3Bb: Gene encoding the *Bacillus thuringiensis* δ-endotoxin 3Bb; PAT: Phosphinotricin-N-acetyltransferases gene from *Streptomyces viridochromogenes*; BAR: Phosphinotricin-N-acetyltransferases gene from *Streptomyces hygroscopicus*; CP4/EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* strain CP4.

Methods	Oligonucleotide names	Oligonucleotide sequences	References
Screening qPCR	PLD F	GCTTAGGGAACAGGGAAGTAAAGTT	Mbongolo Mbella *et al.* (2011)
Screening qPCR	PLD F	GCTTAGGGAACAGGGAAGTAAAGTT	Mbongolo Mbella *et al.* (2011)
	PLD R	CTTAGCATAGTCTGTGCCATCCA	Mbongolo Mbella *et al.* (2011)
Screening qPCR	ADH F	TCTCTTCCTCCTTTAGAGCTACCACTA	Mbongolo Mbella *et al.* (2011)
	ADH R	AATCGATCCAAAGCGAGATGA	Mbongolo Mbella *et al.* (2011)
Screening qPCR	LEC F	AACCGGTAGCGTTGCCAG	Vaïtilingom *et al.* (1999)
	LEC R	AGCCCATCTGCAAGCCTTT	Vaïtilingom *et al.* (1999)
Screening qPCR	CRU F	CAGCTCAACAGTTTCCAAACGA	Mbongolo Mbella *et al.* (2011)
	CRU R	CGACCAGCCTCAGCCTTAAG	Mbongolo Mbella *et al.* (2011)
Screening qPCR	pFMV F	CGAAGACTTAAAGTTAGTGGGCATCT	Broeders *et al.* (2013)
	pFMV R	TTTTGTCTGGTCCCCACAA	Broeders *et al.* (2013)
Screening qPCR	cry3Bb F	CTACCAGTCCTTCCTGAACACC	Broeders *et al.* (2015)
	cry3Bb R	GCGTACTCCTCGATCTTCTTGT	Broeders *et al.* (2015)

Table A.1 *continued*

Methods	Oligonucleotide names	Oligonucleotide sequences	References
Screening qPCR	PAT F	CCGCGGTTTGTGATATCGTT	Barbau-Piednoir *et al.* (2012)
	PAT R	TCTTGCAACCTCTCTAGATCATCAA	Barbau-Piednoir *et al.* (2012)
Screening qPCR	BAR F	CGTCAACCACTACATCGAGACAA	Barbau-Piednoir *et al.* (2012)
	BAR R	GTCCACTCCTGCGGTTCCT	Barbau-Piednoir *et al.* (2012)
Screening qPCR	CP4/EPSPS F	GCATGCTTCACGGTGCAA	Barbau-Piednoir *et al.* (2012)
	CP4/EPSPS R1	TGAAGGACCGGTGGGAGAT	Barbau-Piednoir *et al.* (2012)
	CP4/EPSPS R2	TGAAGGACCTGTGGGAGAT	Barbau-Piednoir *et al.* (2012)
Screening qPCR	p35S F	AAAGCAAGTGGATTGATGTGATA	Barbau-Piednoir *et al.* (2010)
	p35S R	GGGTCTTGCGAAGGATAGTG	Barbau-Piednoir *et al.* (2010)
Screening qPCR	tNOS F	GATTAGAGTCCCGCAATTATACATTTAA	Barbau-Piednoir *et al.* (2010)
	tNOS R	TTATCCTAGKTTGCGCGCTATATTT	Barbau-Piednoir *et al.* (2010)
Screening qPCR	t35S pCAMBIA c-F	CGGGGGATCTGGATTTTAGTA	Fraiture *et al.* (2014)
	t35S pCAMBIA a-R	AGGGTTCCTATAGGGTTTCGCTC	Fraiture *et al.* (2014)
Identification qPCR	3272 maize F	TCATCAGACCAGATTCTCTTTTATGG	Joint Research Centre (2015a)
	3272 maize R	CGTTTCCCGCCTTCAGTTTA	Joint Research Centre (2015a)

Table A.1 *continued*

Methods	Oligonucleotide names	Oligonucleotide sequences	References
	3272 maize P	ACTGCTGACGCGGCCAAACACTG	Joint Research Centre (2015a)
Identification qPCR	98140 maize F	GTGTGTATGTCTCTTTGCTTGGTCTT	Joint Research Centre (2015a)
	98140 maize R	GATTGTCGTTTCCCGCCTTC	Joint Research Centre (2015a)
	98140 maize P	CTCTATCGATCCCCCTCTTTGATAGTTT AAACT	Joint Research Centre (2015a)
Identification qPCR	Bt11 maize F	GCGGAACCCCTATTTGTTTA	Joint Research Centre (2015a)
	Bt11 maize R	TCCAAGAATCCCTCCATGAG	Joint Research Centre (2015a)
	Bt11 maize P	AAATACATTCAAATATGTATCCGCTCA	Joint Research Centre (2015a)
Identification qPCR	DAS 40278-9 maize F	CACGAACCATTGAGTTACAATC	Joint Research Centre (2015a)
	DAS 40278-9 maize R	GAAATTGCGTGACTCAAATTCC	Joint Research Centre (2015a)
	DAS 40278-9 maize P	CCTGCAGGTCGACGGCCGAGTAC	Joint Research Centre (2015a)
Identification qPCR	DAS59122 maize F	GGGATAAGCAAGTAAAAGCGCTC	Joint Research Centre (2015a)
	DAS59122 maize R	CCTTAATTCTCCGCTCATGATCAG	Joint Research Centre (2015a)
	DAS59122 maize P	TTTAAACTGAAGGCGGGAAACGACAA	Joint Research Centre (2015a)
Identification qPCR	GA21 maize F	CTTATCGTTATGCTATTTGCAACTTTAGA	Joint Research Centre (2015a)

Table A.1 *continued*

Methods	Oligonucleotide names	Oligonucleotide sequences	References
	GA21 maize R	TGGCTCGCGATCCTCCT	Joint Research Centre (2015a)
	GA21 maize P	CATATACTAACTCATATCTCTTTCTC AACAGCAGGTGGGT	Joint Research Centre (2015a)
Identification qPCR	MIR 604 maize F	GCGCACGCAATTCAACAG	Joint Research Centre (2015a)
	MIR 604 maize R	GGTCATAACGTGACTCCCTTAATTCT	Joint Research Centre (2015a)
	MIR 604 maize P	AGGCGGGAAACGACAATCTGATCATG	Joint Research Centre (2015a)
Identification qPCR	MIR162 maize F	GCGCGGTGTCATCTATGTTACTAG	Joint Research Centre (2015a)
	MIR162 maize R	TGCCTTATCTGTTGCCTTCAGA	Joint Research Centre (2015a)
	MIR162 maize P	TCTAGACAATTCAGTACATTAAAAACGT CCGCCA	Joint Research Centre (2015a)
Identification qPCR	MON 810 maize F	TCGAAGGACGAAGGACTCTAACGT	Joint Research Centre (2015a)
	MON 810 maize R	GCCACCTTCCTTTTCCACTATCTT	Joint Research Centre (2015a)
	MON 810 maize P	AACATCCTTTGCCATTGCCCAGC	Joint Research Centre (2015a)
Identification qPCR	MON 87427 maize F	CACGTTGAAGGAAAATGGATTG	Joint Research Centre (2015a)
	MON 87427 maize R	TCGCGATCCTCCTCAAAGAC	Joint Research Centre (2015a)
	MON 87427 maize P	AGGGAGTATGTAGATAAATTTTCAAAGC GTTAGACGGC	Joint Research Centre (2015a)

Table A.1 *continued*

Methods	Oligonucleotide names	Oligonucleotide sequences	References
Identification qPCR	MON 87460 maize F	ACGGAAACGGTCGGGTCA	Joint Research Centre (2015a)
	MON 87460 maize R	CCATGTAGATTTCCCGGTTTT	Joint Research Centre (2015a)
	MON 87460 maize P	CGGGACAATATGGAGAAAAAGAAAGAG	Joint Research Centre (2015a)
Identification qPCR	NK603 maize F	ATGAATGACCTCGAGTAAGCTTGTTAA	Joint Research Centre (2015a)
	NK603 maize R	AAGAGATAACAGGATCCACTCAAACACT	Joint Research Centre (2015a)
	NK603 maize P	TGGTACCACGCGACACACTTCCACTC	Joint Research Centre (2015a)
Identification qPCR	T25 maize F	ACAAGCGTGTCGTGCTCCAC	Joint Research Centre (2015a)
	T25 maize R	GACATGATACTCCTTCCACCG	Joint Research Centre (2015a)
	T25 maize P	TCATTGAGTCGTTCCGCCATTGTCG	Joint Research Centre (2015a)
Identification qPCR	TC1507 maize F	TAGTCTTCGGCCAGAATGG	Joint Research Centre (2015a)
	TC1507 maize R	CTTTGCCAAGATCAAGCG	Joint Research Centre (2015a)
	TC1507 maize P	TAACTCAAGGCCCTCACTCCG	Joint Research Centre (2015a)
DNA Walking	p35S-F a (p35S R)	GGGTCTTGCGAAGGATAGTG	Barbau-Piednoir *et al.* (2010)
	p35S-F b	TGTGCGTCATCCCTTACGTCAGT	Fraiture *et al.* (2015b)

Table A.1 *continued*

Methods	Oligonucleotide names	Oligonucleotide sequences	References
	p35S-F c	TATCACATCAATCCACTTGCTTT	Barbau-Piednoir *et al.* (2010)
DNA Walking	p35S-R a (p35S F)	AAAGCAAGTGGATTGATGTGATA	Barbau-Piednoir *et al.* (2010)
	p35S-R b	ACTGACGTAAGGGATGACGCACA	Fraiture *et al.* (2015b)
	p35S-R c	CACTATCCTTCGCAAGACCC	Barbau-Piednoir *et al.* (2010)
DNA Walking	tNOS-F a (tNOS F)	GATTAGAGTCCCGCAATTATACATTTAA	Barbau-Piednoir *et al.* (2010)
	tNOS-F b	TTAATACGCGATAGAAAACAAAAT	Fraiture *et al.* (2015b)
	tNOS-F c	AAATATAGCGCGCAAMCTAGGATAA	Barbau-Piednoir *et al.* (2010)
DNA Walking	tNOS-R a (tNOS R)	TTATCCTAGKTTGCGCGCTATATTT	Barbau-Piednoir *et al.* (2010)
	tNOS-R b	ATTTTGTTTTCTATCGCGTATTAA	Fraiture *et al.* (2015b)
	tNOS-R c	TTAAATGTATAATTGCGGGACTCTAATC	Barbau-Piednoir *et al.* (2010)
DNA Walking	t35S pCAMBIA c-R	TACTAAAATCCAGATCCCCCG	Fraiture *et al.* (2014)
	t35S pCAMBIA b-R	GTGTTGAGCATATAAGAAACCC	Fraiture *et al.* (2014)
	t35S pCAMBIA a-R	AGGGTTCCTATAGGGTTTCGCTC	Fraiture *et al.* (2014)

Table A.2: Results from qPCR analysis using PLD, ADH, LEC, CRU, Sad1, p35S, tNOS and t35S pCAMBIA markers applied on the different samples. The positive and the negative expected signals are indicated by + and −, respectively.

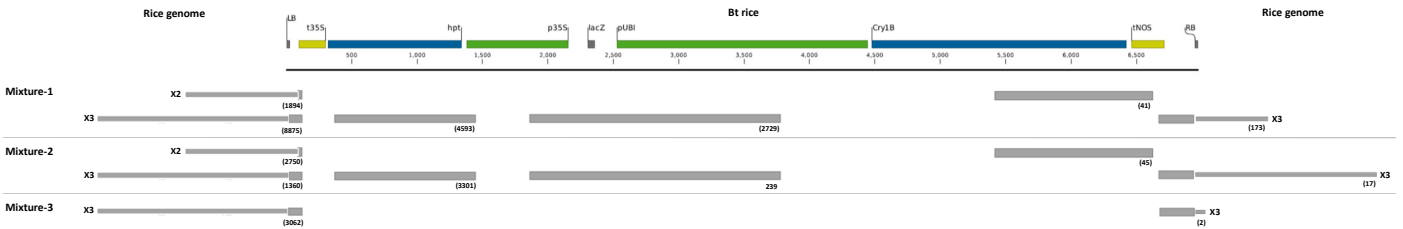
	PLD		ADH		LEC		p35S		tNOS		t35S pCAMBIA	
	Expected	C_t observed	Expected	C_t observed	Expected	C_t observed	Expected	C_t observed	Expected	C_t observed	Expected	C_t observed
Bt rice 100%	+	17.2	−	40	−	40	+	18.8	+	18.6	+	21.8
Bt rice 1%	+	17.3	−	40	−	40	+	26.8	+	26.9	+	27.6
Bt rice 0.1%	+	17.4	−	40	−	40	+	30.1	+	30.9	+	31.9
Bt rice 0.01%	+	17.3	−	40	−	40	+	33.6	+	33.4	+	36
Bt noodles 100%	+	18	−	40	−	40	+	18.8	+	18.7	+	22.9
Bt noodles 1%	+	18.5	−	40	−	40	+	26.6	+	26.9	+	30.1
Mixture-1	+	25.7	+	20.3	−	40	+	25.4	+	25	+	30.9
Mixture-2	+	25.7	+	20.6	+	18.8	+	22.9	+	23.7	+	30.7
Mixture-3	+	35.4	+	27.8	+	25	+	29.6	+	30.3	+	35.3

Table A.3: Number of sequences, for each tested sample, corresponding to GMO or only to plant genomes. The percentages in parentheses were calculated according to the total amount of generated sequences for each sample.

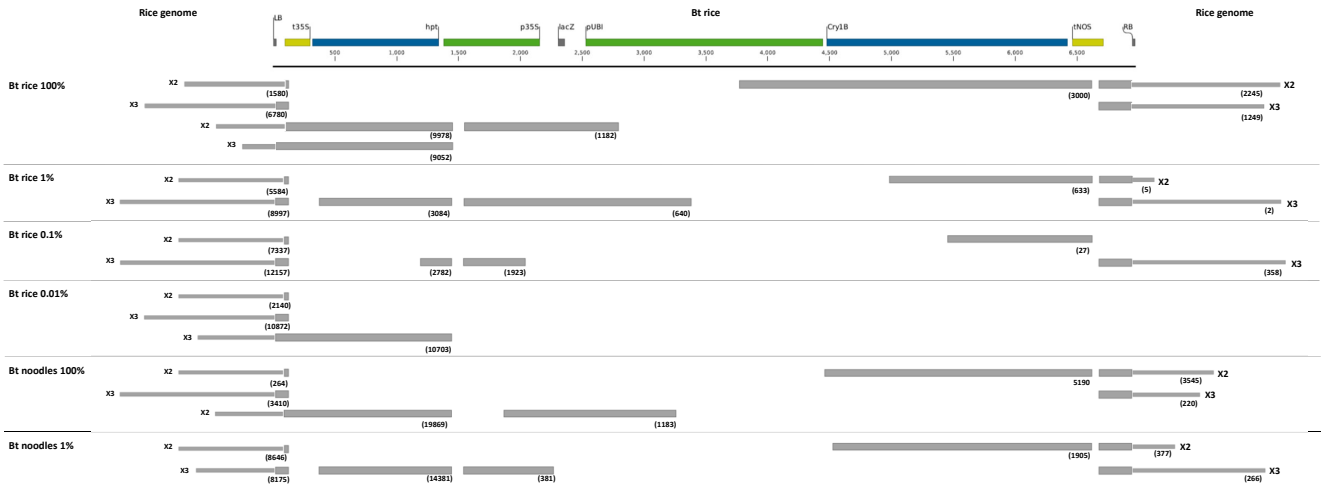
Samples	Number of sequences matching to GMO	Number of sequences matching only to plant genomes
Bt rice 100%	28153	0
	(97.4%)	0%)
Bt rice 1%	18966	5172
	(76.7%)	(20.9%)
Bt rice 0.1%	24514	2128
	(89.5%)	(7.8%)
Bt rice 0.01%	23740	2833
	(86.8%)	(10.4%)
Bt noodles 100%	33772	0
	(98.8%)	0%)
Bt noodles 1%	34309	441
	(92.7%)	(1.2%)
Mixture-1	34582	1259
	(94.4%)	(3.4%)
Mixture-2	16025	2381
	(83.9%)	(12.5%)
Mixture-3	12285	1109
	(87.4%)	(7.9%)
Kuwaiti matrix	30743	2922
	(88.3%)	(8.4%)

A.2 Additional figures

A


Rice genome
Bt rice
Rice genome
LB
t35S
hpt
p35S
lacZ
pUBI
Cry1B
tNOS
RB
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
Mixture-1
X2
(1894)
X3
(8875)
(4593)
(2729)
(41)
X3
(173)
Mixture-2
X2
(2750)
X3
(1360)
(3301)
239
(45)
X3
(17)
Mixture-3
X3
(3062)
X3
(2)

B


Rice genome
Bt rice
Rice genome
LB
t35S
hpt
p35S
lacZ
pUBI
Cry1B
tNOS
RB
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
Bt rice 100%
Bt rice 1%
Bt rice 0.1%
Bt rice 0.01%
Bt noodles 100%
Bt noodles 1%
X2
X3
(1580)
(6780)
(9978)
(1182)
(9052)
(3000)
(2245)
(1249)
(5584)
(8997)
(3084)
(640)
(633)
(5)
(2)
(7337)
(12157)
(2782)
(1923)
(27)
(358)
(2140)
(10872)
(10703)
(264)
(3410)
(19869)
(1183)
5190
(3545)
(220)
(8646)
(8175)
(14381)
(381)
(1905)
(377)
(266)

C

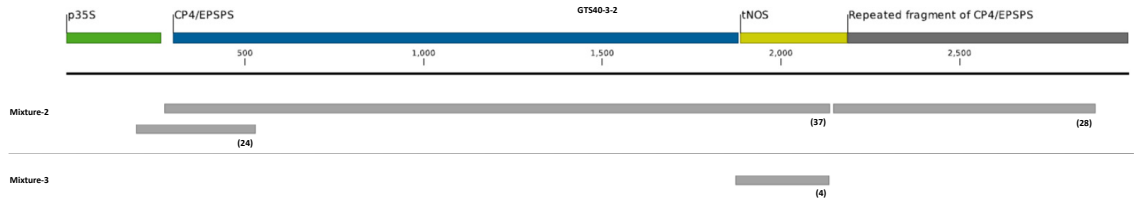


Figure A.1: Characterization of Bt rice (A), MON863 (B) and GTS-40-3-2 (C) events by the DNA walking/NGS strategy for the Mixture-1, Mixture-2 and Mixture-3 samples.
Below the transgenic cassettes, the longest generated amplicons targeting the transgene flanking regions and constructs for each sample are schematically represented by grey rectangles. The number of sequences supporting each rectangle is indicated below the box in parentheses. The region corresponding to chromosome II and III from the rice genome are respectively indicated by X2 and X3. The scale is indicated in bp. LB: left border; t35S: CaMV 35S terminator; hpt: hygromycin phosphotransferase gene; p35S: CaMV 35S promoter; lacZ: LacZ alpha fragment; pUBI: maize ubiquitin promoter; Cry1B: synthetic Cry1B gene; tNOS: *Agrobacterium tumefaciens* nopaline synthase terminator; RB: right border; nptII: neomycin phosphotransferase II gene; p4-AS1: modified CaMV 35S promoter; wtCAB: Wheat major chlorophyll a/b binding protein gene; iAct: Rice Actin intron; Cry3Bb: synthetic Cry3Bb gene; tahsp17: Wheat heat shock protein terminator; CP4/EPSPS: *Agrobacterium tumefaciens* (strain CP4) 5-enolpyruvylshikimate-3-phosphate synthase.

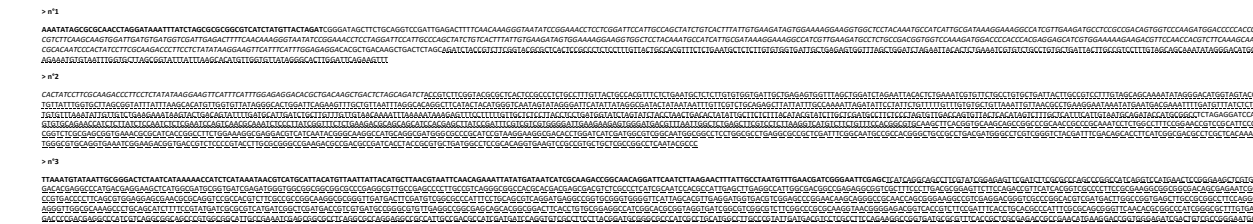


Figure A.2: The longest sequences from the different classes of fragments characterizing unique configurations of elements from the Kuwaiti matrix. CP4/EPSPS (dashed underlined): *Agrobacterium tumefaciens* strain CP4 5-enolpyruvulshikimate-3-phosphate synthase gene; Cry1Ab (in bold and double underlined): synthetic Cry1Ab gene; Cry1F (double underlined): synthetic Cry1F gene; CTP (wave underlined): *Arabidopsis thaliana* chloroplast transit peptide; hsp70 (dotted underlined): maize hsp70 intron; mas1 (in bold and dashed underlined): mannopine synthase region; p35S (in italic): CaMV 35S promoter; pAct (thick underlined): rice Actin promoter; PAT (underlined): phosphinotricin acetyltransferase gene; t35S (in bold and underlined): CaMV 35S terminator; tNOS (in bold): *Agrobacterium tumefaciens* nopaline synthase terminator.

Bibliography

Afroz, A., A. Qureshi, M. Zahur, U. Rashid, and H. Rashid (2012). "Transgenic lines of *Oryza Sativa* (Basmati 2000) expressing Xa21 demonstrate enhanced resistance to bacterial blight". *Plant Molecular Biology Reporter* 30.1, pp. 204–213. DOI: 10.1007/s11105-011-0329-y (p. 90).

AFSCA (2016). *Organismes génétiquement modifiés (OGM)*. URL: http://www.afsca.be/productionvegetale/produitsvegetaux/ogm/ (p. 7).

Ahmad, P., M. Ashraf, M. Younis, X. Hu, A. Kumar, N. A. Akram, and F. Al-Qurainy (2012). "Role of transgenic plants in agriculture and biopharming". *Biotechnology advances* 30.3, pp. 524–540. DOI: 10.1016/j.biotechadv.2011.09.006 (pp. 80, 109, 121).

Akiyama, H., N. Sasaki, K. Sakata, K. Ohmori, A. Toyota, Y. Kikuchi, T. Watanabe, S. Furui, K. Kitta, and T. Maitani (2007). "Indicated detection of two unapproved transgenic rice lines contaminating vermicelli products". *Journal of Agricultural and Food Chemistry* 55.15, pp. 5942–5947. DOI: 10.1021/jf070508m (pp. 96, 102).

Akritidis, P., K. Pasentsis, A. S. Tsaftaris, P. V. Mylona, and A. N. Polidoros (2008). "Identification of unknown genetically modified material admixed in conventional cotton seed and development of an event-specific detection method". *Electronic Journal of Biotechnology* 11.2, pp. 76–83. DOI: 10.2225/vol11-issue2-fulltext-11 (pp. 37, 40, 42, 111).

Allen, R. C., S. Rogelj, S. E. Cordova, and T. L. Kieft (2006). "An immuno-PCR method for detecting *Bacillus thuringiensis* Cry1Ac toxin". *Journal of Immunological Methods* 308.1–2, pp. 109–115. DOI: 10.1016/j.jim.2005.10.006 (pp. 9, 10).

An, S., S. Park, D.-H. Jeong, D.-Y. Lee, H.-G. Kang, J.-H. Yu, J. Hur, S.-R. Kim, Y.-H. Kim, M. Lee, S. Han, S.-J. Kim, J. Yang, E. Kim, S. J. Wi, H. S. Chung, J.-P. Hong, V. Choe, H.-K. Lee, J.-H. Choi, J. Nam, S.-R. Kim, P.-B. Park, K. Y. Park, W. T. Kim, S. Choe, C.-B. Lee, and G. An (2003). "Generation and analysis of end sequence database for T-DNA tagging lines in rice". *Plant Physiology* 133.4, pp. 2040–1047. DOI: 10.1104/pp.103.030478 (p. 37).

Anandhakumar, C., S. Kizaki, T. Bando, G. N. Pandian, and H. Sugiyama (2015). "Advancing small-molecule-based chemical biology with next-generation sequencing technologies". *ChemBioChem* 16.1, pp. 20–38. DOI: 10.1002/cbic.201402556 (pp. 69, 70).

Andersen, M. M., X. Landes, W. Xiang, A. Anyshchenko, J. Falhof, J. T. Østerberg, L. I. Olsen, A. K. Edenbrandt, S. E. Vedel, B. J. Thorsen, P. Sandøe, C. Gamborg, K. Kappel, and M. G. Palmgren (2015). "Feasibility of new breeding techniques for organic farming". *Trends in Plant Science* 20.7, pp. 426–434. DOI: 10.1016/j.tplants.2015.04.011 (pp. 4, 222).

Angers-Loustau, A., M. Petrillo, L. Bonfini, F. Gatto, S. Rosa, A. Patak, and J. Kreysa (2014). "JRC GMO-Matrix: a web application to support genetically modified organisms detection strategies". *BMC Bioinformatics* 15, 417. DOI: 10.1186/s12859-014-0417-8 (pp. 10, 11, 19, 36, 51, 52).

Anklam, E., F. Gadani, P. Heinze, H. Pijnenburg, and G. Van Den Eede (2002). "Analytical methods for detection and determination of genetically modified organisms in agricultural crops and plant-derived food products". *European Food Research and Technology* 214.1, pp. 3–26. DOI: 10.1007/s002170100415 (p. 9).

AOCS, Urbana, USA (2015). *Certified Reference Materials (CRM)*. The American Oil Chemists' Society. URL: http://www.aocs.org/LabServices/content.cfm?ItemNumber=19248 (p. 112).

Araki, M., K. Nojima, and T. Ishii (2014). "Caution required for handling genome editing technology". *Trends in Biotechnology* 32.5, pp. 234–237. DOI: 10.1016/j.tibtech.2014.03.005 (p. 94).

Arumuganathan, K. and E. Earle (1991). "Nuclear DNA content of some important plant species". *Plant Molecular Biology Reporter* 9.3, pp. 208–218. DOI: 10.1007/BF02672069 (pp. 122, 136).

Arun, Ö. Ö., F. Yılmaz, and K. Muratoğlu (2013). "PCR detection of genetically modified maize and soy in mildly and highly processed foods". *PCR detection of genetically modified maize and soy in mildly and highly processed foods* 32.2, pp. 525–531. DOI: 10.1016/j.foodcont.2013.01.023 (p. 132).

Au, K. F., J. G. Underwood, L. Lee, and W. H. Wong (2012). "Improving PacBio long read accuracy by short read alignment". *PLoS ONE* 7.10, e46679. DOI: 10.1371/journal.pone.0046679 (pp. 48, 65, 187).

Ayadi, M., D. Mieulet, D. Fabre, J.-L. Verdeil, A. Vernet, E. Guiderdoni, and K. Masmoudi (2014). "Functional analysis of the durum wheat gene *Td*PIP2;1 and its promoter region in response to abiotic stress in rice". *Plant Physiology and Biochemistry* 79, pp. 98–108. DOI: 10.1016/j.plaphy.2014.02.018 (p. 90).

Azegami, T., H. Itoh, H. Kiyono, and Y. Yuki (2015). "Novel transgenic rice-based vaccines". *Archivum Immunologiae et Therapiae Experimentalis* 63.2, pp. 87–99. DOI: 10.1007/s00005-014-0303-0 (p. 88).

Babekova, R., T. Funk, S. Pecoraro, K.-H. Engel, and U. Busch (2009). "Development of an event-specific real-time PCR detection method for the transgenic Bt rice line KMD1". *European Food Research and Technology* 228.5, pp. 707–716. DOI: 10.1007/s00217-008-0981-0 (pp. 37, 44, 93, 97, 103, 150).

Baek, S.-H., H.-J. Chung, H.-K. Lee, R. D'Souza, Y. Jeon, H.-J. Kim, S.-J. Kweon, and S.-T. Hong (2014). "Treatment of obesity with the resveratrol-enriched rice DJ-526". *Scientific Reports* 4, article number 3879. DOI: 10.1038/srep03879 (p. 91).

Bahrdt, C., A. B. Krech, A. Wurz, and D. Wulff (2010). "Validation of a newly developed hexaplex real-time PCR assay for screening for presence of GMOs in food, feed and seed". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2103–2112. DOI: 10.1007/s00216-009-3380-x (pp. 19, 20, 24, 53, 54, 191).

Ballari, R. V. and A. Martin (2013). "Assessment of DNA degradation induced by thermal and UV radiation processing: implications for quantification of genetically modified organisms". *Food Chemistry* 141.3, pp. 2130–2136. DOI: 10.1016/j.foodchem.2013.05.032 (pp. 132, 141).

Balzergue, S., B. Dubreucq, S. Chauvin, I. Le-Clainche, F. Le Boulaire, R. de Rose, V. Samson F. Biaudet, A. Lecharny, C. Cruaud, J. Weissenbach, M. Caboche, and L. Lepiniec (2001). "Improved PCR-walking for large-scale isolation of plant T-DNA borders". *BioTechniques* 30.3, pp. 496–498 (p. 37).

Barbau-Piednoir, E., A. Lievens, G. Mbongolo-Mbella, N. Roosens, M. Sneyers, A. Leunda-Casi, and M. Van den Bulcke (2010). "SYBR®Green qPCR screening methods for the presence of "35S promoter" and "NOS terminator" elements in food and feed products". *European Food Research and Technology* 230.3, pp. 383–393. DOI: 10.1007/s00217-009-1170-5 (pp. 11, 13, 60, 61, 110, 114, 116, 118, 121, 134, 150–152, 154, 155, 157, 161, 163, 164, 174, 229, 232, 233).

Barbau-Piednoir, E., A. Lievens, E. Vandermassen, E.-G. Mbongolo-Mbella, A. Leunda-Casi, N. Roosens, M. Sneyers, and M. Van den Bulcke (2012). "Four new SYBR®Green qPCR screening methods for the detection of Roundup Ready®, LibertyLink®, and CryIAb traits in genetically modified products". *European Food Research and Technology* 234.1, pp. 13–23. DOI: 10.1007/s00217-011-1605-7 (pp. 121, 174, 229).

Barbau-Piednoir, E., N. Botteldoorn, M. Yde, J. Mahillon, and N. H. Roosens (2013). "Development and validation of qualitative SYBR®Green real-time PCR for detection and discrimination of *Listeria* spp. and *Listeria monocytogenes*". *Applied Microbiology and Biotechnology* 97.9, pp. 4021–4037. DOI: 10.1007/s00253-012-4477-2 (pp. 110, 114).

Barbau-Piednoir, E., P. Stragier, N. Roosens, M. Mazzara, C. Savini, G. Van den Eede, and M. Van den Bulcke (2014). "Inter-laboratory testing of GMO

detection by Combinatory SYBR®Green PCR Screening (CoSYPS)". *Food Analytical Methods* 7.8, pp. 1719–1728. DOI: 10.1007/s12161-014-9837-3 (pp. 13, 14, 60, 61).

Bartlett, J. G., M. A. Smedley, and W. A. Harwood (2014). "Analysis of T-DNA/host-plant DNA junction sequences in single-copy transgenic barley lines". *Biology* 3.1, pp. 39–55. DOI: 10.3390/biology3010039 (p. 37).

Basak, S., N. Z. Ehtesham, B. Sesikeran, and S. Ghosh (2014). "Detection and identification of transgenic elements by fluorescent-PCR-based capillary gel electrophoresis in genetically modified cotton and soybean". *Journal of AOAC International* 97.1, pp. 159–165. DOI: 10.5740/jaoacint.13-027 (pp. 20, 27, 62).

Bendich, A. J. and L. P. Gauriloff (1984). "Morphometric analysis of cucurbit mitochondria: the relationship between chondriome volume and DNA content". *Protoplasma* 119.1, pp. 1–7. DOI: 10.1007/BF01287811 (p. 205).

Ben Saad, R., D. Fabre, D. Mieulet, D. Meynard, M. Dingkuhn, A. Al-Doss, E. Guiderdoni, and A. Hassairi (2012). "Expression of the *Aeluropus littoralis AlSAP* gene in rice confers broad tolerance to abiotic stresses through maintenance of photosynthesis". *Plant, Cell & Environment* 35.3, pp. 626–643. DOI: 10.1111/j.1365-3040.2011.02441.x (p. 90).

Bhat, S., J. Herrmann, P. Armishaw, P. Corbisier, and K. R. Emslie (2009). "Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number". *Analytical and Bioanalytical Chemistry* 394.2, pp. 457–467. DOI: 10.1007/s00216-009-2729-5 (pp. 33, 34, 63).

Bhutani, S., R. Kumar, R. Chauhan, R. Singh, V. K. Chowdhury, J. B. Chowdhury, and R. K. Jain (2006). "Development of transgenic indica rice plants containing potato proteinase inhibitor 2 (Pin2) gene with improved defense against yellow stem borer". *Physiology and Molecular Biology of Plants* 12.1, pp. 43–52 (p. 89).

Block, A., F. Debode, L. Grohmann, J. Hulin, I. Taverniers, L. Kluga, E. Barbau-Piednoir, S. Broeders, I. Huber, M. Van den Bulcke, P. Heinze, G. Berben, U. Busch, N. Roosens, E. Janssen, J. Žel, K. Gruden, and D. Morisset (2013). "The GMOseek matrix: a decision support tool for optimizing the detection of genetically modified plants". *BMC Bioinformatics* 14, 256. DOI: 10.1186/1471-2105-14-256 (pp. 19, 52).

Breitler, J. C., J. M. Vassal, M. Del Mar Catala, D. Meynard, V. Marfà, E. Melé, M. Royer, I. Murillo, B. San Segundo, E. Guiderdoni, and J. Messeguer (2004). "*Bt* rice harbouring *cry* genes controlled by a constitutive or wound-inducible promoter: protection and transgene expression under Mediterranean field conditions". *Plant Biotechnology Journal* 2.5, pp. 417–430. DOI: 10.1111/j.1467-7652.2004.00086.x (pp. 89, 111, 133, 152, 157, 163, 172, 200).

Breyer, D., L. Kopertekh, and D. Reheul (2014). "Alternatives to antibiotic resistance marker genes for *in vitro* selection of genetically modified plants

– scientific developments, current use, operational access and biosafety considerations". *Critical Reviews in Plant Sciences* 33.4, pp. 286–330. DOI: 10.1080/07352689.2013.870422 (p. 85).

Brod, F. C. A., J. P. van Dijk, M. M. Voorhuijzen, A. Z. Dinon, L. H. S. Guimarães, I. M. J. Scholtens, A. C. M. Arisi, and E. J. Kok (2014). "A high-throughput method for GMO multi-detection using a microfluidic dynamic array". *Analytical and Bioanalytical Chemistry* 406.5, pp. 1397–1410. DOI: 10.1007/s00216-013-7562-1 (pp. 33, 34, 63).

Brodmann, P. D., E. C. Ilg, H. Berthoud, and A. Herrmann (2002). "Real-time quantitative polymerase chain reaction methods for four genetically modified maize varieties and maize DNA content in food". *Journal of AOAC International* 85.3, pp. 646–653 (p. 191).

Broeders, S. R. M., S. C. J. De Keersmaecker, and N. H. C. Roosens (2012a). "How to deal with the upcoming challenges in GMO detection in food and feed". *Journal of Biomedicine and Biotechnology*, article ID 402418. DOI: 10.1155/2012/402418 (pp. 4, 7, 11, 12, 52, 81, 110, 116, 121, 122, 131, 132, 135, 150, 171, 191).

Broeders, S., N. Papazova, M. Van den Bulcke, and N. Roosens (2012b). "Development of a Molecular Platform for GMO Detection in Food and Feed on the Basis of "Combinatory qPCR" Technology". *Polymerase Chain Reaction*. Ed. by P. Hernandez-Rodriguez. Rijeka, Croatia: InTech. Chap. 18, pp. 363–404. ISBN: 9789535106128. DOI: 10.5772/37898 (pp. 11, 12, 19, 50–53, 65, 81, 82, 110, 121, 131, 134, 135, 150, 171).

Broeders, S., E. Barbau-Piednoir, E. Vandermassen, F. Debode, M. Mazzara, and N. Roosens (2013). "New SYBR®Green methods targeting promoter sequences used for screening of several GM events pending for authorisation in Europe". *European Food Research and Technology* 236.3, pp. 537–547. DOI: 10.1007/s00217-013-1910-4 (pp. 13, 60, 61, 110, 112, 114, 118, 120, 121, 133, 134, 163, 173, 174, 228).

Broeders, S., M.-A. Fraiture, E. Vandermassen, M. Delvoye, E. Barbau-Piednoir, A. Lievens, and N. Roosens (2015). "New qualitative trait-specific SYBR®Green qPCR methods to expand the panel of GMO screening methods used in the CoSYPS". *European Food Research and Technology* 241.2, pp. 275–287. DOI: 10.1007/s00217-015-2454-6 (pp. 11, 14, 19, 51, 60, 61, 174, 228).

Buermans, H. P. J. and J. T. den Dunnen (2014). "Next generation sequencing technology: advances and applications". *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842.10, pp. 1932–1941. DOI: 10.1016/j.bbadis.2014.06.015 (pp. 40, 47, 49, 64, 65, 69, 70).

Burns, M. J., A. M. Burrell, and C. A. Foy (2010). "The applicability of digital PCR for the assessment of detection limits in GMO analysis". *European*

Food Research and Technology 231.3, pp. 353–362. DOI: 10.1007/s00217-010-1277-8 (pp. 33, 34, 63).

Burrell, A., C. Foy, and M. Burns (2011). "Applicability of three alternative instruments for food authenticity analysis: GMO identification". *Biotechnology Research International*, article ID 838232. DOI: 10.4061/2011/838232 (pp. 26, 27).

Butler, J. R., J. M. Ladowski, G. R. Martens, M. Tector, and A. J. Tector (2015). "Recent advances in genome editing and creation of genetically modified pigs". *International Journal of Surgery* 23.B, pp. 2017–222. DOI: 10.1016/j.ijsu.2015.07.684 (p. 223).

Cambia, Canberra, Australia (2015). *pCAMBIA Legacy Vectors*. URL: http://www.cambia.org/daisy/bioforge_legacy/3724.html (pp. 112, 121).

Cao, Y., G. Wu, Y. Wu, S. Nie, L. Zhang, and C. Lu (2011). "Characterization of the transgenic rice event TT51-1 and construction of a reference plasmid". *Journal of Agricultural and Food Chemistry* 59.16, pp. 8550–8559. DOI: 10.1021/jf201699s (pp. 37, 42, 92, 93, 97, 102, 111).

Cao, J., J. Xu, T. Zhao, D. Cao, X. Huang, P. Zhang, and F. Luan (2014). "Flanking sequence determination and specific PCR identification of transgenic wheat B102-1-2". *Preparative Biochemistry and Biotechnology* 44.3, pp. 257–265. DOI: 10.1080/10826068.2013.812563 (p. 150).

CERA (2012). *GM Crop Database*. Center for Environmental Risk Assessment (CERA), ILSI Research Foundation, Washington D.C. URL: http://cera-gmc.org/GMCropDatabase (pp. 82, 92, 109, 150).

Chaogang, S., W. Jianhua, Z. Guoying, S. Gang, P. Baozhen, L. Juanli, J. Dendi, C. Shenxiang, N. M. Upadhyaya, P. Waterhouse, and G. Zuxun (2003). "Ectopic expression of the spike protein of Rice Ragged Stunt Oryzavirus in transgenic rice plants inhibits transmission of the virus to insects". *Molecular Breeding* 11.4, pp. 295–301. DOI: 10.1023/A:1023418416094 (p. 90).

Chaouachi, M., M. N. Fortabat, A. Geldreich, P. Yot, C. Kerlan, N. Kebdani, C. Audeon, M. Romaniuk, and Y. Bertheau (2008). "An accurate real-time PCR test for the detection and quantification of cauliflower mosaïc virus (CaMV): applicable in GMO screening". *European Food Research and Technology* 227.3, pp. 789–798. DOI: 10.1007/s00217-007-0787-5 (p. 110).

Chaouachi, M., M. S. Zellama, N. Nabi, A. B. Hafsa, and K. Saïd (2014). "Molecular identification of four genetically modified maize (Bt11, Bt176, Mon810 and T25) by duplex quantitative real-time PCR". *Food Analytical Methods* 7.1, pp. 224–233. DOI: 10.1007/s12161-013-9667-8 (pp. 19, 21, 53).

Chen, D.-x., X.-w. Chen, B.-t. Ma, Y.-p. Wang, L.-h. Zhu, and S.-g. Li (2010). "Genetic transformation of rice with *Pi-d2* gene enhances resistance to rice blast fungus *Magnaporthe oryzae*". *Rice Science* 17.1, pp. 19–27. DOI: 10.1016/S1672-6308(08)60100-6 (p. 90).

Chen, L., J. Guo, Q. Wang, G. Kai, and L. Yang (2011a). "Development of the visual loop-mediated isothermal amplification assays for seven genetically modified maize events and their application in practical samples analysis". *Journal of Agricultural and Food Chemistry* 59.11, pp. 5914–5918. DOI: 10.1021/jf200459s (pp. 36, 38, 39, 63).

Chen, M., A. Shelton, and G.-y. Ye (2011b). "Insect-resistant genetically modified rice in China: from research to commercialization". *Annual Review of Entomology* 56, pp. 81–101. DOI: 10.1146/annurev-ento-120709-144810 (pp. 93, 109).

Chen, X., X. Wang, N. Jin, Y. Zhou, S. Huang, Q. Miao, Q. Zhu, and J. Xu (2012a). "Endpoint visual detection of three genetically modified rice events by loop-mediated isothermal amplification". *International Journal of Molecular Sciences* 13.11, pp. 14421–14433. DOI: 10.3390/ijms131114421 (pp. 36, 38, 39, 63, 104).

Chen, Z., Y. Pan, S. Wang, Y. Ding, W. Yanga, and C. Zhu (2012b). "Overexpression of a protein disulfide isomerase-like protein from *Methanothermobacter thermoautotrophicum* enhances mercury tolerance in transgenic rice". *Plant Science* 197, pp. 10–20. DOI: 10.1016/j.plantsci.2012.08.005 (p. 90).

Cheng, Y., M. Zhang, K. Hu, F. Sun, R. Tao, X. Gao, and F. Luan (2014). "Loop-mediated isothermal amplification for the event-specific detection of wheat B73-6-1". *Food Analytical Methods* 7.2, pp. 500–505. DOI: 10.1007/s12161-013-9718-1 (pp. 36, 39, 63).

Choi, S. H., Y. T. Oh, J. Y. Kwon, S. N. Lee, B. D. Han, and K. H. Ryu (2010). "Development of detection system using multiplex PCR and liquid beadarray for stacked genetically modified rice event (LS28×Cry1Ac)". *Journal of the Korean Society for Applied Biological Chemistry* 53.5, pp. 639–646. DOI: 10.3839/jksabc.2010.097 (pp. 31, 32, 62).

Choi, S. H. (2011). "Hexaplex PCR assay and liquid bead array for detection of stacked genetically modified cotton event 281-24-236×3006-210-23". *Analytical and Bioanalytical Chemistry* 401.2, pp. 647–655. DOI: 10.1007/s00216-011-5132-y (pp. 31, 32, 62).

Christou, P., T. L. Ford1, and M. Kofron (1991). "Production of transgenic rice (*Oryza Sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos". *Nature Biotechnology* 9, pp. 957–962. DOI: 10.1038/nbt1091-957 (p. 89).

Chun, Y. J., D. I. Kim, K. W. Park, S.-C. Jeong, S. Park, K. Back, and C.-G. Kim (2013). "Fitness cost and competitive ability of transgenic herbicide-tolerant rice expressing a protoporphyrinogen oxidase gene". *Journal of Ecology and Environment* 36.1, pp. 39–47. DOI: 10.5141/ecoenv.2013.005 (p. 89).

Clarke, W. E., I. A. Parkin, H. A. Gajardo, D. J. Gerhardt, E. Higgins, C. Sidebottom, A. G. Sharpe, R. J. Snowdon, M. L. Federico, and F. L. Iniguez-Luy (2013). "Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L". *PLoS ONE* 8.12, e81992. DOI: 10.1371/journal.pone.0081992 (pp. 47, 64).

CLC bio (2015). *Main page*. URL: http://www.clcbio.com (p. 199).

Coca, M., C. Bortolotti, M. Rufat, G. Peñas, R. Eritja, D. Tharreau, A. M. del Pozo, J. Messeguer, and B. San Segundo (2004). "Transgenic rice plants expressing the antifungal AFP protein from *Aspergillus giganteus* show enhanced resistance to the rice blast fungus *Magnaporthe grisea*". *Plant Molecular Biology* 54.2, pp. 245–259. DOI: 10.1023/B:PLAN.0000028791.34706.80 (p. 90).

Coca, M., G. Peñas, J. Gómez, S. Campo, C. Bortolotti, J. Messeguer, and B. San Segundo (2006). "Enhanced resistance to the rice blast fungus *Magnaporthe grisea* conferred by expression of a *cecropin A* gene in transgenic rice". *Planta* 223.3, pp. 392–406. DOI: 10.1007/s00425-005-0069-z (p. 90).

Cold Spring Harbor Laboratory (2016). *Research: Pacific Biosciences Sequencing*. URL: http://cshl.edu/Research/PacBio.html (p. 175).

Collonnier, C., A. Schattner, G. Berthier, F. Boyer, G. Coué-Philippe, A. Diolez, M.-N. Duplan, S. Fernandez, N. Kebdani, A. Kobilinsky, M. Romaniuk, M. de Beuckeleer, M. de Loose, P. Windels, and Y. Bertheau (2005). "Characterization and event specific-detection by quantitative real-time PCR of T25 maize insert". *Journal of AOAC International* 88.2, pp. 536–546 (pp. 37, 43, 111).

Corbisier, P., S. Trapmann, D. Gancberg, L. Hannes, P. Van Iwaarden, G. Berben, H. Schimmel, and H. Emons (2005). "Quantitative determination of Roundup Ready soybean (*Glycine max*) extracted from highly processed flour". *Analytical and Bioanalytical Chemistry* 383.2, pp. 282–290. DOI: 10.1007/s00216-005-0013-x (p. 121).

Corbisier, P., S. Bhat, L. Partis, V. R. D. Xie, and K. R. Emslie (2010). "Absolute quantification of genetically modified MON810 maize (*Zea mays* L.) by digital polymerase chain reaction". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2143–2150. DOI: 10.1007/s00216-009-3200-3 (pp. 33, 34, 63).

Côté, M.-J., A. J. Meldrum, P. Raymond, and C. Dollard (2005). "Identification of genetically modified potato (*Solanum tuberosum*) cultivars using event specific polymerase chain reaction". *Journal of Agricultural and Food Chemistry* 53.17, pp. 6691–6696. DOI: 10.1021/jf050591i (pp. 37, 111).

Cottage, A., A. Yang, H. Maunders, R. C. de Lacy, and N. A. Ramsay (2001). "Identification of DNA sequences flanking T-DNA insertions by PCR-walking". *Plant Molecular Biology Reporter* 19.4, pp. 321–327. DOI: 10.1007/BF02772830 (p. 37).

Cottenet, G., C. Blancpain, V. Sonnard, and P. F. Chuah (2013). "Development and validation of a multiplex real-time PCR method to simultaneously detect 47 targets for the identification of genetically modified organisms". *Analytical and Bioanalytical Chemistry* 405.21, pp. 6831–6844. DOI: 10.1007/s00216-013-7125-5 (pp. 19, 21–23, 53).

Cullen, D., W. Harwood, M. Smedley, H. Davies, and M. Taylor (2011). "Comparison of DNA walking methods for isolation of transgene-flanking regions in GM potato". *Molecular Biotechnology* 49.1, pp. 19–31. DOI: 10.1007/s12033-010-9371-5 (pp. 37, 111).

Dasgupta, M. G., V. Dharanishanthi, I. Agarwal, and K. V. Krutovsky (2015). "Development of genetic markers in *Eucalyptus* species by target enrichment and exome sequencing". *PLoS ONE* 10.1, e0116528. DOI: 10.1371/journal.pone.0116528 (pp. 47, 64).

Datta, S. K., K. Datta, N. Soltanifar, G. Donn, and I. Potrykus (1992). "Herbicide-resistant Indica rice plants from IRRI breeding line IR72 after PEG-mediated transformation of protoplasts". *Plant Molecular Biology* 20.4, pp. 619–629. DOI: 10.1007/BF00046447.

Datta, S. K. (2004). "Rice biotechnology: a need for developing countries". *AgBioForum* 7.1–2, pp. 31–35 (p. 109).

Davison, J. and Y. Bertheau (2007). "EU regulations on the traceability and detection of GMOs: difficulties in interpretation, implementation and compliance". *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 2, 077. DOI: 10.1079/PAVSNNR20072077 (p. 4).

Davison, J. (2010). "GM plants: science, politics and EC regulations". *Plant Science* 178.2, pp. 94–98. DOI: 10.1016/j.plantsci.2009.12.005 (p. 80).

Decathlon Project (2016). *Decathlon*. URL: http://www.decathlon-project.eu/ (p. 99).

Del Fabbro, C., S. Scalabrin, M. Morgante, and F. M. Giorgi (2013). "An extensive evaluation of read trimming effects on Illumina NGS data analysis". *PLoS ONE* 8.12, e85024. DOI: 10.1371/journal.pone.0085024 (p. 200).

Demeke, T. and G. R. Jenkins (2010). "Influence of DNA extraction methods, PCR inhibitors and quantification methods on real-time PCR assay of biotechnology-derived traits". *Analytical and Bioanalytical Chemistry* 396.6, pp. 1977–1990. DOI: 10.1007/s00216-009-3150-9 (pp. 12, 53).

De Steur, H., J. B. Mogendi, D. Blancquaert, W. Lambert, D. Van Der Straeten, and X. Gellynck (2014). "Genetically modified rice with health benefits as a means to reduce micronutrient malnutrition: global status, consumer preferences and potential health impacts of rice biofortification". *Wheat and Rice in Disease Prevention and Health: Benefits, risks and mechanisms of whole grains in health promotion*. Ed. by R. R. Watson, V. R. Preedy, and S. Zibadi. London, UK: Elsevier Academic Press. Chap. 21, pp. 283–299. ISBN:

9780124017160. DOI: 10.1016/B978-0-12-401716-0.00021-0 (pp. 80, 87, 88, 92, 93).

Devic, M., S. Albert, M. Delseny, and T. J. Roscoe (1997). "Efficient PCR walking on plant genomic DNA". *Plant Physiology and Biochemistry* 35.4, pp. 331–339 (p. 37).

Devos, Y., J. Aguilera, Z. Diveki, A. Gomes, Y. Liu, C. Paoletti, P. du Jardin, L. Herman, J. N. Perry, and E. Waigmann (2014). "EFSA's scientific activities and achievements on the risk assessment of genetically modified organisms (GMOs) during its first decade of existence: looking back and ahead". *Transgenic Research* 23.1, pp. 1–25. DOI: 10.1007/s11248-013-9741-4 (pp. 6, 222).

Di, H., L. Shi, H. Shen, H. Yan, H. Meng, L. Li, M. J. Alam, S. Yamasaki, and L. Ye (2014). "Rapid detection of genetically modified ingredients in soybean products by real-time loop-mediated isothermal amplification". *Journal of Food and Nutrition Research* 2.7, pp. 363–368. DOI: 10.12691/jfnr-2-7-6 (pp. 36, 38, 39, 63).

Dobnik, D., D. Morisset, and K. Gruden (2010). "NAIMA as a solution for future GMO diagnostics challenges". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2229–2233. DOI: 10.1007/s00216-009-3197-7 (pp. 26, 62).

Dobnik, D., B. Spilsberg, A. B. Košir, A. Holst-Jensen, and J. Žel (2015). "Multiplex quantification of 12 European Union authorized genetically modified maize lines with droplet digital polymerase chain reaction". *Analytical Chemistry* 87.16, pp. 8218–8226. DOI: 10.1021/acs.analchem.5b01208 (p. 63).

Dörries, H.-H., I. Remus, A. Grönewald, C. Grönewald, and K. Berghof-Jäger (2010). "Development of a qualitative, multiplex real-time PCR kit for screening of genetically modified organisms (GMOs)". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2043–2054. DOI: 10.1007/s00216-009-3149-2 (pp. 19, 24, 53, 191).

Duan, G., T. Kamiya, S. Ishikawa, T. Arao, and T. Fujiwara (2012). "Expressing S*cACR3* in rice enhanced arsenite efflux and reduced arsenic accumulation in rice grains". *Plant and Cell Physiology* 53.1, pp. 154–163. DOI: 10.1093/pcp/pcr161 (p. 90).

DuBose, A. J., S. T. Lichtenstein, N. Narisu, L. L. Bonnycastle, A. J. Swift, P. S. Chines, and F. S. Collins (2013). "Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene". *Nucleic Acids Research* 41.6, e70. DOI: 10.1093/nar/gks1463 (pp. 47, 64).

Elsevier (2015). *Scopus*. URL: http://www.scopus.com/home.url (p. 121).

Elsevier (2016). *Scopus*. URL: http://www.scopus.com/home.url (p. 86).

EMBL-EBI (2015). *ClustalW2*. URL: http://www.ebi.ac.uk/Tools/msa/clustalw2/ (pp. 112, 120, 121, 136, 166).

EMBnet (2015). *wEMBOSS*. URL: http://wemboss.sourceforge.net/ (pp. 114, 120, 152, 164, 166).

Emslie, K. R., L. Whaites, K. R. Griffiths, and E. J. Murby (2007). "Sampling plan and test protocol for the semiquantitative detection of genetically modified canola (*Brassica napus*) seed in bulk canola seed". *Journal of Agricultural and Food Chemistry* 55.11, pp. 4414–4421. DOI: 10.1021/jf070267i (pp. 9, 10).

Endo, M., K. Osakabe, K. Ono, H. Handa, T. Shimizu, and S. Toki (2007). "Molecular breeding of a novel herbicide-tolerant rice by gene targeting". *The Plant Journal* 52.1, pp. 157–166. DOI: 10.1111/j.1365-313X.2007.03230.x (p. 94).

Endo, M., T. Shimizu, and S. Toki (2012). "Selection of transgenic rice plants using a herbicide tolerant form of the acetolactate synthase gene. Methods and Protocols". *Transgenic Plants: Methods and Protocols*. Ed. by J. A. Dunwell and A. C. Wetten. Vol. 847. Methods in Molecular Biology. Totowa, NJ, USA: Humana Press. Chap. 6, pp. 59–66. ISBN: 9781617795572. DOI: 10.1007/978-1-61779-558-9_6 (p. 89).

Endo, M., T. Shimizu, T. Fujimori, S. Yanagisawa, and S. Toki (2013). "Herbicide-resistant mutations in acetolactate synthase can reduce feedback inhibition and lead to accumulation of branched-chain amino acids". *Food and Nutrition Sciences* 4.5, pp. 522–528. DOI: 10.4236/fns.2013.45067 (p. 94).

ENGL *ad hoc* working group on "unauthorised GMOs" (2011). *Overview on the detection, interpretation and reporting on the presence of unauthorised genetically modified materials*. EUR 25008 EN. European Commission Joint Research Centre. ISBN: 9789279218002. DOI: 10.2788/89665. URL: http://gmo-crl.jrc.ec.europa.eu/doc/2011-12-12%20ENGL%20UGM%20WG%20Publication.pdf (pp. 86, 132).

Ermolli, M., A. Fantozzi, M. Marini, D. Scotti, B. Balla, S. Hoffmann, M. Querci, C. Paoletti, and G. Van den Eede (2006). "Food safety: screening tests used to detect and quantify GMO proteins". *Accreditation and Quality Assurance* 11.1, pp. 55–57. DOI: 10.1007/s00769-005-0027-2 (pp. 9, 10).

Espinoza, C., R. Schlechter, D. Herrera, E. Torres, A. Serrano, C. Medina, and P. Arce-Johnson (2013). "Cisgenesis and intragenesis: new tools for improving crops". *Biological Research* 46.4, pp. 323–331. DOI: 10.4067/S0716-97602013000400003 (p. 202).

European Network of GMO Laboratories (ENGL) (2015). *Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing*. JRC95544. 24 pp. URL: http://www.iss.it/binary/rogm/cont/MPR_Report_Application_20_10_2015.pdf (p. 51).

European Union Reference Laboratory for GM Food and Feed and Feed (2006a). *Event-specific method for the quantification of sugar beet line H7-1 using Real-time PCR: protocol*. Tech. Rep. CRLVL28/04VP (p. 121).

European Union Reference Laboratory for GM Food and Feed and Feed (2006b). *Sampling and DNA extraction of cotton seeds: Report from the Validation of the "CTAB/Genomic-tip 20" method for DNA extraction from ground cotton seeds*. URL: http://gmo-crl.jrc.ec.europa.eu/summaries/281-3006%20Cotton_DNAExtr.pdf (pp. 112, 134, 163, 173).

Fagan, J., B. Schoel, A. Haegert, J. Moore, and J. Beeby (2001). "Performance assessment under field conditions of a rapid immunological test for transgenic soybeans". *International Journal of Food Science & Technology* 36.4, pp. 357–367. DOI: 10.1046/j.1365-2621.2001.00482.x (p. 9).

Fahrenkrug, S. C., A. Blake, D. F. Carlson, T. Doran, A. Van Eenennaam, D. Faber, C. Galli, Q. Gao, P. B. Hackett, N. Li, E. A. Maga, W. M. Muir, J. D. Murray, D. Shi, R. Stotish, E. Sullivan, J. F. Taylor, M. Walton, M. Wheeler, B. Whitelaw, and B. P. Glenn (2010). "Precision genetics for complex objectives in animal agriculture". *Journal of Animal Science* 88.7, pp. 2530–2539. DOI: 10.2527/jas.2010-2847 (p. 223).

Fantozzi, A., M. Ermolli, M. Marini, B. Balla, M. Querci, and G. Van den Eede (2008). "Innovative application of fluorescent microsphere based assay for multiple GMO detection". *Food Analytical Methods* 1.1, pp. 10–17. DOI: 10.1007/s12161-007-9005-0 (pp. 31, 32, 62).

FAO (2015). *FAO Rice Market Monitor (RMM): April 2015*. Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/economic/est/publications/rice-publications/rice-market-monitor-rmm/en/ (p. 80).

Fernandes, T. J., M. B. P. P. Oliveira, and I. Mafra (2013). "Tracing transgenic maize as affected by breadmaking process and raw material for the production of a traditional maize bread, *broa*". *Food Chemistry* 138.1, pp. 687–692. DOI: 10.1016/j.foodchem.2012.10.068 (p. 132).

Fernandez, S., C. Charles-Delobel, A. Geldreich, G. Berthier, F. Boyer, C. Collonnier, G. Coué-Philippe, A. Diolez, M.-N. Duplan, N. Kebdani, M. Romaniuk, M. Feinberg, and Y. Bertheau (2005). "Quantification of the 35S promoter in DNA extracts from genetically modified organisms using real-time polymerase chain reaction and specificity assessment on various genetically modified organisms, part I: operating procedure". *Journal of AOAC International* 88.2, pp. 547–557 (p. 121).

Foti, N., R. Onori, E. Donnarumma, B. De Santis, and M. Miraglia (2006). "Real-time PCR multiplex method for the quantification of Roundup Ready soybean in raw material and processed food". *European Food Research and Technology* 222.1, pp. 209–216. DOI: 10.1007/s00217-005-0107-x (p. 191).

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2014). "An innovative and integrated approach based on DNA walking to identify unauthorised GMOs". *Food Chemistry* 147, pp. 60–69. DOI: 10.1016/j.foodchem.2013.09.112 (pp. 11, 12, 14, 19, 40, 44, 51–55, 57, 61, 80, 81, 91, 96, 98, 104, 107, 131–138, 145, 150, 151, 153, 154, 161, 163, 164, 171, 174, 177, 180, 187, 191, 211, 212, 220, 229, 233).

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, F. Van Nieuwerburgh, D. Deforce, and N. H. Roosens (2015a). "Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products". *Food chemistry* 173, pp. 1259–1265. DOI: 10.1016/j.foodchem.2014.09.148 (pp. 40, 44, 51, 54, 55, 98, 104, 129, 150, 151, 153, 161, 171, 173, 174, 177, 179, 187, 199, 220).

Fraiture, M.-A., P. Herman, L. Lefèvre, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2015b). "Integrated DNA walking system to characterize a broad spectrum of GMOs in food/feed matrices". *BMC Biotechnology* 15, 76. DOI: 10.1186/s12896-015-0191-3 (pp. 40, 44, 47, 51, 54, 55, 57, 65, 98, 101, 104, 147, 171, 172, 174, 177, 179, 180, 187, 220, 232, 233).

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2015c). "Current and new approaches in GMO detection: challenges and solutions". *BioMed Research International*, article ID 392872. DOI: 10.1155/2015/392872 (pp. 7, 51, 65, 80).

Fraiture, M.-A., N. H. C. Roosens, I. Taverniers, M. De Loose, D. Deforce, and P. Herman (2016a). "Biotech rice: current developments and future detection challenges in food and feed chain". *Trends in Food Science & Technology* 52, pp. 66–79. DOI: 10.1016/j.tifs.2016.03.011 (p. 78).

Fraiture, M.-A., S. Broeders, P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2016b). "Advances in the identification of genetically modified foods". *Advances in Food Authenticity Testing*. Ed. by G. Downey. 1st edition. Woodhead Publishing Series in Food Science, Technology and Nutrition. Cambridge, UK: Woodhead Publishing. Chap. 21. ISBN: 9780081002209. Accepted (pp. 50, 82, 98).

Fraley, R. T., S. G. Rogers, and R. B. Horsch (1983). "Use of a chimeric gene to confer antibiotic resistance to plant cells". *Advances in Gene Technology: Molecular Genetics of Plants and Animals*. Miami Winter Symposia. Ed. by K. Downey, R. W. Voellmy, F. Ahmad, and J. Schultz. Vol. 20. Miami, FL, USA: Academic Press, pp. 211–221. ISBN: 9780122214806. DOI: 10.1016/B978-0-12-221480-6.50018-9 (p. 2).

Fritsch, L., R. Fischer, C. Wambach, M. Dudek, S. Schillberg, and F. Schröper (2015). "Next-generation sequencing is a robust strategy for the high-throughput detection of zygosity in transgenic maize". *Transgenic Research* 24.4, pp. 615–623. DOI: 10.1007/s11248-015-9864-x (p. 64).

Fu, K., W. Huang, T. Deng, F. Li, and H. Liu (2015a). "Multiplex PCR assay and liquid bead array for detection of 13 lines genetically modified maize". *Journal of Chinese Institute of Food Science and Technology* 15.1, pp. 188–197. DOI: 10.16429/j.1009-7848.2015.01.028 (pp. 31, 62).

Fu, W., P. Zhu, C. Wang, K. Huang, Z. Du, W. Tian, Q. Wang, H. Wang, W. Xu, and S. Zhu (2015b). "A highly sensitive and specific method for the screening detection of genetically modified organisms based on digital PCR without pretreatment". *Scientific Reports* 5, article number 12715. DOI: 10.1038/srep12715 (p. 63).

Fukayama, H., H. Tsuchida, S. Agarie, M. Nomura, H. Onodera, K. Ono, B.-H. Lee, S. Hirose, S. Toki, M. S. Ku, A. Makino, M. Matsuoka, and M. Miyao (2001). "Significant accumulation of C_4-specific pyruvate, orthophosphate dikinase in a C_3 plant, rice". *Plant Physiology* 127.3, pp. 1136–1146. DOI: 10.1104/pp.010641 (p. 90).

Fukuta, S., Y. Mizukami, A. Ishida, J. Ueda, M. Hasegawa, I. Hayashi, M. Hashimoto, and M. Kanbe (2004). "Real-time loop-mediated isothermal amplification for the CaMV-35S promoter as a screening method for genetically modified organisms". *European Food Research and Technology* 218.5, pp. 496–500. DOI: 10.1007/s00217-003-0862-5 (pp. 33, 38, 63).

García-Cañas, V., C. Simó, C. León, E. Ibáñez, and A. Cifuentes (2011). "MS-based analytical methodologies to characterize genetically modified crops". *Mass Spectrometry Reviews* 30.3, pp. 396–416. DOI: 10.1002/mas.20286 (p. 9).

Garg, A. K., R. J. H. Sawers, H. Wang, J.-K. Kim, J. M. Walker, T. P. Brutnell, M. V. Parthasarathy, R. D. Vierstra, and R. J. Wu (2006). "Light-regulated overexpression of an *Arabidopsis* phytochrome A gene in rice alters plant architecture and increases grain yield". *Planta* 223.4, pp. 627–636. DOI: 10.1007/s00425-005-0101-3 (p. 90).

Genetic Rights Foundation (2015). *Biosafety Scanner*. URL: http://en.biosafetyscanner.org/ (pp. 82, 93, 109).

Genohub (2015). *Main page*. URL: https://genohub.com/ (p. 212).

Gerdes, L., U. Busch, and S. Pecoraro (2012). "GMO*finder* – a GMO screening database". *Food Analytical Methods* 5.6, pp. 1368–1376. DOI: 10.1007/s12161-012-9378-6 (pp. 19, 52).

Gheysen, G., R. Villarroel, and M. Van Montagu (1991). "Illegitimate recombination in plants: a model for T-DNA integration". *Genes & Development* 5, pp. 287–297. DOI: 10.1101/gad.5.2.287 (p. 124).

Giovannoli, C., L. Anfossi, C. Baggiani, and G. Giraudi (2008). "Binding properties of a monoclonal antibody against the Cry1Ab from *Bacillus thuringensis* for the development of a capillary electrophoresis competitive immunoassay". *Analytical and Bioanalytical Chemistry* 392.3, pp. 385–393. DOI: 10.1007/s00216-007-1811-0 (pp. 9, 10).

GMO Compass (2015). *GMO Database*. URL: http://www.gmo-compass.org/eng/gmo/db/ (pp. 82, 150).

Golden Rice Humanitarian Board (2015). *Golden Rice Project*. URL: http://www.goldenrice.org/ (p. 88).

Greilhuber, J. (2005). "Intraspecific variation in genome size in angiosperms: identifying its existence". *Annals of Botany* 95.1, pp. 91–98. DOI: 10.1093/aob/mci004 (p. 205).

Gruère, G. P. and S. Rao (2007). "A review of international labeling policies of genetically modified food to evaluate India's proposed rule". *AgBioForum* 10.1, pp. 51–64 (p. 4).

Gryson, N. (2010). "Effect of food processing on plant DNA degradation and PCR-based GMO analysis: a review". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2003–2022. DOI: 10.1007/s00216-009-3343-2 (p. 132).

Gu, S.-b., Y. Wu, S.-C. Li, L. Li, and J.-b. Yang (2009). "Development of a real-time multiplex PCR system for the quantitative detection of Chinese GM rice". *Journal of the Science of Food and Agriculture* 89.6, pp. 1101–1106. DOI: 10.1002/jsfa.3553 (p. 96).

Guan, X., J. Guo, P. Shen, L. Yang, and D. Zhang (2010). "Visual and rapid detection of two genetically modified soybean events using loop-mediated isothermal amplification method". *Food Analytical Methods* 3.4, pp. 313–320. DOI: 10.1007/s12161-010-9132-x (pp. 36, 38, 39, 63).

Guertler, P., I. Huber, S. Pecoraro, and U. Busch (2012). "Development of an event-specific detection method for genetically modified rice Kefeng 6 by quantitative real-time PCR". *Journal für Verbraucherschutz und Lebensmittelsicherheit* 7.1, pp. 63–70. DOI: 10.1007/s00003-011-0748-6 (pp. 97, 103).

Guo, J., L. Yang, L. Chen, D. Morisset, X. Li, L. Pan, and D. Zhang (2011). "MPIC: a high-throughput analytical method for multiple DNA targets". *Analytical Chemistry* 86.5, pp. 1579–1586. DOI: 10.1021/ac103266w (pp. 20, 26, 27, 62).

Hamels, S., T. Glouden, K. Gillard, M. Mazzara, F. Debode, N. Foti, M. Sneyers, T. E. Nuez, M. Pla, G. Berben, W. Moens, Y. Bertheau, C. Audéon, G. Van den Eede, and J. Remacle (2009). "A PCR-microarray method for the screening of genetically modified organisms". *European Food Research and Technology* 228.4, pp. 531–541. DOI: 10.1007/s00217-008-0960-5 (pp. 26, 29, 62).

Han, X., H. Wang, H. Chen, L. Mei, S. Wu, G. Jia, T. Cheng, S. Zhu, and X. Lin (2013). "Development and primary application of a fluorescent liquid bead array for the simultaneous identification of multiple genetically modified maize". *Biosensors and Bioelectronics* 49, pp. 360–366. DOI: 10.1016/j.bios.2013.05.045 (pp. 31, 32, 62).

Hartung, F. and J. Schiemann (2014). "Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU". *The Plant Journal* 78.5, pp. 742–752. DOI: 10.1111/tpj.12413 (pp. 6, 7, 222).

Hatada, I. and T. Horii (2016). "Genome editing: a breakthrough in life science and medicine". *Endocrine Journal* 63.2, pp. 105–110. DOI: 10.1507/endocrj.EJ15-0716 (p. 223).

He, Y., T. Ning, T. Xie, Q. Qiu, L. Zhang, Y. Sun, D. Jiang, K. Fu, F. Yin, W. Zhang, L. Shen, H. Wang, J. Li, Q. Lin, Y. Sun, H. Li, Y. Zhu, and D. Yang (2011). "Large-scale production of functional human serum albumin from transgenic rice seeds". *Proceedings of the National Academy of Sciences of the United States of America* 108.47, pp. 19078–19083. DOI: 10.1073/pnas.1109736108 (p. 91).

He, Z., X. Xia, S. Peng, and T. A. Lumpkin (2014). "Meeting demands for increased cereal production in China". *Journal of Cereal Science* 59.3, pp. 235–244. DOI: 10.1016/j.jcs.2013.07.012 (pp. 80, 93).

Heide, B. R., S. M. Drømtorp, K. Rudi, E. Heir, and A. L. Holck (2008a). "Determination of eight genetically modified maize events by quantitative, multiplex PCR and fluorescence capillary gel electrophoresis". *European Food Research and Technology* 227.4, pp. 1125–1137. DOI: 10.1007/s00217-008-0828-8 (pp. 20, 28, 62).

Heide, B. R., E. Heir, and A. Holck (2008b). "Detection of eight GMO maize events by qualitative, multiplex PCR and fluorescence capillary gel electrophoresis". *European Food Research and Technology* 227.2, pp. 527–535. DOI: 10.1007/s00217-007-0751-4 (pp. 20, 28, 62).

Hermann, S. R., J. A. C. Miller, S. O'Neill, T. T. Tsao, R. M. Harding, and J. L. Dale (2000). "Single-primer amplification of flanking sequences". *BioTechniques* 29.6, pp. 1176–1180 (p. 111).

Herrera-Estrella, L., A. Depicker, M. Van Montagu, and J. Schell (1983). "Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector". *Nature* 303.5914, pp. 209–213. DOI: 10.1038/303209a0 (p. 1).

Hiei, Y., T. Komari, and T. Kubo (1997). "Transformation of rice mediated by *Agrobacterium tumefaciens*". *Plant Molecular Biology* 35.1, pp. 205–218. DOI: 10.1023/A:1005847615493 (pp. 2, 83, 85).

Hien, D. T., M. Jacobs, G. Angenon, C. Hermans, T. T. Thu, L. V. Son, and N. H. Roosens (2003). "Proline accumulation and Δ^1-pyrroline-5-carboxylate synthetase gene properties in three rice cultivars differing in salinity and drought tolerance". *Plant Science* 165.5, pp. 1059–1068. DOI: 10.1016/S0168-9452(03)00301-7 (p. 90).

High, S. M., M. B. Cohen, Q. Y. Shu, and I. Altosaar (2004). "Achieving successful deployment of *Bt* rice". *Trends in Plant Science* 9.6, pp. 286–292. DOI: 10.1016/j.tplants.2004.04.002 (p. 109).

Höhne, M., C. Santisi, and R. Meyer (2002). "Real-time multiplex PCR: an accurate method for the detection and quantification of 35S-CaMV promoter in genetically modified maize-containing food". *European Food Research and Technology* 215.1, pp. 59–64. DOI: 10.1007/s00217-002-0503-4 (p. 121).

Holck, A., M. Va, L. Didierjean, and K. Rudi (2002). "5'-Nuclease PCR for quantitative event-specific detection of the genetically modified Mon810 MaisGard maize". *European Food Research and Technology* 214.5, pp. 449–454. DOI: 10.1007/s00217-001-0473-y (pp. 37, 43, 111).

Holck, A., B. O. Pedersen, and E. Heir (2010). "Detection of five novel GMO maize events by qualitative, multiplex PCR and fluorescence capillary gel electrophoresis". *European Food Research and Technology* 231.3, pp. 475–483. DOI: 10.1007/s00217-010-1302-y (pp. 20, 27, 62).

Holck, A. L. and B. O. Pedersen (2011). "Simple, sensitive, accurate multiplex quantitative competitive PCR with capillary electrophoresis detection for the determination of genetically modified maize". *European Food Research and Technology* 233.6, pp. 951–961. DOI: 10.1007/s00217-011-1594-6 (pp. 20, 27, 62).

Holme, I. B., T. Wendt, and P. B. Holm (2013). "Intragenesis and cisgenesis as alternatives to transgenic crop development". *Plant Biotechnology Journal* 11.4, pp. 395–407. DOI: 10.1111/pbi.12055 (p. 202).

Holst-Jensen, A., Y. Bertheau, M. de Loose, L. Grohmann, S. Hamels, L. Hougs, D. Morisset, S. Pecoraro, M. Pla, M. Van den Bulcke, and D. Wulff (2012). "Detecting un-authorized genetically modified organisms (GMOs) and derived materials". *Biotechnology Advances* 30.6, pp. 1318–1335. DOI: 10.1016/j.biotechadv.2012.01.024 (pp. 12, 53, 81, 82, 110, 171).

Hoque, M. E., J. W. Mansfield, and M. H. Bennett (2005). "*Agrobacterium*-mediated transformation of Indica rice genotypes: an assessment of factors affecting the transformation efficiency". *Plant Cell, Tissue and Organ Culture* 82.1, pp. 45–55. DOI: 10.1007/s11240-004-6154-3 (pp. 2, 83, 85).

Hsieh, F. and B. S. Luh (1991). "Rice Snack Foods". *Rice: Utilization*. Ed. by B. S. Luh. II vols. New York, NY, USA: Van Nostrand Reinhold. Chap. 10, pp. 225–249. ISBN: 0442004850 (p. 133).

Hu, T., X. Qv, G. Xiao, and X. Huang (2009). "Enhanced tolerance to herbicide of rice plants by over-expression of a glutathione *S*-transferase". *Molecular Breeding* 24, pp. 409–418. DOI: 10.1007/s11032-009-9302-y (p. 89).

Huang, Y., B. Niu, Y. Gao, L. Fu, and W. Li (2010). "CD-HIT Suite: a web server for clustering and comparing biological sequences". *Bioinformatics* 26.5, pp. 680–682. DOI: 10.1093/bioinformatics/btq003 (p. 175).

Huang, X., L. Chen, J. Xu, H.-F. Ji, S. Zhu, and H. Chen (2014). "Rapid visual detection of phytase gene in genetically modified maize using loop-mediated isothermal amplification method". *Food Chemistry* 156, pp. 184–189. DOI: 10.1016/j.foodchem.2014.01.102 (pp. 36, 38, 63).

Huang, L., L. Zheng, Y. Chen, F. Xue, L. Cheng, S. B. Adeloju, and W. Chen (2015). "A novel GMO biosensor for rapid ultrasensitive and simultaneous detection of multiple DNA components in GMO products". *Biosensors and Bioelectronics* 66, pp. 431–437. DOI: 10.1016/j.bios.2014.12.005 (p. 104).

Huber, I., A. Block, D. Sebah, F. Debode, D. Morisset, L. Grohmann, G. Berben, D. Štebih, M. Milavec, J. Žel, and U. Busch (2013). "Development and validation of duplex, triplex, and pentaplex real-time PCR screening assays for the detection of genetically modified organisms in food and feed". *Journal of Agricultural and Food Chemistry* 61.43, pp. 10293–10301. DOI: 10.1021/jf402448y (pp. 19, 21, 24, 53, 191).

Humphrey, B. D., N. Huang, and K. C. Klasing (2002). "Rice expressing lactoferrin and lysozyme has antibiotic-like properties when fed to chicks". *Journal of Nutrition* 132.6, pp. 1214–1218 (p. 90).

Illumina (2015a). *iGenomes*. URL: http://support.illumina.com/sequencing/sequencing_software/igenome.html (p. 200).

Illumina (2015b). *Main page*. URL: http://www.illumina.com/ (p. 212).

Illumina (2016). *Main page*. URL: http://www.illumina.com/ (p. 69).

IRMM, Geel, Belgium (2015). *RM catalogue*. IRMM (Institute for Reference Materials and Measurements). URL: https://ec.europa.eu/jrc/en/reference-materials/catalogue (pp. 112, 133).

ISO 21571 (2005). *ISO 21571:2005 Foodstuffs – Methods of analysis for the detection of genetically modified organisms and derived products – Nucleic acid extraction*. ISO. Geneva, Switzerland: International Organization for Standardization (pp. 112, 134, 163, 173).

ISO 24276 (2006). *ISO 24276:2006 Foodstuffs – Methods of analysis for the detection of genetically modified organisms and derived products – General requirements and definitions*. ISO. Geneva, Switzerland: International Organization for Standardization (p. 132).

Jacchia, S., E. Nardini, N. Bassani, C. Savini, J.-H. Shim, K. Trijatmiko, J. Kreysa, and M. Mazzara (2015). "International ring trial for the validation of an event-specific Golden Rice 2 quantitative real-time polymerase chain reaction method". *Journal of Agricultural and Food Chemistry* 63.20, pp. 4954–4965. DOI: 10.1021/acs.jafc.5b00951 (p. 97).

James, C. and A. F. Krattiger (1996). *Global Review of the Field Testing and Commercialization of Transgenic Plants, 1986 to 1995: The First Decade of Crop Biotechnology*. ISAAA Briefs 1. Ithaca, NY, USA: ISAAA. 31 pp. URL: https://www.isaaa.org/kc/Publications/pdfs/isaaabriefs/Briefs%201.pdf (p. 2).

James, C. (2009). *Global Status of Commercialized Biotech/GM Crops: 2009*. ISAAA Brief 41. Ithaca, NY, USA: ISAAA. ISBN: 978-1-892456-48-6. URL: https://isaaa.org/resources/publications/briefs/41/ (p. 109).

James, C. (2013). *Global Status of Commercialized Biotech/GM Crops: 2013*. ISAAA Brief 46. Ithaca, NY, USA: ISAAA. ISBN: 978-1-892456-55-9. URL: http://www.isaaa.org/resources/publications/briefs/46/ (p. 191).

James, C. (2014). *Global Status of Commercialized Biotech/GM Crops: 2014*. ISAAA Brief 49. Ithaca, NY, USA: ISAAA. ISBN: 978-1-892456-59-1. URL: http://www.isaaa.org/resources/publications/briefs/49/ (pp. 8, 50, 88, 92, 150).

Jang, H.-J., I.-H. Cho, H.-S. Kim, J.-W. Jeon, S.-Y. Hwang, and S.-H. Paek (2011). "Development of a chemiluminometric immunosensor array for on-site monitoring of genetically modified organisms". *Sensors and Actuators B: Chemical* 155.2, pp. 598–605. DOI: 10.1016/j.snb.2011.01.016 (pp. 9, 10).

Ji, J. and J. Braam (2010). "Restriction site extension PCR: a novel method for high-throughput characterization of tagged DNA fragments and genome walking". *PLoS ONE* 5.5, e10577. DOI: 10.1371/journal.pone.0010577 (p. 37).

Joint Research Centre (2011). *Compendium of reference methods for GMO analysis*. EUR 24526 EN. European Commission Joint Research Centre. ISBN: 9789279201745. DOI: 10.2788/16086 (pp. 121, 220).

Joint Research Centre (2014). *JRC GMO-Matrix application*. URL: http://gmo-crl.jrc.ec.europa.eu/jrcgmomatrix/ (pp. 52, 157).

Joint Research Centre (2015a). *GMOMETHODS: EU Database of Reference Methods for GMO Analysis*. Joint Research Centre - European Union Reference Laboratory for GM Food and Feed. URL: http://gmo-crl.jrc.ec.europa.eu/gmomethods/ (pp. 11–18, 20, 51, 54, 131, 174, 229–232).

Joint Research Centre (2015b). *JRC GMO-Amplicons*. Joint Research Centre - European Union Reference Laboratory for GM Food and Feed. URL: http://gmo-crl.jrc.ec.europa.eu/jrcgmoamplicons/ (pp. 19, 54).

Joint Research Centre – European Commission (2005). *Event-specific method for the quantitation of maize line MON 863 using real-time PCR. Protocol*. protocol. URL: http://gmo-crl.jrc.ec.europa.eu/summaries/MON863-WEB-Protocol-Validation.pdf (p. 158).

Jones, D. H. and S. C. Winistorfer (1992). "Sequence specific generation of a DNA panhardle permits PCR amplication of unknown flanking DNA". *Nucleic Acids Research* 20.3, pp. 595–600. DOI: 10.1093/nar/20.3.595 (p. 110).

Kamle, S., A. Ojha, and A. Kumar (2011). "Development of an enzyme linked immunosorbant assay for the detection of Cry2Ab protein in transgenic plants". *GM Crops* 2.2, pp. 118–125. DOI: 10.4161/gmcr.2.2.16191 (pp. 9, 10).

Kamle, S. and S. Ali (2013). "Genetically modified crops: detection strategies and biosafety issues". *Gene* 522.2, pp. 123–132. DOI: 10.1016/j.gene.2013.03.107 (pp. 8, 50, 80).

Kanchiswamy, C. N., M. Malnoy, R. Velasco, J.-S. Kim, and R. Viola (2015). "Non-GMO genetically edited crop plants". *Trends in Biotechnology* 33.9, pp. 489–491. DOI: 10.1016/j.tibtech.2015.04.002 (p. 222).

Kanzaki, H., S. Nirasawa, H. Saitoh, M. Ito, M. Nishihara, R. Terauchi, and I. Nakamura (2002). "Overexpression of the wasabi defensin gene confers enhanced resistance to blast fungus (*Magnaporthe grisea*) in transgenic rice". *Theoretical and Applied Genetics* 105.6, pp. 809–814. DOI: 10.1007/s00122-001-0817-9 (p. 90).

Kathuria, H., J. Giri, H. Tyagi, and K. Akhilesh (2007). "Advances in transgenic rice biotechnology". *Critical Reviews in Plant Sciences* 26.2, pp. 65–103. DOI: 10.1080/07352680701252809 (pp. 2, 82, 83, 109, 121).

Katovich, E. (2012). *The regulation of genetically modified organisms in Latin America: policy implications for trade, biosafety, and development*. University of Minnesota Digital Conservancy. URL: http://purl.umn.edu/140920 (p. 4).

Kawahara, Y., M. de la Bastide, J. P. Hamilton, H. Kanamori, W. R. McCombie, S. Ouyang, D. C. Schwartz, T. Tanaka, J. Wu, S. Zhou, K. L. Childs, R. M. Davidson, H. Lin, L. Quesada-Ocampo, B. Vaillancourt, H. Sakai, S. S. Lee, J. Kim, H. Numa, T. Itoh, C. R. Buell, and T. Matsumoto (2013). "Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data". *Rice* 6, 4. DOI: 10.1186/1939-8433-6-4 (pp. 201, 205).

Khush, G. S. (2013). "Strategies for increasing the yield potential of cereals: case of rice as an example". *Plant Breeding* 132.5, pp. 433–436. DOI: 10.1111/pbr.1991 (p. 80).

Kiddle, G., P. Hardinge, N. Buttigieg, O. Gandelman, C. Pereira, C. J. McElgunn, M. Rizzoli, R. Jackson, N. Appleton, C. Moore, L. C. Tisi, and J. A. H. Murray (2012). "GMO detection using a bioluminescent real time reporter (BART) of loop mediated isothermal amplification (LAMP) suitable for field use". *BMC Biotechnology* 12, 15. DOI: 10.1186/1472-6750-12-15 (pp. 36, 38, 63).

Kim, Y.-S., I.-S. Kim, M.-J. Bae, Y.-H. Choe, Y.-H. Kim, H.-M. Park, H.-G. Kang, and H.-S. Yoon (2013). "Homologous expression of cytosolic dehydroascorbate reductase increases grain yield and biomass under paddy field conditions in transgenic rice (*Oryza sativa* L. *japonica*)". *Planta* 237.6, pp. 1613–1625. DOI: 10.1007/s00425-013-1862-8 (p. 90).

Kircher, M., P. Heyn, and J. Kelso (2011). "Addressing challenges in the production and analysis of illumina sequencing data". *BMC Genomics* 12, 382. DOI: 10.1186/1471-2164-12-382 (p. 205).

Kluga, L., S. Folloni, D. M. Kagkli, A. Bogni, N. Foti, C. Savini, M. Mazzara, G. Van den Eede, and M. Van den Bulcke (2013). "Combinatory SYBR® Green real-time PCR screening approach for tracing materials derived from genetically modified rice". *Food Analytical Methods* 6.2, pp. 361–369. DOI: 10.1007/s12161-012-9493-4 (pp. 96, 102, 110, 150).

Knapp, S., Y. Larondelle, M. Roßberg, D. Furtek, and K. Theres (1994). "Transgenic tomato lines containing *Ds* elements at defined genomic positions as tools for targeted transposon tagging". *Molecular and General Genetics MGG* 243.6, pp. 666–673. DOI: 10.1007/BF00279576 (p. 37).

Kok, E. J., J. Pedersen, R. Onori, S. Sowa, M. Schauzu, A. De Schrijver, and T. H. Teeri (2014). "Plants with stacked genetically modified events: to assess or not to assess?" *Trends in Biotechnology* 32.2, pp. 70–73. DOI: 10.1016/j.tibtech.2013.12.001 (p. 6).

Komori, T., T. Imayama, N. Kato, Y. Ishida, U. Jun, and T. Komari (2007). "Current status of binary vectors and superbinary vectors". *Plant Physiology* 145.4, pp. 1155–1160. DOI: 10.1104/pp.107.105734 (pp. 109, 121, 151).

Köppel, R., F. Zimmerli, and A. Breitenmoser (2010). "Multiplex real-time PCR for the simultaneous detection and quantification of DNA from three transgenic rice species and construction and application of an artificial oligonucleotide as reference molecule". *European Food Research and Technology* 230.5, pp. 731–736. DOI: 10.1007/s00217-010-1213-y (pp. 97, 103).

Köppel, R., A. Sendic, and H.-U. Waiblinger (2014). "Two quantitative multiplex real-time PCR systems for the efficient GMO screening of food products". *European Food Research and Technology* 239.4, pp. 653–659. DOI: 10.1007/s00217-014-2261-5 (pp. 19, 24, 53, 191).

Köppel, R. and T. Bucher (2015). "Rapid establishment of droplet digital PCR for quantitative GMO analysis". *European Food Research and Technology* 241.3, pp. 427–439. DOI: 10.1007/s00217-015-2475-1 (pp. 33, 35, 63, 104).

Köppel, R., T. Bucher, A. Frei, and H.-U. Waiblinger (2015). "Droplet digital PCR versus multiplex real-time PCR method for the detection and quantification of DNA from the four transgenic soy traits MON87769, MON87708, MON87705 and FG72, and lectin". *European Food Research and Technology* 241.4, pp. 521–527. DOI: 10.1007/s00217-015-2481-3 (pp. 19, 24, 33, 35, 53, 63).

Kovalic, D., C. Garnaat, L. Guo, Y. Yan, J. Groat, A. Silvanovich, L. Ralston, M. Huang, Q. Tian, A. Christian, N. Cheikh, J. Hjelle, S. Padgette, and G. Bannon (2012). "The Use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology". *The Plant Genome* 5.3,

pp. 149–163. DOI: 10.3835/plantgenome2012.10.0026 (pp. 46–48, 65, 145, 191, 192, 200).

Krizkova, L. and M. Hrouda (1998). "Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions". *The Plant Journal* 16.6, pp. 673–680. DOI: 10.1046/j.1365-313x.1998.00330.x (p. 124).

Laible, G., J. Wei, and S. Wagner (2015). "Improving livestock for agriculture – technological progress from random transgenesis to precision genome editing heralds a new era". *Biotechnology Journal* 10.1, pp. 109–120. DOI: 10.1002/biot.201400193 (p. 223).

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2". *Nature Methods* 9.4, pp. 357–359. DOI: 10.1038/nmeth.1923 (p. 199).

Lee, D., M. La Mura, T. R. Allnutt, and W. Powell (2009). "Detection of genetically modified organisms (GMOs) using isothermal amplification of target DNA sequences". *BMC Biotechnology* 9, 7. DOI: 10.1186/1472-6750-9-7 (pp. 33, 38, 39, 63).

Leimanis, S., M. Hernández, S. Fernández, F. Boyer, M. Burns, S. Bruderer, T. Glouden, N. Harris, O. Kaeppeli, P. Philipp, M. Pla, P. Puigdomènech, M. Vaitilingom, Y. Bertheau, and J. Remacle (2006). "A microarray-based detection system for genetically modified (GM) food ingredients". *Plant Molecular Biology* 61.1, pp. 123–139. DOI: 10.1007/s11103-005-6173-4 (pp. 26, 29, 62).

Leimanis, S., S. Hamels, F. Nazé, G. M. Mbella, M. Sneyers, R. Hochegger, H. Broll, L. Roth, K. Dallmann, A. Micsinai, J. L. La Paz, M. Pla, C. Brünen-Nieweler, N. Papazova, I. Taverniers, N. Hess, B. Kirschneit, Y. Bertheau, C. Audeon, V. Laval, U. Busch, S. Pecoraro, K. Neumann, S. Rösel, J. van Dijk, E. Kok, G. Bellocchi, N. Foti, M. Mazzara, W. Moens, J. Remacle, and G. Van Den Eede (2008). "Validation of the performance of a GMO multiplex screening assay based on microarray detection". *European Food Research and Technology* 227.6, pp. 1621–1632. DOI: 10.1007/s00217-008-0886-y (pp. 26, 29, 62).

Leoni, C., M. Volpicella, F. De Leo, R. Gallerani, and L. R. Ceci (2011). "Genome walking in eukaryotes". *The FEBS Journal* 278.21, pp. 3953–3977. DOI: 10.1111/j.1742-4658.2011.08307.x (pp. 36, 37, 66, 68, 110, 111, 152).

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows–Wheeler transform". *Bioinformatics* 25.14, pp. 1754–1760. DOI: 10.1093/bioinformatics/btp324 (p. 199).

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup

(2009). “The sequence alignment/map format and SAMtools”. *Bioinformatics* 25.16, pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352 (p. 200).

Li, T., B. Liu, M. H. Spalding, D. P. Weeks, and B. Yang (2012). “High-efficiency TALEN-based gene editing produces disease-resistant rice”. *Nature Biotechnology* 30, pp. 390–392. DOI: 10.1038/nbt.2199 (p. 94).

Li, L., J. Wang, C. Zang, Z. Sui, and Z. Zhao (2013a). “Development of a reference plasmid for the quantitative detection of genetically modified rice Huahui 1”. *Biotechnology Bulletin* 1, pp. 96–101 (p. 103).

Li, L., X. Zhang, Y. Wan, and W. Jin (2013b). “Development of a novel reference plasmid for accurate quantification of genetically modified Kefeng6 rice DNA in food and feed samples”. *BioMed Research International*, article ID 134675. DOI: 10.1155/2013/134675 (pp. 97, 104).

Li, Q., J. Fang, X. Liu, X. Xi, M. Li, Y. Gong, and M. Zhang (2013c). “Loop-mediated isothermal amplification (LAMP) method for rapid detection of *cry1Ab* gene in transgenic rice (*Oryza sativa* L.)” *European Food Research and Technology* 236.4, pp. 589–598. DOI: 10.1007/s00217-013-1911-3 (pp. 36, 38, 63, 104).

Li, F., W. Yan, L. Long, X. Qi, C. Li, and S. Zhang (2014). “Development and application of loop-mediated isothermal amplification assays for rapid visual detection of *cry2Ab* and *cry3A* genes in genetically-modified crops”. *International Journal of Molecular Sciences* 15.9, pp. 15109–15121. DOI: 10.3390/ijms150915109 (pp. 36, 38, 63).

Li, X., Y. Wu, J. Li, Y. Li, L. Long, F. Li, and G. Wu (2015). “Development and validation of a 48-target analytical method for high-throughput monitoring of genetically modified organisms”. *Scientific Reports* 5, 7616. DOI: 10.1038/srep07616 (pp. 33, 34, 63, 104).

Liang, C., J. P. van Dijk, I. M. J. Scholtens, M. Staats, T. W. Prins, M. M. Voorhuijzen, A. M. da Silva, A. C. M. Arisi, J. T. den Dunnen, and E. J. Kok (2014). “Detecting authorized and unauthorized genetically modified organisms containing *vip3A* by real-time PCR and next-generation sequencing”. *Analytical and Bioanalytical Chemistry* 406.11, pp. 2603–2611. DOI: 10.1007/s00216-014-7667-1 (pp. 37, 40, 41, 44, 45, 47, 49, 64, 65, 150, 162, 172, 192).

Lievens, A., M. Petrillo, M. Querci, and A. Patak (2015). “Genetically modified animals: options and issues for traceability and enforcement”. *Trends in Food Science & Technology* 44.2, pp. 159–176. DOI: 10.1016/j.tifs.2015.05.001 (p. 223).

Lipp, M., E. Anklam, and J. W. Stave (2000). “Validation of an immunoassay for detection and quantitation of a genetically modified soybean in food and food fractions using reference materials: interlaboratory study”. *Journal of AOAC International* 83.4, pp. 919–927 (pp. 9, 10).

Lipp, M., A. Bluth, F. Eyquem, L. Kruse, H. Schimmel, G. Van den Eede, and E. Anklam (2001). "Validation of a method based on polymerase chain reaction for the detection of genetically modified organisms in various processed foodstuffs". *European Food Research and Technology* 212.4, pp. 497–504. DOI: 10.1007/s002170000274 (p. 132).

Lipton, C. R., J. X. Dautlick, G. D. Grothaus, P. L. Hunst, K. M. Magin, C. A. Mihaliak, F. M. Rubio, and J. W. Stave (2000). "Guidelines for the validation and use of immunoassays for determination of introduced proteins in biotechnology enhanced crops and derived food ingredients". *Food and Agricultural Immunology* 12.2, pp. 153–164. DOI: 10.1080/095401000404094 (p. 9).

Liu, X., X. Bai, X. Wang, and C. Chu (2007). "OsWRKY71, a rice transcription factor, is involved in rice defense response". *Journal of Plant Physiology* 164.8, pp. 969–979. DOI: 10.1016/j.jplph.2006.07.006 (p. 90).

Liu, M., Y. Luo, R. Tao, R. He, K. Jiang, B. Wang, and L. Wang (2009). "Sensitive and rapid detection of genetic modified soybean (Roundup Ready) by loop-mediated isothermal amplification". *Bioscience, Biotechnology, and Biochemistry* 73.11, pp. 2365–2369. DOI: 10.1271/bbb.80723 (pp. 33, 63).

Lu, C. (2010). "The first approved transgenic rice in China". *GM Crops* 1.3, pp. 113–115. DOI: 10.4161/gmcr.1.3.12377 (p. 93).

Luminex Corporation (2015). *Luminex*. URL: https://www.luminexcorp.com/ (pp. 31, 63).

Lusser, M., C. Parisi, D. Plan, and E. Rodríguez-Cerezo (2011). *New plant breeding techniques: State-of-the-art and prospects for commercial development*. EUR 24760 EN. European Commission Joint Research Centre. ISBN: 9789279197154. DOI: 10.2791/54761. URL: http://ftp.jrc.es/EURdoc/JRC63971.pdf (pp. 94, 100, 221, 222).

Lusser, M., C. Parisi, D. Plan, and E. Rodríguez-Cerezo (2012). "Deployment of new biotechnologies in plant breeding". *Nature Biotechnology* 30.3, pp. 231–239. DOI: 10.1038/nbt.2142 (pp. 221, 222).

Ma, J., Y. P. Wang, S. Ren, Z. Zhang, S. Lu, and P. W. Wang (2014). "Cloning flanking sequence by single-primer PCR in transgenic plants". *Genetics and molecular research: GMR* 13.4, pp. 8403–8410. DOI: 10.4238/2014.October.20.16 (p. 37).

Mackenzie, S. and L. McIntosh (1999). "Higher plant mitochondria". *The Plant Cell* 11.4, pp. 571–585. DOI: http:/âĂŃ/âĂŃdx.âĂŃdoi.âĂŃorg/âĂŃ10.âĂŃ1105/âĂŃtpc.âĂŃ11.âĂŃ4.âĂŃ571 (p. 205).

Mäde, D., C. Degner, and L. Grohmann (2006). "Detection of genetically modified rice: a construct-specific real-time PCR method based on DNA sequences from transgenic Bt rice". *European Food Research and Technology* 224, pp. 271–278. DOI: 10.1007/s00217-006-0467-x (pp. 96, 102).

Majhi, B. B., J. M. Shah, and K. Veluthambi (2014). "A novel T-DNA integration in rice involving two interchromosomal translocations". *Plant Cell Reports* 33.6, pp. 929–944. DOI: 10.1007/s00299-014-1572-0 (pp. 37, 42, 150).

Masuda, H., Y. Ishimaru, M. S. Aung, T. Kobayashi, Y. Kakei, M. Takahashi, K. Higuchi, H. Nakanishi, and N. K. Nishizawa (2012). "Iron biofortification in rice by the introduction of multiple genes involved in iron nutrition". *Scientific Reports* 2, article number 543. DOI: 10.1038/srep00543 (p. 90).

Mazzara, M., S. Cordeil, and G. Van den Eede (2006a). *Report on the Verification of an Event-Specific Detection Method for Identification of Rice GM-event LLRICE601 Using a Real-time PCR Assay*. European Commission Joint Research Centre. URL: http://gmo-crl.jrc.ec.europa.eu/summaries/Verification%20Report%20LLRice601%20event.pdf (pp. 97, 102).

Mazzara, M., E. Grazioli, C. Savini, and G. Van den Eede (2006b). *Event-specific Method for the Quantitation of Rice Line LLRICE62 Using Real-time PCR – Validation Report and Protocol – Sampling and DNA Extraction of Rice*. European Commission Joint Research Centre. ISBN: 9279031295. DOI: 10.2788/31940. URL: http://gmo-crl.jrc.ec.europa.eu/summaries/LLRICE62_validated_Protocol.pdf (pp. 97, 102).

Mazzara, M., C. Paoletti, P. Corbisier, E. Grazioli, S. Larcher, G. Berben, M. De Loose, I. Folch, C. Henry, N. Hess, L. Hougs, E. Janssen, G. Moran, R. Onori, and G. Van den Eede (2013). "Kernel lot distribution assessment (KeLDA): a comparative study of protein and DNA-based detection methods for GMO testing". *Food Analytical Methods* 6.1, pp. 210–220. DOI: 10.1007/s12161-012-9407-5 (pp. 9, 50).

Mbongolo Mbella, E. G., A. Lievens, E. Barbau-Piednoir, M. Sneyers, A. Leunda-Casi, and N. Roosens (2011). "SYBR®Green qPCR methods for detection of endogenous reference genes in commodity crops: a step ahead in combinatory screening of genetically modified crops in food and feed products". *European Food Research and Technology* 232.3, pp. 485–496. DOI: 10.1007/s00217-010-1408-2 (pp. 13, 110, 116, 121, 134, 135, 150, 174, 228).

McKenzie, M. J., V. Mett, and P. E. Jameson (2000). "Modified ELISA for the detection of neomycin phosphotransferase II in transformed plant species". *Plant Cell Reports* 19.3, pp. 286–289. DOI: 10.1007/s002990050014 (pp. 9, 10).

Metzker, M. L. (2010). "Sequencing technologies – the next generation". *Nature Reviews Genetics* 11, pp. 31–46. DOI: 10.1038/nrg2626 (pp. 70, 71).

Michael, T. P. and S. Jackson (2013). "The first 50 plant genomes". *The Plant Genome* 6.2. DOI: 10.3835/plantgenome2013.03.0001in (p. 192).

Milavec, M., D. Dobnik, L. Yang, D. Zhang, K. Gruden, and J. Žel (2014). "GMO quantification: valuable experience and insights for the future". *Analytical and Bioanalytical Chemistry* 406.26, pp. 6485–6497. DOI: 10.1007/s00216-014-8077-0 (pp. 20, 62).

Min, G.-S. and J. R. Powell (1998). "Long-distance genome walking using the long and accurate polymerase chain reaction". *BioTechniques* 24.3, pp. 398–400 (p. 111).

Mirkov, T. E. (2003). "The Molecular Basis of Genetic Modification and Improvement of Crops". *Plants, Genes, and Crop Biotechnology*. Ed. by M. J. Chrispeels and D. E. Sadava. 2nd edition. Sudbury, MA, USA: Jones and Bartlett Publishers. Chap. 6, pp. 124–151. ISBN: 0763715867 (pp. 1, 3).

Morisset, D., D. Dobnik, S. Hamels, J. Žel, and K. Gruden (2008a). "NAIMA: target amplification strategy allowing quantitative on-chip detection of GMOs". *Nucleic Acids Research* 36.18, e118. DOI: 10.1093/nar/gkn524 (pp. 26, 29, 62).

Morisset, D., D. Stebih, K. Cankar, J. Zel, and K. Gruden (2008b). "Alternative DNA amplification methods to PCR and their application in GMO detection: a review". *European Food Research and Technology* 227.5, pp. 1287–1297. DOI: 10.1007/s00217-008-0850-x (pp. 9, 50, 63, 64).

Morisset, D., D. Štebih, M. Milavec, K. Gruden, and J. Žel (2013). "Quantitative analysis of food and feed samples with droplet digital PCR". *PLoS ONE* 8.8, e62583. DOI: 10.1371/journal.pone.0062583 (pp. 33–35, 63).

Morisset, D., P. K. Novak, D. Zupanič, K. Gruden, N. Lavrač, and J. Žel (2014). "GMOseek: a user friendly tool for optimized GMO testing". *BMC Bioinformatics* 15, 258. DOI: 10.1186/1471-2105-15-258 (p. 191).

Mueller, P. R. and B. Wold (1989). "*In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR". *Science* 246.4931, pp. 780–786. DOI: 10.1126/science.2814500 (p. 111).

Murai, N., J. D. Kemp, D. W. Sutton, M. G. Murray, J. L. Slightom, D. J. Merlo, N. A. Reichert, C. Sengupta-Gopalan, C. A. Stock, R. F. Barker, and T. C. Hall (1983). "Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors". *Science* 222.4623, pp. 476–482. DOI: 10.1126/science.222.4623.476 (p. 2).

Myrick, K. V. and W. M. Gelbart (2002). "Universal fast walking for direct and versatile determination of flanking sequence". *Gene* 284.1–2, pp. 125–131. DOI: 10.1016/S0378-1119(02)00384-0 (p. 37).

Nadal, A., A. Coll, J.-L. La Paz, T. Esteve, and M. Pla (2006). "A new PCR-CGE (size and color) method for simultaneous detection of genetically modified maize events". *Electrophoresis* 27.19, pp. 3879–3888. DOI: 10.1002/elps.200600124 (pp. 20, 27, 62).

Nadal, A., T. Esteve, and M. Pla (2009). "Multiplex polymerase chain reaction-capillary gel electrophoresis: a promising tool for GMO screening assay for simultaneous detection of five genetically modified cotton events and species". *Journal of AOAC International* 92.3, pp. 765–772 (pp. 20, 27, 62).

Nakaya, H. I., E. M. Reis, and S. Verjovski-Almeida (2007). "Concepts on Microarray Design for Genome and Transcriptome Analyses". *Nucleic Acids*

Hybridization Modern Applications. Ed. by A. Buzdin and S. Lukyanov. Dordrecht, The Netherlands: Springer. Chap. 11, pp. 265–307. ISBN: 9781402060397. DOI: 10.1007/978-1-4020-6040-3_11 (p. 26).

Navarro, E., G. Serrano-Herasa, M. Castañoa, and J. Solera (2015). "Real-time PCR detection chemistry". *Clinica Chimica Acta* 439, pp. 231–250. DOI: 10.1016/j.cca.2014.10.017 (pp. 10, 51).

Nielsen, C. R., K. G. Berdal, and A. Holst-Jensen (2008). "Anchored PCR for possible detection and characterisation of foreign integrated DNA at near single molecule level". *European Food Research and Technology* 226.5, pp. 949–956. DOI: 10.1007/s00217-007-0616-x (p. 132).

Nill, K. (2002). *Glossary of Biotechnology Terms*. 3rd edition. Boca Raton London New York Washington, D.C: CRC Press LLC. ISBN: 1587161222 (p. 1).

NCBI (2015). *BLAST®*. URL: http://blast.ncbi.nlm.nih.gov/Blast.cgi (pp. 120, 121, 136, 166).

Ohkawa, Y. and H. Ohkawa (2002). *Transgenic Rice and Potato Plants Expressing Human Cytochrome P450S Show Cross-Tolerance to Herbicides by Detoxifying Them*. Food and Fertilizer Technology Center. URL: http://www.fftc.agnet.org/library.php?func=view&id=20110725130014 (p. 89).

Ohri, D. (1998). "Genome size variation and plant systematics". *Annals of Botany* 82.suppl 1, pp. 75–83 (p. 205).

Ohta, M., Y. Hayashi, A. Nakashima, A. Hamada, A. Tanaka, T. Nakamura, and T. Hayakawa (2002). "Introduction of a Na^+/H^+ antiporter gene from *Atriplex gmelini* confers salt tolerance to rice". *FEBS Letters* 532.3, pp. 279–282. DOI: 10.1016/S0014-5793(02)03679-7 (p. 90).

Okuzaki, A. and K. Toriyama (2004). "Chimeric RNA/DNA oligonucleotide-directed gene targeting in rice". *Plant Cell Reports* 22.7, pp. 509–512. DOI: 10.1007/s00299-003-0698-2 (p. 94).

Opel, K. L., D. Chung, and B. R. McCord (2010). "A study of PCR inhibition mechanisms using real time PCR". *Journal of Forensic Sciences* 55.1, pp. 25–33. DOI: 10.1111/j.1556-4029.2009.01245.x (pp. 12, 53).

Pacific Biosciences (2015). *PacBio*. URL: http://www.pacb.com/ (p. 47).

Pacific Biosciences (2016). *PacBio*. URL: http://www.pacb.com/ (pp. 70, 187).

Paine, J. A., C. A. Shipton, S. Chaggar, R. M. Howells, M. J. Kennedy, G. Vernon, S. Y. Wright, E. Hinchliffe, J. L. Adams, A. L. Silverstone, and R. Drake (2005). "Improving the nutritional value of Golden Rice through increased pro-vitamin A content". *Nature Biotechnology* 23, pp. 482–487. DOI: 10.1038/nbt1082 (p. 88).

Pan, A., L. Yang, S. Xu, C. Yin, K. Zhang, Z. Wang, and D. Zhang (2006). "Event-specific qualitative and quantitative PCR detection of MON863 maize based upon the 3'-transgene integration sequence". *Journal of Cereal Science*

43.2, pp. 250–257. DOI: 10.1016/j.jcs.2005.10.003 (pp. 37, 44, 158, 159, 161).

Parisi, C., P. Tillie, and E. Rodríguez-Cerezo (2016). "The global pipeline of GM crops out to 2020". *Nature Biotechnology* 34, pp. 31–36. DOI: 10.1038/nbt.3449 (pp. 1, 81).

Park, S.-B., H.-Y. Kim, and J.-H. Kim (2015). "Multiplex PCR system to track authorized and unauthorized genetically modified soybean events in food and feed". *Food Control* 54, pp. 47–52. DOI: 10.1016/j.foodcont.2015.01.035 (pp. 19, 24, 53).

Parker, J. D., P. S. Rabinovitch, and G. C. Burmer (1991). "Targeted gene walking polymerase chain reaction". *Nucleic Acids Research* 19.11, pp. 3055–3060. DOI: 10.1093/nar/19.11.3055 (p. 111).

Pauwels, K., N. Podevin, D. Breyer, D. Carroll, and P. Herman (2014). "Engineering nucleases for gene targeting: safety and regulatory considerations". *New Biotechnology* 31.1, pp. 18–27. DOI: 10.1016/j.nbt.2013.07.001 (p. 94).

Peano, C., M. C. Samson, L. Palmieri, M. Gulli, and N. Marmiroli (2004). "Qualitative and quantitative evaluation of the genomic DNA extracted from GMO and non-GMO foodstuffs with four different extraction methods". *Journal of Agricultural and Food Chemistry* 52.23, pp. 6962–6968. DOI: 10.1021/jf040008i (p. 132).

Pla, M., A. Nadal, V. Baeten, C. Bahrdt, G. Berben, Y. Bertheau, A. Coll, J. P. van Dijk, D. Dobnik, J. A. Fernandez Pierna, K. Gruden, S. Hamels, A. Holck, A. Holst-Jensen, E. Janssen, E. J. Kok, J. L. La Paz, V. Laval, S. Leimanis, A. Malcevschi, N. Marmiroli, D. Morisset, T. W. Prins, J. Remacle, G. Ujhelyi, and D. Wulff (2012). "New Multiplexing Tools for Reliable GMO Detection". *Genetically Modified and Non-Genetically Modified Food Supply Chains: Co-Existence and Traceability*. Ed. by Y. Bertheau. Oxford, UK: Wiley-Blackwell. Chap. 19. ISBN: 9781444337785. DOI: 10.1002/9781118373781.ch19 (pp. 20, 26, 62).

Prins, T. W., J. P. van Dijk, H. G. Beenen, A. M. A. Van Hoef, M. M. Voorhuijzen, C. D. Schoen, H. J. Aarts, and E. J. Kok (2008). "Optimised padlock probe ligation and microarray detection of multiple (non-authorised) GMOs in a single reaction". *BMC Genomics* 9, 584. DOI: 10.1186/1471-2164-9-584 (pp. 26, 30, 62).

Qian, C. (2013). "Qualitative and Quantitative PCR Detection of Two Transgenic Rice". Master's thesis. Shenyang Agricultural University. URL: http://www.dissertationtopic.net/doc/1666803 (pp. 97, 103).

Querci, M., M. Van den Bulcke, J. Žel, G. Van den Eede, and H. Broll (2010). "New approaches in GMO detection". *Analytical and Bioanalytical Chemistry* 396.6, pp. 1991–2002. DOI: 10.1007/s00216-009-3237-3 (pp. 12, 52).

Quirasco, M., B. Schoel, P. Chhalliyil, J. Fagan, and A. Gálvez (2008). "Real-time and conventional PCR detection of Liberty Link® rice varieties and transgenic soy in rice sampled in the Mexican and American retail markets". *Analytical and Bioanalytical Chemistry* 392.3, pp. 395–404. DOI: 10.1007/s00216-008-2265-8 (pp. 92, 96, 102).

Quynh, T. (2010). *Tram's kochen*. URL: http://tramkochen.blogspot.be/2010/04/banh-canh-bot-gao-cach-2.html (p. 133).

Rai, M. (2006). "Refinement of the Citrus tristeza virus resistance gene (*Ctv*) positional map in *Poncirus trifoliata* and generation of transgenic grapefruit (*Citrus paradisi*) plant lines with candidate resistance genes in this region". *Plant Molecular Biology* 61.3, pp. 399–414. DOI: 10.1007/s11103-006-0018-7 (p. 37).

Randhawa, G. J., M. Singh, D. Morisset, P. Sood, and J. Žel (2013). "Loop-mediated isothermal amplification: rapid visual and real-time methods for detection of genetically modified crops". *Journal of Agricultural and Food Chemistry* 61.47, pp. 11338–11346. DOI: 10.1021/jf4030085 (pp. 36, 38, 63).

Raymond, P., L. Gendron, M. Khalf, S. Paul, K. L. Dibley, S. Bhat, V. R. D. Xie, L. Partis, M.-E. Moreau, C. Dollard, M.-J. Coté, S. Laberge, and K. R. Emslie (2010). "Detection and identification of multiple genetically modified events using DNA insert fingerprinting". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2091–2102. DOI: 10.1007/s00216-009-3295-6 (pp. 37, 40, 42, 43, 111).

Reiting, R., H. Broll, H.-U. Waiblinger, and L. Grohmann (2007). "Collaborative study of a T-nos real-time PCR method for screening of genetically modified organisms in food products". *Journal für Verbraucherschutz und Lebensmittelsicherheit* 2.2, pp. 116–121. DOI: 10.1007/s00003-007-0189-4 (p. 121).

Reiting, R., L. Grohmann, and D. Mäde (2010). "A testing cascade for the detection of genetically modified rice by real-time PCR in food and its application for detection of an unauthorized rice line similar to KeFeng6". *Journal für Verbraucherschutz und Lebensmittelsicherheit* 5.2, pp. 185–188. DOI: 10.1007/s00003-010-0573-3 (pp. 110, 150).

Reiting, R., L. Grohmann, G. Moris, and D. Mäde (2013). "Detection and characterization of an unknown rice event in Basmati rice products". *European Food Research and Technology* 236.4, pp. 715–723. DOI: 10.1007/s00217-013-1928-7 (pp. 96, 102).

Rhoads, A. and K. F. Au (2015). "PacBio sequencing and its applications". *Genomics, Proteomics & Bioinformatics* 13.5, pp. 278–289. DOI: 10.1016/j.gpb.2015.08.002 (pp. 70, 71).

Ricroch, A. E., A. Boisron, and M. Kuntz (2014). "Looking back at safety assessment of GM food/feed: an exhaustive review of 90-day animal feeding

studies". *International Journal of Biotechnology* 13.4, pp. 230–256. DOI: 10.1504/IJBT.2014.068940 (p. 6).

Rieber, N., M. Zapatka, B. Lasitschka, D. Jones, P. Northcott, B. Hutter, N. Jäger, M. Kool, M. Taylor, P. Lichter, S. Pfister, S. Wolf, B. Brors, and R. Eils (2013). "Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies". *PLoS ONE* 8.6, e66621. DOI: 10.1371/journal.pone.0066621 (p. 205).

Rogan, G. J., Y. A. Dudin, T. C. Lee, K. M. Magin, J. D. Astwood, N. S. Bhakta, J. N. Leach, P. R. Sanders, and R. L. Fluchs (1999). "Immunodiagnostic methods for detection of 5-enolpyruvylshikimate-3-phosphate synthase in Roundup Ready® soybeans". *Food Control* 10.6, pp. 407–414. DOI: 10.1016/S0956-7135(99)00083-3 (pp. 9, 10).

Rønning, S. B., M. Vaïtilingom, K. G. Berdal, and A. Holst-Jensen (2003). "Event specific real-time quantitative PCR for genetically modified Bt11 maize (*Zea mays*)". *European Food Research and Technology* 216.4, pp. 347–354. DOI: 10.1007/s00217-002-0653-4 (pp. 37, 42, 111).

Rudenko, G. N., C. M. T. Rommens, H. J. J. Nijkamp, and J. Hille (1993). "Supported PCR: an efficient procedure to amplify sequences flanking a known DNA segment". *Plant Molecular Biology* 21.4, pp. 723–728. DOI: 10.1007/BF00014557 (p. 37).

Rudi, K., I. Rud, and A. Holck (2003). "A novel multiplex quantitative DNA array based PCR (MQDA-PCR) for quantification of transgenic maize in food and feed". *Nucleic Acids Research* 31.11, e62. DOI: 10.1093/nar/gng061 (pp. 26, 29, 30, 62).

Ruffalo, M., T. LaFramboise, and M. Koyutürk (2011). "Comparative analysis of algorithms for next-generation sequencing read alignment". *Bioinformatics* 27.10, pp. 2790–2796. DOI: 10.1093/bioinformatics/btr477 (p. 211).

Ruttink, T., R. Demeyer, E. Van Gulck, B. Van Droogenbroeck, M. Querci, I. Taverniers, and M. De Loose (2010a). "Molecular toolbox for the identification of unknown genetically modified organisms". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2073–2089. DOI: 10.1007/s00216-009-3287-6 (pp. 37, 40, 111, 132, 171, 191).

Ruttink, T., D. Morisset, B. Van Droogenbroeck, N. Lavrač, G. L. M. Van Den Eede, J. Žel, and M. De Loose (2010b). "Knowledge-technology-based discovery of unauthorized genetically modified organisms". *Analytical and Bioanalytical Chemistry* 396.6, pp. 1951–1959. DOI: 10.1007/s00216-009-3218-6 (pp. 81, 86).

Saika, H., A. Oikawa, F. Matsuda, H. Onodera, S. Kazuki, and S. Toki (2011). "Application of gene targeting to designed mutation breeding of high-tryptophan rice". *Plant Physiology* 156.3, pp. 1269–1277. DOI: 10.1104/pp.111.175778 (p. 95).

Sallaud, C., D. Meynard, J. van Boxtel, C. Gay, M. Bès, J. P. Brizard, P. Larmande, D. Ortega, M. Raynal, M. Portefaix, P. B. F. Ouwerkerk, S. Rueb, M. Delseny, and E. Guiderdoni (2003). "Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics". *Theoretical and Applied Genetics* 106.8, pp. 1396–1408. DOI: 10.1007/s00122-002-1184-x (p. 37).

Sallaud, C., C. Gay, P. Larmande, M. Bès, P. Piffanelli, B. Piégu, G. Droc, F. Regad, E. Bourgeois, D. Meynard, C. Périn, X. Sabau, A. Ghesquière, J. C. Glaszmann, M. Delseny, and E. Guiderdoni (2004). "High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics". *The Plant Journal* 39, pp. 450–464. DOI: 10.1111/j.1365-313X.2004.02145.x (p. 37).

Salvo-Garrido, H., S. Travella, L. J. Bilham, W. A. Harwood, and J. W. Snape (2004). "The distribution of transgene insertion sites in barley determined by physical and genetic mapping". *Genetics* 167.3, pp. 1371–1379. DOI: 10.1534/genetics.103.023747 (p. 37).

Sambrook, J. F. and D. W. Russell (2001). *Molecular cloning, a laboratory manual*. 3rd edition. New York, NY, USA: Cold Spring Harbor Laboratory Press. ISBN: 0879695773 (pp. 118, 120, 121, 165).

Samson, F., V. Brunaud, S. Balzergue, B. Dubreucq, L. Lepiniec, G. Pelletier, M. Caboche, and A. Lecharny (2002). "FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants". *Nucleic Acids Research* 30.1, pp. 94–97. DOI: 10.1093/nar/30.1.94 (p. 37).

Samson, M. C., M. Gullí, and N. Marmiroli (2013). "Multiplex real-time PCR assays for simultaneous detection of maize MON810 and GA21 in food samples". *Food Control* 30.2, pp. 518–525. DOI: 10.1016/j.foodcont.2012.08.001 (pp. 19, 23, 24, 53).

Sanger, F., S. Nicklen, and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors". *Proceedings of the National Academy of Sciences of the United States of America* 74.12, pp. 5463–5467 (p. 68).

Santiago-Felipe, S., L. A. Tortajada-Genaro, R. Puchades, and A. Maquieira (2014). "Recombinase polymerase and enzyme-linked immunosorbent assay as a DNA amplification-detection strategy for food analysis". *Analytica Chimica Acta* 811, pp. 81–87. DOI: 10.1016/j.aca.2013.12.017 (pp. 9, 10).

Santos, E., S. Remy, E. Thiry, S. Windelinckx, R. Swennen, and L. Sági (2009). "Characterization and isolation of a T-DNA tagged banana promoter active during *in vitro* culture and low temperature stress". *BMC Plant Biology* 9, 77. DOI: 10.1186/1471-2229-9-77 (p. 37).

SCBD (2015). *The Cartagena Protocol on Biosafety*. Secretariat of the Convention on Biological Diversity. URL: http://bch.cbd.int/protocol/ (p. 8).

SCBD (2016). *Biosafety Clearing-House*. Cartagena Protocol on Biosafety. URL: https://bch.cbd.int/ (pp. 1, 82).

Schadt, E. E., S. Turner, and A. Kasarskis (2010). "A window into third-generation sequencing". *Human Molecular Genetics* 19.R2, R227–R240. DOI: 10.1093/hmg/ddq416 (p. 70).

Schatz, M. C., J. Witkowski, and W. R. McCombie (2012). "Current challenges in *de novo* plant genome sequencing and assembly". *Genome Biology* 13, 243. DOI: 10.1186/gb-2012-13-4-243 (pp. 47, 65).

Schell, J., M. Van Montagu, M. Holsters, P. Zambryski, H. Joos, D. Inzé, L. Herrera-Estrella, A. Depicker, M. De Block, A. Caplan, P. Dhaese, E. Van Haute, J.-P. Hernalsteens, H. De Greve, J. Leemans, R. Deblaere, L. Willmitzer, J. Schröder, and L. Otten (1983). "Ti plasmids as experimental gene vectors for plants". *Advances in Gene Technology: Molecular Genetics of Plants and Animals*. Miami Winter Symposia. Ed. by K. Downey, R. W. Voellmy, F. Ahmad, and J. Schultz. Vol. 20. Miami, FL, USA: Academic Press, pp. 191–209. ISBN: 9780122214806. DOI: 10.1016/B978-0-12-221480-6.50017-7 (p. 2).

Schrader, C., A. Schielke, L. Ellerbroek, and R. Johne (2012). "PCR inhibitors - occurrence, properties and removal". *Journal of Applied Microbiology* 113.5, pp. 1014–1026. DOI: 10.1111/j.1365-2672.2012.05384.x (pp. 12, 53).

Seo, S., I. Mitsuhara, J. Feng, T. Iwai, M. Hasegawa, and Y. Ohashi (2011). "Cyanide, a coproduct of plant hormone ethylene biosynthesis, contributes to the resistance of rice to blast fungus". *Plant Physiology* 155.1, pp. 502–514. DOI: 10.1104/pp.110.162412 (p. 90).

Seruggia, D. and L. Montoliu (2015). "The new CRISPR-Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals". *Transgenic Research* 23.5, pp. 707–716. DOI: 10.1007/s11248-014-9823-y (p. 223).

Shah, J. M., R. Singh, and K. Veluthambi (2013). "Transgenic rice lines constitutively co-expressing *tlp-D*34 and *chi*11 display enhancement of sheath blight resistance". *Biologia Plantarum* 57.2, pp. 351–358. DOI: 10.1007/s10535-012-0291-z (p. 90).

Shan, G., S. K. Embrey, and B. W. Schafer (2007). "A highly specific enzyme-linked immunosorbent assay for the detection of Cry1Ac insecticidal crystal protein in transgenic WideStrike cotton". *Journal of Agricultural and Food Chemistry* 55.15, pp. 5974–5979. DOI: 10.1021/jf070664t (pp. 9, 10).

Shan, Q., Y. Wang, K. Chen, Z. Liang, J. Li, Y. Zhang, K. Zhang, J. Liu, D. F. Voytas, X. Zheng, Y. Zhang, and C. Gao (2013). "Rapid and efficient gene

modification in rice and *Brachypodium* using TALENs". *Molecular Plant* 6.4, pp. 1365–1368. DOI: 10.1093/mp/sss162 (p. 95).

Shan, Q., Y. Zhang, K. Chen, K. Zhang, and C. Gao (2015). "Creation of fragrant rice by targeted knockout of the *OsBADH2* gene using TALEN technology". *Plant Biotechnology Journal* 13.6, pp. 791–800. DOI: 10.1111/pbi.12312 (p. 95).

Shao, N., S.-M. Jiang, M. Zhang, J. Wang, S.-J. Guo, Y. Li, H.-W. Jiang, C.-X. Liu, D.-B. Zhang, L.-T. Yang, and S.-C. Tao (2014). "MACRO: a combined microchip-PCR and microarray system for high-throughput monitoring of genetically modified organisms". *Analytical Chemistry* 86.2, pp. 1269–1276. DOI: 10.1021/ac403630a (pp. 26, 62).

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing". *Nature Biotechnology* 26, pp. 1135–1145. DOI: 10.1038/nbt1486 (pp. 68, 69).

Shukla, V. K., Y. Doyon, J. C. Miller, R. C. DeKelver, E. A. Moehle, S. E. Worden, J. C. Mitchell, N. L. Arnold, S. Gopalan, X. Meng, V. M. Choi, J. M. Rock, Y.-Y. Wu, G. E. Katibah, G. Zhifang, D. McCaskill, M. A. Simpson, B. Blakeslee, S. A. Greenwalt, H. J. Butler, S. J. Hinkley, L. Zhang, E. J. Rebar, P. D. Gregory, and F. D. Urnov (2009). "Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases". *Nature* 459.7245, pp. 437–441. DOI: 10.1038/nature07992 (p. 222).

Shyamala, V. and G. F.-L. Ames (1989). "Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR". *Gene* 84.1, pp. 1–8. DOI: 0.1016/0378-1119(89)90132-7 (p. 110).

Sims, D., I. Sudbery, N. E. Ilott, A. Heger, and C. P. Ponting (2014). "Sequencing depth and coverage: key considerations in genomic analyses". *Nature Reviews Genetics* 15.2, pp. 121–132. DOI: 10.1038/nrg3642 (p. 193).

Sivamani, E., H. Huet, P. Shen, C. A. Ong, A. de Kochko, C. Fauquet, and R. N. Beachy (1999). "Rice plant (*Oryza sativa* L.) containing Rice tungro spherical virus (RTSV) coat protein transgenes are resistant to virus infection". *Molecular Breeding* 5.2, pp. 177–185. DOI: 10.1023/A:1009633816713 (p. 90).

Smith, D. S., P. W. Maxwell, and S. H. De Boer (2008). "Evaluation of neomycin phosphotransferase II and Cry3A as targets for the serological detection of transgenic potato tubers". *American Journal of Potato Research* 85.1, pp. 60–68. DOI: 10.1007/s12230-008-9007-5 (pp. 9, 10).

Song, Q., G. Wei, and G. Zhou (2014). "Analysis of genetically modified organisms by pyrosequencing on a portable photodiode-based bioluminescence sequencer". *Food Chemistry* 154, pp. 78–83. DOI: 10.1016/j.foodchem.2014.01.001 (pp. 41, 45, 64).

Spalinskas, R., M. Van den Bulcke, and A. Milcamps (2013a). "Efficient retrieval of recombinant sequences of GM plants by cauliflower mosaic virus 35S promoter-based bidirectional LT-RADE". *European Food Research and*

Technology 237.6, pp. 1025–1031. DOI: 10.1007/s00217-013-2078-7 (pp. 37, 40, 43, 104).

Spalinskas, R., M. Van den Bulcke, G. Van den Eede, and A. Milcamps (2013b). "LT-RADE: an efficient user-friendly genome walking method applied to the molecular characterization of the insertion site of genetically modified maize MON810 and rice LLRICE62". *Food Analytical Methods* 6.2, pp. 705–713. DOI: 10.1007/s12161-012-9438-y (pp. 37, 43, 104, 111, 131).

Spertini, D., C. Béliveau, and G. Bellemare (1999). "Screening of transgenic plants by amplification of unknown genomic DNA flanking T-DNA". *BioTechniques* 27.2, pp. 308–314 (p. 37).

Sridevi, G., C. Parameswari, N. Sabapathi, V. Raghupathy, and K. Veluthambi (2008). "Combined expression of chitinase and β-1,3-glucanase genes in *indica* rice (*Oryza sativa* L.) enhances resistance against *Rhizoctonia solani*". *Plant Science* 175.3, pp. 283–290. DOI: 10.1016/j.plantsci.2008.04.011 (p. 90).

Stave, J. W. (2002). "Protein immunoassay methods for detection of biotech crops: applications, limitations, and practical considerations". *Journal of AOAC International* 82.3, pp. 780–786 (pp. 9, 10).

Stein, A. J. and E. Rodríguez-Serezo (2009). *The global pipeline of new GM crops: implications of asynchronous approval for international trade*. EUR 23846 EN. European Commission Joint Research Centre. ISBN: 9789279126031. DOI: 10.2791/12087. URL: http://ftp.jrc.es/EURdoc/JRC51799.pdf (pp. 81, 109, 110, 126, 150).

Strain, M. C., S. M. Lada, T. Luong, S. E. Rought, S. Gianella, V. H. Terry, C. A. Spina, C. H. Woelk, and D. D. Richman (2013). "Highly precise measurement of HIV DNA by droplet digital PCR". *PLoS ONE* 8.4, e55943. DOI: 10.1371/journal.pone.0055943 (pp. 33, 63).

Strauss, S. H. and J. K. Sax (2016). "Ending event-based regulation of GMO crops". *Nature Biotechnology* 34, pp. 474–477. DOI: 10.1038/nbt.3541 (p. 223).

Su, C., J. Xie, X. Wang, and Y. Peng (2011). "Integrated structure and event-specific real-time detection of transgenic *cry1Ac/SCK* rice Kefeng 6". *European Food Research and Technology* 232.2, pp. 351–359. DOI: 10.1007/s00217-010-1399-z (pp. 37, 42, 89, 97, 103, 111).

Sudhakar, D., X. Fu, E. Stoger, S. Williams, J. Spence, D. P. Brown, M. Bharathi, J. A. Gatehouse, and P. Christou (1998). "Expression and immunolocalisation of the snowdrop lectin, GNA in transgenic rice plants". *Transgenic Research* 7.5, pp. 371–378. DOI: 10.1023/A:1008856703464 (p. 89).

Suzuki, K., L. Yang, and F. Takaiwa (2012). "Transgenic rice accumulating modified cedar pollen allergen Cry j 2 derivatives". *Journal of Bioscience and Bioengineering* 113.2, pp. 249–251. DOI: 10.1016/j.jbiosc.2011.10.005 (p. 90).

Székács, A., É. Lauber, E. Takács, and B. Darvas (2010). "Detection of Cry1Ab toxin in the leaves of *MON 810* transgenic maize". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2203–2211. DOI: 10.1007/s00216-009-3384-6 (pp. 9, 10).

Taheri, A., S. J. Robinson, I. Parkin, and M. Y. Gruber (2012). "Revised selection criteria for candidate restriction enzymes in genome walking". *PLoS ONE* 7.4, e35117. DOI: 10.1371/journal.pone.0035117 (p. 37).

Tan, G., Y. Gao, M. Shi, X. Zhang, S. He, Z. Chen, and C. An (2005). "SiteFinding-PCR: a simple and efficient PCR method for chromosome walking". *Nucleic Acids Research* 33.13, e122. DOI: 10.1093/nar/gni124 (p. 37).

Tan, T., Z. Jintao, and C. Chen (2011). "The impact of commercialization of GM rice in China". *American-Eurasian Journal of Agricultural & Environmental Sciences* 10.3, pp. 296–199 (p. 109).

Tan, G. Y., T. G. Nan, W. Gao, Q. X. Li, J. J. Cui, and B. M. Wang (2013). "Development of monoclonal antibody-based sensitive sandwich ELISA for the detection of antinutritional factor cowpea trypsin inhibitor". *Food Analytical Methods* 6.2, pp. 614–620. DOI: 10.1007/s12161-012-9459-6 (pp. 9, 10).

Tang, K., X. Sun, Q. Hu, A. Wu, C.-H. Lin, H.-J. Lin, R. M. Twyman, P. Christou, and T. Feng (2001). "Transgenic rice plants expressing the ferredoxin-like protein (AP1) from sweet pepper show enhanced resistance to *Xanthomonas oryzae* pv. *oryzae*". *Plant Science* 160.5, pp. 1035–1042. DOI: 10.1016/S0168-9452(01)00351-X (p. 90).

Tang, N., H. Zhang, X. Li, J. Xiao, and L. Xiong (2012). "Constitutive activation of transcription factor OsbZIP46 improves drought tolerance in rice". *Plant Physiology* 158.4, pp. 1755–1768. DOI: 10.1104/pp.111.190389 (p. 90).

Taverniers, I., P. Windels, M. Vaïtilingom, A. Milcamps, E. Van Bockstaele, G. Van den Eede, and M. De Loose (2005). "Event-specific plasmid standards and real-time PCR methods for transgenic Bt11, Bt176, and GA21 maize and transgenic GT73 canola". *Journal of Agricultural and Food Chemistry* 53.8, pp. 3041–3052. DOI: 10.1021/jf0483467 (pp. 37, 42, 111).

Terada, R., M. Nakajima, M. Isshiki, R. J. Okagaki, S. R. Wessler, and K. Shimamoto (2000). "Antisense *Waxy* genes with highly active promoters effectively suppress *Waxy* gene expression in transgenic rice". *Plant and Cell Physiology* 41.7, pp. 881–888. DOI: 10.1093/pcp/pcd008 (p. 90).

The Council Of The European Communities (1990). "Council Directive 90/220/EEC of 23 April 1990 on the deliberate release into the environment of genetically modified organisms". *Official Journal of the European Communities* 33.L117, pp. 15–27 (p. 93).

The European Commission (2011a). "Commission Regulation (EU) No 619/2011 of 24 June 2011 laying down the methods of sampling and analysis for the

official control of feed as regards presence of genetically modified material for which an authorisation procedure is pending or the authorisation of which has expired". *Official Journal of the European Union* 54.L166, pp. 9–15 (pp. 50, 52, 118, 131).

The European Commission (2011b). "Commission Implementing Decision No 2011/884/EU of 22 December 2011 on emergency measures regarding unauthorised genetically modified rice in rice products originating from China and repealing Decision 2008/289/EC". *Official Journal of the European Union* 54.L343, pp. 140–148 (pp. 82, 110).

The European Commission (2015a). *RASFF - Food and Feed Safety Alerts*. URL: http://ec.europa.eu/food/safety/rasff/index_en.htm (p. 81).

The European Commission (2015b). *RASFF Portal*. URL: https://webgate.ec.europa.eu/rasff-window/portal/?event=SearchForm&cleanSearch=1 (pp. 110, 126).

The European Commission (2016). *Genetically Modified Organisms. EU Register of authorised GMOs*. URL: http://ec.europa.eu/food/dyna/gm_register/index_en.cfm (pp. 6, 82, 92).

The European Parliament and The Council Of The European Union (2001). "Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC". *Official Journal of the European Communities* 44.L106, pp. 1–28 (pp. 4, 93, 131, 150).

The European Parliament and The Council Of The European Union (2003a). "Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed". *Official Journal of the European Union* 46.L268, pp. 1–23 (pp. 4, 93, 109, 131, 150, 191).

The European Parliament and The Council Of The European Union (2003b). "Regulation (EC) No 1830/2003 of the European Parliament and of the Council of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC". *Official Journal of the European Union* 46.L268, pp. 24–28 (pp. 4, 109, 131, 150, 191).

Theuns, I., P. Windels, S. De Buck, A. Depicker, E. Van Bockstaele, and M. De Loose (2002). "Identification and characterization of T-DNA inserts by T-DNA fingerprinting". *Euphytica* 123.1, pp. 75–84. DOI: 10.1023/A:1014415619527 (pp. 37, 111).

Thudi, M., Y. Li, S. A. Jackson, G. D. May, and R. K. Varshney (2012). "Current state-of-art of sequencing technologies for plant genomics research". *Briefings in Functional Genomics* 11.1, pp. 3–11. DOI: 10.1093/bfgp/elr045 (p. 69).

Tian, X., J. Zheng, S. Hu, and J. Yu (2006). "The rice mitochondrial genomes and their variations". *Plant Physiology* 140.2, pp. 401–410. DOI: 10.âĂŃ1104/âĂŃpp.âĂŃ105.âĂŃ070060 (p. 205).

Triglia, T., M. G. Peterson, and D. J. Kemp (1988). "A procedure for I amplification of DNA segments that lie outside the boundaries of known sequences". *Nucleic Acids Research* 16.16, 8186 (p. 110).

Trinh, Q., H. Shi, W. Xu, J. Hao, Y. Luo, and K. Huang (2012a). "Loop-linker PCR: an advanced PCR technique for genome walking". *IUBMB Life* 64.10, pp. 841–845. DOI: 10.1002/iub.1069 (pp. 37, 42, 43, 111).

Trinh, Q., W. Xu, H. Shi, Y. Luo, and K. Huang (2012b). "An A-T linker adapter polymerase chain reaction method for chromosome walking without restriction site cloning bias". *Analytical Biochemistry* 425.1, pp. 62–67. DOI: 10.1016/j.ab.2012.02.029 (p. 37).

Trinh, Q., P. Zhu, H. Shi, W. Xu, J. Hao, Y. Luo, and K. Huang (2014). "A-T linker adapter polymerase chain reaction for determining flanking sequences by rescuing inverse PCR or thermal asymmetric interlaced PCR products". *Analytical Biochemistry* 466, pp. 24–26. DOI: 10.1016/j.ab.2014.07.022 (p. 37).

Turkec, A., S. J. Lucas, B. Karacanli, A. Baykut, and H. Yuksel (2016). "Assessment of a direct hybridization microarray strategy for comprehensive monitoring of genetically modified organisms (GMOs)". *Food Chemistry* 194, pp. 399–409. DOI: 10.1016/j.foodchem.2015.08.030 (p. 62).

Twyman, R. M., E. Stöger, A. Kohli, T. Capell, and P. Christou (2002). "Selectable and Screenable Markers for Rice Transformation". *Testing for Genetic Manipulation in Plants*. Ed. by J. F. Jackson and H. F. Linskens. Vol. 22. Molecular Methods of Plant Analysis. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg. Chap. 1, pp. 1–17. ISBN: 9783642077302. DOI: 10.1007/978-3-662-04904-4_1 (p. 2).

U.S. Food and Drug Administration (FDA) (2015). *FDA Has Determined That the AquAdvantage Salmon is as Safe to Eat as Non-GE Salmon*. URL: http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm472487.htm (p. 223).

Vaïtilingom, M., H. Pijnenburg, F. Gendre, and P. Brignon (1999). "Real-time quantitative PCR detection of genetically modified Maximizer maize and Roundup Ready soybean in some representative foods". *Journal of Agricultural and Food Chemistry* 47.12, pp. 5261–5266. DOI: 10.1021/jf981208v (pp. 121, 174, 228).

Van den Bulcke, M., A. Lievens, E. Barbau-Piednoir, G. MbongoloMbella, N. Roosens, M. Sneyers, and L. A. Casi (2010). "A theoretical introduction to "Combinatory SYBR®Green qPCR Screening" a matrix-based approach for the detection of materials derived from genetically modified plants".

Analytical and Bioanalytical Chemistry 396.6, pp. 2113–2123. DOI: 10.1007/s00216-009-3286-7 (pp. 19, 52, 121, 174, 183).

Vega, E. D. and M. L. Marina (2014). "Characterization and study of transgenic cultivars by capillary and microchip electrophoresis". *International Journal of Molecular Sciences* 15.12, pp. 23851–23877. DOI: 10.3390/ijms151223851 (pp. 20, 62).

Verma, V., S. Sharma, S. V. Devi, S. Rajasubramaniam, and I. Dasgupta (2012). "Delay in virus accumulation and low virus transmission from transgenic rice plants expressing *Rice tungro spherical virus* RNA". *Virus Genes* 45.2, pp. 350–359. DOI: 10.1007/s11262-012-0787-9 (p. 90).

Volpicella, M., C. Leoni, A. Costanza, I. Fanizza, A. Placido, and L. R. Ceci (2012a). "Genome walking by next generation sequencing approaches". *Biology* 1.3, pp. 495–507. DOI: 10.3390/biology1030495 (p. 66).

Volpicella, M., C. Leoni, I. Fanizza, S. Rius, R. Gallerani, and L. R. Ceci (2012b). "Genome walking by Klenow polymerase." *Analytical Biochemistry* 430.2, pp. 200–202. DOI: 10.1016/j.ab.2012.08.008 (pp. 66, 110).

von Götz, F. (2010). "See what you eat – broad GMO screening with microarrays". *Analytical and Bioanalytical Chemistry* 396.6, pp. 1961–1967. DOI: 10.1007/s00216-009-3204-z (pp. 26, 62).

Wahler, D., L. Schauser, J. Bendiek, and L. Grohmann (2013). "Next-generation sequencing as a tool for detailed molecular characterisation of genomic insertions and flanking regions in genetically modified plants: a pilot study using a rice event unauthorised in the EU". *Food Analytical Methods* 6.6, pp. 1718–1727. DOI: 10.1007/s12161-013-9673-x (pp. 46, 47, 65, 104, 145, 192).

Waiblinger, H.-U., B. Ernst, A. Anderson, and K. Pietsch (2008). "Validation and collaborative study of a P35S and T-nos duplex real-time PCR screening method to detect genetically modified organisms in food products". *European Food Research and Technology* 226.5, pp. 1221–1228. DOI: 10.1007/s00217-007-0748-z (pp. 13, 19, 21, 53, 121, 191).

Waltz, E. (2016). "Gene-edited CRISPR mushroom escapes US regulation". *Nature* 532.7599, p. 293. DOI: 10.1038/nature.2016.19754. URL: http://www.nature.com/news/gene-edited-crispr-mushroom-escapes-us-regulation-1.19754 (p. 222).

Wang, Y. and S. Johnston (2007). "The status of GM rice R&D in China". *Nature Biotechnology* 25, pp. 717–718. DOI: 10.1038/nbt0707-717 (p. 81).

Wang, H., J. Fang, C. Liang, M. He, Q. Li, and C. Chu (2011a). "Computation-assisted SiteFinding-PCR for isolating flanking sequence tags in rice". *BioTechniques* 51.6, pp. 421–423 (p. 37).

Wang, W.-X., T.-H. Zhu, F.-X. Lai, and Q. Fu (2011b). "Event-specific qualitative and quantitative detection of transgenic rice Kefeng-6 by characterization of the transgene flanking sequence". *European Food Research*

and Technology 232.2, pp. 297–305. DOI: 10.1007/s00217-010-1389-1 (pp. 37, 40, 44, 81, 97, 103, 110, 111).

Wang, W.-x., T.-h. Zhu, F.-x. Lai, and Q. Fu (2012). "Event-specific qualitative and quantitative detection of transgenic rice Kefeng-8 by characterization of the transgene flanking sequence". *European Food Research and Technology* 234.3, pp. 477–484. DOI: 10.1007/s00217-011-1654-y (pp. 37, 42, 97, 103, 109, 111).

Wang, X., X. Chen, J. Xu, P. Wang, and W. Shena (2014). "Multiplex event-specific qualitative polymerase chain reaction for detecting three transgenic rice lines and application of a standard plasmid as a quantitative reference molecule". *Analytical Biochemistry* 464, pp. 1–8. DOI: 10.1016/j.ab.2014.07.004 (pp. 97, 102, 103).

Wang, C., R. Li, S. Quan, P. Shen, D. Zhang, J. Shi, and L. Yang (2015a). "GMO detection in food and feed through screening by visual loop-mediated isothermal amplification assays". *Analytical and Bioanalytical Chemistry* 407.16, pp. 4829–4834. DOI: 10.1007/s00216-015-8652-z (pp. 36, 38, 63).

Wang, M., Y. Liu, C. Zhang, J. Liu, X. Liu, L. Wang, W. Wang, H. Chen, C. Wei, X. Ye, X. Li, and J. Tu (2015b). "Gene editing by co-transformation of TALEN and chimeric RNA/DNA oligonucleotides on the rice *OsEPSPS* gene and the inheritance of mutations". *PLoS ONE* 10.4, e0122755. DOI: 10.1371/journal.pone.0122755 (p. 94).

Wendt, T., F. Doohan, and E. Mullins (2012). "Production of *Phytophthora infestans*-resistant potato (*Solanum tuberosum*) utilising *Ensifer adhaerens* OV14". *Transgenic Research* 21.3, pp. 567–578. DOI: 10.1007/s11248-011-9553-3 (p. 85).

Whitworth, K. M., R. R. R. Rowland, C. L. Ewen, B. R. Trible, M. A. Kerrigan, A. G. Cino-Ozuna, M. S. Samuel, J. E. Lightner, D. G. McLaren, A. J. Mileham, K. D. Wells, and R. S. Prather (2016). "Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus". *Nature Biotechnology* 34, pp. 20–22. DOI: 10.1038/nbt.3434 (p. 223).

Willems, S., M.-A. Fraiture, D. Deforce, S. C. J. De Keersmaecker, M. De Loose, T. Ruttink, P. Herman, F. Van Nieuwerburgh, and N. Roosens (2016). "Statistical framework for detection of genetically modified organisms based on next generation sequencing". *Food Chemistry* 192, pp. 788–798. DOI: 10.1016/j.foodchem.2015.07.074 (pp. 40, 46–49, 64, 65, 99, 104, 177, 187, 189, 221).

Windels, P., I. Taverniers, A. Depicker, E. Van Bockstaele, and M. De Loose (2001). "Characterisation of the Roundup Ready soybean insert". *European Food Research and Technology* 213.2, pp. 107–112. DOI: 10.1007/s002170100336 (pp. 37, 42, 111).

Windels, P., S. Bertrand, A. Depicker, W. Moens, E. Van Bockstaele, and M. De Loose (2003a). "Qualitative and event-specific PCR real-time detection

methods for StarLink maize". *European Food Research and Technology* 216.3, pp. 259–263. DOI: 10.1007/s00217-002-0652-5 (pp. 37, 42).

Windels, P., S. De Buck, E. Van Bockstaele, M. De Loose, and A. Depicker (2003b). "T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences". *Plant Physiology* 133.4, pp. 2061–2068. DOI: 10.1104/pp.103.027532 (pp. 37, 111).

Wolt, J. D., K. Wang, and B. Yang (2016). "The regulatory status of genome-edited crops". *Plant Biotechnology Journal* 14.2, pp. 510–518. DOI: 10.1111/pbi.12444 (pp. 94, 222).

Wu, G., Y. Wu, S. Nie, L. Zhang, L. Xiao, Y. Cao, and C. Lu (2010). "Real-time PCR method for detection of the transgenic rice event TT51-1". *Food Chemistry* 119.1, pp. 417–422. DOI: 10.1016/j.foodchem.2009.08.031 (pp. 97, 102).

Wu, H., W. Wu, Z. Chen, W. Wang, G. Zhou, T. Kajiyama, and H. Kambara (2011). "Highly sensitive pyrosequencing based on the capture of free adenosine 5' phosphosulfate with adenosine triphosphate sulfurylase". *Analytical Chemistry* 83.9, pp. 3600–3605. DOI: 10.1021/ac2000785 (pp. 41, 64).

Xia, H., B.-R. Lu, K. Xu, W. Wang, X. Yang, C. Yang, J. Luo, F. Lai, W. Ye, and Q. Fu (2011). "Enhanced yield performance of *Bt* rice under target-insect attacks: implications for field insect management". *Transgenic Research* 20.3, pp. 655–664. DOI: 10.1007/s11248-010-9449-7 (p. 109).

Xu, W., K. Huang, H. Zhao, and Y. Luo (2005). "Application of immunoaffinity column as cleanup tool for an enzyme linked immunosorbent assay of phosphinothricin-*N*-acetyltransferase detection in genetically modified maize and rape". *Journal of Agricultural and Food Chemistry* 53.11, pp. 4315–4321. DOI: 10.1021/jf050218a (pp. 9, 10).

Xu, W., K. Huang, Z. Liang, A. Deng, Y. Yuan, F. Guo, and Y. Luo (2009). "Application of stepwise ammonium sulfate precipitation as cleanup tool for an enzyme-linked immunosorbent assay of glyphosate oxidoreductase in genetically modified rape of gt73". *Journal of Food Biochemistry* 33.5, pp. 630–648. DOI: 10.1111/j.1745-4514.2009.00242.x (pp. 9, 10).

Xu, J., J. Cao, D. Cao, T. Zhao, X. Huang, P. Zhang, and F. Luan (2013a). "Flanking sequence determination and event-specific detection of genetically modified wheat B73-6-1". *Acta Biochimica et Biophysica Sinica* 45.5, pp. 416–421. DOI: 10.1093/abbs/gmt016 (pp. 37, 44, 150).

Xu, J., Q. Zheng, L. Yu, R. Liu, X. Zhao, G. Wang, Q. Wang, and J. Cao (2013b). "Loop-mediated isothermal amplification (LAMP) method for detection of genetically modified maize T25". *Food Science & Nutrition* 1.6, pp. 432–438. DOI: 10.1002/fsn3.68 (pp. 36, 39, 63).

Xu, C., L. Li, W. Jin, and Y. Wan (2014). "Recombinase polymerase amplification (RPA) of CaMV-35S promoter and *nos* terminator for rapid

detection of genetically modified crops". *International Journal of Molecular Sciences* 15.10, pp. 18197–18205. DOI: 10.3390/ijms151018197 (p. 104).

Yang, L., J. Chen, C. Huang, Y. Liu, S. Jia, L. Pan, and D. Zhang (2005a). "Validation of a cotton-specific gene, *Sad1*, used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic cottons". *Plant Cell Reports* 24.4, pp. 237–245. DOI: 10.1007/s00299-005-0929-9 (p. 121).

Yang, L., S. Xu, A. Pan, C. Yin, K. Zhang, Z. Wang, Z. Zhou, and D. Zhang (2005b). "Event specific qualitative and quantitative polymerase chain reaction detection of genetically modified MON863 maize based on the 5'-transgene integration sequence". *Journal of Agricultural and Food Chemistry* 53.24, pp. 9312–9318. DOI: 10.1021/jf051782o (pp. 37, 40, 44, 111, 158, 159, 161).

Yang, L., C. Wang, A. Holst-Jensen, D. Morisset, Y. Lin, and D. Zhang (2013). "Characterization of GM events by insert knowledge adapted re-sequencing approaches". *Scientific Reports* 3, article number 2839. DOI: 10.1038/srep02839 (pp. 46–48, 65, 89, 98, 104, 145, 192, 193, 200).

Yao, J.-L., S. Tomes, and A. P. Gleave (2013). "Transformation of apple (*Malus* × *domestica*) using mutants of apple *acetolactate synthase* as a selectable marker and analysis of the T-DNA integration sites". *Plant Cell Reports* 32.5, pp. 703–714. DOI: 10.1007/s00299-013-1404-7 (p. 37).

Yarasi, B., V. Sadumpati, C. P. Immanni, D. R. Vudem, and V. R. Khareedu (2008). "Transgenic rice expressing *Allium sativum* leaf agglutinin (ASAL) exhibits high-level resistance against major sap-sucking pests". *BMC Plant Biology* 8, 102. DOI: 10.1186/1471-2229-8-102 (p. 89).

Ye, X., S. Al-Babili, A. Klöti, J. Zhang, P. Lucca, P. Beyer, and I. Potrykus (2000). "Engineering the provitamin A (β-Carotene) biosynthetic pathway into (carotenoid-free) rice endosperm". *Science* 287.5451, pp. 303–305. DOI: 10.1126/science.287.5451.303 (p. 90).

Young, L., J. Hammerlindl, V. Babic, J. McLeod, A. Sharpe, C. Matsalla, F. Bekkaoui, L. Marquess, and H. M. Booker (2015). "Genetics, structure, and prevalence of FP967 (CDC Triffid) T-DNA in flax". *SpringerPlus* 4, 146. DOI: 10.1186/s40064-015-0923-9 (pp. 46, 47, 65).

Yu, S., W. Wang, and B. Wang (2012). "Recent progress of salinity tolerance research in plants". *Russian Journal of Genetics* 48.5, pp. 497–505. DOI: 10.1134/S1022795412050225 (pp. 109, 121).

Yuanxin, Y., A. Chengcai, L. Li, G. Jiayu, T. Guihong, and C. Zhangliang (2003). "T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends". *Nucleic Acids Research* 31.12, e68. DOI: 10.1093/nar/gng068 (p. 37).

Zahradnik, C., C. Kolm, R. Martzy, R. L. Mach, R. Krska, A. H. Farnleitner, and K. Brunner (2014). "Detection of the 35S promoter in transgenic maize via

various isothermal amplification techniques: a practical approach". *Analytical and Bioanalytical Chemistry* 406.27, pp. 6835–6842. DOI: 10.1007/s00216-014-7889-2 (pp. 36, 38, 63).

Zapiola, M. L., C. K. Campbell, M. D. Butler, and C. A. Mallory-Smith (2008). "Escape and establishment of transgenic glyphosate-resistant creeping bentgrass *Agrostis stolonifera* in Oregon, USA: a 4-year study". *Journal of Applied Ecology* 45.2, pp. 486–494. DOI: 10.1111/j.1365-2664.2007.01430.x (p. 110).

Zhang, X., Z. Yuan, X. Guo, J. Li, Z. Li, and Q. Wang (2008). "Expression of *Chlamydophila psittaci* MOMP heat-labile toxin B subunit fusion gene in transgenic rice". *Biologicals* 36.5, pp. 296–302. DOI: 10.1016/j.biologicals.2008.04.002 (p. 91).

Zhang, M., N. Huo, Y. Liu, Y. Qiu, J. Ao, F. Luan, Q. Li, and X. Gao (2012). "Event-specific detection of genetically modified wheat B73-6-1 based on the 3'-flanking sequence". *European Food Research and Technology* 235.6, pp. 1149–1159. DOI: 10.1007/s00217-012-1848-y (pp. 37, 42, 104, 131).

Zhang, M., Y. Liu, L. Chen, S. Quan, S. Jiang, D. Zhang, and L. Yang (2013a). "One simple DNA extraction device and its combination with modified visual loop-mediated isothermal amplification for rapid on-field detection of genetically modified organisms". *Analytical Chemistry* 85.1, pp. 75–82. DOI: 10.1021/ac301640p (pp. 36, 38, 39, 63, 64).

Zhang, P., J. Xu, Q. Zheng, F. Luan, J. Cao, and H. Hou (2013b). "Flanking sequence determination and event specific detection of transgenic wheat B72-8-11b strain". *Applied Biochemistry and Biotechnology* 169.5, pp. 1523–1530. DOI: 10.1007/s12010-012-9989-9 (pp. 37, 42).

Zhang, H., J. Zhang, P. Wei, B. Zhang, F. Gou, Z. Feng, Y. Mao, L. Yang, H. Zhang, N. Xu, and J.-K. Zhu (2014a). "The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation". *Plant Biotechnology Journal* 12.6, pp. 797–807. DOI: 10.1111/pbi.12200 (p. 95).

Zhang, L., H. Chen, Y. Li, Y. Li, S. Wang, J. Su, X. Liu, D. Chen, and X. Chen (2014b). "Evaluation of the agronomic performance of atrazine-tolerant transgenic *japonica* rice parental lines for utilization in hybrid seed production". *PLoS ONE* 9.10, e108569. DOI: 10.1371/journal.pone.0108569 (p. 89).

Zhang, M., Y. Yu, X. Gao, K. Zhang, F. Luan, Y. Zhu, and B. Qu (2015). "Event-specific quantitative detection of genetically modified wheat B72-8-11 based on the 3' flanking sequence". *European Food Research and Technology* 240.4, pp. 775–782. DOI: 10.1007/s00217-014-2383-9 (pp. 37, 42, 150).

Zhao, T., C.-y. Lin, and Z.-c. Shen (2011). "Development of transgenic glyphosate-resistant rice with G6 gene encoding 5-enolpyruvylshikimate-3-

phosphate synthase". *Agricultural Sciences in China* 10.9, pp. 1307–1312. DOI: 10.1016/S1671-2927(11)60123-5 (p. 89).

Zheng, S.-J., B. Henken, E. Sofiari, E. Jacobsen, F. A. Krens, and C. Kik (2001). "Molecular characterization of transgenic shallots (*Allium cepa* L.) by adaptor ligation PCR (AL-PCR) and sequencing of genomic DNA flanking T-DNA borders". *Transgenic Research* 10.3, pp. 237–245. DOI: 10.1023/A:1016633410041 (p. 37).

Zhou, L. and J. A. Holliday (2012). "Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture". *BMC Genomics* 13, 703. DOI: 10.1186/1471-2164-13-703 (pp. 47, 64).

Zhu, H., X. Zhao, J. Jia, J. Sun, and K. Zhao (2008). "A specific qualitative and real-time PCR detection of MON863 maize based on the 5'-transgene integration sequence". *Journal of Cereal Science* 48.3, pp. 592–597. DOI: 10.1016/j.jcs.2008.01.004 (pp. 158, 159, 161, 166).

Zhu, X., L. Chen, P. Shen, J. Jia, D. Zhang, and L. Yang (2011). "High sensitive detection of Cry1Ab protein using a quantum dot-based fluorescence-linked immunosorbent assay". *Journal of Agricultural and Food Chemistry* 59.6, pp. 2184–2189. DOI: 10.1021/jf104140t (pp. 9, 10).

Zimmermann, A., J. Lüthy, and U. Pauli (2000). "Event specific transgene detection in Bt11 corn by quantitative PCR at the integration site". *LWT - Food Science and Technology* 33.3, pp. 210–216. DOI: 10.1006/fstl.2000.0637 (pp. 37, 42).

Zuniga-Soto, E., E. Mullins, and B. Dedicova (2015). "Ensifer-mediated transformation: an efficient non-Agrobacterium protocol for the genetic modification of rice". *SpringerPlus* 4, 600. DOI: 10.1186/s40064-015-1369-9 (p. 85).

Curriculum vitae

Marie-Alice Fraiture was born on September 25th, 1987 in Bastogne, Belgium. In 2005 she enrolled at the Université Catholique de Louvain (UCL) where she obtained a bachelor's degree in Biology in 2009, followed by a master's degree in Biochemistry and Molecular and Cellular Biology in 2011. During her master, she also did an internship of four months at the Plant Sciences Division in the University of Nottingham (Great-Britain) in collaboration with Prof. Malcolm Bennett and Dr. Ute Voß. Since then she worked for three months in the familial nursery specialized in the production of Christmas tree. In May 2012 she started on her PhD project UGMmonitor in collaboration with Prof. Dieter Deforce (UGent), Prof. Marc De Loose (ILVO) and Dr. Nancy Roosens (WIV-ISP). Through this PhD project, she was first author (ORCID: 0000-0002-2527-1785) of seven peer-reviewed publications and two national publications as well as co-author of two peer-reviewed publications and one national publication. She also participated in one international congress and two national congresses.

List of publications

Publications in internationally peer-reviewed journals

Fraiture, M.-A., P. Herman, N. Papazova, M. De Loose, D. Deforce, T. Ruttink, and N. H. Roosens. "An integrated strategy combining DNA walking and NGS to detect GMO". *Food Chemistry*, submitted.

Fraiture, M.-A., N. H. C. Roosens, I. Taverniers, M. De Loose, D. Deforce, and P. Herman (2016). "Biotech Rice: current developments and future detection challenges in food and feed chain". *Trends in Food Science & Technology*, 52, pp. 66–79. DOI: 10.1016/j.tifs.2016.03.011.

Willems, S.* **M.-A. Fraiture***, D. Deforce*, S. C. J. De Keersmaecker, M. De Loose, T. Ruttink, P. Herman, F. Van Nieuwerburgh*, and N. H. Roosens* (2016). "Statistical framework for detection of genetically modified organisms based on next generation sequencing". *Food Chemistry* 192, pp. 788–798. DOI: 10.1016/j.foodchem.2015.07.074.

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2015). "Current and new approaches in GMO detection: challenges and solutions". *BioMed Research International*, article ID 392872. DOI: 10.1155/2015/392872.

Fraiture, M.-A., P. Herman, L. Lefèvre, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2015). "Integrated DNA walking system to characterize a broad spectrum of GMOs in food/feed matrices". *BMC Biotechnology* 15, 76. DOI: 10.1186/s12896-015-0191-3.

*Equally contributed

Broeders, S.*, **M.-A. Fraiture***, E. Vandermassen, M. Delvoye, E. Barbau-Piednoir, A. Lievens, and N. Roosens (2015). "New qualitative trait-specific SYBR®Green qPCR methods to expand the panel of GMO screening methods used in the CoSYPS". *European Food Research and Technology* 241.2, pp. 275–287. DOI: 10.1007/s00217-015-2454-6.

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, F. Van Nieuwerburgh, D. Deforce, and N. H. Roosens (2015). "Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products". *Food Chemistry* 173, pp. 1259–1265. DOI: 10.1016/j.foodchem.2014.09.148.

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens (2014). "An innovative and integrated approach based on DNA walking to identify unauthorised GMOs". *Food Chemistry* 147, pp. 60–69. DOI: 10.1016/j.foodchem.2013.09.112.

Publications in books

Fraiture, M.-A., S. Broeders, P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. Roosens (2016). "Advances in the identification of genetically modified foods". *Advances in Food Authenticity Testing*. Ed. by G. Downey. 1st edition. Woodhead Publishing Series in Food Science, Technology and Nutrition. Cambridge, UK: Woodhead Publishing. Chap. 21. ISBN: 9780081002209. Accepted.

Publications in other journals and reports

Fraiture, M.-A., P. Herman, N. Papazova, and N. Roosens (2015). "UGM characterization using DNA walking strategy". *LabInfo* No. 14, pp. 32–34.

Willems, S.*, **M.-A. Fraiture***, S. C. J. De Keersmaecker, and N. H. Roosens (2015). "Next generation sequencing to identify GMO in food and feed products". *LabInfo* No. 13, pp. 18–20.

Fraiture, M.-A., P. Herman, G. Berben, F. Debode, E. Janssen, I. Taverniers, E. M. De Loose, and N. Roosens (2013). "New techniques and methods developed by the Belgian NRL-GMO to identify unauthorized GMOs in the UGMMONITOR project". *LabInfo* No. 9, pp. 25–27.

Communications at national conferences

Fraiture, M.-A.. "Strategies to detect unauthorized GMO in the food and feed chain". *Operational Directorate Expertise, Service Provision and Customer Relations*. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, June 1, 2016.

Fraiture, M.-A.. "UGMMONITOR project". *Fourth UGMmonitor steering committee*. Brussels, Belgium, May 20, 2016.

Fraiture, M.-A.. "Development of a molecular biology platform and a database for the detection of UGMs in the food and feed chain in Europe". *Tuesday Seminar*. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, May 10, 2016.

Fraiture, M.-A.. "Detection of unauthorized GMO in the food and feed chain by using NGS". *ILVO Genomics Symposium*. Institute for Agricultural and Fisheries Research (ILVO). Melle, Belgium, April 21, 2016.

Fraiture, M.-A.. "Detection of unauthorized GMO in the food and feed chain". UGent. Ghent, Belgium, March 23, 2016.

Fraiture, M.-A.. "UGMMONITOR project". *Yearly T&V PhD Day*. Institute for Agricultural and Fisheries Research (ILVO). Melle, Belgium, November 19, 2015.

Fraiture, M.-A.. "NGS technologies for GMO detection". *Workshop on novel technologies for detection of allergens and related parameters in food*. Institute for Agricultural and Fisheries Research (ILVO). Merelbeke, Belgium, June 11, 2015.

Fraiture, M.-A.. "UGMMONITOR project". *Third UGMmonitor steering committee*. Brussels, Belgium, April 28, 2015.

Fraiture, M.-A.. "UGMMONITOR project". *Yearly T&V PhD Day*. Institute for Agricultural and Fisheries Research (ILVO). Melle, Belgium, December 1, 2014.

Fraiture, M.-A.. "Relevance of new high-tech tools in GMOs detection". *NRL-GMO communication group*. Brussels, Belgium, November 21, 2014.

Fraiture, M.-A.. "An innovative and integrated approach based on DNA walking to identify unauthorized GMOs: NGS adaptation?". *GA-genomics steering committee*. Institute for Agricultural and Fisheries Research (ILVO). Melle, Belgium, October 13, 2014.

Fraiture, M.-A.. "UGMMONITOR project". *Second UGMmonitor steering committee*. Brussels, Belgium, April 25, 2014.

Fraiture, M.-A.. "Development of a screening marker for the detection of unauthorised GMOs". *NRL-GMO Theoretical workshop*. Brussels, Belgium, November 28, 2013.

Fraiture, M.-A.. "An innovative integrated approach based on DNA walking to identify unauthorized GMOs". *Trends in Food Analysis VII*. Koninklijke Vlaamse Chemische Vereniging (KVCV), Sectie Voeding. Merelbeke, Belgium, September 19, 2013.

Fraiture, M.-A.. "Development of a molecular biology platform and a database for the detection of UGMs in the food and feed chain in Europe". *Tuesday Seminar*. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, May 28, 2013.

Fraiture, M.-A.. "UGMMONITOR project". Operational Directorate Expertise, Service Provision and Customer Relations. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, May 8, 2013.

Fraiture, M.-A.. "UGMMONITOR project". *First UGMmonitor steering committee*. Brussels, Belgium, April 22, 2013.

Fraiture, M.-A.. "UGMMONITOR project". *Chinese delegation at WIV-ISP*. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, May 14, 2013.

Fraiture, M.-A.. "UGMMONITOR project". *Labmeeting of the Platform Biotechnology and Molecular Biology*. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, November 5, 2012.

Fraiture, M.-A.. "UGMMONITOR project". *Tunisian delegation at WIV-ISP*. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, September 14, 2012.

Communications at international conferences

Fraiture, M.-A.. "An innovative integrated approach based on DNA walking to identify unauthorized GMOs". *19th ENGL Plenary Meeting*. European Network of GMO Laboratories (ENGL). Ispra, Italy, June 19, 2013.

Posters presented at national conferences

Haegeman, A., T. De Mulder, C. De Tender, **M.-A. Fraiture**, I. Kirov, G. Luypaert, F. van Parijs, E. Veeckman, C. Verwimp, S. Baeyen, J. Debode, E. De Keyser, J. De Riek, L. Devriese, K. Goossens, S. Maes, H. Muylle, N. Peiren, G. Rasschaert, I. Taverniers, E. Van Coillie, L. Vandaele, K. Van Laere, S. De Campeneere, M. De Loose, M. Heyndrickx, M. Maes, J. Robbens, I. Roldán-Ruiz, J. Van Huylenbroeck and T. Ruttink Tom (2015). "Towards population genomics and metagenomics in agriculture and marine environments". *BIG N2N Annual Symposium*. Bioinformatics Institutes Ghent from nucleotides to networks. Ghent, Belgium, May 21, 2015.

Abstracts presented at national conferences

Fraiture, M.-A., P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. Roosens (2013). "An innovative integrated approach based on DNA walking to identify unauthorized GMOs". *Trends in Food Analysis VII*. Koninklijke Vlaamse Chemische Vereniging (KVCV), Sectie Voeding. Merelbeke, Belgium, September 19, 2013.

Acknowledgements

I'm totally aware that a PhD thesis is not the work of only one person but the results of a good collaboration between several actors. Therefore, I would like to thank all these persons without whom none of this would have been possible.

First of all, my special thanks go to all my promoters, Prof. Dr. Dieter Deforce, Dr. Ir. Nancy Roosens and Prof. Dr. Marc De Loose, which gave me the great opportunity to work on this PhD project. It was a wonderful scientific and personal experience. Thanks a lot to all of you for your guidance, your support, your confidence, your precious time and your advices highly helpful to continuously improve my work. I have really appreciated to work with you.

Dieter, I'm very grateful that you have accepted to be my promoter, especially because I understand that it is not always simple to work with a PhD student which works not specifically on your research field as well as which works not physically in your laboratory. I would like also to thank you (and Inge) for the help provided regarding all the academic process.

Nancy, thank you very much for your encouragement, your energy and your enthusiasm. It is largely thanks to you that we did it! For me, you were the ideal mentor helping me to grow as a better scientist.

Marc, I am thankful for your valuable and constructive suggestions for positively evolving both this work and myself. Thank you very much also for your understanding.

Second, I would like to thank all the members of my jury, Prof. Dr. Tom Coenye, Prof. Dr. Sarah De Saeger, Prof. Dr. Ann Depicker, Dr. Ir. Esther Kok and Dr. Philippe Herman, for the time that they dedicated to me as well as for your critical evaluation of this work. Thank you very much for your interesting questions, comments, advices and suggestions allowing to improve this work.

Third, millions thanks goes to all my colleagues and ex-colleagues (Sylvia, Sigrid, Nina, Philippe, Elodie, Véronique, Xavier, Aurélie, Armel, Els, Maud, Loïc,

Dirk, Marianne, Marjolein, Lies, Marieke, Bart, Tom, Isabel, Annique, Mieke, Jurgen, Cindy, Sam, Sander, Assia, Kevin, Ivy, Linda, Axel, Céline, Bert, Qiang, Alfredo, Joost and Raf) for all the funny and great moments, the nice working atmosphere and the fruitful scientific discussions.

Finally, I am very grateful to my family and friends for their encouragement and support. In addition, Gauthier, thank you very much for your kindness, your help, your patience, you constant motivation, your encouragement, your support, your understanding and making my life colourful. Last but not least, I would like also to warmly thank two of my strongest supporters, Merlot and Dudu.

Thanks to all of you,
Marie-Alice

FACULTY OF PHARMACEUTICAL SCIENCES
DEPARTMENT OF PHARMACEUTICS
LABORATORY OF PHARMACEUTICAL BIOTECHNOLOGY
Ottergemsesteenweg 460
B-9000 GENT, BELGIUM
MarieAlice.Fraiture@UGent.be
www.ugent.be/fw