Development of multiplex PCR assays for the detection of genetically modified plants

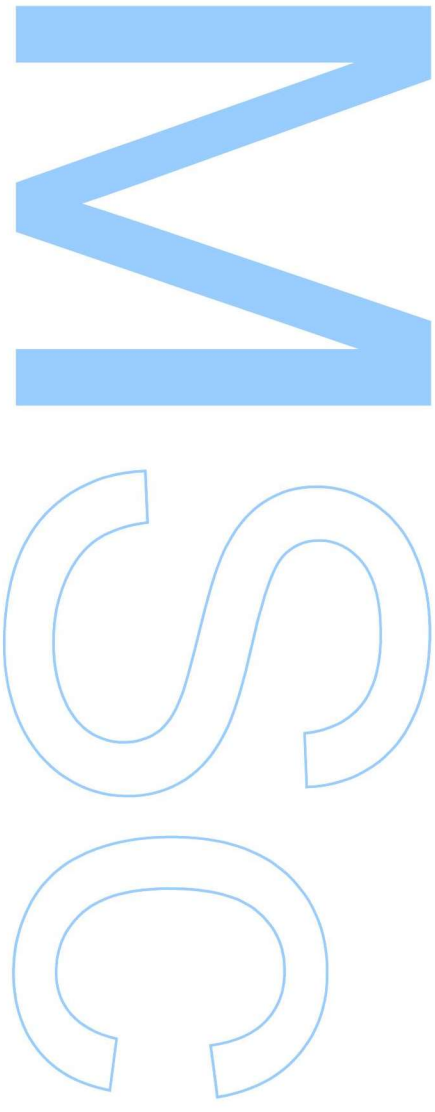
Filipa Isabel Menezes Moreira

Mestrado em Genética Forense
Departamento de Biologia
2015

Orientador
Doutor Filipe Pereira, Investigador,
CIIMAR - Centro Interdisciplinar de Investigação Marinha e Ambiental

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,

Porto, _____/_____/_____



Agradecimentos

Ao Filipe, meu orientador, por ter aceitado orientar uma das fases mais importantes do meu percurso académico e por ter sido uma pessoa incansável! Obrigada por toda a preocupação, todo o empenho, todos os ensinamentos, toda a orientação, todo o apoio, toda a paciência, todas as sugestões e principalmente por seres a ótima pessoa que és. Nunca me senti sozinha nesta luta, pois sabia que podia sempre contar contigo. A ciência ganha com pessoas como tu, que ajudam os outros com todo o entusiasmo e altruísmo.

Ao Prof. Doutor António Amorim, diretor do Mestrado em Genética Forense da Faculdade de Ciências da Universidade do Porto, por todo o trabalho realizado em benefício deste mestrado.

Ao corpo docente do Mestrado em Genética Forense por todos os ensinamentos dados que me foram úteis durante a realização desta dissertação, e que o serão certamente para o futuro.

Ao grupo de Genética Populacional do IPATIMUP, por me terem recebido e por estarem sempre prontos a ajudar. Obrigada também a alguns membros de outros grupos que ajudaram sempre que necessário.

Aos meus Pais, Manuel e Cândida, e ao meu irmão Pedro, por todo o apoio, não só ao longo desta etapa, mas ao longo da minha vida. Obrigada por terem ajudado a formar a pessoa que sou hoje e me fazerem sentir segura e feliz.

À restante família, por todo o apoio e ajuda sempre que mais necessitei.

Aos meus amigos, por todos os momentos de amizade, por todas as gargalhadas dadas e pelos momentos de incentivo.

Abstract

The terms genetically modified (GM) or transgenic plant describes an organism that contains a gene or genes that have been introduced artificially in the genome in order to improve product quality, pest resistance and agronomic traits. The number of cultivated GM crops has been growing in recent years despite the intense discussion about the benefits or damage that these organisms may have on humans and ecosystems. For this reason, there is great interest in developing effective methods for the identification of GM plants in different stages of the chain of cultivation, processing and distribution. The detection of a GM plant in a sample is carried out by targeting the most common genetic elements (promotors, protein-coding regions or terminators) used in GM constructs or inserts. The P35S and the T-nos elements are the typical targets of screening approaches, but a variety of new authorized and unauthorized transgenic plants have been developed without such elements. This work describes the development of multiplex PCR assays for detection of seven transgenic elements that cover almost all known GM plants: P35S, T-nos, bar, ctp2-cp4epsps, P35S-pat, Cry1Ab and FMV35S. We started by testing 20 target regions by singleplex PCR defined by 28 PCR primers collected from the literature or designed by us. The regions that were successfully amplified by singleplex PCR were then combined in five multiplex PCRs, two of them including a chloroplast DNA (cpDNA) *trn*L gene as internal control for plant DNA. The target regions were designed to yield amplicons with different lengths to be easily discriminated by conventional and capillary electrophoresis. Our multiplex PCRs successfully detected the presence of transgenic elements in 17 reference samples from GM plants. Our method has the advantage of targeting different elements simultaneously, avoiding false-negative results and decreasing the cost of the screening assay. The use of target regions of small length allows the analysis of forensic samples with degraded DNA, as the detection of illegal GMO crops or to verify the labelling of food products.

Keywords: genetically modified organisms, transgenic elements, multiplex PCR

Resumo

Os termos geneticamente modificado (GM) ou planta transgénica referem-se a um organismo que contém um gene ou genes introduzidos artificialmente no genoma, de forma a melhorar a qualidade do produto, conferir resistência a pesticidas e melhorar as características agronómicas. O número de plantas GM tem vindo a crescer nos últimos anos, apesar da intensa discussão acerca dos benefícios ou malefícios que estes organismos possam ter nos humanos e nos ecossistemas. Por esta razão, existe um grande interesse no desenvolvimento de métodos eficazes para a identificação de plantas GM nos diferentes estágios da cadeia de cultivo, processamento e distribuição. A deteção de uma planta transgénica é feita através da identificação dos elementos genéticos mais comuns (promotores, regiões codificadoras de proteínas e terminadores) inseridos no genoma da planta**.** Os elementos designados por P-35S e o T-Nos são os alvos típicos numa fase de screening, mas uma variedade de novas plantas transgénicas, autorizadas e não-autorizadas, foram desenvolvidas sem esses elementos. Este trabalho descreve o desenvolvimento de um PCR multiplex para a deteção de sete elementos transgénicos incluídos em quase todas as plantas GM: P35S, T-nos, bar, ctp2-cp4epsps, P35S-pat, Cry1Ab e FMV35S. Começamos por testar 20 regiões alvo por PCR singleplex, definidas por 28 primers recolhidos da literatura ou desenhados por nós. As regiões que foram amplificadas com sucesso por PCR singleplex, foram então combinadas em cinco PCR´s multiplex, dois deles incluindo um gene *trn*L do DNA de cloroplasto (cpDNA) como controlo interno para o DNA da planta. As regiões alvo foram desenhadas para abranger amplicões de diferentes tamanhos para facilmente serem discriminados através de uma eletroforese convencional ou capilar. Os nossos PCR´s multiplex detetaram a presença de elementos transgénicos em 17 amostras de referência de plantas transgénicas. O nosso método tem a vantagem de detectar simultaneamente diferentes regiões alvo, evitando falsos negativos e fazendo baixar o custo do ensaio de screening. O uso de regiões alvo de menor tamanho permite a análise de amostras forenses com DNA degradado, na deteção de cultivos ilegais de plantas GM ou para verificar a rotulagem dos produtos alimentares.

Palavras-chave: organismos geneticamente modificados, elementos transgénicos, PCR multiplex

Índice

Tables

Figures

Figure 13. Sequence alignment of the ctp2- cp4epsps element from sample GM05 (MON1445 cotton) and the reference sequence used by us. The sequencing revealed that sample GM05 has several polymorphism in relation to the reference sequence in the cp4epsps element that do not change the amino acid.

Figure 14. Screening of samples GM01 to GM11 with multiplex nº1 including the target regions for ctp2-cp4epsps (88 bp), T-nos (125), P35S (158), bar (176) and P35S-pat (208). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

Figure 15. Screening of samples GM01 to GM11 with multiplex nº2 including the target regions for P35S (82), P35S-pat (102), T-nos (125), bar (176) and ctp2-cp4epsps (200). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

Figure 16. Screening of samples GM01 to GM11 with multiplex nº 3 including the target regions for P35S (82), P35S-pat (102), T-nos (125), cp4epsps (145), bar (176) and ctp2-cp4epsps (200), Cry1Ab (226) and FMV35S (252). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

Figure 17. Chloroplast DNA (cpDNA) trnL [tRNA-Leu (UAA)] intron in maize and cotton. a) Agarose gel electrophoresis with *trn*L amplified products in three maize and three cotton samples. B) Multiple sequence alignment of *trn*L intron in three maize and three cotton samples. The grey regions indicate insertion/deletion polymorphic sites.

Figure 18. Screening of samples GM01 to GM07 and GM09 to GM11 with multiplex nº 6 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

Figure 19. Screening of samples GM12 to GM20 multiplex nº 6 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. PC – Positive control; NC - negative control; L - 100 bp DNA ladder.

Figure 20. Screening of samples GM01 to GM07 and GM09 to GM11 with multiplex nº 7 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230), FMV35S (252 bp) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

Figure 21. Screening of samples GM12 to GM20 with multiplex nº 7 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230), FMV35S (252 bp) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. PC - Positive Control; NC - negative control; L - 100 bp DNA ladder.

Figure 22. Discrimination of all target regions included in multiplexes 6 and 7 by conventional and automated capillary electrophoresis (QIAxcel system). LMW peak represents the 15-bp internal control on the capillary electrophoresis. A 100-bp DNA ladder was used on the agarose electrophoreses.

Abbreviations

adh1	*alcohol dehydrogenase 1* gene
bp	Base pair(s)
Bt	*Bacillus thuringiensis*
° C	Celsius degree
CaMV	*Cauliflower mosaic virus*
cpDNA	chloroplast DNA
ctp2	Chloroplast transit peptide
DNA	Deoxyribonucleic acid
ELISA	enzyme-linked immunosorbant assay
epsps	*5-enolpyruvulshikimate-3-phosphate synthase*
EU	European Union
FMV	*Figwort mosaic virus*
gDNA	genomic DNA
GM	Genetically modified
GMO	Genetically modified organism
indel	Insertion/deletion
mRNA	Messenger RNA
µg	Microgram
µL	Microliter
µM	Micromolar
mL	Milliliter
ng	Nanogram
NGS	Next generation sequencing
nt	Nucleotide
ORF	open-reading frame
P35S	35S promoter
pat	*Phosphinothricin N-acetyltransferase* gene
PCR	Polymerase chain reaction
rDNA	recombinant DNA
RNA	Ribonucleic acid
Tm	melting temperature
T-nos	nos terminator
***trn*L**	tRNA-Leu (UAA) intron
tRNA	Transfer RNA

Introduction

Genetically modified (GM) or transgenic plants are those that have undergone changes in their genetic structure by means of a biotechnological process of recombination, which allows the introduction of new features in a species (Parekh 2004; Žel *et al.* 2012). This process is referred to as the transformation event and involves DNA isolation from different organisms, modification of a defined DNA region and transfer of DNA into the genome of the target organism that becomes a GM organism or GMO (Miraglia *et al.* 2004).

GM plants have been developed in an attempt to improve food quality and solve plant disease and weed management problems associated with commercial agriculture (Table 1). The number of cultivated GM plants has been rising each year and are now widespread in many parts of the world (Peterson *et al.* 2000; Crawley *et al.* 2001). The growing area of GM crops have increased from 1.7 million hectares in 1996 to over 181.5 million hectares in 2014, representing a more than 100-fold increase since the commercialization began in 1996 (James 2014). The number of countries growing GM crops also increased from 6 in 1996 to 28 in 2014. More than 18 million farmers grew GM plants around the world, being 94.1% small and resource-poor farmers from developing countries. The country with the largest area of biotech crops in 2014 was the USA with over 73 million hectares of GM maize, soybean, cotton, canola, sugar beet, alfalfa, papaya and squash. The next countries in the rank of GM crop production are Brazil (42.2 million hectares) and Argentina (24.3 million hectares) due to the production of GM soybean, maize and cotton. Only five European Union (EU) countries planted GM crops (Spain, Portugal, Romania, Czech Republic and Slovakia), all of them growing only the *Bacillus thuringiensis* (Bt) maize. In Portugal, less than 50,000 hectares of Bt maize are produced per year (James 2014), mainly in Alentejo and Ribatejo (Brookes 2008).

The four principal crops grow all over the world are soybean, cotton, maize and canola. When considering the four crops together, nearly half (49%) of the planted area worldwide was occupied with GM plants (James 2014). The biotech crops had a positive impact on farm income by combining an enhanced productivity and efficiency gains (Brookes & Barfoot 2015). For example, an economic gains at the farm level of ~116.9 billion US dollars were generated globally by biotech crops from 1996 to 2012, of which 58% were due to reduced production costs such as less ploughing, fewer pesticide sprays and less labor, and 42% due to substantial yield gains of 377 million tons. GM plants are contributing to sustainability by reducing the use of pesticides and saving on fossil fuels, increasing efficiency of water usage and are a land-saving

technology due to the higher productivity helping to preclude deforestation and protect biodiversity in forests and in other habitats (Brookes & Barfoot 2015). GM crops also contribute to alleviate poverty and hunger in developing countries by increasing the income of small farmers (James 2013).

Table 1. Examples of the first GM plants commercialized in the 90s. Adapted from Gachet et al. 1998.

Crops	Property	Country	Company
Corn	Insect resistance	USA, Canada, Japan, EU	Ciba-Geigy, Monsanto, Mycogen, Sandoz, Northrup King
	Herbicide tolerance	USA, EU, Switzerland	de Kalb, AgrEvo, Plant Genetic Systems
	Male sterility	USA	Plant Genetic Systems
Soy bean	Herbicide tolerance	USA, Canada, Japan, EU, Argentina, Switzerland	AgrEvo, Monsanto
Tomato	Ripening slower	USA, GB	Agritope, Calgene, DNA Plant Technology, Monsanto, Zeneca
Potato	Insect resistance	USA, Canada, Japan	Monsanto
Papaya	Virus resistance	USA	Cornell U.
Rape seed	Male sterility	EU	Plant Genetic Systems
	Herbicide tolerance	Japan, Canada, USA	AgrEvo, Monsanto, Plant Genetic Systems
	High level of lauric acid	Canada	Calgene
	Composition of oil	USA	Calgene
Cotton	Insect resistance	USA, Mexico, Australia, Japan	Monsanto
	Herbicide tolerance	USA	Calgene/Rhone-Poulenc, Du Pont, Monsanto
Tobacco	Herbicide tolerance	EU	SEITA

For centuries, plant breeders converted wild plants into domesticated crops by pure breeding or hybridization with the aim of improving the yield and quality of crops and to provide crops with built-in protection against insects and diseases (Purugganan & Fuller 2009). These conventional processes are often inefficient, time consuming and can only be achieved within reproductively compatible species. The tools of modern biotechnology allow plant breeders to introduce new genes in a plant species in a more precise and selective manner. The new traits added by genetic engineering are often similar to those incorporated through conventional breeding methods. Several transformation methods have routinely been used for the introduction of target DNA into plant cells and its subsequent integration into the genome. The introduction of DNA into plant cells can be biological or physical (Parekh 2004). In biological methods, genetic sequences coding for proteins that result in diverse traits are introduced in plants by *Agrobacterium*-mediated transformation, using the natural ability of *Agrobacterium tumefaciens* to transfer synthetic DNA sequences containing genes of interest to plant cells (Gelvin 2003). Particle bombardment, microinjection, sonication

and electric current pulse are examples of physical methods of plant transformation (Gupta & Ram 2004).

The introduced transgenic sequence or recombinant DNA (rDNA) differentiates the GM plant from its non-GM counterpart at the DNA level and is used as a target for the DNA-based detection of the GM plants (Žel *et al.* 2012). The rDNA sequence is introduced in a unique place in the plant genome resulting in a so-called GM event. The genetic modification usually involves insertion of a synthetic combination of several DNA sequences (the insert or construct) into the genome of the organism to be modified. These DNA sequences are usually taken from other naturally occurring organisms. The construct is usually composed of three elements (Figure 1):

1) Gene promoter. The promoter sequences are located in the 5' region adjacent to the transcriptional start site of a gene, defining where transcription by RNA polymerase begins. It functions as an on/off switch for reading of the inserted/altered gene.
2) Coding-gene sequence or open-reading frame. The sequence of a gene coding for a specific selected feature to be expressed in the GM plant.
3) Gene terminator. The terminator elements are usually at the end of a gene and cause transcription to stop.

However, gene constructs often include other elements that control and stabilize the function of the gene or facilitate the combination of the various elements of the construct (Miraglia *et al.* 2004).

Figure 1. Schematic representation of a genetic construct used for plant transformation. The DNA sequence to be delivered to the plant genome usually includes the promoter, the terminator and the gene encoding the trait of interest.

The most successful traits that have been introduced into crop plants by genetic modification are herbicide tolerance and insect resistance. Herbicide tolerance is the dominant trait deployed in many species, such as maize, rice, cotton, canola, chicory, soybean and tobacco. The herbicide tolerant crops tolerate specific herbicides with a broad-spectrum activity killing the surrounding weeds. Plant species (e.g., maize, cotton, tomato and potato) with insect-resistance traits have been successfully developed, reducing the use of insecticides in the field. Other traits that have been

used include modification of oil content (canola, soybean), delayed ripening (tomato, melon), viral resistance (papaya, squash, potato), among others (Gupta & Ram 2004).

The most common genetic elements used in plant transformation are indicated in Table 2. The *Cauliflower mosaic virus* (*CaMV*) 35S promoter (P35S or CaMV35S) is the most widely used constitutive promoter in GM plants. The *CaMV* infects plants of the cabbage family by inserting a circular genome of about 8,000 bp into the host cells, which is then entirely transcribed to give a 35S RNA molecule or in part to give a 19S RNA molecule (Odell *et al.* 1985; Benfey & Chua 1990). The CaMV35S promoter is responsible for transcription of the entire viral genome, resulting in high levels of gene expression in dicotyledonous plants. It is less effective in monocotyledonous plants, especially in cereals. The *Figwort mosaic virus* 34S or 35S promoter (FMV35S) possesses a promotor activity similar to CaMV35S (Sanger *et al.* 1990) and has been used in several GM crops.

The most frequently used terminator in approved GM plants is the *nos* terminator derived from *Agrobacterium tumefaciens* (T-nos), isolated from the *nopaline synthase* gene. The terminator sequence ensures that the transgene expression is stopped and processed properly.

An example of a gene coding for a specific selected feature to be expressed in GM plants is *the glufosinate ammonium tolerance* gene (known as *bar* from bialaphos resistance), derived from the common soil bacterium *Streptomyces hygroscopicus* (Thompson et al. 1987). *S. hygroscopicus* and *Streptomyces viridochromogenes* produce a natural herbicide called bialaphos. The herbicide tolerance *bar* gene encode for an enzyme (phosphinothricin acetyltransferase or PAT) that convert herbicides containing glufosinate ammonium to a non-toxic compound. The glufosinate used in several herbicides (e.g., Basta®, Rely®, Finale® and Liberty®) inhibit an enzyme called glutamine synthetase, which is involved in the synthesis of glutamine and in ammonia detoxification. The glufosinate blocks the activity of the glutamine synthetase by mimicking the enzyme's substrate glutamate which is used to produce glutamine. The reduction in glutamine levels and the increase in ammonia concentration in plant tissues leads to cell membrane disruption and cessation of photosynthesis resulting in plant withering and death. The PAT enzyme acts by catalysing the acetylation of glufosinate (phosphinothricin) so that it no longer inhibits the enzyme glutamine synthase and, thus, eliminating its herbicidal activity.

The *pat* gene derived from a gram-positive spore-forming soil bacterium, *S. viridochromogenes,* also codes for the PAT enzyme and is used in GM constructs to confer tolerance to glufosinate-containing herbicides (Wohlleben *et al.* 1988; Wehrmann *et al.* 1996).

The herbicide resistance genes are often fused with a strong promotor in the transgenic constructs. For example, the junction of the CaMV P35S promoter and the synthetic *pat* gene are used in many GM crops (P35S-pat).

The *5-enolpyruvulshikimate-3-phosphate synthase* (*epsps*) gene from *A. tumefaciens* strain CP4 (also known as *cp4epsps*) encodes the EPSPS enzyme responsible for tolerance to glyphosate-containing herbicides (Pollegioni *et al.* 2011). The process of herbicide resistance of GM plants with the *epsps* is different from those GMO with *bar*/*pat* constructs. Glyphosate inactivates the enzyme EPSPS, which is part of the shikimate pathway involved in the biosynthesis of the aromatic amino acids tyrosine, phenylalanine and tryptophan, as well as other aromatic compounds (Maeda & Dudareva 2012). Plants treated with glyphosate herbicides die by lacking the aromatic amino acids. However, the *epsps* from *A. tumefaciens* encodes an EPSPS enzyme that is not affected by glyphosate, which allows GM plants to bypass the action of the herbicide, rather than breaking the herbicide down. GM plants containing the *epsps* gene allow the use glyphosate-containing herbicides for weed control, which will kill the weeds but not the transgenic plant (Pollegioni *et al.* 2011).

The coding region of *epsps* is inserted in GM plants fused with a chloroplast transit peptide (ctp) coding region (ctp2-cp4epsps) in order to target the EPSPS to the chloroplast, the site of aromatic amino acid biosynthesis (Bruce 2000). A pre-protein is imported into the chloroplasts where the ctp is cleaved and degraded, releasing the mature EPSPS protein. The ctp coding region from *Petunia hybrida* and the chloroplast transit peptide 2 (ctp2) from *Arabidopsis thaliana* are often used in transgenic constructs.

The *cry1Ab* gene (*cry* stands for crystal) codes for a delta-endotoxin known as "Bt-toxin" that confers resistance to larvae of lepidopteran insects, such as moths and butterflies (Evans 2004). The *cry1Ab* is derived from the widespread gram-positive soil bacterium *B. thuringiensis*, which produces proteins during sporogenesis that selectively affect particular groups of insects. Different versions of the protein classified into groups (CryI–CryIV) are produced by different strains of the bacterium. The Cry proteins bind to the specific sites of the cells in the insect digestive tract, forming pores that disrupt midgut ion flow, causing gut paralysis and ultimately leading to insect death (Gill *et al.* 1992). The *Cry* genes have now been introduced in maize, cotton, sugar beet and other crops, being generally referred to as Bt varieties.

Table 2. List of the most common genetic elements used in genetically modified (GM) plants.

Abbreviation	Genetic element	Type	Species of origin
P35S	Cauliflower Mosaic Virus (CaMV) 35S promoter	Promotor	Cauliflower mosaic virus
T-nos	*nos* terminator	Terminator	*Agrobacterium tumefaciens*
bar	*Glufosinate ammonium tolerance gene*	Open-reading frame	*Streptomyces hygroscopicus*
ctp2-cp4epsps	Chloroplast transit peptide (ctp2) + *5-enolpyruvulshikimate-3-phosphate synthase* gene (epsps)	Open-reading frame	*Arabidopsis thaliana* (ctp2) + *A. tumefaciens strain CP4 (epsps)*
P35S-pat	CaMV P-35S promoter + synthetic Phosphinothricin N-acetyltransferase gene (pat)	Promotor + open-reading frame	Cauliflower Mosaic Virus + *Streptomyces viridochromogenes*
Cry1Ab	delta-endotoxin gene	Open-reading frame	*Bacillus thuringiensis*
FMV35S	Figwort mosaic virus 35S promoter	Promotor	Figwort mosaic virus

Detection of a GM plant can be achieved by detecting a DNA, RNA or a protein that is related with the genetic modification introduced in the GMO of interest (Žel *et al.* 2012). It is possible to use the recombinant proteins produced by the GM plant as detection targets by immunological methods or other techniques (Ahmed 2002). The most common protein-based assays are immunoassays where the target proteins are detected by specific antibodies coupled to a colorimetric detection system. For example, western blot and enzyme-linked immunosorbant assay (ELISA) techniques have been used for GM detection (Stave 1999). However, detection methods targeting proteins present some limitations compared to DNA-based detection methods. For example, the recombinant protein may not be expressed in certain plant tissues, most proteins are more unstable than DNA and there is a high cost associated with the development of specific antibodies (Ahmed 2002). Methods based on the detection of RNA are also rarely used for GMO detection. The RNA may not be present in all tissues and has a high propensity to degradation by ribonucleases. In most techniques, the RNA has to be converted to complementary DNA by reverse transcription, which increases the cost of the analysis. Therefore, most methods for GMO detection rely on the analyses of DNA, which has the advantage of being able to be amplified by polymerase chain reaction (PCR), which is more sensitive than conventional protein detection tests. Moreover, detection of amplified PCR products (i.e., amplicons) can be done using conventional molecular genetic techniques. DNA is also a very stable molecule that can be recovered from degraded biological products, such as processed food products. Finally, studying the DNA allows the identification of the precise GM event since the genetic modifications are done at the DNA level (Gachet *et al.* 1998).

The most commonly used DNA-based methodology for GMO testing is the PCR (Gachet *et al.* 1998; Holst-Jensen *et al.* 2003). The PCR has the advantage of allowing the amplification of small amounts of target DNA, and results can be obtained qualitatively or quantitatively. The detection and identification of the PCR amplified targets can be done using conventional gel electrophoresis, capillary gel electrophoresis [e.g., (Nadal *et al.* 2006; Heide *et al.* 2008)], array hybridization [e.g., (Morisset *et al.* 2008a; Hamels *et al.* 2009)], electrochemical sensors [e.g., (Kumar & Kang 2007; Sun *et al.* 2008)], among others. Alternatives to PCR have also been proposed, including isothermal amplification, direct detection of genomic DNA by electrochemical sensors, cDNA analysis by microarray and direct hybridization of genomic DNA to microarrays (Morisset *et al.* 2008b; Holst-Jensen 2009). The quantification of DNA targets is usually done by real-time PCR, where the copy number of the transgenic element detected is correlated to a common plant marker, allowing

the determination of the percentage of the GMO in the sample tested (Cankar *et al.* 2006; Gašparič *et al.* 2010).

Both labelling and traceability of GM plants are important issues that are considered in trade and regulation. The EU legislation on GMOs has the main objective to protect health and environment, so the application of GMO technology is strictly regulated in the EU (Lynch & Vogel 2001). Submission and validation of GMO detection methods are an integral part of the EU regulatory approval process for GMOs and only GMOs that have undergone appropriate assessments of the risks to health and the environment can enter the market. Methods based on the PCR are mandatory in the European Union (EU) for GMO testing. Traceability systems document the history of a product and may serve the purpose of both marketing and health protection. Regarding the labelling of food products, the threshold level below which GMO labeling in the EU is not required is currently 0.9%. The control of the labelling of foodstuffs is based on the detection of the foreign DNA sequences and follows a strict regulation. If the content is above the threshold, the product has to be labeled. The mandatory-labeling legislation allow consumers to have a choice in selecting the foods they feel comfortable with. The correct detection of GM materials is of forensic relevance not only due to strict legislation regarding the labelling of food products but also due to the type of materials from which DNA has to be extracted. For example, transgenic constructs have to be identified in DNA extracted from products like corn germ, flour, pasta, corn flakes, cookies, baked products, sugars derived from corn starch, soy cream or milk (liquid or lyophilized), tofu, meat products, lecithin and even oil. It is therefore necessary to develop DNA-based detection kits able to detect low quality and/or quantity DNA samples. Moreover, the sample size of the material to be analyzed must be sufficient to ensure a statistically representative sample (Ahmed 2002). Therefore, the implementation of legal regulations requires appropriate sampling protocols and analytical methodologies that allow for accurate determination of the content of GM material within a food and feed sample.

Testing of GM plants is usually performed in a stepwise manner, including different phases: screening of transgenic elements including or not the use of construct-specific methods, the use of event-specific elements for a definite identification of the GM plant and quantification of the GM to define whether its content is above the labeling threshold set in legislation. The testing usually starts with the screening of a part of the transgenic sequence that is present in many GMOs, such as regulatory sequences of promoters or protein-coding sequences (Žel *et al.* 2012). The aim of this preliminary screening for the presence or absence of a GMO is to cover as many GMOs as possible, while at the same time reducing the cost of analysis.

Construct-specific methods that usually span over two or more genetic elements, such as a promoter and a protein-coding gene sequences, can also be used. However, methods designed specifically for a certain construct will only allow their detection, limiting the range of the method. If a screening result is negative, the analysis is concluded and no action on the sample tested is required. The identification of the individual GMOs in the sample is performed with event-specific methods that target the nucleotide sequence at the junction between the host genome and the transgenic sequence. However, event-specific methods only allow the detection of well characterized GMOs, and transgenic plants cannot be identified through them, since the transgenic construct is randomly inserted in any part of the genome. An extensive list of event-specific PCR methods can be found in previous works (Holst-Jensen 2009; Žel et al. 2012).

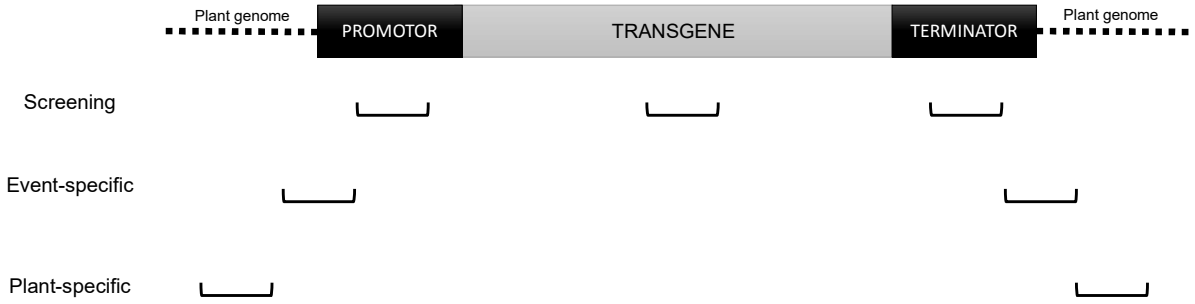


Figure 2. Different approaches for the detection of genetically modified organisms.

The detection of the transgenic elements for the screening step can be performed individually by singleplex PCR, but this process can be time consuming and expensive due to the need of performing several independent reactions. One of the possibilities for improving efficiency of GMO diagnostics is to analyze several targets simultaneously. Multiplex PCR is a variant of singleplex PCR in which more than one target sequence can be amplified by including more than one pair of primers in the same reaction (Edwards & Gibbs 1994). Multiplex PCR has the potential to produce considerable savings of time and effort in the laboratory by simultaneously amplifying multiple sequences in a single reaction, without compromising the results. Multiplex PCR has been successfully applied in many areas of species identification and has been also successfully applied in molecular diagnosis of pathogenic bacteria, fungi and viruses (Henegariu *et al.* 1997; Elnifro *et al.* 2000). The use of multiplex for detection of GM plants is usually associated with real-time PCR. For example, a duplex real-time PCR screening method for the detection of P35S, T-nos was validated in an inter-

laboratory collaborative study (Waiblinger *et al.* 2008). A pentaplex real-time PCR screening assays for P35S, T-nos, ctp2-cp4epsps, bar, and pat was also recently developed (Huber *et al.* 2013). The major limitation of multiplex real-time PCRs is the restricted number of target regions that can be simultaneously analyzed in a single reaction. The maximum number of target regions that can be simultaneously analyzed by real-time PCR in a single reaction is currently six (Bahrdt *et al.* 2010). In order to overcome this limitation, real-time PCRs methods are run on 96-well plates with different pairs of primers and probes. In recent years, Next Generation Sequencing (NGS) has been developed allowing massive sequencing of DNA fragments. A statistical framework for detection of GM plants by NGS has been recently proposed (Willems *et al.* 2016). Although NGS has the advantage of not requiring a priori knowledge of the transgene sequence, it still lacks the necessary precision to be applied to forensic casework (Bandelt & Salas 2012). The large size of plant genomes and the need of a bioinformatics support also pose problems for the implementation of NGS methods in GM testing.

Moreover, the detection of almost all GM plants developed to date only requires the screening of a few transgenic elements (P35S, T-nos, etc.), which are already well characterized. The use of expensive and time-consuming real-time PCR approaches can also be avoided when quantification is not required. At this regard, it has been recently shown that the detection of only five DNA target sequences (P35S, T-nos, bar, ctp2-cp4epsps and P35S-pat) can be used as a universal screening approach for at least 81 GM plant events (Waiblinger *et al.* 2010). The experimental verification of presence or absence of the screening elements was done by singleplex PCR. The work of Waiblinger et al. prove that it is not necessary to introduce new sequencing technologies in the routine detection of GMOs. Furthermore, a 'GMOseek matrix:' was recently developed to facilitate the selection of a minimum set of tests that can cover a maximum number of GMOs, reducing the number of event-specific tests to be performed. The GMOseek matrix is a tabulated database in which the rows represent the GMOs and the columns the targeted sequence elements (Block *et al.* 2013). Each GMO can be described by its specific combination of the respective target sequences.

In this dissertation we further extend the pioneer work of Waiblinger et al. by combining in multiplex PCR the five elements that identify most GMOs developed to date. Moreover, we used the GMOseek matrix to select and include additional genetic elements that are common in many GM plants, increasing the scope of the multiplex PCRs. Some of our multiplex PCRs were also designed to include a chloroplast DNA (cpDNA) target region as internal control for plant DNA. The target regions were

designed to yield amplicons with different lengths to be discriminated by conventional and capillary electrophoresis.

Objectives

Our main objective is to develop multiplex PCR assays to detect genetically modified (GM) plants. The assays will be designed for screening of the most common transgenic sequences used in GM plant events. The multiplex PCR will target multiple regions simultaneously in order to avoid false-negative results and reduce the cost of analyses. The assays will be optimized for both conventional and capillary gel electrophoresis.

In particular, our specific aims are:

1) Identify the genetic elements most used in transgenic plants to be targeted by screening assays.
2) Collect information from the literature and public databases on PCR primers for amplification of the selected transgenic elements used in GM plants.
3) Design new PCR primers in order to obtain amplicons with specific lengths that could be included in multiplex PCRs.
4) Develop an internal control for the presence of plant DNA to be used in multiplex PCRs.
5) Determine the efficiency of PCR primers by singleplex PCR in reference DNA samples.
6) Sequence the obtained amplicons of target regions to confirm the correct PCR amplification.
7) Combine the selected PCR primers in different multiplex PCR assays, including the internal control for plant DNA to rule out failure of amplification in the absence positive amplification.
8) Test the multiplex PCR assays in reference DNA samples by conventional and capillary gel electrophoresis.

Materials and Methods

Selection of genetic elements used in transgenic plants

We started by selecting the genetic elements most commonly used in genetically modified (GM) plants. Our selection includes five target sequences (Table 2) previously described as being used in at least 81 GM plant events (Waiblinger *et al.* 2010): Cauliflower mosaic virus (CaMV) 35S promoter (P35S), *nos* terminator derived from *Agrobacterium tumefaciens* (T-nos), *glufosinate ammonium tolerance* gene from *Streptomyces hygroscopicus* (bar), junction of the chloroplast-transitpeptide (ctp2) from *Arabidopsis thaliana* and the *5-enolpyruvulshikimate-3-phosphate synthase* (*epsps*) gene from *A. tumefaciens strain CP4* (ctp2-cp4epsps) and the junction of the CaMV P-35S promoter and the synthetic *phosphinothricin N-acetyltransferase* (*pat*) gene (P35S-pat). Primers for individual detection of the cp4epsps and pat elements with no junction to other elements were also included in some multiplex PCRs. We also include the gene for a delta-endotoxin from *Bacillus thuringiensis* (Cry1Ab) and the Figwort mosaic virus 35S promoter (FMV35S) in some of our multiplex PCRs (Table 2).

We then decided to select a reference sequence for each genetic element to facilitate the annotation of PCR primers and the validation of sequenced PCR products. The sequences of the transgenic elements were downloaded from the NCBI Entrez Nucleotide database (http://www.ncbi.nlm.nih.gov/) by searching with the name of the target region and the species (e.g., *nos* terminator AND *Agrobacterium tumefaciens*) or using accession numbers retrieved from published works (Table 3). All downloaded sequences from each element were aligned using the default parameters of the Muscle tool (Edgar 2004) implemented in Geneious Pro v5.3 software (Biomatters Ltd., Auckland, New Zealand). The alignments were manually curated to remove sequences with a short length, with nucleotide ambiguities or wrong annotations. The curated alignments were then realigned as previously described. The reference sequence to be used in each element was selected from the multiple sequence alignment taking into account different parameters (Table 3). For example, we used the complete reference genome sequences obtained from the NCBI's Reference Sequence (RefSeq) database identified with a 'NC_' accession prefix for the Cauliflower mosaic virus (for the P35S element) and Figwort mosaic virus (for the FMV35S element). The largest sequence with a clear annotation for the *bar* gene belonging to *S. hygroscopicus* was used as reference for the bar element. The *Oryza sativa* reference sequence for the T-nos element was selected according to Waiblinger *et al.* 2010. In the case of Cry1Ab, we

selected a sequence from a synthetic construct with a clearly annotated full-length Cry1Ab gene with the accession number AX392802.1. The reference for the P35S-pat junction was also obtained from a synthetic construct used in Waiblinger *et al.* 2010 and includes a 19 nt polylinker separating the 35S promoter from the *pat* gene. However, some P35S-pat constructs used in transgenic plants have larger sequences separating the two elements. For example, the Bt11 modified maize line has a 179 nt enhance element derived from alcohol dehydrogenase 1 (*adh1*) gene separating the two elements. In this case, we used the sequence of a synthetic construct (AX392802.1) as reference. The reference sequence for the ctp2-cp4epsps element was constructed by concatenating the following sequences: *Zea mays* transgenic partial *ctp2* and *cp4epsps* sequences to define the junction region (FN550387.1) and a 228 nt *ctp2* flanking region from the expression vector pMON100407 (JN400385.1) and the *Glycine max* transgenic *cp4epsps* gene (AB209952.1) to have larger flanking regions for primer design (Table 3).

Table 3. Reference sequences of the genetic elements used to design the multiplex PCRs.

Genetic element	Accession number	Source	Reference
P35S	NC_001497.1	Cauliflower mosaic virus	(Franck *et al.* 1980)
T-nos	EU880444.1	*Oryza sativa* Indica Group	(Waiblinger *et al.* 2010; Wu *et al.* 2010)
bar	JQ293091.1	*Streptomyces hygroscopicus*	GenBank direct submission
ctp2-cp4epsps	FN550387.1 + JN400385.1 + AB209952.1	Concatenated sequence	(Waiblinger *et al.* 2010; Preuss *et al.* 2012)
P35S-pat[1]	DL476427.1	Synthetic construct	(Narva 2008; Waiblinger *et al.* 2010)
P35S-pat[2]	AY629236.1	Synthetic construct	(Hernández *et al.* 2005)
Cry1Ab	AX392802.1	Synthetic construct	(Carozzi *et al.* 2002)
FMV35S	NC_003554.1	Figwort mosaic virus	(Richins *et al.* 1987)

[1] P35S-pat construct with a 19 nt polylinker separating the 35S promoter from the pat gene.
[2] P35S-pat construct with a 179 nt enhance element separating the 35S promoter from the pat gene.

Design of PCR primers

We started by collecting information on PCR primers for amplification of transgenic elements used in GM plants. We mainly focused our search on three publications describing methods for detection of transgenic plants, from which several primers were obtained (James *et al.* 2003; Waiblinger *et al.* 2010; Guo *et al.* 2011). In addition, primers were also retrieved from the EU Database of Reference Methods for GMO Analysis (http://gmo-crl.jrc.ec.europa.eu/gmomethods/) by searching for the

genetic element and selecting the most relevant methods according to the website (Table 4). Nevertheless, it was necessary to design new primers in order to obtain amplicons with specific lengths that could be included in multiplex PCRs. Several criteria were taken into account when designing the primers using the reference sequence selected for each element. For instance, primers were designed with predicted melting temperature (Tm) between 59 to 62 °C and avoiding potential self-annealing sites and hairpin formation as determined in the OligoCalc website (Kibbe 2007).

We also designed a pair of PCR primers to be used in some multiplex PCRs as internal control for the presence of plant DNA. The primers amplify a region of about 70 to 100 bp in the chloroplast DNA (cpDNA) *trn*L [tRNA-Leu (UAA)] intron located at positions 49,415 to 49,492 of the *Nicotiana tabacum* cpDNA reference sequence (NC_001879.2). We redesigned the primers previously described (Taberlet *et al.* 2007) and named “g” (three bases at the 3’ end were added) and “h” (one base at the 5’ end was deleted) in order to obtain a more balanced Tm between primers (Table 4).

The following nomenclature was implemented to designate all primers used in this study: ‘GM + abbreviation of the genetic element + 5’ primer position in reference sequence + F (forward primer) or R (reverse primer). For example, ‘GMtnos39F’ indicates a forward primer for amplification of region in the T-nos element starting at position 39 of the reference sequence (Table 4).

The predicted amplicons were selected to allow significant size differences for a clear electrophoretic separation. Potential cross-reactivity between all primer pairs was tested using the AutoDimer program (Vallone & Butler 2004). No interactions were found for the default Score of 7, which is determined by combining the number of Watson-Crick base pairs (+1) with mismatches (-1). We also set the Score to a lower value (4) and only a few weak interactions were found. Therefore, all primers were used in the multiplex PCRs.

Table 4. List of PCR primers selected for detection of the genetic elements used in GM plants. Only the primers with a code in the third column defining amplicons highlighted in blue were used in the present study.

Target	Original name[1]	Code for used primers	Primer sequence (5´- 3´)	Predicted amplicon length (nt)							Tm (°C)	Length (nt)
ctp2-cp4epsps	HANS- F	GMctp2_163F	GGGATGACGTTAATTGGCTCTG	88	249	200	417				62.1	22
	HANS- R	GMcp4_250R	GGCTGCTTGCACCGTGAAG								61.6	19
cp4epsps (alone)	CP4- GUO- F	GMcp4_266F	GCAAATCCTCTGGCCTTTCC	146							60.5	20
	CP4- GUO- R	GMcp4_334R	CTTGCCCGTATTGATGACGTC								61.2	21
	QL-ELE-00-019-F		GCATGCTTCACGGTGCAA	108							56.3	18
	QL-ELE-00-019-R		TGAAGGACCGGTGGGAGAT								59.5	19
	GMcp4_362R	GMcp4_362R	GTTTCACCGCTCGCGAGAC								61.6	19
	GMcp4_435F	GMcp4_435F	GATCCGTAAGGAAGGCGACA	145							60.5	20
	GMcp4_579R	GMcp4_579R	GCTGTCGAAATCGTAGACCC								60.5	20
T-nos	GUO2011-F	GMtnos39F	GAATCCTGTTGCCGGTCTTG	180		125					60.5	20
	GUO2011-R	GMtnos218R	TTATCCTAGTTTGCGCGCTA								56.4	20
	T-Nos HANS F		CATGTAATGCATGACGTTATTTATG	84							59.2	25
	T-Nos HANS R		TTGTTTTCTATCGCGTATTAAATGT								57.6	25
	PERMI2002- F2		GTAATGCATGACGTTATTTATGAGA		109						59.2	25
	PERMI2002- F		TTAAGATTGAATCCTGTTGCCG	192							58.4	22
	PERMI2002- R		TAATTTATCCTAGTTTGCGCGC								58.4	22
	James2003-R	GMtnos163R	GCGGGACTCTAATCATAAAAACC								60.9	23
	Xu2006-F		TGAATCCTGTTGCCGGTCTT	138							58.4	20
	Xu2006-R		AAATGTATAATTGCGGGACTCTAATC								61.7	26
	QL-ELE-00-009-F		GCATGACGTTATTTATGAGATGGG	118							62	24
	QL-ELE-00-009-R		GACACCGCGCGCGATAATTTATCC								66.9	24
	QL-ELE-00-018-F		GATTAGAGTCCCGCAATTATACATTTAA	69							62.7	28
	QL-ELE-00-018-R		TTATCCTAGTTTGCGCGCTATATTT								60.9	25
P35S	P-35S-F HANS	GMca35s7240F	GCCTCTGCCGACAGTGGT	82					158	275	60.8	18
	P-35S-R HANS	GMca35s7321R	AAGACGTGGTTGGAACGTCTTC								62.1	22
	QL-ELE-00-005-F		GCTCCTACAAATGCCATCA	195							55	19
	QL-ELE-00-005 -R		GATAGTGGGATTGTGCGTCA								58.4	20
	QL-ELE-00-017-F		AAAGCAAGTGGATTGATGTGATA	75	192						57.6	23
	QL-ELE-00-017-R	GMca35s7397R	GGGTCTTGCGAAGGATAGTG								60.5	20
	QT-ELE-00-001-F		CGTCTTCAAAGCAAGTGGATTG	79		199					60.1	22

	QT-ELE-00-001-R		TCTTGCGAAGGATAGTGGGATT								60.1	22
	QL-ELE-00-004-F		CCACGTCTTCAAAGCAAGTGG	123							61.2	21
	QL-ELE-00-004-R		TCCTCTCCAAATGAAATGAACTTCC								62.5	25
	Xu2006-F		AGACTGGCGAACAGTTCATACAGA	187							63.6	24
	Xu2006-R		CAATGGAATCCGAGGAGGT								57.5	19
							202					
P35S-pat	P35S-pat-F HANS (in 35S)	GM35s-pat1872F	AAGTTCATTTCATTTGGAGAGGACA	102							60.9	25
	P35S-pat-R HANS (in pat)	GM35s-pat1973R	CGGCCATATCAGCTGCTGTAG								63.2	21
								208				
pat (alone)	HUBER-F		CGCGGTTTGTGATATCGTTAAC		108						60.1	23
	QL-ELE-00-021F		CCGCGGTTTGTGATATCGTT	109							58.4	20
	QL-ELE-00-021R/HUBER-R	GMpat2079R	TCTTGCAACCTCTCTAGATCATCAA								62.5	25
	QT-ELE-00-002F (KVM-5)	GMpat2106F	TTGAGGGTGTTGTGGCTGGTA	68		303					61.2	21
	QT-ELE-00-002R (KVM-6)		TGTCCAATCGTAAGCGTTCCT								59.5	21
	Xu2006-F		GACAGAGCCACAAACACCACAA	144							62.1	22
	Xu2006-R		CAATCGTAAGCGTTCCTAGCCT								62.1	22
	PERMI2002- F		GAAGGCTAGGAACGCTTACG	263							60.5	20
	PERMI2002- R		GCCAAAAACCAACATCATGC								56.4	20
	GMpat2408R	GMpat2408R	GCCAAAAACCAACATCATGCC								59.5	21
bar	GMbar308F	GMbar308F	GCTCCACGCTCTACACCC			181					60.8	18
	Bar-F GUO	GMbar93F	AAGCACGGTCAACTTCCGTAC	396	176		230	300			61.2	21
	Bar-R GUO	GMbar448R	AAACCCACGTCATGCCAGTTC								61.2	21
	GMbar268R	GMbar268R	TCGACTCGGCCGTCCAGT								60.8	18
	GMbar322R	GMbar322R	TGTAGAGCGTGGAGCCCAG								61.6	19
	GMbar392R	GMbar392R	TCGTTGGGCAGCCCGATG								60.8	18
	HANS-F		ACAAGCACGGTCAACTTCC	60							57.5	19
	HANS-R		GAGGTCGTCCGTCCACTC								60.8	18
	Xu2006-F		GCTCCACGCTCTACACCCAC	181							64.6	20
	Xu2006-R		AAACCCACGTCATGCCAGTT								58.4	20
	QL-ELE-00-22F		CGTCAACCACTACATCGAGACAA	69							62.9	23
	QL-ELE-00-22R		GTCCACTCCTGCGGTTCCT								61.6	19
Cry1Ab	GUO- E-Cry-F		CCGCACCCTGAGCAGCAC	189							62.9	18

	GUO- E-Cry-R		GGTGGCACGTTGTTGTTCTGA								61.2	21
	JAMES2003F	GMcryIAB_3F	GGACAACAACCCCAACATCAAC	152	226						62.1	22
	JAMES2003R		GCACGAACTCGCTCAGCAG								61.6	19
	QT-ELE-00-003F		CCCATCGACATCAGCCTGAGC	129							65.3	21
	QT-ELE-00-003R		CAGGAAGGCGTCCCACTGGC								66.6	20
	GMcryIAB228R	GMcryIAB228R	GAAGGCGTCCCACTGGC								59.8	17
FMV35S	GUO-F	GMfig35S_6489F	TCAGTCCAAAGCCTCAACAAG	252							59.5	21
	GUO-R	GMfig35S_6740R	CCTAACAATTCTGCACCATTCC								60.1	22
	QL-ELE-00-015-F		CAAAATAACGTGGAAAAGAGCT	78							56.4	22
	QL-ELE-00-015-R		TCTTTTGTGGTCGTCACTGC								58.4	20
	QL-ELE-00-010-F		AAGCCTCAACAAGGTCAG	196							53.8	18
	QL-ELE-00-010-R		CTGCTCGATGTTGACAAG								53.8	18
	Xu2006-F		CAGCATTCCAGATTGGGTTCA	172							59.5	21
	Xu2006-R		CTTTTGGCTAATGGTTTGGAGAC								60.9	23
cpDNA *trn*L	trnLG_F	trnLG_F	GGGCAATCCTGAGCCAAATC	78							60.5	20
	trnLH_R	trnLH_R	CATTGAGTCTCTGCACCTATC								59.5	21

[1] The original name of the primer includes the first author of the publication where the primer was obtained (Permingeat *et al.* 2002; James *et al*. 2003; Xu *et al.* 2006; Waiblinger *et al.* 2010; Guo *et al.* 2011; Huber *et al.* 2013). The primers with a numeric code starting in 'QL' or ' QT' were retrieved from the EU Database of Reference Methods for GMO Analysis and those highlighted in red were designed by us.

Samples

Certified reference materials for GM plants were purchased from Eurofins GeneScan (Freiburg, Germany) and Sigma-Aldrich (Sintra, Portugal). The samples are described in Table 5. The control samples obtained from Eurofins GeneScan were all of genomic DNA (gDNA) extracted from 100% (w/w) GMO material, with the exception of the rice sample (GM08) that was derived from plasmid DNA.

The samples ordered from Sigma-Aldrich were of DNA or powder. The DNA samples of six GM events were part of two sets of DNA standards, one for maize (NK603, GA21 and CBH-3511 Starlink) and another for rapeseed or canola (GT73 Roundup Ready™, LibertyLink™ Falcon GS40/90 and MS8xRf3 Rapeseed). The DNA has been extracted from plant material as described in the official protocols of the Sigma-Aldrich company and tested in several validation studies. The extracted DNA was quantified and diluted to form a 1% solution.

The powder samples were produced and certified under the responsibility of the Institute for Reference Materials and Measurements of the European Commission's Joint Research Centre (EC-JRC-IRMM). The dried maize powder (GM12) has been produced from whole seeds of heterozygous 98140 maize of the variety TF0750C and non-modified maize of the variety X1083A, in a mixture representing 10% of GM maize. The dried potato powder (GM16) is composed of milled, dried powder from purely EH92-527-1 potatoes (100% GMO). The soya seed powder (GM17) has been produced from whole kernels of event 356043 soya and non-modified near-isogenic soya in a mixture representing 10% of GM soya (Table 5).

Genetic elements present in the GM samples

We identified the genetic elements introduced into the 20 GM samples using the GMOseek matrix (Block *et al.* 2013). The GMOseek database is a comprehensive, open-access tabulated database for the presence or absence of 247 genetic elements within an array of 328 GMO events. We searched the database using the GMO name or unique identifier for different classes of GMO elements, such as promoters, terminators, open-reading frame (ORF) segments, miscellaneous elements (e.g. vector elements, enhancers) and junctions between elements in the transgene constructs (Table 6).

Table 5. List of samples of certified genetically modified plants used in this work.

Sample code	Description	Catalogue Number	Source	GMO	Unique Identifier	Species	Type	Value (% GMO)	Amount per Unit
GM01	Bt176 corn	5211500701	Eurofins GeneScan	BT176	SYN-EV176-9	maize	genomic DNA	100	0.15 ml (100cp/µl)
GM02	Bt11 corn	5211501001	Eurofins GeneScan	BT11	SYN-BTØ11-1	maize	genomic DNA	100	0.15 ml (100cp/µl)
GM03	T25 corn	5211500801	Eurofins GeneScan	T25	ACS-ZMØØ3-2	maize	genomic DNA	100	0.15 ml (100cp/µl)
GM04	MON531 cotton	5211512301	Eurofins GeneScan	MON531	MON-ØØ531-6	cotton	genomic DNA	100	0.15 ml (100cp/µl)
GM05	MON1445 cotton	5211511301	Eurofins GeneScan	MON1445	MON-Ø1445-2	cotton	genomic DNA	100	0.15 ml (100cp/µl)
GM06	LLCotton25 cotton	5211512401	Eurofins GeneScan	LLCotton25	ACS-GHØØ1-3	cotton	genomic DNA	100	0.15 ml (100cp/µl)
GM07	Sunup™ papaya (event 55-1)	5211514801	Eurofins GeneScan	Sunup™ papaya	CUH-CP551-8	papaya	genomic DNA	100	0.15 ml (100cp/µl)
GM08	pGSE219 LL62 rice	5211508501	Eurofins GeneScan	LL62	ACS-OSØØ2-5	rice	plasmid DNA		0.15 ml (100cp/µl)
GM09	Roundup Ready soy	5211500601	Eurofins GeneScan	GTS 40-3-2	MON-Ø4Ø32-6	soybean	genomic DNA	100	0.15 ml (100cp/µl)
GM10	MON89788 soy	5211511901	Eurofins GeneScan	MON89788	MON-89788-1	soybean	genomic DNA	100	0.15 ml (100cp/µl)
GM11	H7-1 sugarbeet	5211507001	Eurofins GeneScan	H7-1	KM-ØØØ71-4	sugar beet	genomic DNA	100	0.15 ml (100cp/µl)
GM12	98140 maize	# ERMBF427D-1G	SIGMA-ALDRICH	event 98140	DP-Ø9814Ø-6	maize	dried maize powder	10	1g
GM13	NK603 Maize	# 69407-1SET-F	SIGMA-ALDRICH	NK603	MON-ØØ6Ø3-6	maize	genomic DNA	1	250±50 ng DNA / 10 µl
GM14	GA21 Maize	# 69407-1SET-F	SIGMA-ALDRICH	GA21	MON-ØØØ21-9	maize	genomic DNA	1	250±50 ng DNA / 10 µl
GM15	CBH351 "StarLink" Maize	# 69407-1SET-F	SIGMA-ALDRICH	CBH351	ACS-ZX004-3	maize	genomic DNA	1	250±50 ng DNA / 10 µl
GM16	EH92-527-1 potato	ERMBF421B-0.5G	SIGMA-ALDRICH	EH92-527-1 potato	BPS-25271-9	potato	dried potato powder	100	1g
GM17	356043 soya	ERMBF425D	SIGMA-ALDRICH	356043 soya	DP-356Ø43-5	soybean	soya seed powder	10	1g
GM18	GT73 Roundup Ready™	# 55231-1KT-F	SIGMA-ALDRICH	GT73	MON-ØØØ73-7	rapeseed	genomic DNA	1	150 ng
GM19	LibertyLink™ Falcon GS40/90	# 55231-1KT-F	SIGMA-ALDRICH	GS40/90	ACS-BN-Ø1Ø-4	rapeseed	genomic DNA	1	150 ng
GM20	MS8xRf3 Rapeseed	# 55231-1KT-F	SIGMA-ALDRICH	MS8xRf3	ACS-BNØØ5-8 x ACS-BNØØ3-6	rapeseed	genomic DNA	1	150 ng

Table 6. Genetic elements introduced into the 20 GM samples analysed in this work according to the GMOseek matrix.

Sample code	Description	Promotor														
		p35s	**pPCDK**	**pPEPC**	**pFMV**	**pFMV/TSF1**	**tsf1**	**pALS**	**pUbiZM1**	**pNOS**	**pKit3**	**pSCP1**	**pSsuAra**	**pTA29**	**pActin1**	**pCBI**
GM01	Bt176 corn	+	+	+												
GM02	Bt11 corn	+														
GM03	T25 corn	+														
GM04	MON531 cotton	+														+
GM05	MON1445 cotton	+			+											
GM06	LLCotton25 cotton	+														
GM07	Sunup™ papaya (event 55-1)	+								+						
GM08	pGSE219 LL62 rice	+														
GM09	Roundup Ready soy	+														
GM10	MON89788 soy				+	+	+									
GM11	H7-1 sugarbeet				+											
GM12	98140 maize	+						+	+							
GM13	NK603 Maize	+													+	
GM14	GA21 Maize														+	
GM15	CBH351 "StarLink" Maize	+														
GM16	EH92-527-1 potato									+						
GM17	356043 soya	+									+	+				
GM18	GT73 Roundup Ready™				+											
GM19	LibertyLink™ Falcon GS40/90	+														
GM20	MS8xRf3 Rapeseed												+	+		
	Total	14	1	1	4	1	1	1	1	2	1	1	1	1	2	1

Table 6. (cont.)

Sample code	Description	Terminator						
		t35S	T-nos	tE9	tpinII	tg7	tALS (Glycin max)	t7S
GM01	Bt176 corn	+	+					
GM02	Bt11 corn		+					
GM03	T25 corn	+						
GM04	MON531 cotton		+					+
GM05	MON1445 cotton		+	+				
GM06	LLCotton25 cotton		+					
GM07	Sunup™ papaya (event 55-1)	+	+					
GM08	pGSE219 LL62 rice	+						
GM09	Roundup Ready soy		+					
GM10	MON89788 soy			+				
GM11	H7-1 sugarbeet			+				
GM12	98140 maize				+			
GM13	NK603 Maize		+					
GM14	GA21 Maize		+					
GM15	CBH351 "StarLink" Maize	+	+					
GM16	EH92-527-1 potato		+					
GM17	356043 soya				+		+	
GM18	GT73 Roundup Ready™			+				
GM19	LibertyLink™ Falcon GS40/90	+						
GM20	MS8xRf3 Rapeseed		+			+		
	Total	6	12	4	2	1	1	1

Table 6. (cont.)

Sample code	Description	Open-reading frame (ORF)																					
		lacZ	bla	cry1Ab	cry1Ac	crylaB/crylac	cry9C	pat	nptII	aad	GUS	PRSV/CMV CP	PRSV CP	cp4-epsps	ALS (ZM)	gat4621	m epsps	bxn	gox	barnase	barstar	gox247	bar
GM01	Bt176 corn	+	+	+		+																	+
GM02	Bt11 corn		+	+		+		+															
GM03	T25 corn		+					+															
GM04	MON531 cotton				+				+	+													
GM05	MON1445 cotton								+	+				+					+				
GM06	LLCotton25 cotton																						+
GM07	Sunup™ papaya (event 55-1)								+		+	+	+										
GM08	pGSE219 LL62 rice																						+
GM09	Roundup Ready soy													+									
GM10	MON89788 soy													+									
GM11	H7-1 sugarbeet									+				+									
GM12	98140 maize														+	+							
GM13	NK603 Maize													+									
GM14	GA21 Maize																+						
GM15	CBH351 "StarLink" Maize		+				+																+
GM16	EH92-527-1 potato								+														
GM17	356043 soya															+							
GM18	GT73 Roundup Ready™													+					+			+	
GM19	LibertyLink™ Falcon GS40/90							+															
GM20	MS8xRf3 Rapeseed																			+	+		+
	Total	1	4	2	1	2	1	3	4	3	1	1	1	6	1	2	1	0	2	1	1	1	5

Table 6. (cont.)

Sample code	Description	Miscellaneous elements																			
		intron9	ori-pUC	IVS2	IVS6	TiplasmidDNA/ /RightBoreder	TiplasmidDNA/ /LeftBoreder	ori322/rop	oriV	CTP1	CTP2	CTP4	hsp70	hsp70intron	ract1int	ColE1-ori	pUC19	RuBisCo	CTP/cab22L	CTP/OTP	ori322
GM01	Bt176 corn	+	+																		
GM02	Bt11 corn			+	+																+
GM03	T25 corn		+																		
GM04	MON531 cotton					+		+	+												
GM05	MON1445 cotton					+			+	+	+										+
GM06	LLCotton25 cotton					+	+														
GM07	Sunup™ papaya (event 55-1)					+	+		+							+					
GM08	pGSE219 LL62 rice		+			+											+				
GM09	Roundup Ready soy		+									+									
GM10	MON89788 soy																				
GM11	H7-1 sugarbeet								+		+										+
GM12	98140 maize																				
GM13	NK603 Maize										+		+	+	+						
GM14	GA21 Maize																			+	
GM15	CBH351 "StarLink" Maize		+														+		+		
GM16	EH92-527-1 potato																				
GM17	356043 soya																				
GM18	GT73 Roundup Ready™					+	+			+	+										
GM19	LibertyLink™ Falcon GS40/90					+	+														
GM20	MS8xRf3 Rapeseed					+	+														
	Total	1	5	1	1	8	5	1	4	2	4	1	1	1	1	1	2	0	1	1	3

Table 6. (cont.)

Sample code	Description	Junctions								
		p35S-pat	p35S-IVS2	pEPC-cryLAB	p35S-nptII	35S-bar	CTP2-CP4EPSPS	pNOS-nptII	pTA29-barnase	pSSUara-bar
GM01	Bt176 corn			+		+				
GM02	Bt11 corn	+	+							
GM03	T25 corn	+								
GM04	MON531 cotton				+					
GM05	MON1445 cotton				+		+			
GM06	LLCotton25 cotton					+				
GM07	Sunup™ papaya (event 55-1)							+		
GM08	pGSE219 LL62 rice					+				
GM09	Roundup Ready soy									
GM10	MON89788 soy						+			
GM11	H7-1 sugarbeet						+			
GM12	98140 maize									
GM13	NK603 Maize						+			
GM14	GA21 Maize									
GM15	CBH351 "StarLink" Maize					+				
GM16	EH92-527-1 potato							+		
GM17	356043 soya									
GM18	GT73 Roundup Ready™						+			
GM19	LibertyLink™ Falcon GS40/90	+								
GM20	MS8xRf3 Rapeseed						+		+	+
	Total	3	1	1	2	4	6	2	1	1

Sample preparation and DNA extraction

Total DNA was extracted from the powder samples by using the protocol described by Doyle (Doyle 1987, 1990) with adaptations suggested by others (John 1992; Lodhi *et al.* 1994; Porebski *et al.* 1997). The protocol started with the addition of 500 µL of CTAB (cetyltrimethylammonium bromide) Lysis buffer (AppliChem GmbH, Germany) and 0,2% 2-mercaptoethanol (Sigma-Aldrich) to approximately 0.5 cm^3 of powder, followed by an incubation at 60°C for 1h. We then added 575µL of phenol (AppliChem) to the samples in Light Phase Lock Gel (5 Prime, Germany) tubes, which were then centrifuged for 3 min at 14000g. The aqueous phase was then removed to a new tube with 575µL of chloroform-isoamyl alcohol (24:1) (AppliChem) for a new centrifugation for 3 min at 14000g. The aqueous phase was removed and mixed with 1000µL of ethanol 96% and kept at -20°C for 30 min. The solution was then centrifuged for 15 min at 14000g at 4°C. After disregarding the supernatant, we added 1000µL of ethanol 70% and centrifuged for 5 min at 14000g. The supernatant was disregarded and the tubes were dried to precipitate the DNA. The pellet was ressuspened in 50µL of water and stored at -20°C. The DNA extractions were performed in isolated work areas, separated from PCR amplified products. Negative control extractions were used to screen for contaminants entering the process at any stage.

Singleplex and multiplex PCR amplifications

All PCR primer pairs were initially tested in singleplex PCR in a total volume of 10 µL as follows: 5 µL of Multiplex PCR Master Mix (Qiagen, Valencia, CA), 1 µL of forward primer (2 µM), 1 µL of reverse primer (2 µM), 2 µL of DNase, RNase- and protease-free water (5 Prime) and 1 µL of the template DNA sample. The thermocycling conditions were: initial step at 95 °C for 15 min; 35 cycles at 94 °C for 30 s, 55 °C for 90 s and 72 °C for 1min; with a final extension at 72 °C for 10 min. PCRs were performed in a BioRad ThermoCycler (BioRad Laboratories Inc., Hercules, CA, USA). The selected primers were combined in the same proportion for multiplex PCRs. The PCR setup and thermocycling conditions used for multiplex PCRs were the same as in singleplex reactions, with the exception of the use of 1 µL of the primers mix (each at 2uM) and 3 µL of water (5 Prime) for a final volume of 10 µL. We tested the sensitivity of the multiplex PCRs using gradients of primer annealing temperatures between 46 °C to 66 ° C. Slight alterations to these PCR conditions were used in some cases as indicated in the results and discussion section.

Electrophoretic separation and sequencing

The PCR products in singleplex or multiplex PCR were separated by electrophoresis on a 2% agarose gel containing GelRed (Biotium, USA) at 140 V for 50 minutes. The fragments were visualized under UV light in a Molecular Imager® ChemiDocTM XRS Imaging System (Bio Rad, USA) with Quantity One software v4.6.9 (Bio Rad).

We also analysed some of the PCR products by capillary electrophoresis using the QIAxcel system (Qiagen). The PCR products were run using the DNA High Resolution Gel Cartridge (Qiagen) under method OM700.mtd for 10 s at 5 kV voltage for sample injection and 700 s and 3 kV voltage for fragment separation. The results were analysed using QX Biocalculator Fast Analysis Software (Qiagen).

The samples used to test the primers by singleplex PCR were sequenced in both directions with the same primers used for PCR in order to confirm the target sequence. A total of 1.5 µL of PCR product was purified with ExoSAP-IT® (USB, Affimetrix, USA) according to the manufacturer's recommendations. The sequencing reaction was performed by combining 2.5 µL of purified PCR product, 2 µL of Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and 0.5 µL of primer (2 µM). Thermal cycler conditions were: 96 °C for 2 min, 35 cycles at 96 °C for 15 s, 50 °C for 9 s and 60 °C for 2 min and one final hold at 60 °C for 10 min. Sequencing reaction products were purified using Sephadex™ G-50 Fine DNA Grade columns (GE Healthcare, United Kingdom) according to the manufacturer's recommendations. Purified samples were added to 12 µL of HI-DI formamide (Applichem, Germany). Sequencing was performed in a Genetic Analyzer 3130xl sequencer (Applied Biosystems), according to the manufacturer's recommendations. Sequence analysis was performed using Sequencing Analysis software v5.2 (Applied Biosystems) and Geneious v5.3.

Results and Discussion

The screening of samples for transgenic elements is the most used method to detect GM plants. Our initial aim was to test different combinations of transgenic elements in multiplex PCR. We chose to use the five target sequences identified by Waiblinger et al. 2010, which are sufficient to detect the presence of most authorised and unauthorised GM plants. In addition to the two most common elements used in GM plans (P35S and T-nos), Waiblinger et al. 2010 identified the bar gene and the ctp2-cp4epsps and P35S-pat junctions as important elements for a broad-range screening approach. The P35S (*n* = 181) and T-nos (*n* = 178) are the most frequent elements present in the 328 GMO events described in the GMOseek matrix (Block *et al.* 2013). The bar (*n* = 34), ctp2-cp4epsps (*n* = 48) and P35S-pat (*n* = 73) are less common among GM plant events (Table 7). Using our set of 20 reference samples as an example, it is also clear that the P35S (n= 14) and T-nos (*n* = 12) are the most frequent elements (Table 6 and 7). The bar gene (*n* = 5) and the ctp2-cp4epsps (*n* = 6) and P35S-pat (*n* = 3) junctions were only found in a subset of our samples.

We also used the information available on the GMOseek matrix to select other elements that might provide additional information about the possible presence or absence of different GM plant events. The distribution of elements in the 328 GMO events described in the GMOseek matrix allowed us to identify the Cry1Ab gene and FMV35S promotor as useful targets for a screening analysis. The Cry1Ab occurs in 64 events out of 328, being more common than the bar and ctp2-cp4epsps elements (Table 7). The FMV35S promotor occurs in 33 out of 328 events. According to the GMOseek matrix, the Cry1Ab is present in two and the FMV35S promotor in four of our reference samples (Table 6 and 7).

Table 7. Distribution of the seven transgenic elements in the GMOseek matrix (Block *et al.* 2013) and in our set of 20 reference samples of GM plants.

Elements	**Number of samples with the element**	
	Complete GMOseek matrix (*n* = 328)	Our set of reference samples (*n* = 20)
P35S	181	14
T-nos	178	12
bar	34	5
ctp2-cp4epsps	48	6
P35S-pat	73	3
Cry1Ab	64	2
FMV35S	33	4

We compiled a list with 74 PCR primers for amplification of the seven transgenic elements selected for the multiplex PCRs. The primers were obtained from public databases and peer-reviewed publications or were designed by us (Table 4). We designed nine primers in order to obtain amplicons with different lengths for the different elements, allowing their incorporation in multiplex PCRs. For example, the primers obtained from publications or databases for the Cry1Ab gene had lengths that overlapped with other elements. Therefore, we designed a reverse primer (GMcryIAB228R) that combined with the forward primer GMcryIAB_3F generates an amplicon with unique length (226 bp) among all elements (Table 4). The annealing temperatures of the 74 primers ranged from 53.8 to 66.9 °C, with a mean value of 60.4 °C. The mean length of these primers was 21.1 nucleotides, varying from 17 to 28 nucleotides.

We then selected 28 primers located in the seven transgenic elements to test in the laboratory (Table 4 and Figures 3 to 9). These primers can be combined in different ways so that a maximum number of 20 target regions can be amplified (Figure 10). The average length of the amplicons defined by these primers, when considering the reference genome sequences, was 202.2 bp, ranging from 82 to 417 bp. In some cases, the primer located in an element can be used to amplify only a section of that element or a junction including the element. For example, the forward primer GMca35s7240F can be combined with the reverse GMca35s7321R to amplify an 82-bp region in the P35S promotor or with the reverse GM35s-pat1973R located in the pat gene to amplify the P35S-pat junction. For this reason, we were able to design and test primers not only for the seven elements listed in Table 2, but also for the amplification of the cp4epsps and pat elements alone without being part of junctions (Table 4).

Figure 3. Schematic representation of the location of PCR primers selected for the P35S element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

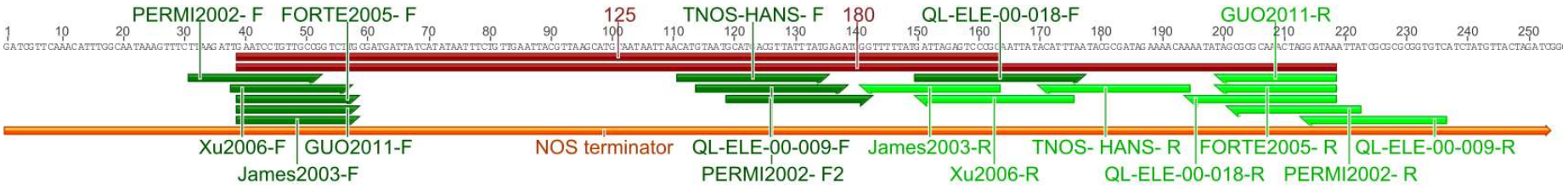


Figure 4. Schematic representation of the location of PCR primers selected for the T-nos element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

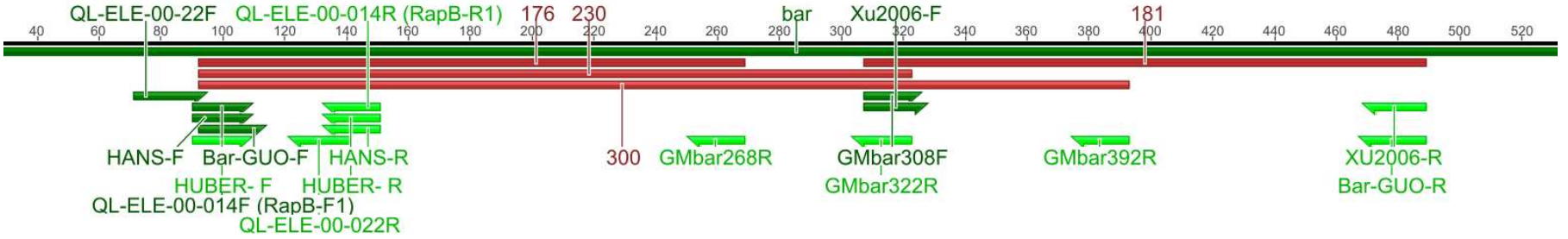


Figure 5. Schematic representation of the location of PCR primers selected for the bar element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

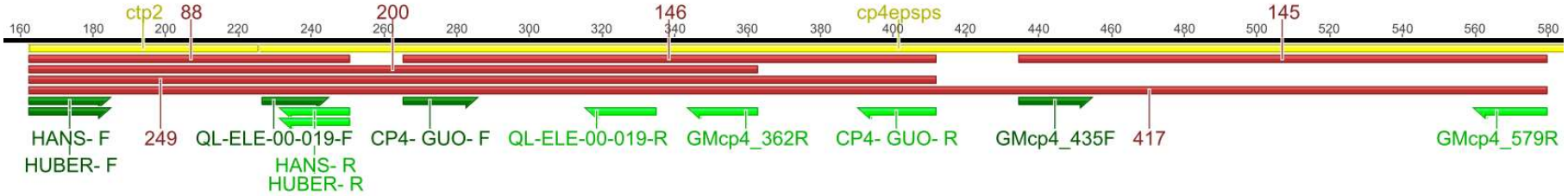


Figure 6. Schematic representation of the location of PCR primers selected for the ctp2-cp4epsps element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

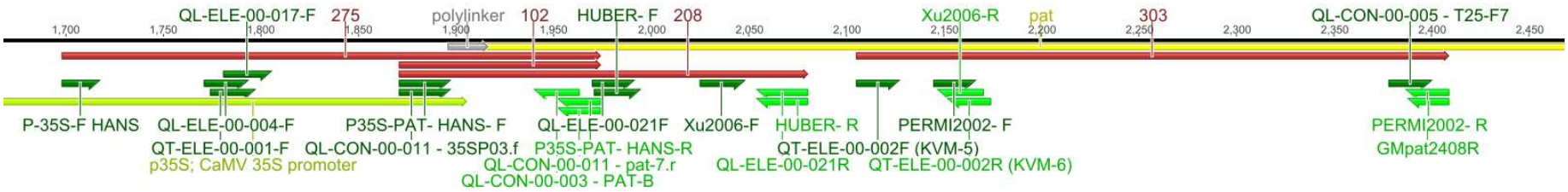


Figure 7. Schematic representation of the location of PCR primers selected for the P35S-pat element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

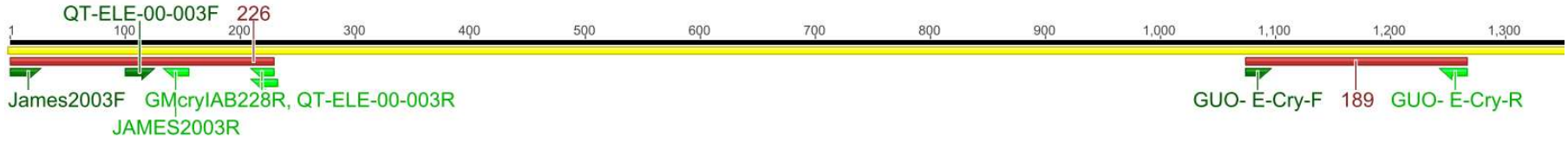


Figure 8. Schematic representation of the location of PCR primers selected for the Cry1Ab element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

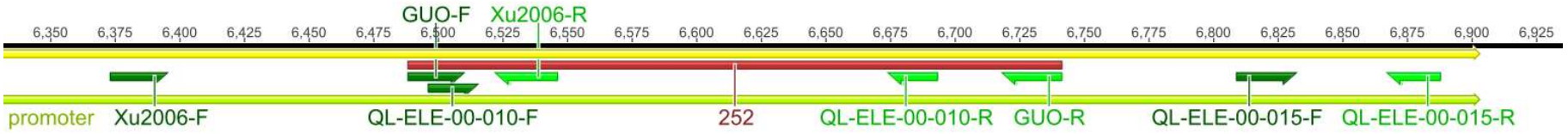


Figure 9. Schematic representation of the location of PCR primers selected for the FMV35S element on the reference genomic sequence. The forward (dark green) and reverse (light green) primers are indicated by arrows. Some of the amplified regions are indicated by red bars identified by their length.

Element	Primers	Samples					
417 - ctp2-cp4epsps	GMctp2_163F GMcp4_579R	GM05	GM10	GM11	GM13	GM18	GM20
303 - pat (alone)	GMpat2106F GMpat2408R	GM02	GM03	GM19			
300 - bar	GMbar93F GMbar392R	GM01	GM06	GM08	GM15	GM20	
275 - P35S-pat	GMcaP35s7240F GM35s-pat1973R	GM02	GM03	GM19			
252 - FMV35S	GMfig35S_6489F GMfig35S_6740R	GM05	GM10	GM11	GM18		
249 - ctp2-cp4epsps	GMctp2_163F GMcp4_334R	GM05	GM10	GM11	GM13	GM18	GM20
230 - bar	GMbar93F GMbar322R	GM01	GM06	GM08	GM15	GM20	
226 - Cry1Ab	GMcrylAB_3F GMcrylAB228R	GM01	GM02				
208 - P35S-pat	GM35s-pat1872F GMpat2079R	GM02	GM03	GM19			
200 - ctp2-cp4epsps	GMctp2_163F GMcp4_362R	GM05	GM10	GM11	GM13	GM18	GM20
181- bar	GMbar308F GMbar448R	GM01	GM06	GM08	GM15	GM20	
180 – T-nos	GMtnos39F GMtnos218R	GM01 GM09 GM21	GM02 GM13	GM04 GM14	GM05 GM15	GM06 GM16	GM07 GM20
176 - bar	GMbar93F GMbar268R	GM01	GM06	GM15	GM20		
158 - P35S	GMca35s7240F GMca35s7397R	GM01 GM07 GM17	GM02 GM08 GM19	GM03 GM09 GM21	GM04 GM12	GM05 GM13	GM06 GM15
146 – cp4 (alone)	GMcp4_266F GMcp4_334R	GM05	GM09	GM10	GM11	GM13	GM18
145 - cp4 (alone)	GMcp4_435F GMcp4_579R	GM05	GM09	GM10	GM11	GM13	GM18
125 - T-nos	GMtnos39F GMtnos163R	GM01 GM09 GM21	GM02 GM13	GM04 GM14	GM05 GM15	GM06 GM16	GM07 GM20
102 - P35S-pat	GM35s-pat1872F GM35s-pat1973R	GM02	GM03	GM19			
88 - ctp2-cp4epsps	GMctp2_163F GMcp4_250R	GM05	GM10	GM11	GM13	GM18	GM20
82 - P35S	GMcaP35s7240F GMcaP35s7321R	GM01 GM07 GM17	GM02 GM08 GM19	GM03 GM09 GM21	GM04 GM12	GM05 GM13	GM06 GM15

Figure 10. List of target regions of the elements used in GM plants, including the primers used for PCR amplification. The reference DNA samples expected to have the GM element according to the GMOseek matrix are also indicated.

We started by testing the primer pairs by singleplex PCR in two reference DNA samples expected to have the transgenic element (Figure 11). The DNA was amplified using an annealing temperature of 55 °C during 35 PCR cycles. Thirteen regions out of 20 were successfully amplified for the two DNA samples. The regions cp4 (145 bp), Cry1Ab (226 bp), cp4 (146 bp) and T-nos (180 bp) yielded amplifications in only one of the samples. No amplification was observed in both samples for markers bar (300 bp), ctp2-cp4epsps (417 bp) and ctp2-cp4epsps (249 bp).

Overall, the observed band lengths on agarose gels are in agreement with the expected sizes predicted in the reference genomic sequences (Figure 11). The sequencing of all products amplified by singleplex PCRs confirmed that all target regions were being correctly amplified (data not show). In any case, we observed a few cases of extra amplifications or unexpected amplicon lengths.

The two primer pairs for the P35S element yielded PCR products with the expected length (~82 and ~152), but an additional band was observed in sample GM04 for both target regions (Figure 11). The sequencing of the amplified product in sample GM04 allowed us to discover that a section of the P35S promotor with at least 105 bp (where some of our primers are located) is duplicated upstream the complete P35S promotor (Figure 12). The two regions are separated by a sequence with 6 bp (GGTCCG). The upstream P35S fragment includes the binding region of the forward primer used for the target regions with the expected length of 82 and 158 bp. The additional band results from the PCR between the forward primer located in the upstream P35S fragment with the reverse primers located in the complete P35S promotor, yielding 337 bp (when the primers for the 82-bp region are used) and 413 bp (when the primers for the 158-bp region are used). The reason for the presence of this duplication in sample GM04 (MON531 cotton) is unknown and was never described before to our knowledge.

One of the two T-nos markers (125 bp) was amplified in the two DNA samples, while the target region with an expected length of 180 bp was not amplified in sample GM01. In this case, a week amplification was observed but for a larger amplicon with approximately 210 bp (Figure 11).

Three of the four bar regions tested by singleplex PCR (176, 181 and 230 bp) were successfully amplified in both samples, with all amplicons having the expected length. Only the bar marker with an expected length of 300 bp was not amplified (Figure 11).

We tested four combinations of primers for the ctp2-cp4epsps junction (Figures 6 and 10). The two largest markers with an expected length of 249 and 417 bp were not amplified, whereas amplifications were observed for the two shortest regions with

~88 and 200 bp (Figure 11). The lack of amplification in the two target regions might be related with the reverse primer, since the same forward primer was used in the largest regions and the marker with ~200 bp (Figure 6). The sequencing revealed that sample GM05 has a cp4epsps element with several polymorphism in relation to the reference sequence (Figure 12), while the sequence obtained for sample GM10 and GM09 (sequenced for the cp4epsps alone) had no differences in relation to the reference indicated in Table 3. Most of the polymorphisms observed in sample GM05 are located in the third codon position and do not change the amino acid (Figure 13), suggesting that a different bacterial strain of *A. tumefaciens* was used to provide the *epsps* gene for the construct used in GM05 sample.

The two primer pairs tested for the cp4epsps alone (i.e., not including the ctp2-cp4epsps junction) only amplified for the GM09 sample with the expected length. The reference sample GM05 was not amplified for the cp4epsps with the tested primers. We exclude the possibility of a problem with the DNA extraction of sample GM05 (e.g., contaminants or low DNA concentration) because it was successfully amplified with primers for other target regions (Figure 11). The absence of amplification is most likely related with the different cp4epsps genes used in sample GM05 that might prevent the binding of the primers (Figure 13).

The three pairs of primers tested for the P35S-pat element were successfully amplified in both samples. The sequencing of the amplicons confirmed that sample GM03 has a P35S-pat construct with a short polylinker (~19 bp) separating the 35S promoter from the pat gene. A larger amplicon was observed in sample GM02 in the three pairs of primers (Figure 11). The sequencing of the amplified products confirmed that sample GM02 (Bt11 corn) has a P35S-pat construct with a 179 nt enhance element separating the 35S promoter from the pat gene, as previously suggested (Waiblinger *et al.* 2010). According to the annotations of the reference sequence AY629236.1 (Table 3), the enhancer element is derived from the *alcohol dehydrogenase 1* (*adh1*) gene.

The only pair of primers tested for the pat element alone (i.e., not amplifying the P35S-pat junction) yielded an amplified product with the expected length, although an unspecific amplification with approximately 150 bp was observed in both samples (Figure 11).

The pair of primers for amplification of the FMV35S promotor originated an amplification with the expected length, while the single region tested for the Cry1Ab element was only amplified in one of the two samples (Figure 11).

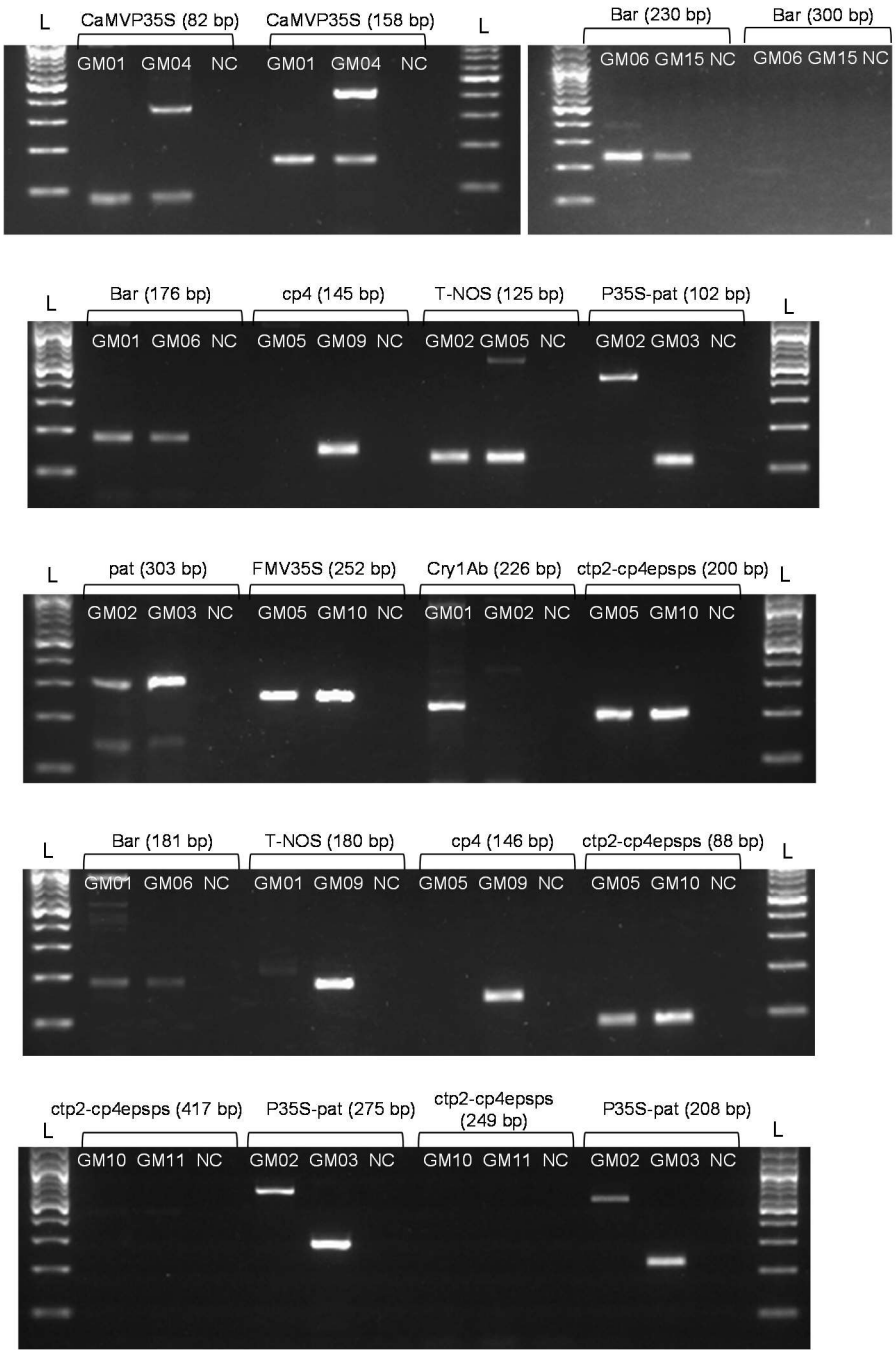


Figure 11. Singleplex PCR amplifications using the primers designed for the seven transgenic elements used in GM plants. Two DNA samples were used to test all designed PCR primer pairs. NC - negative control; L -100 bp DNA ladder.

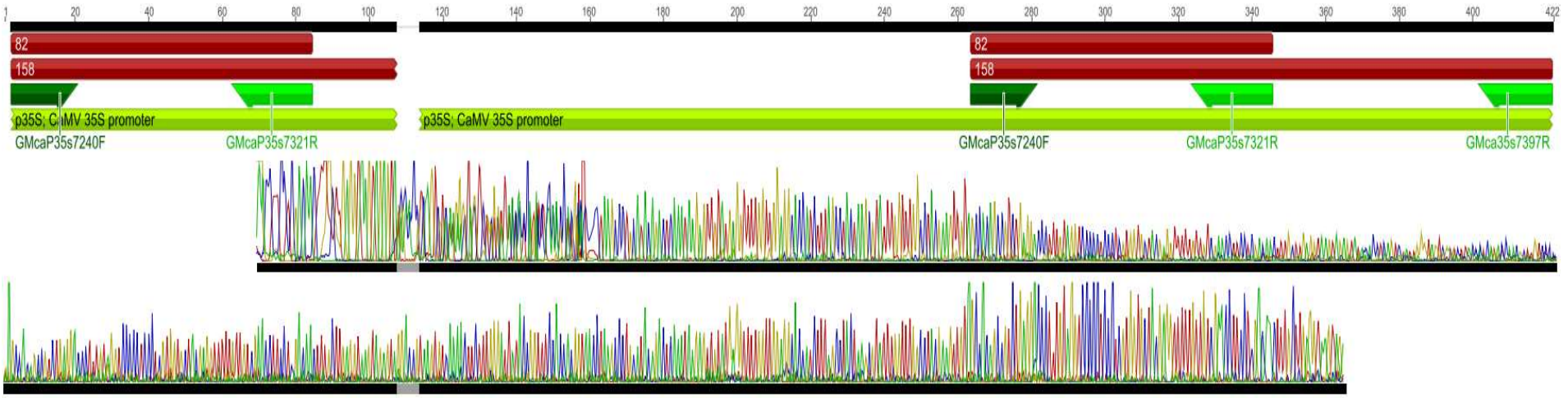


Figure 12. Duplicated region found in sample GM04 (MON531 cotton) for the P35S element. A fragment of the P35S promotor with at least 105 bp is duplicated upstream the complete P35S promotor. The upstream P35S fragment includes the binding region of the forward primer (dark green arrow) used for the target regions with the expected length of 82 and 158 bp (red annotations). The electropherograms depict the sequence resulting from the PCR between the forward primer located in the upstream P35S fragment with the reverse primers located in the complete P35S promotor.

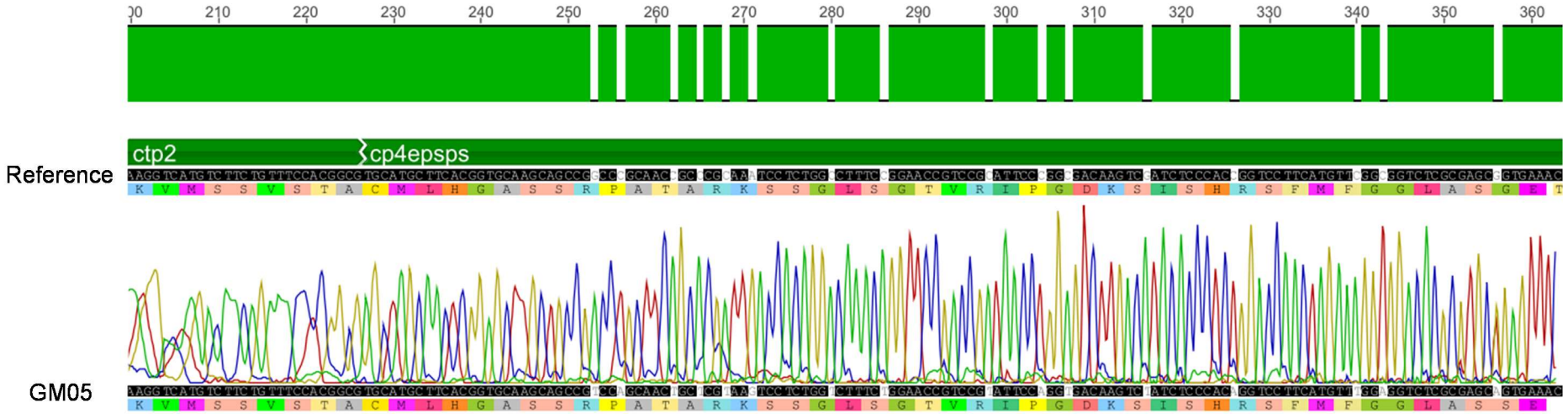


Figure 13. Sequence alignment of the ctp2-cp4epsps element from sample GM05 (MON1445 cotton) and the reference sequence used by us. The sequencing revealed that sample GM05 has several polymorphism in relation to the reference sequence in the cp4epsps element that do not change the amino acid.

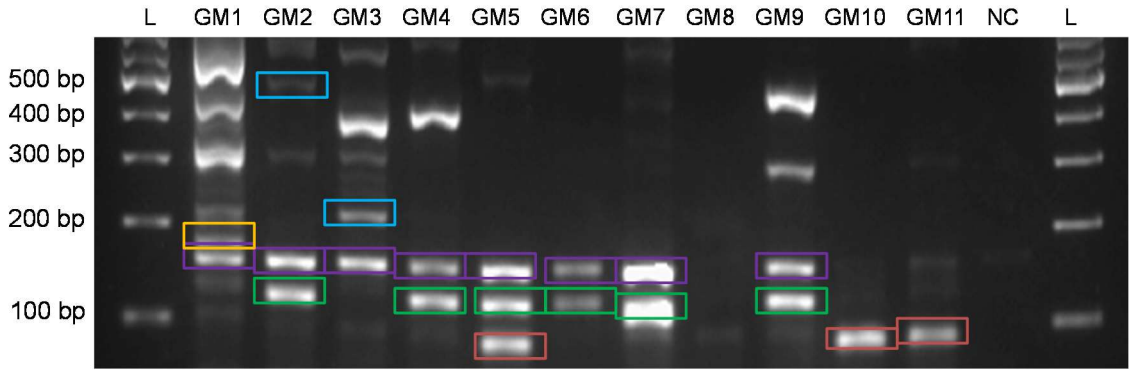
We then combined in different multiplex PCRs the primers pairs that worked well in singleplex PCR. These preliminary tests were done by combining all primers in equal proportions.

The first multiplex (nº 1) included the target regions for ctp2-cp4epsps (88 bp), T-nos (125), P35S (158), bar (176) and P35S-pat (208) and was used to amplify samples GM01 to GM11 (Figure 14). These five elements used in multiplex nº1 are those identified by Waiblinger et al. as being able to detect most GM events.

No amplification was observed for sample GM08 in this and other multiplex PCRs. This lack of amplification may be due to the fact that it is the only sample of a plasmid DNA (pGSE219 LL62 rice), which may not include the complete inserts. Considering the elements believed to be present in these samples according to the GMOseek matrix (Table 6), the expected amplified products were obtained for samples GM02, GM03, GM04, GM05, GM07, GM09, GM10 and GM11. Overall, the amplicons are easily discriminated in a 2% agarose gel.

In the case of sample GM01, no clear band was observed for the T-nos element (~125 bp), although two faint bands are visible between 100 and 150 bp (Figure 14). The smear observed for this sample may indicate that too much DNA template is being used, which may be affecting the amplification of the T-nos region. A dilution of the GM01 DNA will be tested in the future. The P35S-pat amplicon is larger than the observed in sample GM03 due to the enhance element inserted between the 35S promoter and the pat gene. The bar gene was not amplified in sample GM06, although a clear band was obtained in the singleplex PCR with the same primer pair and sample (Figure 11). An inhibitory interaction between the primers included in the multiplex may explain the absence of amplification. The primer pair for the bar (176 bp) will have to be redesigned if this multiplex has to be used in the future.

Some extra bands above 300 bp are observed in samples GM01, GM03 and GM09, which may result from amplification between primers of different elements in the inserted constructs. Some plasmids used to produce GM plants include several genes near to each other. The excision and sequencing of the extra bands will be done in the future to identify the cause of the additional amplifications. The extra band observed in sample GM04 with ~400bp is due to the duplication of a section of the P35S promotor identified by us (Figure 12).

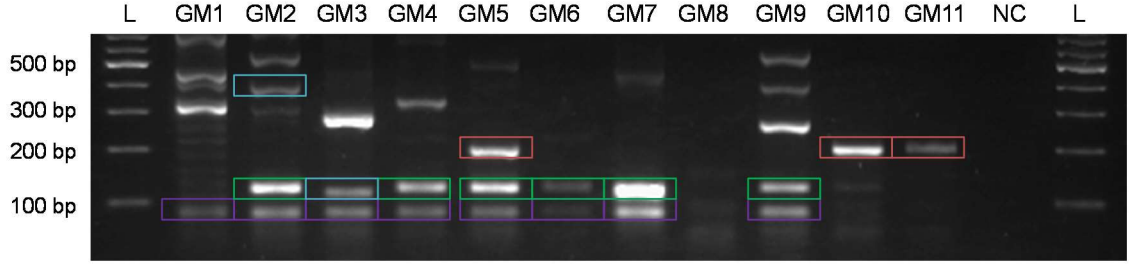


208 - P35S-pat
176 - bar
158 - P35S
125 - T-nos
88 - ctp2-cp4epsps

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat
GM01	+	+		+	
GM02	+	+			+
GM03	+				+
GM04	+	+			
GM05	+	+	+		
GM06	+	+		+	
GM07	+	+			
GM09	+	+			
GM10			+		
GM11			+		

Figure 14. Screening of samples GM01 to GM11 with multiplex nº1 including the target regions for ctp2-cp4epsps (88 bp), T-nos (125), P35S (158), bar (176) and P35S-pat (208). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

The multiplex PCR nº 2 tested by us included the same elements of multiplex nº1 (the five core elements of Waiblinger et al.), but with some of them amplified with different primer pairs (Figure 15): P35S (82), P35S-pat (102), T-nos (125), bar (176) and ctp2-cp4epsps (200). Only the T-nos (125) and bar (176) are the same used in multiplex nº1. In general, successful amplifications were observed in most samples, with a clear band separation. The expected amplifications were observed in samples GM02, GM03, GM04, GM05, GM07, GM09, GM10 and GM11 (Figure 15). We observed the same problem of an excess of template DNA in sample GM01, which presents several unspecific amplifications. As in multiplex nº 1, the extra band in GM04 is due to the duplication of the P35S promotor and sample GM06 lacks the bar region.

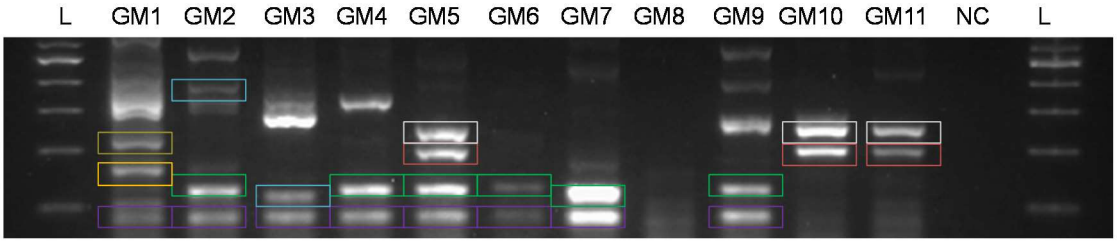


200 - ctp2-cp4epsps
176 - bar
125 - T-nos
102 - P35S-pat
82 - P35S

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat
GM01	+	+		+	
GM02	+	+			+
GM03	+				+
GM04	+	+			
GM05	+	+	+		
GM06	+	+		+	
GM07	+	+			
GM09	+	+			
GM10			+		
GM11			+		

Figure 15. Screening of samples GM01 to GM11 with multiplex nº2 including the target regions for P35S (82), P35S-pat (102), T-nos (125), bar (176) and ctp2-cp4epsps (200). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

We then decided to include three additional primer pairs for amplification of the cp4epsps alone, Cry1Ab and FMV35S, in addition to the five elements used in multiplexes nº 1 and nº 2 (Figure 16). The multiplex nº 3 was initially tested in samples GM01 to GM11 and includes primers for P35S (82), P35S-pat (102), T-nos (125), cp4epsps (145), bar (176) and ctp2-cp4epsps (200), Cry1Ab (226) and FMV35S (252). As previously observed in multiplexes nº 1 and 2, all bands are easily discriminated by conventional electrophoresis (Figure 16). The expected amplicons were observed in samples GM03, GM04, GM07, GM09, GM10 and GM11. The excess of DNA was evident in sample GM01 and may explain the absence of the T-nos element. The Cry1Ab was not amplified in GM02, similar to what was observed by singleplex PCR (Figure 11). Sample GM05 has four of the five expected bands, lacking the cp4epsps alone region as previously observed by singleplex PCR (Figure 11). Nevertheless, the four bands are easily discriminated under the tested conditions. The bar region was not observed in sample GM06 as in previous multiplexes.



252 - FMV35S
226 - Cry1Ab
200 - ctp2-cp4epsps
176 - bar
145 - cp4epsps (alone)
125 - T-nos
102 - P35S-pat
82 - P35S

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat	FMV35S	Cry1Ab	cp4epsps
GM01	+	+		+			+	
GM02	+	+			+		+	
GM03	+				+			
GM04	+	+						
GM05	+	+	+			+		+
GM06	+	+		+				
GM07	+	+						
GM09	+	+						+
GM10			+			+		+
GM11			+			+		+

Figure 16. Screening of samples GM01 to GM11 with multiplex nº 3 including the target regions for P35S (82), P35S-pat (102), T-nos (125), cp4epsps (145), bar (176) and ctp2-cp4epsps (200), Cry1Ab (226) and FMV35S (252). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

We then decided to design multiplex PCRs with a pair of primers as internal control for the presence of plant DNA. The internal control serves to rule out failure of amplification in cases where the target sequence is not detected. We designed primers to amplify a region of the chloroplast DNA (cpDNA) *trn*L [tRNA-Leu (UAA)] intron (Table 4). Our first attempt to include the *trn*L control in multiplex PCRs revealed a considerable difference in length in cotton samples (GM04, GM05 and GM06), which presented a *trn*L region of about 110 bp (Figure 17). This length was larger than the expected considering the reference sequences used by us with 78 bp. We sequenced the tRNA-Leu (UAA) intron using external primers described by Taberlet et al. and found several insertions in the *trn*L of cotton that explain the size difference observed in gels (Figure 17).

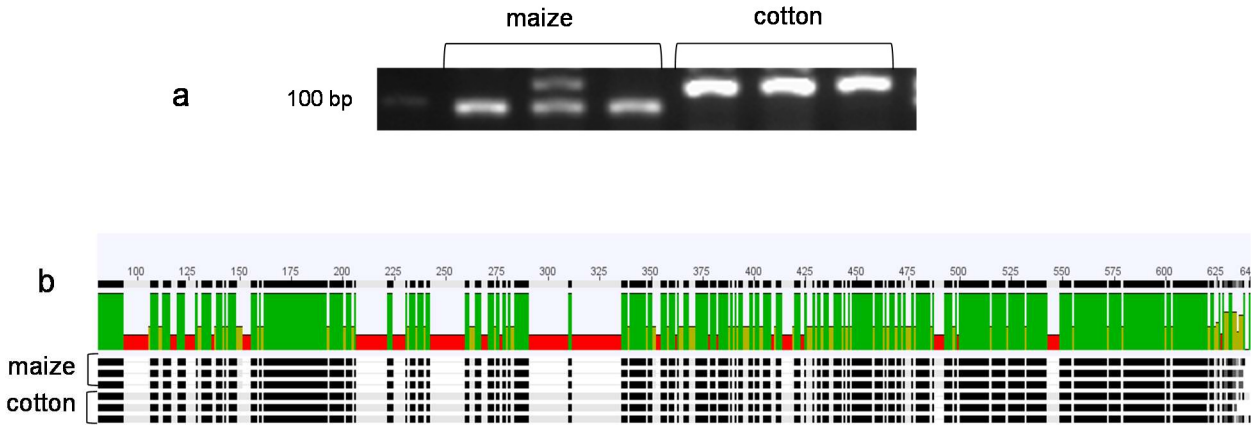


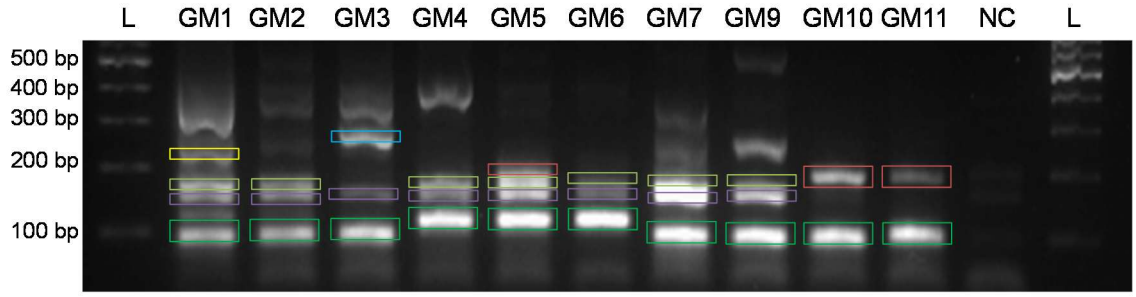
Figure 17. Chloroplast DNA (cpDNA) *trn*L [tRNA-Leu (UAA)] intron in maize and cotton. a) Agarose gel electrophoresis with *trn*L amplified products in three maize and three cotton samples. B) Multiple sequence alignment of *trn*L intron in three maize and three cotton samples. The grey regions indicate insertion/deletion polymorphic sites.

The unexpected large size of *trn*L in cotton lead to an overlap with the shorter target regions included in multiplexes nº 4 and 5 (data not shown). Therefore, we have redesigned the multiplexes with no target regions for GM elements shorter than 120 bp to avoid an overlap with *trn*L amplicons.

The multiplex nº 6 was designed to include the *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230) and P35S-pat (275). We started by testing the multiplex in samples GM01 to GM07 and GM09 to GM11 (we exclude the plasmid DNA GM08). The *trn*L was successfully amplified in all samples (Figure 18). All expected elements were detected in samples GM01, GM03, GM04, GM05, GM07, GM09, GM10 and GM11. There were only two cases of missing amplifications. The

P35S-pat was not amplified in GM02 possibly due to the large target region to be amplified (above 500 bp) resulting from the enhance element separating the 35S promoter from the pat gene. The other missing region was the bar gene in sample GM06, as in previous multiplexes (Figure 18).

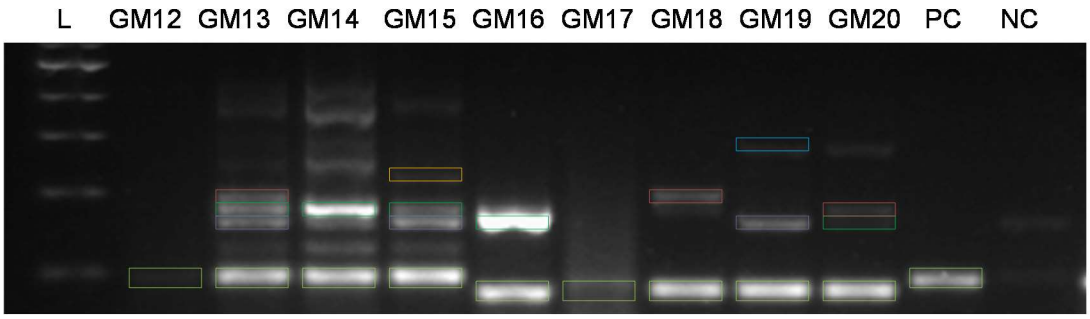
The remaining samples GM12 to GM20 were then analyzed with multiplex nº 6 (Figure 19). Samples GM12 (dried maize powder) and GM17 (soya seed powder) were not amplified for any target region, with only faint bands in the *trn*L control. The lack of amplification is probably due to a low DNA concentration resulting from extracting DNA from powder, suggesting that a higher concentration has to be used in further tests. The *trn*L control was successfully amplified in the remaining samples, including a positive control DNA sample extracted from a maize leave. With the exception of the bar gene in sample GM20, all target regions were successfully amplified in the remaining samples. Nevertheless, week amplifications were observed in some of the largest amplicons (Figure 19).



275 - P35S-pat
230 - bar
200 - ctp2-cp4epsps
180 - T-nos
158 - P35S
78 - *trn*L

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat
GM01	+	+		+	
GM02	+	+			+
GM03	+				+
GM04	+	+			
GM05	+	+	+		
GM06	+	+		+	
GM07	+	+			
GM09	+	+			
GM10			+		
GM11			+		

Figure 18. Screening of samples GM01 to GM07 and GM09 to GM11 with multiplex nº 6 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.

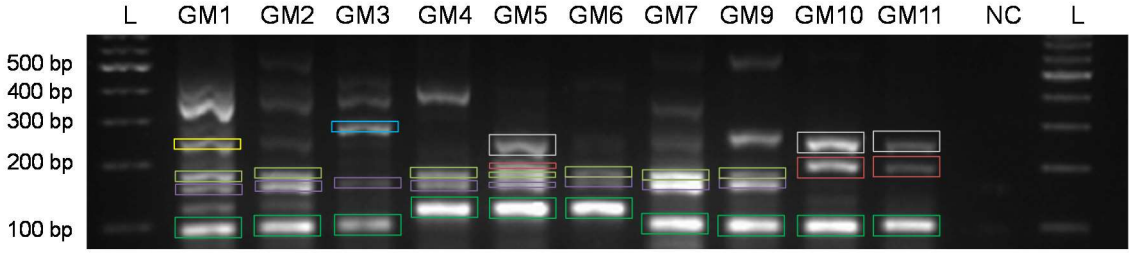


275 - P35S-pat
230 - bar
200 - ctp2-cp4epsps
180 - T-nos
158 - P35S
78 - *trn*L

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat
GM12	+				
GM13	+	+	+		
GM14		+			
GM15	+	+		+	
GM16		+			
GM17	+				
GM18			+		
GM19	+				+
GM20		+	+	+	

Figure 19. Screening of samples GM12 to GM20 multiplex nº 6 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. PC – Positive control; NC - negative control; L - 100 bp DNA ladder.

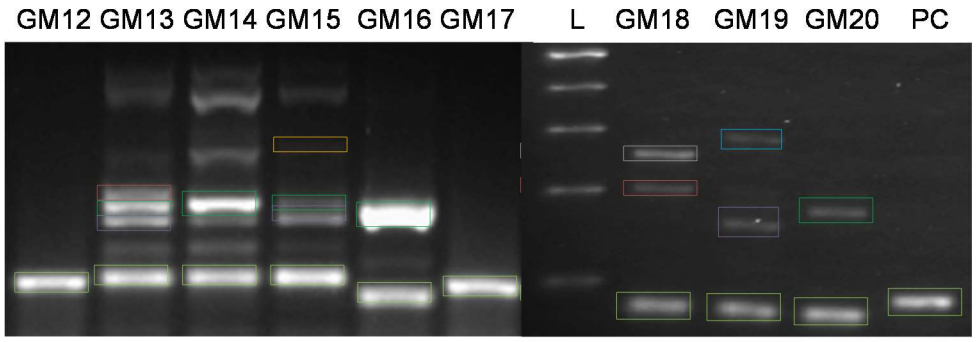
We added the FMV35S (252 bp) target region to the previous multiplex and tested it with the same set of samples (multiplex nº 7). The results were similar to those observed in multiplex nº 6 (Figures 20 and 21). The FMV35S (252 bp) region was amplified in all samples (*n* = 4) it was expected to be present.



275 - P35S-pat
252 - FMV35S
230 - bar
200 - ctp2-cp4epsps
180 - T-nos
158 - P35S
78 - *trn*L

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat	FMV35S
GM01	+	+		+		
GM02	+	+			+	
GM03	+				+	
GM04	+	+				
GM05	+	+	+			+
GM06	+	+		+		
GM07	+	+				
GM09	+	+				
GM10			+			+
GM11			+			+

Figure 20. Screening of samples GM01 to GM07 and GM09 to GM11 with multiplex nº 7 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230), FMV35S (252 bp) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. NC - negative control; L - 100 bp DNA ladder.



275 - P35S-pat
252 - FMV35S
230 - bar
200 - ctp2-cp4epsps
180 - T-nos
158 - P35S
78 - *trn*L

Sample	P35S	T-nos	ctp2-cp4epsps	bar	P35S-pat	FMV35S
GM12	+					
GM13	+	+	+			
GM14		+				
GM15	+	+		+		
GM16		+				
GM17	+					
GM18			+			+
GM19	+				+	
GM20		+	+	+		

Figure 21. Screening of samples GM12 to GM20 with multiplex nº 7 including the target regions for *trn*L (~78 to 100 bp), P35S (158), T-nos (180), ctp2-cp4epsps (200), bar (230), FMV35S (252 bp) and P35S-pat (275). The elements believed to be present in these samples according to the GMOseek matrix are indicated in the table. PC - Positive Control; NC - negative control; L - 100 bp DNA ladder.

The sample GM05 (MON1445 cotton) is the one with more elements (P35S, T-nos, ctp2-cp4epsps, FMV35S and cp4epsps alone), but still it does not include all elements that can be detected by multiplexes 6 and 7. Because it is important to determine if all target regions can be discriminated by conventional electrophoresis, we combined the amplified PCR products of all target regions included in multiplexes 6 and 7 (Figure 22). The amplicons were easily separated on conventional and capillary electrophoresis, suggesting that they can be successfully used in a multiplex configuration.

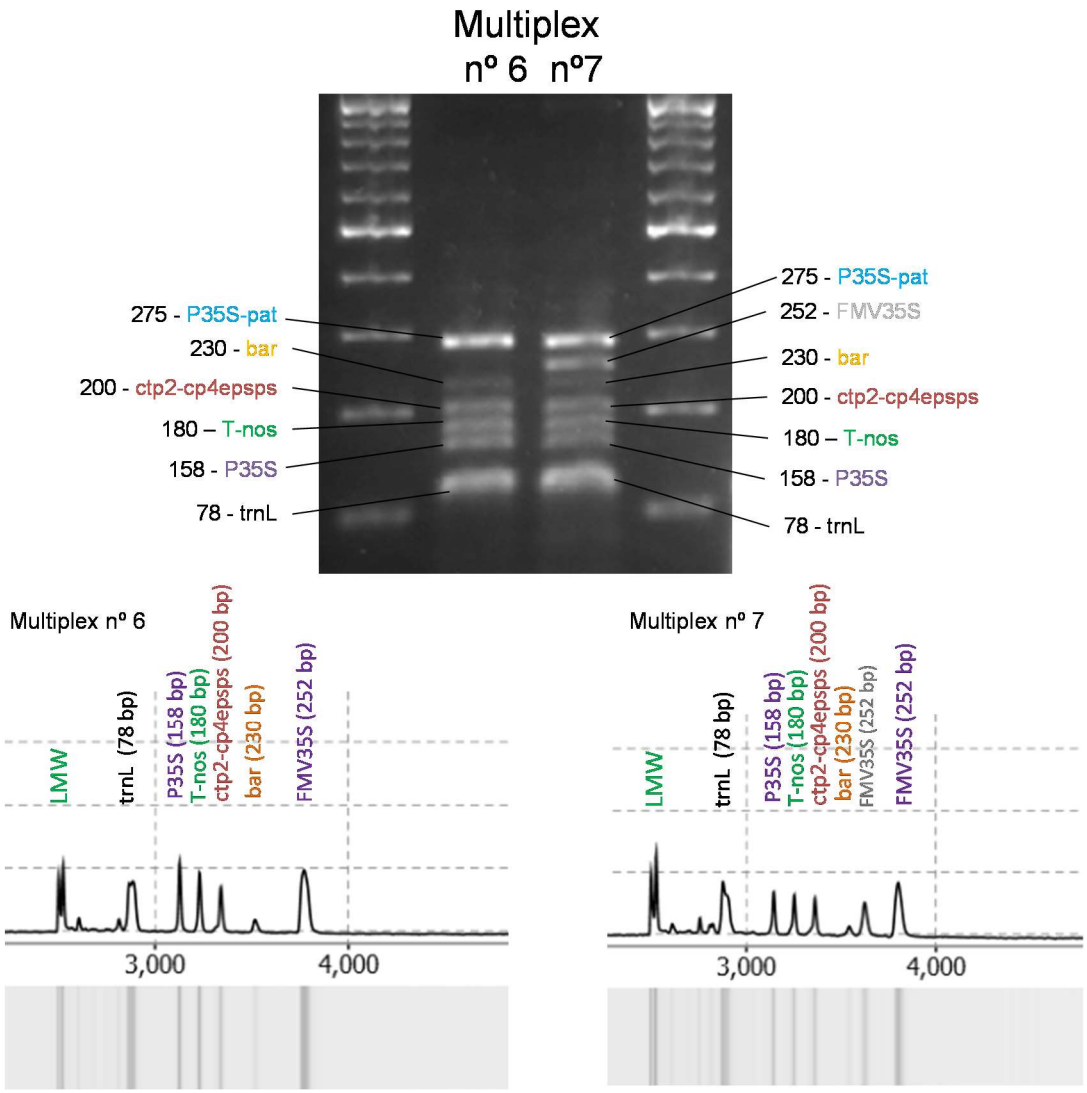


Figure 22. Discrimination of all target regions included in multiplexes 6 and 7 by conventional and automated capillary electrophoresis (QIAxcel system). LMW peak represents the 15-bp internal control on the capillary electrophoresis. A 100-bp DNA ladder was used on the agarose electrophoreses.

Conclusion

The use of accurate methods for detection of GM plants is an important aspect for an effective control on food and feed products, particularly taking into account the growing number of GM crops being developed. Detection of GM plants constitutes an important part of several countries' legislations in the traceability and labelling of the GM products, which is important in view of the free choice of the consumer (Ahmed 2002; Parekh 2004; Žel *et al.* 2012).

The screening is usually the first step when searching for genetic modifications, and further testing can be done to subsequently identify and quantify the potential GM event(s). The screening step should span the widest possible range of GM events that can be encountered on the market. However, most screening tests currently available are based on the detection of the P5S promoter and T-nos terminator, lacking important elements used on GM plants. Moreover, some of the available methods lack detection accuracy and are expensive and time-consuming. It is therefore of paramount importance to develop methods for the simultaneous identification of different transgenic elements to avoid time- and cost-intensive multiple event-specific tests.

The multiplex PCRs described here can be used for the detection of the most common transgenic elements used in GM plants. In addition to the five elements proposed by Waiblinger et al. as a universal screening method (P35S, T-nos, bar, ctp2-cp4epsps and P35S-pat), we included in some PCR multiplexes additional elements that can increase the range of detection (Cry1Ab, FMV35S, cp4epsps and pat). The use of these additional elements in screening assays allows the coverage of several GMO events globally approved for commercialization. The workflow consists of DNA extraction and multiplex PCR amplification followed by electrophoresis for amplicon detection. The size difference between all amplified fragments is sufficient to safely achieve unambiguous identification by electrophoresis.

The reference samples used to test the reproducibility of the multiplex PCRs were correctly identified as having GM materials. The only exception was two powder samples, most likely due to the use of a low DNA concentration as template in the PCR. The samples will be tested under different conditions in future experiments. We also observed that some target regions were not amplified as expected. The absence of amplification observed in some elements might have different reasons. The combination of primers might not be suitable for PCR amplification under the protocol used in this preliminary test. The failure might be due to factors such as template nucleotide composition, formation of secondary structures in the template or primers, primer interactions, etc. Moreover, there might be polymorphisms in the primer-binding

sites in the samples tested, as we discover in the ctp2-cp4epsps element from sample GM05 (MON1445 cotton), or the inserted element might be incomplete in the sample lacking the primer-binging site. It is also possible that the element is not present in the reference samples in disagreement with the GMOseek matrix. Further validation studies are necessary to identify the causes of the missing amplifications.

Our preliminary results suggest that only minor changes are necessary for the implementation of the multiplex PCR. For example, some primers could be redesigned to increase the specificity of some target regions. The screening of additional samples from more species and food products is also recommended to guarantee the specificity of our approach. Future work should also include testing the multiplex PCRs under gradients of annealing temperatures to determine the optimal conditions for amplification. We will also perform sensitivity tests in order to determine the minimum amount of template DNA necessary for detecting the transgenic elements.

One of the most notorious advantages of our method is the screening of several target regions simultaneously, which presents a clear advantage over methods targeting a single region. In cases where one or more target regions fail to amplify, a correct GM detection is still possible based on the information from the remaining regions (when the GM event has more than one element).

The use of target regions of small length (<270 bp) allows the analysis of samples with degraded DNA, such as processed food samples. This feature is an advantage for detection of illegal GMO crops or to verify the labelling of food products.

The detection can be performed in a single reaction by using conventional laboratory equipment. We tested a mixture of the different target regions in order to verify if the amplified regions to be included in the multiplex can easily be separated on conventional and capillary electrophoresis. The amplified products were clearly separated in both electrophoretic systems, suggesting that they can be successfully used in a multiplex configuration. These results suggest that the method can be adopted by different laboratories for an efficient GM plant screening.

In conclusion, we confirmed previous studies that demonstrate the advantages of using multiplex PCR for species identifications. The simplicity of the assays and the fact that the detection is based on the amplification of several target regions are the two main advantages. The identification can be obtained using conventional agarose electrophoresis, with no need for expensive sequencing or real-time PCR apparatus. These multiplex PCR assays are sensitive and specific, although more tests are required to obtain greater precision in sensitivity and specificity. The direct detection of multiple elements in a single reaction is an improvement for the screening of samples. Our multiplex PCRs can serve as a complement to techniques routinely used in

laboratories or as a screening technique prior to other analyses, by reducing costs and increasing the accuracy of the detection.

References

Ahmed F.E. (2002) Detection of genetically modified organisms in foods. *TRENDS in Biotechnology* **20**, 215-23.

Bahrdt C., Krech A., Wurz A. & Wulff D. (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**, 2103-12.

Bandelt H.-J. & Salas A. (2012) Current next generation sequencing technology may not meet forensic standards. *Forensic Science International: Genetics* **6**, 143-5.

Benfey P.N. & Chua N.-H. (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* **250**, 959.

Block A., Debode F., Grohmann L., Hulin J., Taverniers I., Kluga L., Barbau-Piednoir E., Broeders S., Huber I. & Van den Bulcke M. (2013) The GMOseek matrix: a decision support tool for optimizing the detection of genetically modified plants. *BMC bioinformatics* **14**, 256.

Brookes G. (2008) The impact of using GM insect resistant maize in Europe since 1998. *International journal of biotechnology* **10**, 148-66.

Brookes G. & Barfoot P. (2015) Global income and production impacts of using GM crop technology 1996–2013. *GM crops & food*, 00-.

Bruce B.D. (2000) Chloroplast transit peptides: structure, function and evolution. *Trends in cell biology* **10**, 440-7.

Cankar K., Štebih D., Dreo T., Žel J. & Gruden K. (2006) Critical points of DNA quantification by real-time PCR–effects of DNA extraction method and sample matrix on quantification of genetically modified organisms. *BMC biotechnology* **6**, 37.

Carozzi N., Rabe S., Miles P., Warren G. & De Haan P. (2002) Novel insecticidal toxins derived from Bacillus thuringiensis crystal proteins. *Patent WO0215701*.

Crawley M., Brown S., Hails R., Kohn D. & Rees M. (2001) Biotechnology: Transgenic crops in natural habitats. *Nature* **409**, 682-3.

Doyle J.J. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem bull* **19**, 11-5.

Doyle J.J. (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-5.

Edgar R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-7.

Edwards M.C. & Gibbs R.A. (1994) Multiplex PCR: advantages, development, and applications. *Genome Research* **3**, S65-S75.

Elnifro E.M., Ashshi A.M., Cooper R.J. & Klapper P.E. (2000) Multiplex PCR: optimization and application in diagnostic virology. *Clinical microbiology reviews* **13**, 559-70.

Evans S.L. (2004) Producing proteins derived from genetically modified organisms for toxicology and environmental fate assessment of biopesticides. In: *The GMO Handbook* (pp. 53-83. Springer.

Franck A., Guilley H., Jonard G., Richards K. & Hirth L. (1980) Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21**, 285-94.

Gachet E., Martin G., Vigneau F. & Meyer G. (1998) Detection of genetically modified organisms (GMOs) by PCR: a brief review of methodologies available. *Trends in food science & Technology* **9**, 380-8.

Gašparič M.B., Tengs T., La Paz J.L., Holst-Jensen A., Pla M., Esteve T., Žel J. & Gruden K. (2010) Comparison of nine different real-time PCR chemistries for qualitative and quantitative applications in GMO detection. *Analytical and bioanalytical chemistry* **396**, 2023-9.

Gelvin S.B. (2003) Agrobacterium-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiology and molecular biology reviews* **67**, 16-37.

Gill S.S., Cowles E.A. & Pietrantonio P.V. (1992) The mode of action of Bacillus thuringiensis endotoxins. *Annual review of entomology* **37**, 615-34.

Guo J., Yang L., Chen L., Morisset D., Li X., Pan L. & Zhang D. (2011) MPIC: a high-throughput analytical method for multiple DNA targets. *Analytical chemistry* **83**, 1579-86.

Gupta M. & Ram R. (2004) Development of genetically modified agronomic crops. In: *The GMO Handbook* (pp. 219-41. Springer.

Hamels S., Glouden T., Gillard K., Mazzara M., Debode F., Foti N., Sneyers M., Nuez T.E., Pla M. & Berben G. (2009) A PCR-microarray method for the screening of genetically modified organisms. *European Food Research and Technology* **228**, 531-41.

Heide B.R., Drømtorp S.M., Rudi K., Heir E. & Holck A.L. (2008) Determination of eight genetically modified maize events by quantitative, multiplex PCR and fluorescence capillary gel electrophoresis. *European Food Research and Technology* **227**, 1125-37.

Henegariu O., Heerema N., Dlouhy S., Vance G. & Vogt P. (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* **23**, 504-11.

Hernández M., Rodríguez-Lázaro D., Zhang D., Esteve T., Pla M. & Prat S. (2005) Interlaboratory transfer of a PCR multiplex method for simultaneous detection of four genetically modified maize lines: Bt11, MON810, T25, and GA21. *Journal of agricultural and food chemistry* **53**, 3333-7.

Holst-Jensen A. (2009) Testing for genetically modified organisms (GMOs): Past, present and future perspectives. *Biotechnology advances* **27**, 1071-82.

Holst-Jensen A., Rønning S.B., Løvseth A. & Berdal K.G. (2003) PCR technology for screening and quantification of genetically modified organisms (GMOs). *Analytical and bioanalytical chemistry* **375**, 985-93.

Huber I., Block A., Sebah D., Debode F.d.r., Morisset D., Grohmann L., Berben G., Štebih D., Milavec M. & Žel J. (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**, 10293-301.

James C. (2013) *Global status of commercialized biotech/GM crops: 2013*. International Service for the Acquisition of Agri-biotech Applications (ISAAA) Ithaca, NY, USA.

James C. (2014) *Global status of commercialized biotech/GM crops: 2014*. International Service for the Acquisition of Agri-biotech Applications (ISAAA) Ithaca, NY, USA.

James D., Schmidt A.-m., Wall E., Green M. & Masri S. (2003) Reliable detection and identification of genetically modified maize, soybean, and canola by multiplex PCR analysis. *Journal of agricultural and food chemistry* **51**, 5829-34.

John M.E. (1992) An efficient method for isolation of RNA and DNA from plants containing polyphenolics. *Nucleic acids research* **20**, 2381.

Kibbe W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic acids research* **35**, W43-W6.

Kumar K.S. & Kang S.H. (2007) Ultra-fast simultaneous analysis of genetically modified organisms in maize by microchip electrophoresis with LIF detector. *Electrophoresis* **28**, 4247-54.

Lodhi M.A., Ye G.-N., Weeden N.F. & Reisch B.I. (1994) A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. *Plant Molecular Biology Reporter* **12**, 6-13.

Maeda H. & Dudareva N. (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annual review of plant biology* **63**, 73-105.

Miraglia M., Berdal K., Brera C., Corbisier P., Holst-Jensen A., Kok E., Marvin H., Schimmel H., Rentsch J. & Van Rie J. (2004) Detection and traceability of

genetically modified organisms in the food production chain. *Food and Chemical Toxicology* **42**, 1157-80.
Morisset D., Dobnik D., Hamels S., Žel J. & Gruden K. (2008a) NAIMA: target amplification strategy allowing quantitative on-chip detection of GMOs. *Nucleic acids research* **36**, e118-e.
Morisset D., Stebih D., Cankar K., Zel J. & Gruden K. (2008b) Alternative DNA amplification methods to PCR and their application in GMO detection: a review. *European Food Research and Technology* **227**, 1287-97.
Nadal A., Coll A., Paz L., Esteve T. & Pla M. (2006) A new PCR-CGE (size and color) method for simultaneous detection of genetically modified maize events. *Electrophoresis* **27**, 3879-88.
Narva K.E., Olson, P.D., Zhang, J., Sanders, C.D., Wang, J., Zhong, G., Moellenbeck, D., Meyer, S.E., Luckring, A.K., Bing, J.W., Jr, R.F.C., Hakimi, S.M., Gupta, M., Hondred, D., Krone, T.L., Locke, M.E.H. (2008) CORN EVENT DAS-59122-7 AND METHODS FOR DETECTION THEREOF.
Odell J.T., Nagy F. & Chua N.-H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter.
Parekh S.R. (2004) *The GMO handbook: genetically modified animals, microbes, and plants in biotechnology*. Springer Science & Business Media.
Permingeat H.R., Reggiardo M.I. & Vallejos R.H. (2002) Detection and quantification of transgenes in grains by multiplex and real-time PCR. *Journal of agricultural and food chemistry* **50**, 4431-6.
Peterson G., Cunningham S., Deutsch L., Erickson J., Quinlan A., Raez-Luna E., Tinch R., Troell M., Woodbury P. & Zens S. (2000) The risks and benefits of genetically modified crops: a multidisciplinary perspective. *Conservation Ecology* **4**, 13.
Pollegioni L., Schonbrunn E. & Siehl D. (2011) Molecular basis of glyphosate resistance–different approaches through protein engineering. *FEBS Journal* **278**, 2753-66.
Porebski S., Bailey L.G. & Baum B.R. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* **15**, 8-15.
Preuss S.B., Meister R., Xu Q., Urwin C.P., Tripodi F.A., Screen S.E., Anil V.S., Zhu S., Morrell J.A. & Liu G. (2012) Expression of the Arabidopsis thaliana BBX32 gene in soybean increases grain yield. *PloS one* **7**, e30717-e.
Purugganan M.D. & Fuller D.Q. (2009) The nature of selection during plant domestication. *Nature* **457**, 843-8.
Richins R.D., Scholthof H.B. & Shepherd R.J. (1987) Sequence of figwort mosaic virus DNA (caulimovirus group). *Nucleic acids research* **15**, 8451-66.
Sanger M., Daubert S. & Goodman R.M. (1990) Characteristics of a strong promoter from figwort mosaic virus: comparison with the analogous 35S promoter from cauliflower mosaic virus and the regulated mannopine synthase promoter. *Plant molecular biology* **14**, 433-43.
Stave J.W. (1999) Detection of new or modified proteins in novel foods derived from GMO–future needs. *Food control* **10**, 367-74.
Sun W., Zhong J., Qin P. & Jiao K. (2008) Electrochemical biosensor for the detection of cauliflower mosaic virus 35 S gene sequences using lead sulfide nanoparticles as oligonucleotide labels. *Analytical biochemistry* **377**, 115-9.
Taberlet P., Coissac E., Pompanon F., Gielly L., Miquel C., Valentini A., Vermat T., Corthier G., Brochmann C. & Willerslev E. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic acids research* **35**, e14-e.
Thompson C.J., Movva N.R., Tizard R., Crameri R., Davies J.E., Lauwereys M. & Botterman J. (1987) Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus. *The EMBO journal* **6**, 2519.

Vallone P.M. & Butler J.M. (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* **37**, 226-31.

Waiblinger H.-U., Ernst B., Anderson A. & Pietsch K. (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**, 1221-8.

Waiblinger H.-U., Grohmann L., Mankertz J., Engelbert D. & Pietsch K. (2010) A practical approach to screen for authorised and unauthorised genetically modified plants. *Analytical and bioanalytical chemistry* **396**, 2065-72.

Wehrmann A., Van Vliet A., Opsomer C., Botterman J. & Schulz A. (1996) The similarities of bar and pat gene products make them equally applicable for plant engineers. *Nature biotechnology* **14**, 1274-8.

Willems S., Fraiture M.-A., Deforce D., De Keersmaecker S.C., De Loose M., Ruttink T., Herman P., Van Nieuwerburgh F. & Roosens N. (2016) Statistical framework for detection of genetically modified organisms based on Next Generation Sequencing. *Food chemistry* **192**, 788-98.

Wohlleben W., Arnold W., Broer I., Hillemann D., Strauch E. & Punier A. (1988) Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from Streptomyces viridochromogenes Tü494 and its expression in Nicotiana tabacum. *Gene* **70**, 25-37.

Wu G., Wu Y., Nie S., Zhang L., Xiao L., Cao Y. & Lu C. (2010) Real-time PCR method for detection of the transgenic rice event TT51-1. *Food chemistry* **119**, 417-22.

Xu J., Miao H., Wu H., Huang W., Tang R., Qiu M., Wen J., Zhu S. & Li Y. (2006) Screening genetically modified organisms using multiplex-PCR coupled with oligonucleotide microarray. *Biosensors and Bioelectronics* **22**, 71-7.

Žel J., Milavec M., Morisset D., Plan D., Van den Eede G. & Gruden K. (2012) *How to reliably test for GMOs*. Springer.


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO