

This is a copy of the original manuscript, published on the website of the authors.

Non-prejudiced detection and characterization of genetic modifications

Running head: non-prejudiced GMO detection

Torstein Tengs¹, Anja Bråthen Kristoffersen^{2,3}, Haibo Zhang^{1,4}, Knut G. Berdal¹, Marie Løvoll¹ & Arne Holst-Jensen^{1*}

¹ - National Veterinary Institute, Section for Food Bacteriology and GMO,
PO Box 750 Sentrum, 0106 Oslo, Norway

² - National Veterinary Institute, Section for Epidemiology,
PO Box 750 Sentrum, 0106 Oslo, Norway

³ - University of Oslo, Department of Informatics,
PO Box 1080, Blindern, 0316 Oslo, Norway

⁴ - School of Life Science and Biotechnology, Shanghai Jiao Tong University,
800 Dongchuan Road, Shanghai 200240, China

*To whom correspondence should be addressed

Phone : +47-23216243

Fax : +47-23216202

Email: arne.holst-jensen@vetinst.no

This is a copy of the original manuscript, published on the website of the authors.

ABSTRACT

The application of gene technology is becoming widespread much thanks to the rapid increase in technology, resource and knowledge availability. Consequently, the diversity and number of genetically modified organisms (GMOs) that may find their way into the food chain or the environment, intended or unintended, is rapidly growing. From a safety point of view the ability to detect and characterize in detail any GMO, independent of publicly available information, is fundamental. Pre-release risk assessments of GMOs are required in most jurisdictions, and are usually based on application of technologies with limited ability to detect unexpected rearrangements and insertions. We present an array-based approach to address these problems and show with three examples (GTS 40-3-2 Roundup Ready and event A5547-127 soybean as well as T25 Liberty Link Maize) that the method can detect and characterize GMOs with high accuracy while making very few prior assumptions about the actual genetic modifications or constructs in question. Based on the array results, a simple polymerase chain reaction (PCR)-scheme is also described that will enable the user to characterize the inserted sequences to DNA sequence level. The method may provide the biotechnology developers and risk regulators with a useful tool to improve pre-market risk assessments as well as seed producers and other food chain and environmental stakeholders with a platform to improve their ability to detect and characterize GMOs.

Key words: DNA sequence; genetically modified organisms; transgene insert; unauthorised GMO; unknown GMO

This is a copy of the original manuscript, published on the website of the authors.

INTRODUCTION

There are many ways in which the heritable characteristics of a species can be improved. Selective breeding has led to the plethora of phenotypically distinct breeds of dogs, and the Belgian Blue cattle is an example of selective breeding where the genetic factors involved have been characterized in detail (Grobet et al. 1997). Other strategies rely on the mutagenic effect of radiation or chemical treatments in order to generate varieties that again can be screened and bred (for examples, see (Cecchini et al. 1998) and (Maple and Moller 2007)).

Most of the commercially available genetically modified organisms (GMOs) have been generated using more targeted and controlled molecular biology techniques. When a GMO is generated, the coding capacity of a recipient genome is altered. This is usually done by the introduction of a genetic construct. Such constructs can be pieced together using bacterial vector systems combined with cutting and pasting of genetic elements, and the final cassette will include such components as selective markers, trait genes and appropriate promoter and terminator elements. Selective markers are used in the bacterial part of the system in order to ensure that the correct genetic construct is carried by the bacterial clone chosen to propagate the plasmids that will eventually be used in the ultimate transformation, and a separate selection marker is sometimes needed to select for the final transgenic individual. The trait gene inserted is the element that encodes the genetic trait one wishes to introduce whereas the selection marker simplifies the process of eliminating individuals where the genetic cassette has not been inserted or where it has been inserted in a locus that does not give high levels of expression. In some cases, the trait gene itself encodes a gene product that will enable direct selection so that a separate selection marker is not necessary in the final step of the transformation. An example of this is the NK603 (Roundup Ready) Maize where resistance against the herbicide Roundup is the trait gene, and where expression of this gene also was used to allow selection of transgenic plants (see <http://www.agbios.com/docroot/decdocs/02-269-007.pdf>).

A wide array of methods has been developed for the stable introduction of genetic constructs in eukaryotic systems. Depending on the species, any number of strategies can be used to ensure that the genetic cassette becomes a covalently integrated part of the recipient nuclear or organellar genome. For animal cell lines, viral vectors, biolistic guns, electroporation, heat shock or carrier molecule such as cationic lipids can be used to induce transport of high molecular DNA across the cellular membrane (Bonetta 2005). For plants, the two most commonly used approaches are biolistic bombardment and *Agrobacterium-*

This is a copy of the original manuscript, published on the website of the authors.

mediated transformation. For biolistic transformation, a small particle (usually gold or some other sort of metal) is covered with DNA containing the constructs of interest and the DNA is physically shot into the recipient cell nucleus (Taylor and Fauquet 2002). When using *Agrobacterium*, one takes advantage of *Agrobacterium tumefaciens*' natural ability to transform plant cells; upon infection, natural strains of *A. tumefaciens* can transfer a segment of DNA called T-DNA (transfer DNA) from a tumor-inducing (Ti) plasmid to the plant chromosomal DNA and cause Crown Gall disease (Chilton et al. 1977). It was suggested early on that by changing the DNA sequence of the T-DNA, specific gene constructs rather than the natural T-DNA can be introduced into plant cells using this system (Joos et al. 1983) and there are now many sophisticated approaches to this (Komari et al. 2006).

Both with regards to hectareage and variety diversity, the amount of genetically modified crops grown is increasing (James 2008). In 2008, the number of countries planting biotech crops reached 25 (with 55 % of the world's population), and there was an overall increase of 9.4 % in GMO hectareage worldwide. Globally, 65.8 million hectares of herbicide tolerant soybean was grown and virtually all Argentinean soybean hectareage (18.1 million hectares) was dedicated to production of Roundup Ready Soybean (RRS) (James 2008).

In order to regulate import and export, assess safety, monitor farming and comply with international as well as national regulations, it is of great importance to monitor the presence and distribution of GMOs. Many different methods and strategies have been developed that focus on the detection and quantification of GMOs that have been approved (Hernandez et al. 2005; Rodriguez-Lazaro et al. 2007), but given the widespread use of GMOs, it is also important to be aware of the possibility that unknown, insufficiently characterized or unexpected GMOs might be present in the food chain. Recent examples of this include the contamination of Chinese and US rice supply by unapproved GM variants Bt63 and LL601 (Mäde et al. 2006; Akiyama et al. 2007; Vermij 2006), the distributions of unapproved Bt10 corn from the US (Macilwain 2005), the undocumented genetic alterations found in RRS (Windels et al. 2001), the presence of transgenic DNA constructs in native maize grown in the remote mountains of Mexico (Pineyro-Nelson et al. 2009) and the escape of GM creeping bentgrass (*Agrostis stolonifera*) from field trials in USA (Zapiola et al. 2008).

It is also possible that in the future harmful GMOs might be released, and to this end we have developed a microarray-based method for the detection of a large number of genetic elements that can be expected to be found in genetically altered plants. More than 2.7 million basepairs of sequence data relevant for plant biotechnology was targeted with a set of more

This is a copy of the original manuscript, published on the website of the authors.

than 93,000 probes. We show that we can detect and characterize to DNA sequence level both soy and maize GMOs with high accuracy. The method makes very few assumptions and seems to work well even when analyzing large genomes.

MATERIALS AND METHODS

Array design and synthesis

Using the same concept as described in (Tengs et al. 2007), high density microarrays were designed from sequences downloaded from publicly available databases. Briefly, all sequences annotated as ‘vectors’ in GenBank were downloaded, and from this database, all sequences containing at least one (partial) version of the Cauliflower Mosaic Virus (CaMV) P35S promoter were used. CaMV P35S is the most commonly used genetic element in genetic transformation of plants (Cummins et al. 2000; Feinberg et al. 2005; Fernandez et al. 2005) and was in this context treated as a marker for DNA sequences relevant for plant biotechnology. This approach resulted in a broad collection of sequences, and in addition various databases were mined to find full-length sequence data from the majority of trait genes listed in the GM database at the AGBIOS website (<http://www.agbios.com/main.php>). Trait genes included the *vip3A* gene and *cry* genes from *Bacillus thuringiensis*, the 5-enolpyruvylshikimate-3-phosphate synthase gene from *Agrobacterium tumefaciens* and other commercially used gene sequences. In addition, we included a series of genes that might be candidates for bioterrorism-related biotechnology, such as the toxin genes *ace*, *zot* and *ctx* from *Vibrio cholerae*, genes involved in ricin production in *Ricinus communis*, the anthrax lethal factor gene from *Bacillus anthracis* etc. (Supplementary Material 1).

Analyzing this collection of 343 sequences, it became obvious that there was a high degree of redundancy when looking at the genetic elements and sequence motifs. With the same approach as in (Tengs et al. 2007), probes were selected that would give maximum coverage for our set of sequences given the number of probes we could fit on our array (105,000 per array on a two-array slide). We chose a probe overlap scheme where the majority of the probes were tiled with 7 basepair intervals throughout both strands of the array design sequences. Arrays were synthesized by Agilent Technologies, Inc (Santa Clara, CA, USA) and contained a total of 93,128 unique 60-basepair probes tiled throughout both strands of 2,757,516 basepairs of target sequence. More than 99.99 % of the probes overlapped with other probes with 53 basepairs or more on both strands.

This is a copy of the original manuscript, published on the website of the authors.

Array hybridization and data analyses

Maize DNA was purchased from The American Oil Chemists' Society (AOCS; Urbana, IL, USA). Non-GM maize (AOCS 0306-C) was used as a reference and T25 (AOCS 0306-H; also referred to as P35S/bar or Liberty Link Maize event T25; OECD unique identifier [UI] = ACS-ZM003-2) was used as sample. Non-GM soybean reference DNA (AOCS 0707-A) and event A5547-127 soybean (UI = ACS-GM006-4) DNA (AOCS 0707-B) was also acquired from AOCS. Roundup Ready Soybean (event GTS 40-3-2; UI = MON-04032-6) seeds were obtained from Dr. Greg Conko (Competitive Enterprise Institute, Washington DC, USA).

The DNA was processed using standard Agilent reagents and protocols for array CGH (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis Protocol, Version 5.0). Briefly, 1 μ g of sample and reference DNA was digested using *RsaI* and *AluI*. After random priming-based labeling using Cy3 (sample) and Cy5 (reference), the samples were purified and mixed with hybridization reagents. Slides were hybridized for 40 hours at 65 °C, and signals intensities were exported using the program Feature Extraction, version 9.5.3.1 (Agilent Technologies, Inc).

The $\log_2 S$ was plotted against $\log_2 R$ for all probes (where S = sample signal and R = reference signal). Noise levels differed between arrays, and the following strategy for defining a noise cutoff was used: all probes with a $\log_2 R$ signal level above 10 were assumed to represent true signals and a region around the diagonal representing plus/minus three standard deviations of the $\log_2 S/\log_2 R$ ratio was calculated (for probes where $\log_2 R > 10$). The 75% quantile of the $\log_2 R$ outside this region was defined as a moderate noise level. All probes with $\log_2 R$ signals below this quantile were defined as noise. A horizontal line could then be drawn at the crossing point between the diagonal and a vertical line representing this noise cutoff (Figure 1). Finally, probes above this horizontal line were either defined as positive ($\log_2 S/\log_2 R$ above diagonal plus two standard deviations) or negative (below diagonal plus two standard deviations).

All the probes were successively mapped back onto the 343 sequences that the array design was based upon. Sequence windows could be extracted based on the positive/negative status of the individual probes. A detected sequence window was defined as a region containing at least 5 probes with more than 80 % scored as positive. The window expansion stopped whenever the fraction of positive probes fell below 80 % or when 5 or more consecutive negative probes were encountered. Sequence regions corresponding to trailing negative probes were trimmed off, and 25 bases from the start and end of the windows were

This is a copy of the original manuscript, published on the website of the authors.

also removed. This latter step was included to reduce the possibility of having sequence windows flanked by probes that were only partial matches, as it was observed that some non-target probes that had more than ~30 perfectly matching (consecutive) bases could give rise to false positives.

PCR confirmation of detected windows

The windows detected were sorted according to length (Supplementary Material 2). Starting with the longest window reported, sequences were checked for similarity and overlap using 'Blast 2 sequences' (Tatusova and Madden 1999). Sequences were also screened for repetitiveness using Dot Plots (<http://www.vivo.colostate.edu/molkit/dnadot/>), and the three longest windows with non-overlapping sequences were used for primer design using the software Primer Express Version 2.0.0 (Applied Biosystems Inc, Foster City, CA, USA). Minimal amplicon length was set to 200 basepairs and maximum length was set to length of window (see Table 1 for primer sequences).

PCR amplifications were performed using approximately 100 ng of template DNA (wildtype and GMO) in 50 µl reactions and the following PCR cycle: 10 minutes at 95 °C (denaturation/polymerase activation), 35 cycles of 25 seconds at 96 °C (denaturation)/30 seconds at 60 °C (annealing)/35 seconds at 72 °C (elongation) and a final elongation step of 7 minutes at 72 °C. Final concentration of each primer was 500 nM and each dNTP was at 25 µM concentration. All reactions except for T25 primer pair 2F/202R were done using 1 unit of AmpliTaq Gold DNA polymerase, 1 × GeneAmp PCR Gold Buffer and [MgCl₂] 1.5 mM (Applied Biosystems Inc, Foster City, CA, USA). The 2F/202R primer pair was used with 'Taq DNA Polymerase DNA-free' and 1 × 'Reaction buffer complete' (AppliChem, Darmstadt, Germany).

Sequencing of the event A5547-127 insert

Primers were also designed based on the windows detected in the event A5547-127 soybean experiment (Table 1). The largest window found was quite large (1,781 basepairs), and this sequence was split into two 800+ basepair fragments before three primer pairs were designed as described above (two primer pairs for the longest fragment and one pair for the second largest, non-overlapping window). In addition, reverse-complementary versions of all six primers were also synthesized. Instead of performing PCRs with the corresponding primer pairs, these primers were used in a combinatorial way to amplify regions flanked by the

This is a copy of the original manuscript, published on the website of the authors.

original PCR target regions. All primer combinations except for F/R versions of the same primer (for instance 50F/50R) were tested and products run on agarose gels. When present, PCR products from the first round of amplification reactions were purified using ExoSAP-IT (USB Europe GmbH, Staufen, Germany) and sequenced using a 3130xl-Avant Genetic Analyzer (Applied Biosystems) and the BigDye Terminator v3.1 Cycle Sequencing Kit according to manufacturer's recommendations. Sequences were assembled using the software BioEdit v7.0.9.0 (Hall 1999). Sequencing reactions were performed using the original PCR primers in addition to oligos designed for primer walking across the amplicons. Finally, two individual primers were designed in order to amplify the 5' and 3' ends of the GMO cassette (used in combination with internal primers; primers PM_F and PM_R in Table 1)).

RESULTS

The background noise levels were found to be 8.13, 7.48 and 8.31 for the RRS, T25 and event A5547-127 experiments, respectively. The standard deviation also differed between the three arrays and was estimated to be 0.45, 0.22 and 0.43 (Figure 1).

Because of the sequence redundancy in the vector database, many identical windows as well as windows completely encompassed by other windows were reported. Redundant sequences were removed and sequence data for both our constructs (T25 and RRS) were downloaded from the GMO Detection Method Database (GMDD; <http://gmdd.shgmo.org/>) (Dong et al. 2008). Our probes and windows were compared against these reference sequences using BLAST (Altschul et al. 1990). Both arrays had true positive rates above 90 %, and a large number of sequence windows were detected (Table 2). Probes could be mapped back to the target sequences, and the majority of both constructs had good coverage by the array probes (Figure 2). For RRS, most of the windows had bitscores close to the theoretical maximum given the length of the sequence windows detected, but for T25, a fraction of the windows matched the reference sequence downloaded from GMDD more poorly (Table 2).

When going through the literature on T25, we noticed that in the earliest publication describing the transformation cassette it is indicated that the inserted sequence comprises DNA beyond the construct that is described in the GMDD sequence database. It appears that the sequence is flanked by (among other fragments), several hundred basepairs of 'initial' and 'terminal' part of the pUC18 cloning vector (Collonnier et al. 2005). This likely explains the

This is a copy of the original manuscript, published on the website of the authors.

low BLAST scores seen, as the reference sequence used by us for comparison was incomplete.

The amplicons produced by PCR (Figure 3) were of the correct size, and their identities were confirmed using single-pass sequencing (data not shown). A faint band was observed in both the no-template and the wildtype DNA reaction for the T25 primer pair 2F/202R. This was probably due to the fact that the 2F/202R target sequence is a vector backbone sequence (pUC18), and these types of sequences are common contaminants of DNA polymerases (see e.g. (Holst-Jensen et al. 2003)). Multiple 'DNA free' polymerases were tested, and we also set up the reaction with great care in an accredited diagnostics lab using fresh reagents. We concluded that the most likely source of contamination was the DNA polymerases used in spite of them being marketed as free of DNA.

Although no direct sequence information was available for the A5547-127 insert, some features had been described (see for instance the entry for this event in GMDD; <http://gmdd.shgmo.org/event/view/39>). The insert was generated by linearization of a circular vector through restriction enzyme cutting in the *bla* (beta-lactamase) selection marker gene. We thus decided to bidirectionally characterize the insert until we reached the *bla* fragments. A total of 3,210 basepairs of assembled sequences was generated (EMBL/GenBank accession number GQ497217) and a construct map could be generated (Figure 2).

DISCUSSION

We have used tiling arrays to characterize the inserted GMO cassette for three different scenarios: a well characterized event (GTS 40-3-2 [RRS] soybean), a partially characterized event (T25 maize) and an event where no sequence information was available (event A5547-127 soybean). Based on the information acquired from the array analysis, primers can be designed either for the purpose of designing detection assays or amplification of inserted sequence for further characterization (sequencing). For determination of flanking regions, an anchor PCR-based approach (Nielsen et al. 2008) may be used and possible unintended insertion events might also be detected (Windels et al. 2001).

Several incidents of illegal release or unintended contamination of food crops over the last decade illustrate that the food chain and the environment are vulnerable to release of unknown GMOs. The majority of the reported incidents concern GM events authorized in one (exporting) jurisdiction or for one type of use (e.g. feed) but not authorized in another (importing) jurisdiction or for another type of use (e.g. food). Examples include CBH 351

This is a copy of the original manuscript, published on the website of the authors.

Starlink, MIR604, DAS 59122 and MON88017 maize, 55-1/63-1 papaya and A2704-12 soybean (http://ec.europa.eu/food/food/rapidalert/index_en.htm). However, a few of the reported incidents concern GM events that were not authorized in any jurisdiction, e.g. Event 32 and Bt10 maize (<http://gmo-crl.jrc.ec.europa.eu/E32update.htm>; Macilwain 2005), LL601 (Vermij 2006) and Bt63 rice (Mäde et al. 2006; Akiyama et al. 2007). The presence of illegal unauthorized GMO is a global challenge and is not limited to plants or plant derived products (<http://www.gm-inspectorate.gov.uk/gmfish/GMfish2006.cfm>; (Anonymous 2004); Rehbein and Bogerd 2007).

European legislation requires authorization of any GMO or derived material prior to release or use in food or feed (European Commission 2001; European Commission 2003). Both the EU and USA currently have zero tolerance for unauthorized GMOs. Due to labeling requirements if the concentration of authorized GMOs or derived products exceed a threshold of 0.9% and is technically unavoidable (European Commission 2003), GMO detection, identification and quantification methods are well established in Europe. Presence of unauthorized GMOs is a major concern in Europe, Japan and the USA (Mano et al. 2009; Vermij 2006); <http://www.coextra.eu>). This is not solely a consequence of the immediate impact on trade and economy (European Commission 2007; Krueger and Le Buanec 2008) but also of the possible risks to health and the environment (Holst-Jensen 2008).

The methods used today for GMO detection and characterization are almost exclusively based on PCR. PCR can be used as a screening tool where GMO sequences such as commonly used antibiotic resistance markers or trait genes are targeted, it can be used to target specific genetic constructs, or primers can be designed to amplify the border between the inserted sequence and the genomic flank. The latter types of assays are referred to as event-specific, and this category of assays are considered to be the most specific (Holst-Jensen et al. 2003). To ensure maximum selectivity, a fluorescent probe such as TaqMan or MGB can also be included, or amplified products can be labeled and separated using capillary gel electrophoresis (Nadal et al. 2006). In some cases, PCR products are hybridized to arrays to increase throughput and increase the chances of detecting a broad range of GMOs e.g. (Hamels et al. 2009; Leimanis et al. 2008; Xu et al. 2006; Leimanis et al. 2006) and there are also other, more sophisticated methods that might alleviate some of the problems associated with multiplex PCR (see for instance (Chaouachi et al. 2008)).

Although a broad range of different PCR assays can be used to detect GMO events with equal sensitivity and specificity (Andersen et al. 2006), PCR is inherently limited by

This is a copy of the original manuscript, published on the website of the authors.

primer design and the problem of running multiple different amplifications in a reasonable number of reaction volumes. Although multiplexing seems to be able to handle a fair number of distinct amplification reactions in a single reaction tube (Heide et al. 2008), there will be an upper limit to how many different amplifications that can take place in a single reaction vessel using conventional PCR. Since primer design is an essential part of any PCR-based method, these approaches will thus have limitations when it comes to unknown GMOs. We believe that our array-based approach makes fewer assumptions than the currently available PCR methods when working with GMOs where construct information is scarce or completely lacking.

For our GMO tiling arrays to work, elements corresponding to parts of the target construct must have been sequenced and the sequence must have been included in the array design. Although our current design lacks some elements that might be desirable for optimal results when looking at certain commercial GMOs, such as the phosphinotricin-N-acetyltransferase (PAT gene) used in T25, we believe our probes cover a broad range of what can be expected to be found in unknown/poorly characterized GMOs. Using the PAT gene further as an example, information about genetic elements that are not part of the array design can also easily be obtained if flanking regions are detected (as seen in our Event A5547-127 results). The main reason why a relatively small number of probes can cover many different GMO constructs with high probe density is the high degree of recycling of DNA sequences seen in gene construct designs. Our arrays cover both strands of 2,757,516 basepairs of sequences. With a tiling of one probe per 7 bases, this would have yielded $2 \times 2,757,516 / 7 = 787,861$ probes if there was no sequence motif redundancy in our sequence database. The fact that the number of probes needed is only about 12% of this figure substantiates our claim that even though particular trait genes might evade detection, other genetic elements (promoters, terminators, vector backbone, polylinker regions etc.) are likely to be known and have publicly available sequence data associated with them. The vast majority of GMOs developed are generated so that they contain at least one of the most commonly applied genetic elements after transformation (e.g. the Cauliflower Mosaic Virus 35S promoter, an herbicide tolerance or insect resistance trait gene or a pUC-derived vector element). The cost associated with both array design and production has dropped dramatically in recent years and the number of features that can be included on a single array has increased, so we believe that future generations of arrays might give even more complete coverage of target sequences.

This is a copy of the original manuscript, published on the website of the authors.

The method will have some limits when it comes to analysis of GMO material that is not 100% pure. Results from analogous studies using microarrays for gene-copy number changes indicates that presence/absence of a locus (homozygous deletion) requires material with higher than 90% purity (Bhattacharjee et al. 2001). Our previous generation of (single-color) arrays and sample processing protocols were hampered by problems associated with large genomes and poor probe performance, but with this next generation of arrays, we believe that we have solved both of these problems and we hope that this technology will aid us and others when working with GMOs. In addition to the characterization of unknown GMOs, we also believe that the method might be useful as a tool for checking newly developed GMO for unintended transformation events such as fragments of allochthonous DNA scattered in the genome, as seen in for instance SunUp papaya (Ming et al. 2008). The approach should also be useful when working with genetically modified organisms other than plants, such as animals, fungi or bacteria.

ACKNOWLEDGEMENT

The authors would like to thank Ana Heslanyeth Barragan Polania for help with array processing and Jon Bohlin and Ole Christian Lingjærde for assistance with data analysis. This work was financially supported by the European Commission through the Sixth Framework Program, integrated project Co-Extra (<http://www.coextra.eu>; contract FOOD-2005-CT-007158) and by the Research Council of Norway (grants 170363/D10 and 178288/I10). This is gratefully acknowledged.

REFERENCES

- Akiyama H, Sasaki N, Sakata K, Ohmori K, Toyota A, Kikuchi Y, Watanabe T, Furui S, Kitta K, Maitani T (2007) Indicated detection of two unapproved transgenic rice lines contaminating vermicelli products. *J Agric Food Chem* 55:5942-5947
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410
- Andersen CB, Holst-Jensen A, Berdal KG, Thorstensen T, Tengs T (2006) Equal performance of TaqMan, MGB, molecular beacon, and SYBR green-based detection assays in detection and quantification of Roundup Ready soybean. *J Agric Food Chem* 54:9658-9663
- Anonymous (2004) Pig feed blunder. *New Scientist* 2436:4
- Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98:13790-13795
- Bonetta L (2005) The inside scoop-evaluating gene delivery methods. *Nature Methods* 2:875
- Cecchini E, Mulligan BJ, Covey SN, Milner JJ (1998) Characterization of gamma irradiation-induced deletion mutations at a selectable locus in *Arabidopsis*. *Mutat Res* 401:199-206
- Chaouachi M, Chupeau G, Berard A, McKhann H, Romaniuk M, Giancola S, Laval V, Bertheau Y, Brunel D (2008) A high-throughput multiplex method adapted for GMO detection. *J Agric Food Chem* 56:11596-11606
- Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP, Nester EW (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell* 11:263-271
- Collonnier C, Schattner A, Berthier G, Boyer F, Coue-Philippe G, Diolez A, Duplan MN, Fernandez S, Kebdani N, Kobilinsky A, Romaniuk M, de Beuckeleer M, De Loose M, Windels P, Bertheau Y (2005) Characterization and event specific-detection by quantitative real-time PCR of T25 maize insert. *J AOAC Internat* 88:536-546
- Cummins J, Ho MW, Ryan A (2000) Hazardous CaMV promoter? *Nat Biotechnol* 18:363
- Dong W, Yang LT, Shen KL, Kim B, Kleter GA, Marvin HJP, Guo R, Liang WQ, Zhang DB (2008) GMDD: a database of GMO detection methods. *BMC Bioinformatics* 9:260
- European Commission (2003) Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. *Official Journal of the European Union* L268:1-23
- European Commission. Economic impact of unapproved GMOs on EU feed imports and livestock production. Directorate-General for Agriculture and Rural Development. 2007.
- European Commission (2001) Directive 2001/18/EC of the European Parliament and the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. *Official Journal of the European Communities* L106:1-38
- Feinberg M, Fernandez S, Cassard S, Charles-Delobel C, Bertheau Y (2005) Quantitation of 35S promoter in maize DNA extracts from genetically modified organisms using real-time polymerase chain reaction, part 2: Interlaboratory study. *J AOAC Internat* 88:558-573
- Fernandez S, Charles-Delobel C, Geldreich A, Berthier G, Boyer F, Collonnier C, Coue-Philippe G, Diolez A, Duplan MN, Kebdani N, Romaniuk M, Feinberg M, Bertheau Y (2005) Quantification of the 35S promoter in DNA extracts from genetically modified

- organisms using real-time polymerase chain reaction and specificity assessment on various genetically modified organisms, part I: Operating procedure. *J AOAC Internat* 88:547-557
- Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Menissier F, Massabanda J, Fries R, Hanset R, Georges M (1997) A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet* 17:71-74
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95-98
- Hamels S, Glouden T, Gillard K, Mazzara M, Debode F, Foti N, Sneyers M, Nuez TE, Pla M, Berben G, Moens W, Bertheau Y, Audeon C, Van den Eede G, Remacle J (2009) A PCR-microarray method for the screening of genetically modified organisms. *Europ Food Res Technol* 228:531-541
- Heide BR, Dromtorp SM, Rudi K, Heir E, Holck AL (2008) Determination of eight genetically modified maize events by quantitative, multiplex PCR and fluorescence capillary gel electrophoresis. *Europ Food Res Technol* 227:1125-1137
- Hernandez M, Rodriguez-Lazaro D, Ferrando A (2005) Current methodology for detection, identification and quantification of genetically modified organisms. *Curr Analyt Chem* 1:203-221
- Holst-Jensen A (2008) GMO testing: trade, labeling or safety first? *Nat Biotechnol* 26:858-859
- Holst-Jensen A, Ronning SB, Lovseth A, Berdal KG (2003) PCR technology for screening and quantification of genetically modified organisms (GMOs). *Anal Bioanal Chem* 375:985-993
- James, C. (2008). Global status of commercialized biotech/GM crops: 2008. ISAAA Brief No. 39. (Ithaca, New York, USA.: ISAAA).
- Joos H, Timmerman B, Montagu MV, Schell J (1983) Genetic analysis of transfer and stabilization of *Agrobacterium* DNA in plant cells. *EMBO J* 2:2151-2160
- Komari T, Takakura Y, Ueki J, Kato N, Ishida Y, Hiei Y (2006) Binary vectors and super-binary vectors. *Methods Mol Biol* 343:15-41
- Krueger R and Le Buanec B (2008) Action needed to harmonize regulation of low-level presence of biotech traits. *Nat Biotechnol* 26:161-162
- Leimanis S, Hamels S, Naze F, Mbella GM, Sneyers M, Hochegger R, Broll H, Roth L, Dallmann K, Micsinai A, La Paz JL, Pla M, Brunen-Nieweler C, Papazova N, Taverniers I, Hess N, Kirschneit B, Bertheau Y, Audeon C, Laval V, Busch U, Pecoraro S, Neumann K, Rosel S, van Dijk J, Kok E, Bellocchi G, Foti N, Mazzara M, Moens W, Remacle J, Van den Eede G (2008) Validation of the performance of a GMO multiplex screening assay based on microarray detection. *Europ Food Res Technol* 227:1621-1632
- Leimanis S, Hernandez M, Fernandez S, Boyer F, Burns M, Bruderer S, Glouden T, Harris N, Kaeppli O, Philipp P, Pla M, Puigdomenech P, Vaitilingom M, Bertheau Y, Remacle J (2006) A microarray-based detection system for genetically modified (GM) food ingredients. *Plant Mol Biol* 61:123-139
- Macilwain C (2005) US launches probe into sales of unapproved transgenic corn. *Nature* 434:423
- Mäde D, Degner C, Grohmann L (2006) Detection of genetically modified rice: a construct-specific real-time PCR method based on DNA sequences from transgenic Bt rice. *Europ Food Res Technol* 224:271-278
- Mano J, Shigemitsu N, Futo S, Akiyama H, Teshima R, Hino A, Furui S, Kitta K (2009) Real-time PCR array as a universal platform for the detection of genetically modified

- crops and its application in identifying unapproved genetically modified crops in Japan. *J Agric Food Chem* 57:26-37
- Maple J and Moller SG (2007) Mutagenesis in *Arabidopsis*. *Methods Mol Biol* 362:197-206
- Ming R, Hou S, Feng Y, Yu Q, onne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Perez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991-996
- Nadal A, Coll A, La Paz JL, Esteve T, Pla M (2006) A new PCR-CGE (size and color) method for simultaneous detection of genetically modified maize events. *Electrophoresis* 27:3879-3888
- Nielsen CR, Berdal KG, Holst-Jensen A (2008) Anchored PCR for possible detection and characterisation of foreign integrated DNA at near single molecule level. *Europ Food Res Technol* 226:949-956
- Pineyro-Nelson A, Van Heerwaarden J, Perales HR, Srratos-Hernandez JA, Rangel A, Hufford MB, Gepts P, Garay-Arroyo A, Rivera-Bustamante R, varez-Buylla ER (2009) Transgenes in Mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. *Mol Ecol* 18:750-761
- Rehbein H, Bogerd J (2007) Identification of genetically modified Zebrafish (*Danio rerio*) by protein- and DNA-analysis. *Journal für Verbraucherschutz und Lebensmittelsicherheit* 2:122-125
- Rodriguez-Lazaro D, Lombard B, Smith H, Rzezutka A, D'Agostino M, Helmuth R, Schroeter A, Malorny B, Miko A, Guerra B, Davison J, Kobilinsky A, Hernandez M, Bertheau Y, Cook N (2007) Trends in analytical methodology in food safety and quality: monitoring microorganisms and genetically modified organisms. *Trends Food Sci Technol* 18:306-319
- Tatusova TA and Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247-250
- Taylor NJ and Fauquet CM (2002) Microparticle bombardment as a tool in plant science and agricultural biotechnology. *DNA Cell Biol* 21:963-977
- Tengs T, Kristoffersen AB, Berdal KG, Thorstensen T, Butenko MA, Nesvold H, Holst-Jensen A (2007) Microarray-based method for detection of unknown genetic modifications. *BMC Biotechnol* 7:91
- Vermij P (2006) Liberty Link rice raises specter of tightened regulations. *Nat Biotechnol* 24:1301-1302
- Windels P, Taverniers I, Depicker A, Van Bockstaele E, De Loose M (2001) Characterisation of the Roundup Ready soybean insert. *Europ Food Res Technol* 213:107-112
- 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* 22:71-77

This is a copy of the original manuscript, published on the website of the authors.

Zapiola ML, Campbell CK, Butler MD, Mallory-Smith CA (2008) Escape and establishment of transgenic glyphosate-resistant creeping bentgrass *Agrostis stolonifera* in Oregon, USA: a 4-year study. *Journal of Applied Ecology* 45:486-494

This is a copy of the original manuscript, published on the website of the authors.

FIGURE LEGENDS

Figure 1. Example of signal plot (GTS 40-3-2 Roundup Ready soybean [RRS] experiment). Blue circles indicate probes that match the RRS construct. The outer grey lines indicate the diagonal plus/minus three standard deviations of the difference in \log_2 signal for probes with \log_2 reference signal above 10. The black horizontal line represents the background noise level while the diagonal black line is the diagonal (central grey line) plus two standard deviations.

Figure 2. GMO cassettes of GTS 40-3-2 (RRS) and event A5547-127 soybean and T25 maize analyzed with arrays. Blue regions indicate probe coverage with complete overlap for both strands. Red hatch marks indicate false negative probes with orientations according to strandness. Primers have also been indicated (Table 1). Construct labeling terminology was adopted from GMDD (Dong et al. 2008):

35S - Cauliflower Mosaic Virus (CaMV) P35S promoter.

CTP4 - chloroplast transit peptide element 4.

EPSPS - 5-enolpyruvyl shikimate-3-phosphate synthetase.

NOS - nopaline synthase terminator.

EPSPStrunc – truncated EPSPS.

CaMV – fragment of CaMV genome.

PUC18 – pUC18 cloning vector.

PUC19 – pUC19c cloning vector.

PAT – phosphinothricin-N-acetyltransferase.

T35S - CaMV 35S terminator.

bla – beta-lactamase.

Figure 3. PCR products run on 1.1 % agarose gel with pUC Mix Marker 8 (Fermentas International Inc, Ontario, Canada). Lanes 1-3; DNA from GTS 40-3-2 soybean (RRS) amplified with primers designed from the three longest sequence windows detected in the RRS sample (21F/21R, 41F/241R and 10F/210R). Wildtype DNA was used as template in reactions 4-6. The last six lanes have corresponding amplicons from T25 maize (primers 169F/369R, 116F/316R and 2F/202R). Note the relative faintness of the band seen for wildtype DNA in lane 10 compared to the band for T25 in lane seven.

TABLES

Table 1. Primers used to amplify detected sequence windows.

Target	Name*	Sequence (3'-5')
RRS	21F	CGCAAATCCTCTGGCCTTT
	221R	GATGATCCAGGTGTGCGCCTT
	41F	CTTAAGATTGAATCCTGTTGCCG
	241R	GCGCGCGATAATTTATCCTAGTT
	10F	GGAAAAGGAAGGTGGCTCCT
	210R	GATTGTGCGTCATCCCTTACG
T25	169F	AAAAGCATCTTACGGATGGCA
	369R	CGTCGTTTGGTATGGCTTCA
	116F	ACCTCCTCGGATTCCATTGC
	316R	TTGAAGACGTGGTTGGAACGT
	2F	CCTGACGGGCTTGTCTGCT
	202R	GCACATTTCCCCGAAAAGTG
Event A5547-127	50F	ACCCAGAAACGCTGGTGAAA
	50R	TTTACCAGCGTTTCTGGGT
	250F	GGGCAAGAGCAACTCGGTC
	250R	GACCGAGTTGCTCTTGCCC
	398F	CGGGTTGGACTCAAGACGAT
	398R	ATCGTCTTGAGTCCAACCCG
	598F	GCAGGGTCGGAACAGGAGA
	598R	TCTCCTGTTCCGACCCTGC
	68F	CCGTAAGACTGGCGAACAGT
	68R	ACTGTTCGCCAGTCTTTACGG
	519F	TCCTTCGCAAGACCCTTCC
	519R	GGAAGGGTCTTGCGAAGGA
	PM_F**	AGTTGGCCGCAGTGTTATCAC
	PM_R**	GCGGATAAAGTTGCAGGACC

* - numbers represent nucleotide position within the detected sequence windows.

** - primers for amplifying terminal regions.

Table 2. Array performance.

	RRS	T25
Total number of probes	93128	93128
Number of probes that match construct	474	716
True positives	439 (92.6%*)	653 (91.2%)
False positives	2070 (2.22%**)	3929 (4.2%)
Number of unique windows detected***	87	137
Window length	31-1179	31-1756
Average BLAST score (percent of max)****	90.8	63.7

* - relative to number of probes that match construct

** - relative to total number of probes on the array

*** - windows where sequence is not identical to or totally encompassed by another positive window.

**** - for each window length detected, a perfectly matching sequence of the same length was extracted from the array target sequence. These sequences were subjected to BLASTn analysis against their target sequence, and bitscores obtained were used as a theoretical maximum for the detected windows.

This is a copy of the original manuscript, published on the website of the authors.

Additional material:

Supplementary material 1 is a file containing the sequences that were used as a basis for probe design. <Insert URL for readers to download file here>.

Supplementary material 2 is a file containing the sequence windows deducted from the array hybridisation analysis of GTS 40-3-2 and A5547-127 soybean and T25 maize, respectively. <Insert URL for readers to download file here>.