

# Detection and Traceability Technologies to Underpin the GM Inspectorate

T. R. Allnutt, K. Roper, C. Thomas, S. Hugo, G. Kerins  
GM Enforcement Team, CSL, York, YO41 1LZ, UK

## Executive summary

1. The work described in this report has been undertaken for Defra by the GM Enforcement team at the Central Science Laboratory (CSL), Sand Hutton, York. This team provides research support to Defra's GM Inspectorate (GMI), also based at CSL, in addition to undertaking other applied research and development work in the field of GM detection methodology, gene flow and risk assessment.
2. Defra GM Policy, Science and Regulation Unit commissioned the four projects described here to support development of sound, science-based and expedient enforcement of the GM deliberate release legislation by the GM Inspectorate. The report details the results of research carried out in 2003 – 2005 in four discrete but complementary areas; collectively the results provide Defra and CSL with confidence that we have the ability to detect, identify and characterise GM events, whether they are authorised or unauthorised, known or unknown.
3. The GM Inspectorate has authority under part VI of the Environmental Protection Act 1990 for inspection and enforcement of the release and marketing of genetically modified organisms (GMOs) in England<sup>\*,†</sup>. The National Assembly for Wales has given separate authority in respect of inspection and enforcement in Wales. CSL has been responsible for these inspection and enforcement functions in respect of deliberate release of GMOs since June 2000.
4. The role of the GMI is to ensure that the legislation concerning the escape or deliberate release of GMOs is complied with; in support of this the Inspectorate's work falls into two main areas – 1) ensuring compliance with consents issued for deliberate release of GMOs and 2) audits of seed producers and importers to ensure they are taking appropriate steps to prevent the adventitious presence of

---

\* In the European Union (EU) the release of GMOs is restricted under EU Council Directive 2001/18/EC on the deliberate release into the environment of genetically modified organisms, this covers releases for small scale experimental purposes and placing on the market. In Great Britain Directive 2001/18/EC has been implemented by the Environmental Protection Act 1990 (part VI) and regulations made under that Act, in respect of England the Genetically Modified Organisms (Deliberate Release) Regulations 2002.

† Regulation (EC) 1829/2003 (April 2004) creates a harmonised EU-wide approach for authorisation of placing GM food and feeds on the market, this may include authorisation for cultivation on a commercial scale. Regulation (EC) 1830/2003 (April 2004) introduced legally enforceable requirements for labelling and traceability of GMOs.

GM events in conventional seed<sup>‡</sup>. Within these two areas the GMI must also identify where potential breaches of consent conditions or other enforcement issues may have occurred and investigate these on a case-by-case basis.

5. The work of GM inspectors is largely field-based with inspections of deliberate release sites during trial and post-trial periods, audits of consent holders to ensure their management of deliberate release consents is appropriate, and paper-based audits of companies importing and/or producing conventional seeds. The remit of the GM Inspectorate does not extend to routine sampling and testing. However, samples may be taken where inspectors have evidence that there has been inadequate control of consent conditions (for example where ‘volunteer’ plants have not been controlled or seeds have been inappropriately shed), where there are concerns that the incorrect GMO has been released under experimental conditions, or where inspectors have good reason to believe that seed that is being marketed may contain a GMO that is not authorised, not labelled or is inappropriately labelled. GM Inspectors themselves do not undertake GM testing, but are reliant upon the support of skilled diagnosticians and researchers who provide the necessary skills base to undertake molecular and other analyses of samples at the request of the GMI. For initial routine analysis of samples the GMI employs the services of CSL’s commercial GM testing unit, which provides a UKAS accredited service for routine diagnosis of GMOs in food, feed and seeds. Where this service is unable to provide definitive answers to a suspected GM presence, the investigation is passed to researchers in CSL’s GM Enforcement team to resolve.
6. Scenarios in which such ‘forensic’ research would need to be employed are, for example, where the presence of a genetic element/s has been confirmed in a sample, but it cannot be attributed to an authorised or a known GM event or to commonly encountered sources of environmental contamination. In such circumstances the ability to characterise the event in terms of DNA sequence is crucial to establishing whether the GM legislation has been breached. Such characterisation is also needed to underpin any enforcement action that may need to be taken and to ensure that potential harm to human or animal health or the environment that may occur as a result of the release of an unauthorised or unknown GMO is limited. Looking to the future, commercial cultivation of GM crops is a realistic prospect for the UK and the ability to demonstrate that the coexistence legislation can be enforced will prove invaluable to the Regulatory Authorities in gaining confidence from the farming community and other stakeholders.
7. The four areas of work that are described cover the following areas:
  - i) Improved PCR primer sets for plant GM event identification.
  - ii) A rapid protocol for isolation and DNA sequencing of GM event flanks.
  - iii) Investigation of effect of GM genotype on real-time PCR GM quantification.

---

<sup>‡</sup> There is currently no legislation regarding thresholds for adventitious presence of GMOs in conventional seeds, but these are expected in the near future; these statutory thresholds will be enforceable and Defra will require access to technologies to resolve any issues that arise in connection with enforcement. In the absence of specific legislation, tolerance for GMOs in seed is dictated by the requirements of EU Directive 2001/18/EC.

- iv) Molecular methods for oilseed rape variety identification and GM traceability.
8. A description of the rationale and the objectives for each discrete research area is provided together with key results and recommendations arising from each project. The specific methods employed, which are a key outcome of this work, are detailed as standard operating procedures (SOPs) in Appendix 2.
  9. Results are reported according to the objectives set out in the original project descriptions. Key outputs include: i) a set of 96 PCR primers capable of identifying 90% of GM elements in released events; ii) a rapid protocol for the amplification, isolation and DNA sequencing of GM event flanking sequences; iii) detailed data on the performance of real-time PCR in relation to the variable copy number of GM events in different maize kernel tissues; and iv) a new DNA marker system for the identification of oilseed rape varieties for use in the study of GM coexistence and traceability. The completion of this work provides Defra and the GM Inspectorate with assurance that methodologies are available to characterise GM events that have been created utilising conventional technology, but which cannot be immediately characterised as a known GM event.
  10. We describe the development and use of matrix PCR technology in which large numbers of primers are utilised to build up a 'fingerprint' of a GMO to provide information on the structure of the event. As mentioned above, a suite of primers has been developed together with a protocol for their use that will enable identification of 90% of GM elements in currently released events. In addition, using matrix PCR technology in combination with routine screening for known events, the presence of mixtures or stacked events can also be readily identified. In a separate project the development of protocols for isolation and sequencing of regions flanking GM events utilising suppression PCR provides the ability to establish whether an event is unequivocally a specific GM event, and to establish that a single GM event is not, for example, due to environmental contamination (although further work is required to improve sensitivity for this application). It also enables us to ascertain information about the stability of a GM event and any hazard this might pose.
  11. The introduction of GM thresholds for labelling (Regulation (EC) 1830/2003) has led to much discussion surrounding discrepancies in the expression of quantitative measurements of GM content in food/feed and in seed, the former traditionally expressing content as % GM DNA, while the latter expresses content as % GM seed. These two terms are not always interchangeable because the genotype of seed is variable. In the work described here we have taken maize as an example of seed that is highly genotypically variable and used real-time PCR to investigate the relative contribution of kernal tissues to quantification of GM in the whole kernal. Investigations carried out as part of this work have enabled us to establish that it is possible to use real-time PCR to resolve discrepancies between % GM seed and % GM DNA, and that it is possible to analyse maize kernal tissues separately to determine the genotypes of GM and non-GM parents. These techniques could prove invaluable where a conventional crop or seed lot appears to have been 'contaminated' by a GM crop and compensation claims are being sought, and provides Defra with assurance that it would be possible to provide evidence to support or refute such claims.

12. Considering coexistence and traceability issues further, the ability to distinguish between sources of adventitious GM presence in a crop, GM introgression from volunteers and/or sources of gene flow from adventitious presence will be essential to the resolution of claims of contamination, accordingly methods must be available to resolve such issues rapidly and at reasonable cost. In this report results are presented on the use of short interspersed nuclear retrotransposons as molecular markers in PCR-based assays to support rapid, reproducible and robust identification of a number of oilseed rape varieties.
13. This work has generated results that, with some further work, will be appropriate for detailed publication in peer-reviewed scientific papers, it is also expected that the techniques described can be widely disseminated to benefit GM enforcement at the European level via the European Network of GMO Laboratories (ENGL). The research has identified a number of areas where further investigation would be desirable, in particular to make the techniques readily applicable for routine use. For example, the PCR primer sets and new PCR methods developed in this work need to be applied to a wider range of GM plants and materials (including food products). It is hoped that this can be achieved through work with the European Central Reference Laboratory (CRL) which will in future co-ordinate validation of GM detection and quantification methods.
14. Undertaking the work has also identified other areas that warrant further research, these include: (i) more detailed analyses on the source of unexpected GM elements found in PCR tests, whether they are from bacterial contamination or previously unknown GM structures; (ii) application of rapid protocols for isolation and DNA sequencing of GM event flanks to low target concentration materials; and (iii) development of oilseed rape variety identification molecular markers to single locus systems.
15. The studies reported here should provide considerable assurance to Defra that the team of researchers that comprise the GM enforcement team at CSL have developed a broad range of tools and techniques that will support the GM Inspectorate when complex GM enforcement issues arise, whether these are in connection with contamination of conventional seeds with GM seeds or conventional crops with GM crops. Together with other work ongoing at CSL in the field of analysis of the statistics of combined sampling and testing for GMOs, it is clear that a solid, science focussed skills base underpins CSL's GM enforcement work. However, it is essential to recognise that the biotechnology industry is not static and that different GM events and techniques are constantly being developed and new GMOs being released, and that to keep pace with these developments ongoing investment is required.

## Contents

Introduction.....	6
1. Improved PCR Primer Sets for Plant GM Event Identification.....	6
2. A Rapid Protocol for Isolation and DNA Sequencing of GM Event Flanks.....	12
3. Investigation of the Effect of GM Genotype on Real-time PCR GM Quantification.....	16
4. Molecular Methods for Oilseed Rape Variety Identification and GM Traceability.....	23
Conclusions.....	29
References.....	31
Appendix 1. PCR Primer Sequences.....	34
Appendix 2. Standard Operating Procedures.....	36

## Introduction

The ability to detect, identify and characterise Genetically Modified (GM) events in terms of DNA sequence is crucial for the enforcement of EU legislation regarding coexistence, and marketing of GM organisms. The molecular methods presented in this report aim to address specific issues arising from the work of the GM Inspectorate for England and Wales at CSL, but they are also expected to be applied more widely via the European Network of GMO Laboratories (ENGL).

The molecular methods developed in this project fall into four separate sections:

1. Improved polymerase chain reaction (PCR) primer sets for plant GM event identification.
2. A rapid protocol for isolation and DNA sequencing of GM event flanks.
3. Investigation of effect of GM genotype on real-time PCR GM quantification.
4. Molecular methods for oilseed rape (OSR) variety identification and GM traceability.

The rationale and objectives of each section are detailed below along with key results and recommendations arising from each area of research. The specific methods employed, which are a key outcome of this work, are detailed as standard operating procedures (SOPs) in Appendix 2.

### 1. Improved PCR primer sets for plant GM event identification.

In recent years, as plant biotechnology has expanded, the number and diversity of molecular tests for GM detection, quantification and identification have also increased. The aims of detection, quantification and identification methodologies have been kept predominately separate, i.e. it is impractical to incorporate large numbers of specific GM event tests into quantification assays without loss of resolution and similarly it is not practical to incorporate quantification elements into PCR screens with large numbers of element specific primers. Therefore, at this time there is no single test that can be used to detect, identify and quantify GM events. However, using current PCR technology, multiple testing schemes can be devised to fulfil these aims while minimizing costs and time required.

Several PCR approaches have been devised to maximise the number of GM events that can be screened for simultaneously. These can involve large panels of separate PCR reactions with different primer pairs, each designed to detect a specific GM element (gene, promoter, terminator or other GM DNA sequence). Alternatively, several 'element specific' primer pairs can be added to the same reaction

(multiplexed). These approaches can give a reliable indication of which elements are present but they cannot unequivocally identify a GM event or reveal when a mixture of GM events may be present.

The scope of this research was to provide improved PCR detection methods for GM events in soybean, maize, and oilseed rape (OSR). In order to achieve this, new PCR primers were designed for 90% of GM DNA elements\* currently released internationally (Bruderer & Leitner, 2003). Extensive searches of DNA sequence databases and GM inspectorate information was performed to identify elements for which more than one sequence variant had been released. Alignments of these sequences were carried out using CLUSTAL (Thompson *et al.*, 1994) in order to identify conserved regions which allowed the design of primers which would amplify all variants. Figure 1 shows, as an example, an alignment of *pat* and *bar* sequences. However, for some elements a single primer pair could not produce positive signals from all variants (e.g. *CRY* genes) and in these cases several combinations of primers were tested. For one element, *gox*, sequence information was not available due to intellectual property constraints and new primers could not be designed. Elements which contained similarity to endogenous plant sequences, e.g. P-SsuAra promoter and T-E9 were not used as they can potentially produce positive signals from non-GM plant DNA. The new primer set produced from this work will be employed to expand the detection capability of current PCR panels in use at CSL.

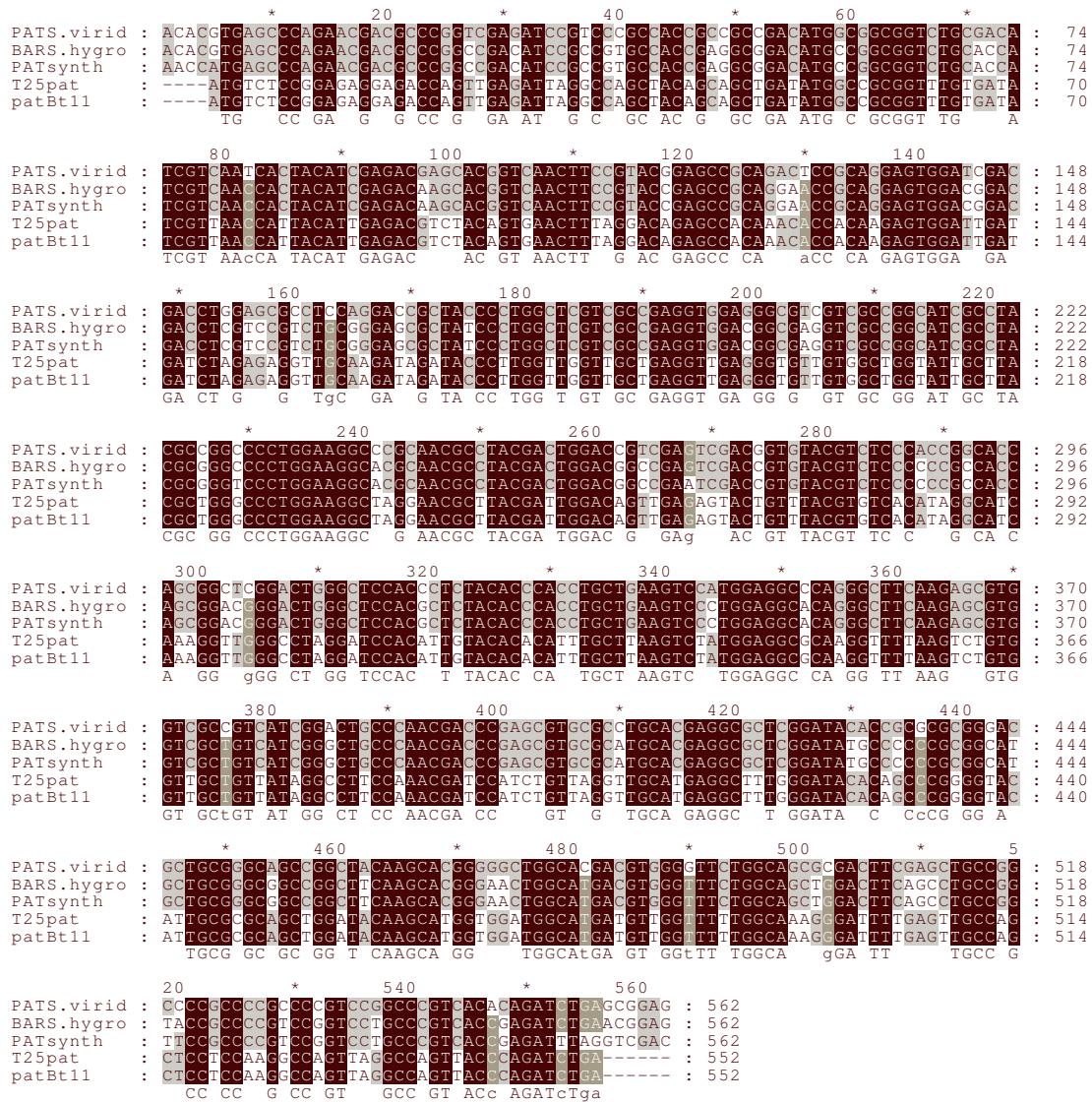
False-positive control primers were designed to amplify endogenous elements contained in *Agrobacterium* (Weller *et al.*, 2002), cauliflower mosaic virus (CAMV) and *Escherichia coli* (Hsu & Tsen, 2001). These species contain sequences frequently used in GM constructs and can therefore give false positive PCR results if they contaminate test DNA. Specific genes (reverse transcriptase for CAMV, *mdh* for *E. coli* and *FLIG* for *Agrobacterium* spp.) other than those used in GM constructs were therefore used as the targets for these primers.

Appendix 1 gives a complete list of the primers that gave positive, reliable results and were therefore used in further study. A generic PCR protocol was used for this work, in which initially only annealing temperature was varied. After testing many primers a temperature which gave reliable amplification for the largest number of primers was used. This was determined to be 55 °C. Several brands of *Taq* polymerase and reaction mixes were used during this work. The most cost-effective (in terms of time-saving and actual unit cost) and reliable of which was found to be RedTaq Readymix PCR mix (Sigma Aldrich, UK). This mix was therefore used in all subsequent PCR reactions. See Appendix 2, SOP 1 for the general PCR protocol used for standard and matrix PCRs, and SOP 3 for DNA extraction method used for all work, including sections 2 - 4 below.

---

\* Note, 'elements' is used to describe DNA sequences which may be genes, parts of genes, promoters, terminators or non-functional sequences.

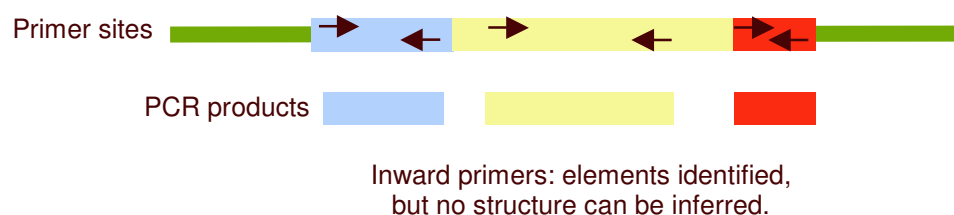
Figure 1. Sequence alignment of *pat* and *bar* genes used for primer design. PATS.virid = wild type *pat* gene; BARS.hygro = wild type *bar* gene; PATsynth = synthetic *pat* gene; T25pat = *pat* gene used in T25 maize; patBt11 = *pat* gene used in Bt11 maize. Fully conserved nucleotides are shown as black columns.



In addition to the design of primers for amplification of specific GM elements (inward primers), this project has provided primers that will amplify PCR products between adjacent GM elements (outward primers). These were designed as the reverse complement sequence of their corresponding inward primers. Using outward primers in GM identification PCRs has the advantage of providing information not only on which elements are present in a test sample but also how they are arranged in the GM construct (see Figure 2). The inferred arrangement of elements can then be used to identify the GM event's structure with a high degree of confidence without prior detailed information. In order to test for all possible combinations of adjacent elements, a combinatorial matrix of outward primer pairs was used. In order to minimise the necessary number of PCRs, separate PCR matrices were designed for different species. Table 1 shows the PCR matrix used for OSR. Figures 3a & 3b show the agarose gel of PCR products amplified by OSR matrix PCRs for event MS8 and a graphical matrix representation of the unique result obtained for different OSR events.

Figure 2. Diagram of the principle of using outward GM element primers to identify elements and construct structure. Primers and their orientation are shown as arrows. The coloured blocks represent different GM elements in a GM construct inserted into plant DNA (green).

i) Inward primers



ii) Outward primers

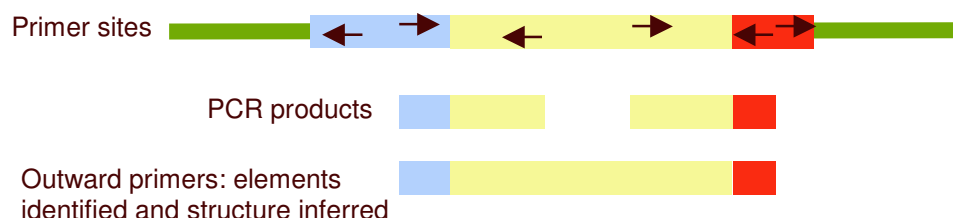


Table 1. Outward primer PCR matrix used for OSR. 1 = P-35S; 2 = T-35S; 4 = Barnase; 5 = Barstar; 10 = nptII; 12 = pat/bar; 15 = P-TA29; 16 = T-g7; 17 = T-nos; 17B = P-nos; 18 = T-ocs.

<b>1+2</b>	<b>1+17B</b>	<b>2+17</b>	<b>4+17</b>	<b>5+17B</b>	<b>12+15</b>	<b>15+18</b>
<b>1+4</b>	<b>1+18</b>	<b>2+17B</b>	<b>4+17B</b>	<b>5+18</b>	<b>12+16</b>	<b>16+17</b>
<b>1+5</b>	<b>2+4</b>	<b>2+18</b>	<b>4+18</b>	<b>10+12</b>	<b>12+17</b>	<b>16+17B</b>
<b>1+10</b>	<b>2+5</b>	<b>4+5</b>	<b>5+10</b>	<b>10+15</b>	<b>12+17B</b>	<b>16+18</b>
<b>1+12</b>	<b>2+10</b>	<b>4+10</b>	<b>5+12</b>	<b>10+16</b>	<b>12+18</b>	<b>17+17B</b>
<b>1+15</b>	<b>2+12</b>	<b>4+12</b>	<b>5+15</b>	<b>10+17</b>	<b>15+16</b>	<b>17+18</b>
<b>1+16</b>	<b>2+15</b>	<b>4+15</b>	<b>5+16</b>	<b>10+17B</b>	<b>15+17</b>	<b>17B+18</b>
<b>1+17</b>	<b>2+16</b>	<b>4+16</b>	<b>5+17</b>	<b>10+18</b>	<b>15+17B</b>	

Figure 3a. Agarose gel of matrix PCR products for OSR event MS8. 1 = Barnase/T-nos; 2 = Barnase/P-TA29; 3 = Barnase/T-g7; 4 = bar/T-g7; 5 = P-TA29/T-nos.

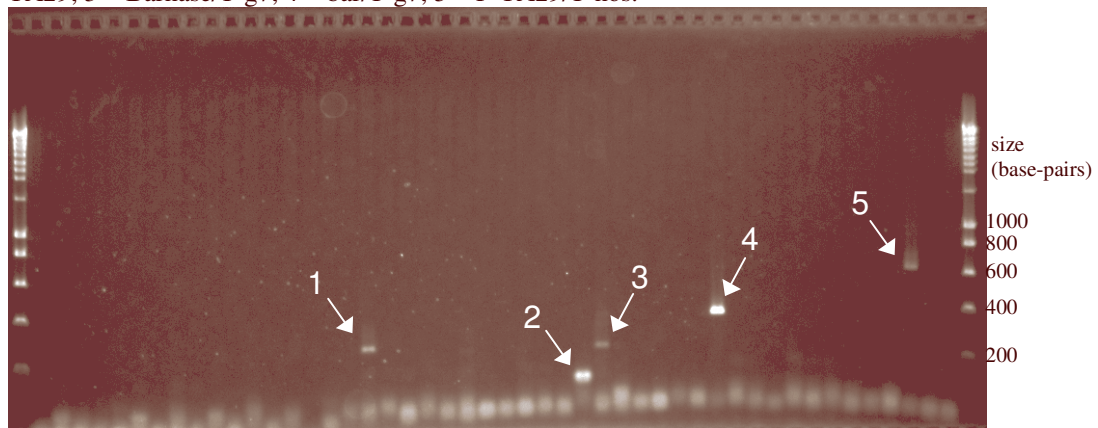
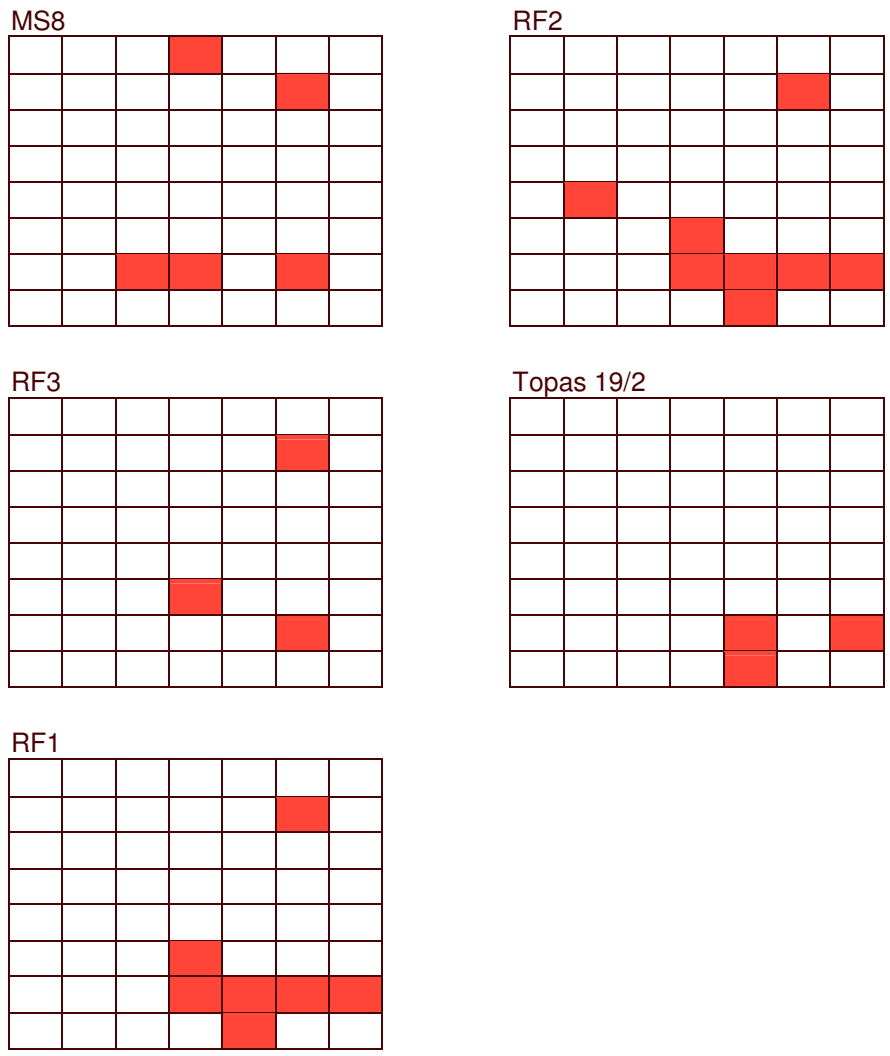
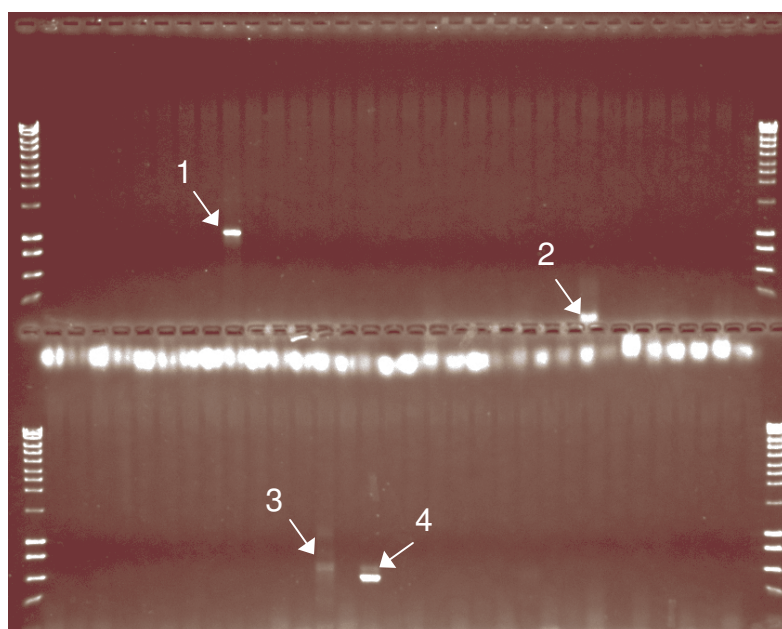


Figure 3b. Unique pattern of positive PCRs (red box) obtained for different OSR GM events with the OSR PCR matrix. Each cell corresponds to the primer combinations shown in Table 1.



Generally, outward primer PCR products have produced the predicted products for the GM event tested but expected products can be absent (e.g. P-35S and *bar* from Topas19/2). Where these absences have occurred, they have been consistent between replicated PCRs and DNA samples. It is likely that such negative results are due to unforeseen sequence variants or deletions / truncations from the target sequences. In other cases, additional, unexpected, products have been amplified. Figure 4 shows an agarose gel of matrix PCRs for Roundup-Ready soybean. In this matrix an unexpected product between an ampicillin resistance gene (beta-lactamase) and pUC plasmid backbone was produced. These sequences have not previously been reported in studies of this GM event. Further PCRs and DNA sequencing will determine if they are part of the known GM insertion or an independent event present in the sample. A product was also obtained for primers to the *epsps* gene alone (i.e. adjacent to itself). This results from the adjacent duplication of part of this gene in the GTS40-3-2 event (Windels *et al.*, 2001). These results highlight the usefulness of the outward matrix PCRs for detecting unexpected GM structures.

Figure 4. Agarose gel of outward matrix PCRs for Roundup-ready soybean. 1 = P-35S/*epsps*; 2 = amp/pUC; 3 = *epsps/epsps*; 4 = *epsps/T-nos*.



Matrix PCR results have shown that the method can simultaneously be used to give a unique 'fingerprint' for different GM events and give information on the structure of the event. In order to include all possible combinations or adjacent GM elements a large number of individual PCR reactions is required. This makes the method unsuitable for initial screening for GM but it would be a good choice for identifying GM events, or their fragments, in the absence of event specific methods (which employ a primer sited near the ends of a GM construct and the other primer in the flanking plant DNA), e.g. tests for 'unauthorised' GM events. Event specific methods, if used in isolation could fail to detect where a mixture or stack of GM events is present, either because an insufficient range of event specific tests is used, or because the admixed event is unauthorised and of unknown composition. The matrix PCR described here could therefore be used in conjunction with event specific tests to help confirm that the event identified is the only one present in a sample.

## 2. A rapid protocol for isolation and DNA sequencing of GM event flanks.

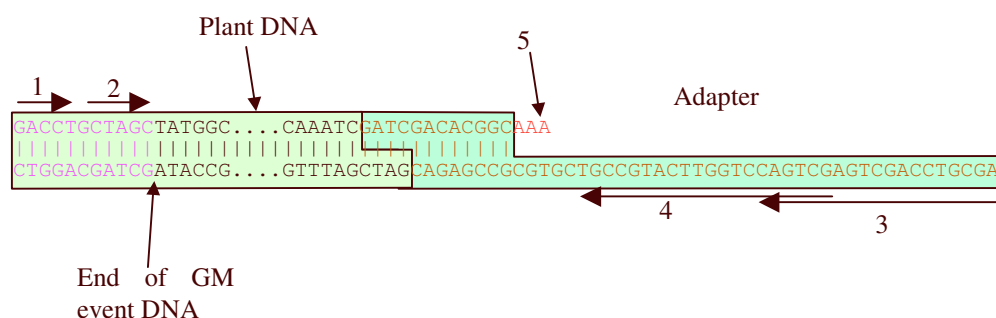
The flanking sequence of GM constructs (i.e. the plant DNA immediately adjacent to inserted GM constructs) is an important resource with four main uses:

1. Provides identification of a GM event. With current technology no two independent transformation events can be inserted into precisely the same locus (without a very small probability of it occurring by chance, at least  $10^{12}:1$ ). The junction can therefore be used to design PCR tests that are unique to the event in question.
2. Provides information for risk assessment of the transformation event. By searching databases with the flank sequence, the likelihood that the disruption of a transformed locus may cause undesirable effects can be assessed.
3. Enables stability of the transformation to be assessed. Flanking sequences can be placed on physical maps for the species in question and any subsequent disruption of the locus by recombination or other means can be detected by its altered map location. Recent advances in mapping technology (e.g. HAPPY mapping (Thangavelu *et al.*, 2003)) should allow such assessments on a routine basis in the future.
4. An issue that arises frequently in GM detection PCR tests is the occurrence of 'single markers' where only one GM element is detected in apparent isolation from other elements expected in GM events. For example, positive p35S PCR test result but no other positive PCR in a multi-element screen. Such occurrences are difficult to resolve and are usually addressed by circumstantial evidence that the single marker arises from bacterial or viral contamination. The protocols developed here can potentially resolve such issues by rapidly providing flanking sequence of the detected single marker, hence identifying its source as either a GM construct or wild type bacterial or viral contamination.

The successful isolation and sequencing of GM flanks requires the combination of two methodologies. First, a PCR step to target and amplify the flank. Second, a method to isolate or clone the PCR product and give a reliable target for DNA sequencing.

The suppression PCR (SPCR) method (Spertini *et al.*, 1999) has been adapted to rapidly obtain GM construct - plant DNA flanking sequences for known and unknown GM events. SPCR is an adapter mediated method for amplification of sequences flanking a known sequence. Sample DNA is digested with a restriction enzyme which gives 5' ends complementary to a synthetic adapter sequence. The adapter (Figure 4) is designed so that it suppresses self-amplification. Only sequences containing the known specific sequence should therefore be amplified. A problem using SPCR with plant genomes is their large size. As genome size increases, the likelihood that the specific primer or adapter primer will randomly amplify a non-specific (background) product increases. In order to reduce the background signal, nested PCRs are performed: a primary PCR using the adapter-ligated DNA sample and a secondary PCR using a dilution of the primary PCR with primers nested internally to the primary PCR. Nested PCR primers specific to the ends of several GM events were designed including Roundup-Ready soybean, T25 maize, MS8 × RF3 OSR and Topas19/2 OSR.

Figure 4. Diagram showing the design and principles of the SPCR adapter. The adapter is shown ligated to a 'GATC' restriction site. 1 & 3 = primary PCR primer sites; 2 & 4 = secondary PCR primer sites; 5 = mismatched 'AAA' sequence prevents the short adapter strand from acting as a PCR primer. Amplification of the intended product can only proceed once primer 1 has been extended in the first round of PCR and replicated a strand complementary to the long single stranded sequence of the adapter, hence providing the primer binding site for primer 3. GM specific primers are longer than those shown. A very short length of plant DNA is shown - much larger flanks are usually obtained.



Despite these precautions to reduce background amplification, initial experiments using SPCR to amplify the flank of the Roundup-ready soybean GM event gave non-specific products (Figure 5). This problem was also observed in OSR and maize. To further increase the specificity and yield of expected GM flank products, a labelled oligonucleotide approach was used. The GM specific primary PCR primer was 5' labelled with a biotin residue. Primary PCR products incorporating the specific biotin labelled primer were then bound to streptavidin paramagnetic particles (PMPs) and washed to remove all other PCR products. A small aliquot of the PMPs was then added to the secondary PCR. Figure 5 shows the increased specificity obtained with this method.

Six GM event sequences have been obtained using SPCR (Table 2). These include flanking plant sequences (Figures 6 & 7) and sequence adjacent to the unexpected *amp* fragment obtained by matrix PCRs. This SPCR product maintained identity to pUC-type plasmid and did not contain any plant sequences. The origin and locus of this sequence remains unknown, however, further SPCR could be used to 'walk' further along this sequence until plant sequence is identified, or the product is identified as plasmid contamination. Comparison of SPCRs in Figure 6 (OSR) and 7 (maize) illustrates the increased specificity problems with increasing genome size. The maize genome is approximately twice the size of the OSR genome (Arumuganathan & Earle, 1991), consequently only one restriction digest / adapter ligation of T25 maize gave a clear SPCR product. The optimised methods resulting from this work are given in Appendix 2, SOP 2.

Figure 5. Effect of biotin labelled selection of primary SPCR products on reaction specificity. Sample DNA (5% RR CRM, Fluka) was digested with *Bst*YI. Lane A = biotin label included. Lane B = biotin label not included. Band 1 = expected 607 bp RR P-35S plant DNA flank. Band 2 = unknown PCR artefact.

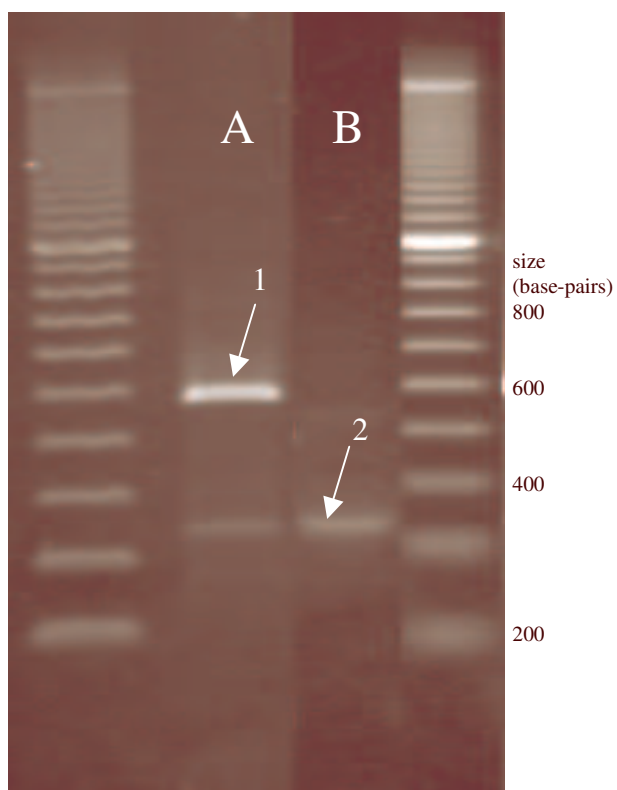


Table 2. Sequenced SPCR products to date.

GM event	Sequence	Length (bp)	Notes
Maize T25	T-35S adjacent to plant flank	460	Short of plant DNA. Further SPCRs required
OSR RF2	T-g7 plant flank	450	
OSR Topas 19/2	P-35S adjacent to plant flank	475	Short of plant DNA. Further SPCRs required
OSR MS8	T-g7 plant flank	600	
Soybean RR	P-35S plant flank	578	200 bp longer than published flank
Soybean RR	Unknown <i>amp</i> element	240	

Figure 6. SPCR products from OSR GM lines MS8, RF3, MS1, RF1, RF2 & Topas19/2 digested with *Bst*YI. MS1 failed to give any SPCR product.

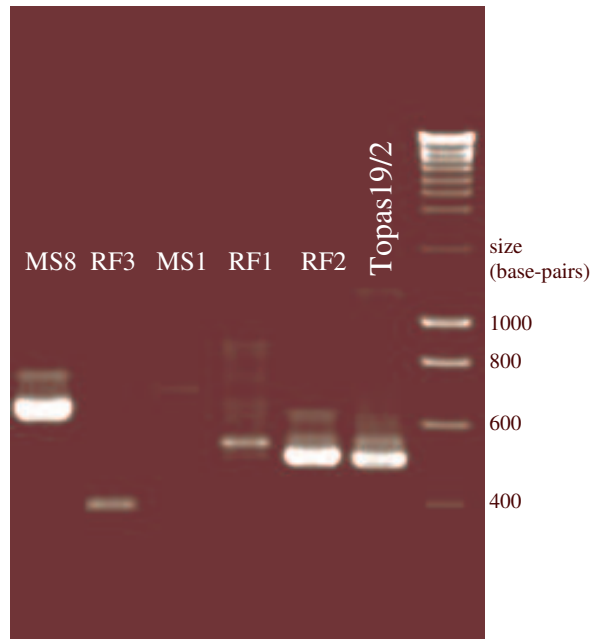
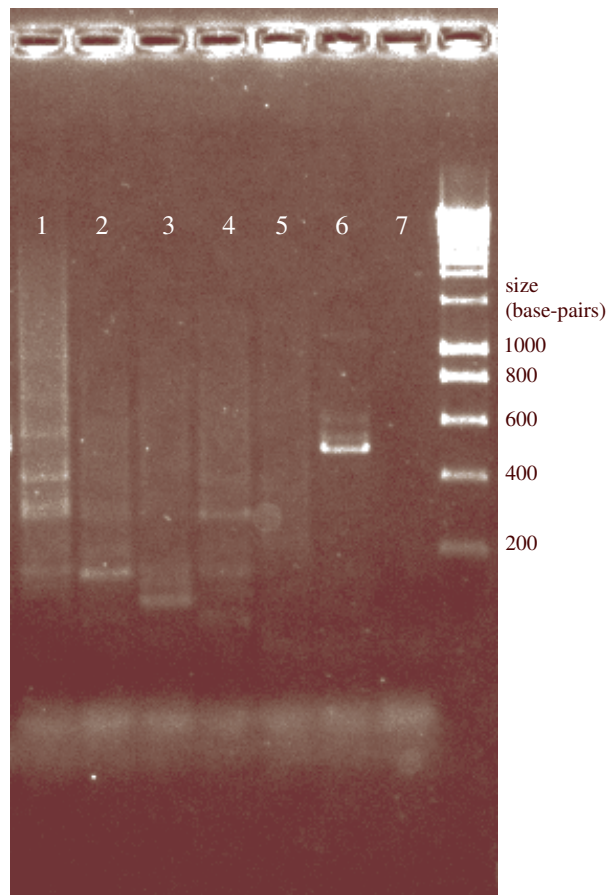


Figure 7. SPCR products from Maize GM line T25 digested with *Bam*HI, *Cla*I, *Bcl*II, *Bgl*III, *Bst*YI., *Msp*I and *Taq*I, lanes 1-7 respectively. Only *Msp*I gave a reliable SPCR product.



SPCR products were isolated by excision directly from ethidium stained agarose gels followed by gel extraction using the commercial kit, Nucleospin Extract II (Machery Nagel). Appendix 2, SOP 4 gives details of the isolation and cloning procedure. All DNA sequences were obtained from the commercial service of the University of Dundee (<http://www.dnaseq.co.uk>). The entire process, from receipt of an unknown GM sample to DNA sequence requires approximately 5 working days to complete. The process has been tested to a GM concentration in the initial sample of 5% (using Roundup-Ready certified reference flour (FLUKA)). However, the process has not been tested with lower concentrations of GM, where their isolation by SPCR may be problematic.

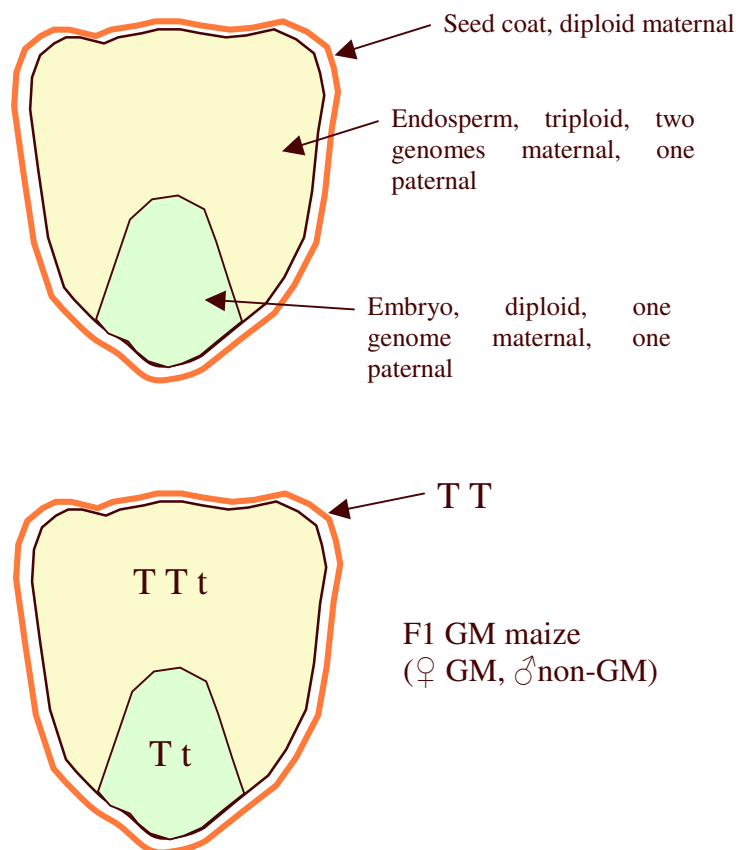
### 3. Investigation of the Effect of GM Genotype on Real-time PCR GM Quantification.

Much discussion has recently concerned current EU legislation and its associated recommendations (Regulation EC1829/2003; EC1830/2003). In particular, the interpretation and definition of the GM thresholds under this legislation have been contentious. 1829:2003 recommendations state that thresholds and quantitative measurements of GM in food and feed should be expressed in terms of %DNA (% GM haploid genomes). However, seed industry methods usually express GM (and other impurity measurements) as % seed. These two measurements will not be equivalent in many cases, predominantly because of variable genotype of GM seed.

It should be noted that the phrase *genotype* is used here rather than *copy number* as used in several EC documents. *Genotype* refers to the number of alleles present in a genome due to its zygosity and ploidy. *Copy number* is a term used in genetics to describe the number of replicate versions of a sequence throughout a genome. Copy number is therefore a very different issue from genotype and it can result from effects such as multiple insertions of GM constructs during transformation, whereas the GM genotype is determined only by the genetics of a particular species and the type of cross that produced the plant studied.

Some crop species are more likely to present these difficulties than others. Soybean, for example is always diploid, has no unusual genetics in its seed and is always produced by selfing. Its GM seed is therefore always homozygous and should provide equivalent % GM seed and % GM DNA measurements for a particular GM event. Conversely, maize kernels contain several different genotypes and are produced by out crossing so their GM genotype is variable and therefore % GM seed and % GM DNA measurements are unlikely to be equivalent. Figure 8 illustrates the complexity involved in maize kernel genotypes. The seed coat is diploid and identical genetically to the maternal plant and in F1 maize seed (produced from a maternal GM elite line) it is therefore homozygous GM. The endosperm of maize is formed from the fusion of two egg nuclei and one sperm cell nucleus and is triploid. In F1 GM maize this results in two GM alleles and one non-GM allele in all cells of the endosperm. The embryo is the result of normal gametic fusion and is diploid with one maternal and one paternal nucleus in each cell. In F1 GM maize, the embryo (and resulting plant) is therefore heterozygous GM.

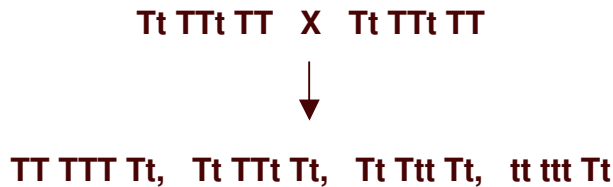
Figure 8. Simplified diagram of maize kernel showing genotypes of each tissue and example for F1 GM maize: T = GM allele, t = wild-type (non-GM allele).



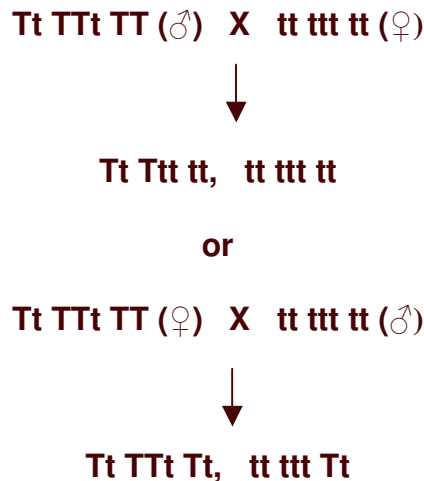
If this complex genetic make-up in maize kernels was constant then a simple conversion factor could possibly be used to equate %GM seed with %GM DNA. However, different parental crosses and different directions of cross will result in very different proportions of GM in each of the kernel tissues. Figure 9 illustrates the genetics of some different GM maize crosses. The second cross is shown in two different directions which affects the genotype of the endosperm and seed coat. The final cross shown ( $Tt\ TTt\ TT$  (♂)  $\times$   $t\ t\ t\ t$  (♀)) is the most likely to occur in the field, i.e. F1 GM pollen (normal GM crop) fertilising non-GM maternal plants. An important point is that the proportion of GM DNA in the total kernel is not just a product of the genotype of each tissue but the relative contribution of each tissue to the total amplifiable DNA.

Figure 9. Genotypes of maize kernel crosses. Genotypes are written in the form xx yyy zz, where x is embryo, y is endosperm and z is seed coat alleles.

**1. F1 GM maize selfed**



**2. F1 GM maize crossed with non-GM maize**



In this work real-time PCR (TaqMan) assays were used to investigate the genotype of each kernel tissue and their relative contribution compared to the same assay on DNA extracts from the total (simulated) kernel. Plants and seeds containing the GM event T25 (Aventis, now Bayer Crop Science) were used in this work. Two panels of 12 kernels were analysed. The first consisting of F1 seed and the second consisting of F3 seed from a cross between F2 parents of known genotype ( $tt \text{ ♂ X } Tt \text{ ♀}$ ). Each kernel was dissected into seed coat, endosperm and embryo tissues and DNA extracted separately from each tissue type (SOP3, Appendix 2). A fourth DNA solution was made for each kernel, consisting of a mix of an equal volume (20  $\mu$ l) of the DNA extraction from each tissue, this solution was therefore a simulated DNA solution theoretically equivalent to a DNA extraction from the whole kernel (there is no reason to suspect that different tissues will perform differently when extracted separately or mixed in the whole kernel). TaqMan PCR (SOP5, Appendix 2) was performed on each sample to obtain an estimate of: (i) total amplifiable DNA using an endogenous control DNA target, *cdc*, (Weekes *et al.*, 2005) and (ii) T25 concentration (% DNA).

The mean *cdc* amounts were calculated in arbitrary units using a calibration curve constructed from a dilution series of T25 maize DNA in soybean DNA (100% to 0%). The mean of all *cdc* values for each tissue was divided by the mean total *cdc* value to calculate the relative proportion of amplifiable DNA contributed by each kernel tissue (Table 3). Embryo and endosperm were found to have very similar contributions to total extracted amplifiable DNA, whereas the seed coat had a negligible contribution. This is to be expected, given that the seed coat is largely dead tissue and its DNA therefore degraded (Giuliani *et al.*, 2002).

Table 3. Calculation of relative proportion of amplifiable DNA extracted from kernel tissues: *cdc* is expressed in arbitrary units. Standard deviation of the value is given in brackets.

Tissue type	Mean <i>cdc</i> amount (s.d.)	proportion of total DNA (s.d.)
Embryo	4593 (1802)	0.417 (0.204)
Endosperm	6314 (1837)	0.573 (0.230)
Seed coat	109 (91)	0.01 (0.01)
Total	11016	

Table 4 shows the results of determination of GM (T25) concentration in tissues of maize kernels from the two panels (F1 and F3) by TaqMan PCR. The relative concentration (%) was calculated by a standard curve of the difference between  $C_T$  values for T25 and *cdc* ( $\Delta C_T$ ). The genotype of each seed has been inferred from the ratio of the tissue GM concentrations, the likely genotype of the parent plants and the known direction of the cross (in F3 kernels). F1 kernels were, as expected, all found to have the same genotype. It is most likely that this seed results from the usual industry practice of tassel removal from the seed producing plants followed by pollination by alternating rows of paternal plants. In this case, due to the positive GM in the seed coat, we can deduce that the maternal parent was homozygous GM (TT) and the pollen donor homozygous non-GM (tt). We would expect the seed coat of such seed to be the maternal genotype (TT = 100% GM) but only ~50% was observed. This may indicate that the determination of GM concentration at the low amplifiable DNA levels found in the seed coat was unreliable. However, because the seed coat has a very low contribution to the total DNA (1%) this does not significantly affect any of the calculations below. F3 kernels were the progeny of a single controlled cross between a non-GM (T25 F2 null segregant) paternal plant and a GM maternal plant (suspected homozygous GM from PCR test - result not shown). As expected, the 12 progeny analysed segregated 1:1 (n = 12, 5:7,  $\chi^2$  not significant) with two genotypes: Tt TTt tt and tt tt Tt.

Table 4. %GM from maize kernel tissues.

Seed No.	Embryo	Endosperm	Seed coat	Mixed	Inferred genotype
F1 Panel	52.08	45.66	66.78	47.98	Tt TTt Tt
	45.65	55.49	69.62	56.20	"
	48.32	55.73	57.45	55.34	"
	44.03	52.74	56.25	54.30	"
	43.09	53.98	53.01	52.98	"
	46.65	54.70	66.91	53.86	"
	44.23	53.91	66.39	50.66	"
	54.28	57.75	55.84	55.75	"
	45.67	57.70	61.76	49.38	"
	46.15	58.15	33.56	51.23	"
	46.93	57.65	72.10	51.53	"
	48.86	56.05	59.99	49.85	"
F3 Panel	0	0	42.92	0	tt ttt Tt
	44.37	53.54	54.53	51.16	Tt TTt Tt
	0	0	47.19	0	tt ttt Tt
	46.67	54.89	55.08	47.02	Tt TTt Tt
	0	0	54.81	0	tt ttt Tt
	0	0	19.36	7.97	tt ttt Tt
	44.77	53.88	40.86	51.88	Tt TTt Tt
	0	0	33.77	0	tt ttt Tt
	0	0	57.24	0	tt ttt Tt
	44.93	53.31	63.71	49.95	Tt TTt Tt
	54.47	58.98	62.81	59.54	Tt TTt Tt
	0	0	56.58	0	tt ttt Tt

Using these data, it is possible to compare observed to expected proportions for GM concentration in embryo and endosperm. For F1 and F3 kernels a proportion (embryo / endosperm) of 0.75 was expected. Observed values were 0.85 (s.d. = 0.08) for both panels. This is a significant deviation ( $P < 0.05$ ) from the expected value and therefore indicates an unknown, but consistent, effect.

The proportions of the concentration of GM from each tissue, combined with the measured amounts of amplifiable DNA from each tissue can be used to estimate a projected value for the total %GM DNA we would expect from whole kernels. Using the following equation:

$$T = (a.x) + (b.y) + (c.z) \quad \text{Equation 1.}$$

where T is the expected total %GM DNA; a, b and c are the observed GM concentrations; and x, y, and z are the proportions of the amounts of amplifiable DNA from embryo, endosperm and seed coat respectively (Table 3). We can then compare these values to that obtained from the equal mixture of DNA solutions from the three tissues, see Table 5, giving mean values for all the Tt TTt Tt genotypes. Observed values were very similar to those expected, which indicates that Equation 1 can be

used to predict the total %GM DNA that would be obtained from other kernel genotypes. We can therefore predict the % GM DNA that would be observed in bulks of kernels resulting from different crosses. Table 6 summarises these calculations for two of the crosses to occur in agriculture: an F1 GM crop crossing with itself and an F1 GM crop out crossing with a non-GM crop. These results show the %GM DNA that would be obtained on a bulk of kernels from these crosses. They are therefore the product of the proportion of amplifiable DNA from each kernel tissue, the genotype of each tissue, and the proportions of different kernel genotypes resulting from the crosses. The F1 crop, crossed with itself results in %GM DNA = 49.38, whereas a % GM seed on this bulk would = 100%. Similarly, for the F1 GM crop out crossed to non-GM (one direction only with GM pollen), %GM DNA = 19.88, whereas % GM seed = 50%. It should be noted that these values refer *on average* to any results that are likely to be a product of these crosses, i.e. F1 'selfed' kernels would form the bulk of a GM harvest and any GM detected that has arisen from such a source would give a GM DNA value 49% that of a % GM seed value. Any GM derived from pollen flow from F1 crops to a non-GM crop would give a GM DNA value 20% that of a % GM seed value.

Table 5. Observed mean ('mixed' DNA solutions) and expected values (Equation 1) for whole kernel % GM DNA.

Genotype	Tissue	Mean % GM DNA
Tt TTt Tt	Embryo	47.04
	Endosperm	54.92
	Seed coat	55.36
	Total observed % GM DNA	51.91
	Total expected % GM DNA	51.64
tt ttt Tt	Embryo	0
	Endosperm	0
	Seed coat	44.5
	Total observed % GM DNA	1.138
	Total expected % GM DNA	0.445

Table 6. Predicted % GM DNA values compared to % GM seed values for bulks of kernels from two different crosses. The bulk mean value is shown in bold. Note that from the first cross the %GM seed value would be 100% and from the second cross it would be 50%.

Cross	Progeny	Tissue contribution			Total % GM DNA
		embryo (0.417)	endosperm (0.573)	seed coat (0.01)	
F1 GM crop Tt x Tt	TT TTT Tt	41.7	57.3	0.05	99.005
	Tt TTt Tt	20.85	37.8	0.05	58.67
	Tt Ttt Tt	20.85	18.9	0.05	39.81
	tt ttt Tt	0	0	0.05	0.05
					<b>49.38</b>
F1 GM with non-GM Tt x tt	Tt Ttt tt	20.85	18.9	0	39.76
	tt ttt tt	0	0	0	0
					<b>19.88</b>

It can be concluded that real-time PCR can be used to resolve issues on the discrepancy between % GM seed and % GM DNA measurements. At least in the case of this GM line in fodder maize, despite a complex genetic system and the problem of origination from different crosses, the %GM DNA value for any bulk kernel sample will be approximately 50% of the % GM seed value. There remains a need to increase the accuracy of measurements presented here (as illustrated by poor accuracy at low GM levels in seed coat samples). This work has only analysed one GM maize fodder variety and factors are likely to be very different for other maize varieties and crop types with varying sizes of embryo and endosperm. However, with further work it may be possible to develop simple correlations between embryo and endosperm mass and the conversion between % seed and % GM values. This work has also shown that separate analysis of maize kernel tissues can be used to determine the genotypes of GM and non-GM parents and the direction of any cross that may have produced the GM-containing kernels.

#### 4. Molecular methods for oilseed rape variety identification and GM traceability.

The molecular identification of an OSR variety can provide invaluable information for resolving problems that may arise in GM crop coexistence and GM traceability including the following:

- Ability to distinguish between sources of adventitious GM presence. When an adventitious GM event is detected in a seed or grain lot, the seeds containing the event can be compared to the marketed GM variety / varieties. If molecular markers of varieties other than the original(s) are detected then it is likely that the presence has arisen from a gene-flow event, i.e. leading to the GM and non-original variety markers in the same seed or plant. Conversely, if no other variety markers are detected in the GM seed, then it is not likely to have arisen from pollen (gene flow). In this way the most likely source of a GM presence (field or post-harvest) can be identified.
- Ability to distinguish GM introgression from volunteers. In the same way as the previous solution, the co-occurrence (or not) of GM and variety markers can tell us whether a suspected source of volunteers is from seed drop or the result of gene flow.
- Ability to trace source of geneflow adventitious presence. If a GM presence is deemed likely to be due to geneflow, the particular variety that originally received the event can be identified and therefore seed / grain lots traced back to the candidate fields which contained the identified variety and where the geneflow could have occurred.

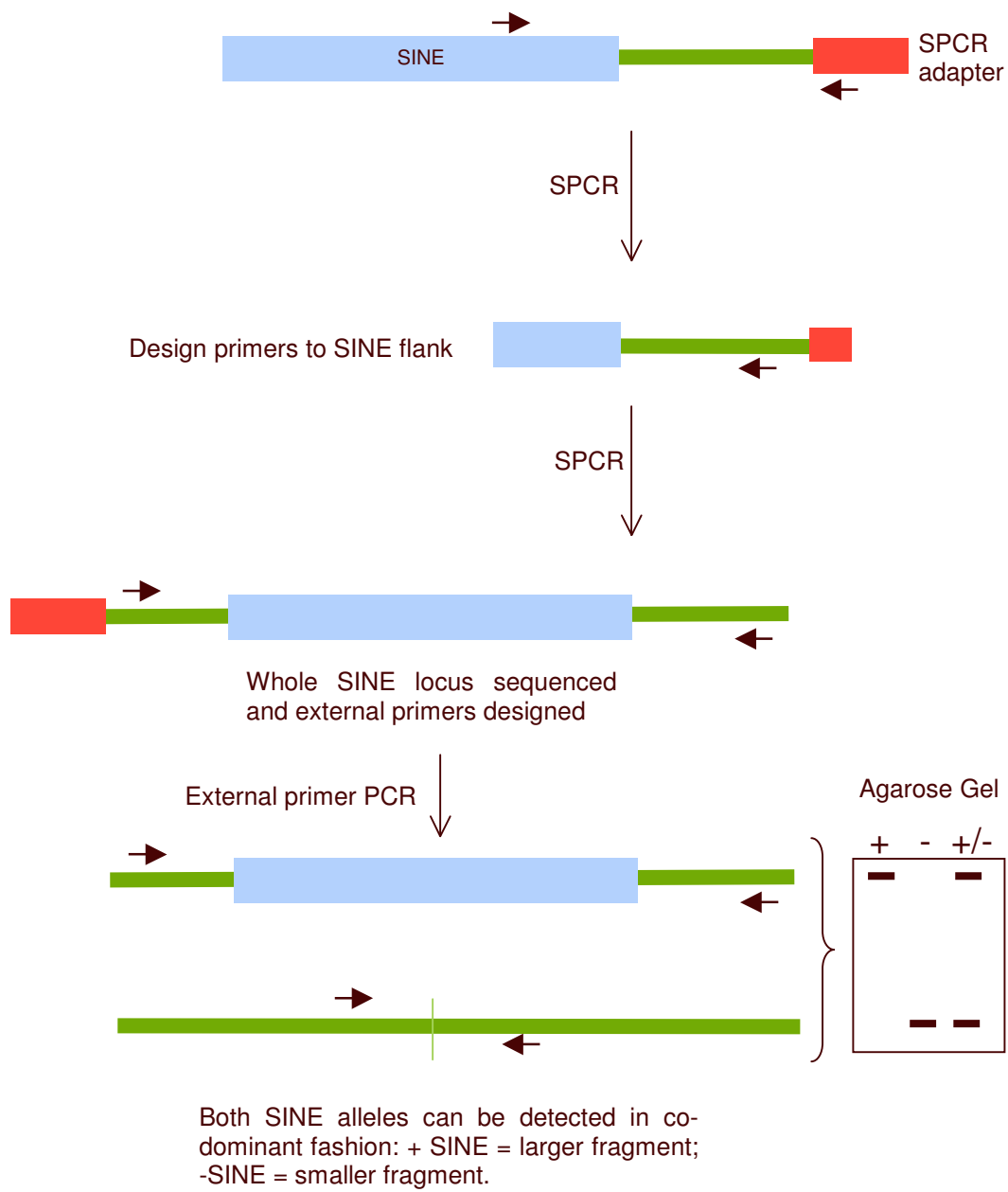
Currently there are over 150 cultivars of oilseed rape (*Brassica napus*) on the UK's national list of plant varieties and seeds (anon. 2004). The main way these are distinguished is through distinctness, uniformity and stability (DUS) testing, carried out by the Department for Environment, Food and Rural Affairs (DEFRA). These field tests are carried out over a minimum of two years, and seek to discover if the variety is clearly distinguished in one or more important characteristics from other varieties currently present in the community, whether the plants of which the variety is composed are genetically similar as regards their characteristics and if, after successive propagations or multiplications, the variety remains true to the description of its essential characteristics. Once a variety has been certified by DUS testing the plant breeder maintains the cultivar against further future testing. However DUS testing cannot be used in the timescales likely to be required to resolve GM issues as they arise. For this reason, the development of molecular variety markers is necessary.

Many different systems, both protein and DNA based, for molecular markers in *Brassica* have been developed. For recent examples, see Yu *et al.* (2005), Liu *et al.* (2005), Suwabe *et al.* (2004) and Varshney *et al.* (2004). However, these published approaches largely concentrate on genetic trait characterisation for plant breeding purposes and not variety identification. Ideally a molecular marker system for variety identification should be: (i) easily reproducible, i.e. robust and transferable between labs; (ii) co-dominant (to facilitate identification of hybrids); (iii) reveal sufficient polymorphism to provide unique marker profiles for different varieties. Existing markers systems, such as microsatellites could provide these properties but they generally require high development costs. Random DNA systems (e.g. random

amplified polymorphic DNAs, RAPDs; amplified fragment length polymorphisms, AFLPs) will not provide co-dominant markers and are not easily reproducible. Specific genes or sequences can be targeted for PCR amplification and their polymorphism revealed (e.g. cleaved amplified polymorphisms, CAPS), but these are unlikely to yield sufficient polymorphism among recently derived genotypes. In this work we have chosen to use the flanking restriction length fragments of a targeted retrotransposon that is likely to be highly polymorphic for the development of variety markers by using the same PCR methods as SPCR. A similar approach has been used successfully for *Pisum* variety identification (Vershinin *et al.*, 2003; Pearce *et al.*, 2000; Ellis *et al.*, 1998).

Retrotransposons are eukaryotic transposable elements, which consist of two main groups, long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. These groups are both divided into two further classes. The LTR retrotransposons consist of Ty1-copia, and Ty3-gypsy groups, and the non-LTR retrotransposons contain the short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs). All of these groups are found with high copy numbers in plants, especially in those with large genome sizes (Kumar & Bennetzen, 1993). S1 elements are a family of SINE retrotransposons initially discovered in *Brassica napus* (Deragon *et al.*, 1994) and have since been closely studied and characterised, and found to contain three subfamilies within the species (Lenoir *et al.*, 1997). A general consensus sequence for the S1 family has been published (Deragon *et al.*, 1996) which has made it possible for general PCR primers to be designed which will amplify a large proportion of SINE retrotransposons in the OSR genome. The small size of SINEs (~170 bp) together with their mode of insertion polymorphism makes it possible for single-locus SINEs to be isolated and converted into co-dominant markers which require only one pair of primers situated outside the insertion site, see Figure 10 for a diagram of this principle. The work presented here aimed to test PCR primers for their ability to amplify a large number of SINE flanking sequences using SPCR and to demonstrate the ability of these multi-locus dominant markers to distinguish OSR varieties, with a view to their future development as single-locus primer sets.

Figure 10. Principle of SINE SPCR and marker isolation.



Twenty OSR varieties were selected from the Home-Grown Cereals Authority (HGCA) recommended lists as representative of varieties currently grown in the UK (seed kindly supplied by National Institute of Agricultural Botany, NIAB). Table 6 gives the details of the study samples. Seven OSR varieties were duplicated, using seed samples supplied by NIAB on two separate occasions (2002 and 2004). Four varieties were spring OSR and 16 were winter OSR. *Arabidopsis thaliana*, four *Brassica oleracea* and one *Brassica rapa* samples were also included as 'outgroups' to provide a relative genetic distance defining the 'exterior' of any OSR clades / groups. OSR plants were grown in a glasshouse and leaf samples used for DNA extractions (SOP 3, Appendix 2).

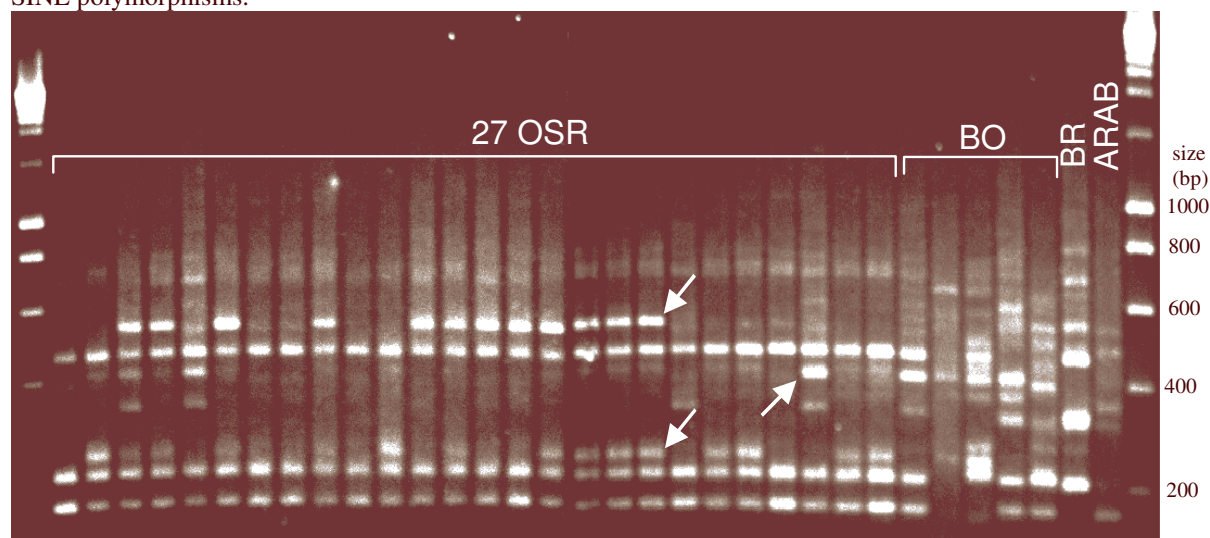
Table 6. *Brassica* samples used in SINE analyses.

Sample Ref.	Species, Variety	Spring or Winter
009.01	<i>B. napus</i> canberra	Winter
009.02	<i>B. napus</i> cohort	Winter
009.03	<i>B. napus</i> concept	Spring
009.04	<i>B. napus</i> disco	Winter
009.05	<i>B. napus</i> dorothy	Spring
009.06	<i>B. napus</i> élan	Winter
009.07	<i>B. napus</i> escort	Winter
009.08	<i>B. napus</i> fortress	Winter
009.09	<i>B. napus</i> herald	Winter
009.10	<i>B. napus</i> lipton	Winter
009.11	<i>B. napus</i> madrigal	Winter
009.12	<i>B. napus</i> mendel	Winter
009.13	<i>B. napus</i> mendel	Winter
009.14	<i>B. napus</i> pronto	Winter
009.15	<i>B. napus</i> pronto	Winter
009.16	<i>B. napus</i> recital	Winter
009.17	<i>B. napus</i> recital	Winter
009.18	<i>B. napus</i> royal	Winter
009.19	<i>B. napus</i> royal	Winter
009.20	<i>B. napus</i> senator	Spring
009.21	<i>B. napus</i> shannon	Winter
009.22	<i>B. napus</i> shannon	Winter
009.23	<i>B. napus</i> spirit	Winter
009.24	<i>B. napus</i> tambora	Spring
009.25	<i>B. napus</i> winner	Winter
009.26	<i>B. napus</i> winner	Winter
010.01	<i>A. thaliana</i> columbia	-
016.01	<i>B. oleracea</i> v. alboglabra	-
016.02	<i>B. oleracea</i> v. green duke	-
016.03	<i>B. oleracea</i> v. nedcha	-
016.04	<i>B. oleracea</i> v. gower	-

Primers were designed for SPCR (SOP 2, Appendix 2) using the consensus sequence of Deragon *et al.*, (1996). The nested pair for SPCR was: SINE5, TACAGRCAMAGGCTGGCGCC (1' SPCR); and SINE6, CTGGCGCCGGGCCTA

GGTGG (2' SPCR). These primed the amplification of the 3' flank of the SINE elements. A total of 17 polymorphic SINE markers were obtained from 4 restriction digests: *Bst*YI, *Bgl*II, *Bam*HI and *Cla*I. No markers were polymorphic between sample duplicates. Figure 11 shows an agarose gel of SINE SPCR products from a *Bgl*II digestion. Eleven markers have been excised from agarose gels, cloned and sequenced. All cloned sequences were confirmed to be of *Brassica* SINE origin by BLAST searches.

Figure 11. Agarose gel of *Bgl*II digestion SPCR SINE products. 27 OSR samples are shown (including duplicates) and *B. oleracea* (BO), *B. rapa* (BR) and *A. thaliana* (ARAB). Arrows indicate three scored SINE polymorphisms.



Binary marker scores were used to calculate a pair-wise simple distance among all samples. This distance matrix was then used to construct a neighbour-joining tree using the computer package, PHYLIP (Felsenstein, 1989) showing the genetic relationships among the samples (Figure 12). Two pairs of OSR varieties were not distinguished by the 17 SINE markers used: Winner / Shannon; and Royal / Recital. winter and spring OSR varieties were clearly divided into two clades. The winter clade consisted of a very closely related group (with short branch lengths) with Fortress and Lipton as outliers. Spring varieties showed larger within clade genetic distances than inter varieties and were paraphyletic with *B. oleracea*, *B. rapa* and *A. thaliana*. This suggests that at larger genetic distances (higher taxonomic levels) the SINE markers may not accurately resolve (evolutionary) relationships because their rate of evolution is too rapid. However, it is well suited to resolving closely related varieties. It should be noted, that within OSR, it is not expected that new transposition events will play a significant role. The observed differences between varieties are due to segregation of a common 'pool' of variable SINE markers in their parental genotypes through the selective breeding process.

This study has demonstrated that SINE sequences can provide a useful degree of polymorphism for variety identification and that a small number is required in order to distinguish common varieties. Compared to other commonly used marker systems, this approach is relatively cheap and fast, requiring basic molecular biology equipment with no special staining or electrophoretic techniques. It is expected that the design of further primers, adapters and use of further restriction digests can

increase the resolution of the method to distinguish all recommended OSR varieties. Once isolated as co-dominant markers, SINE markers will provide more power and the ability to easily detect hybrids of two or more varieties.

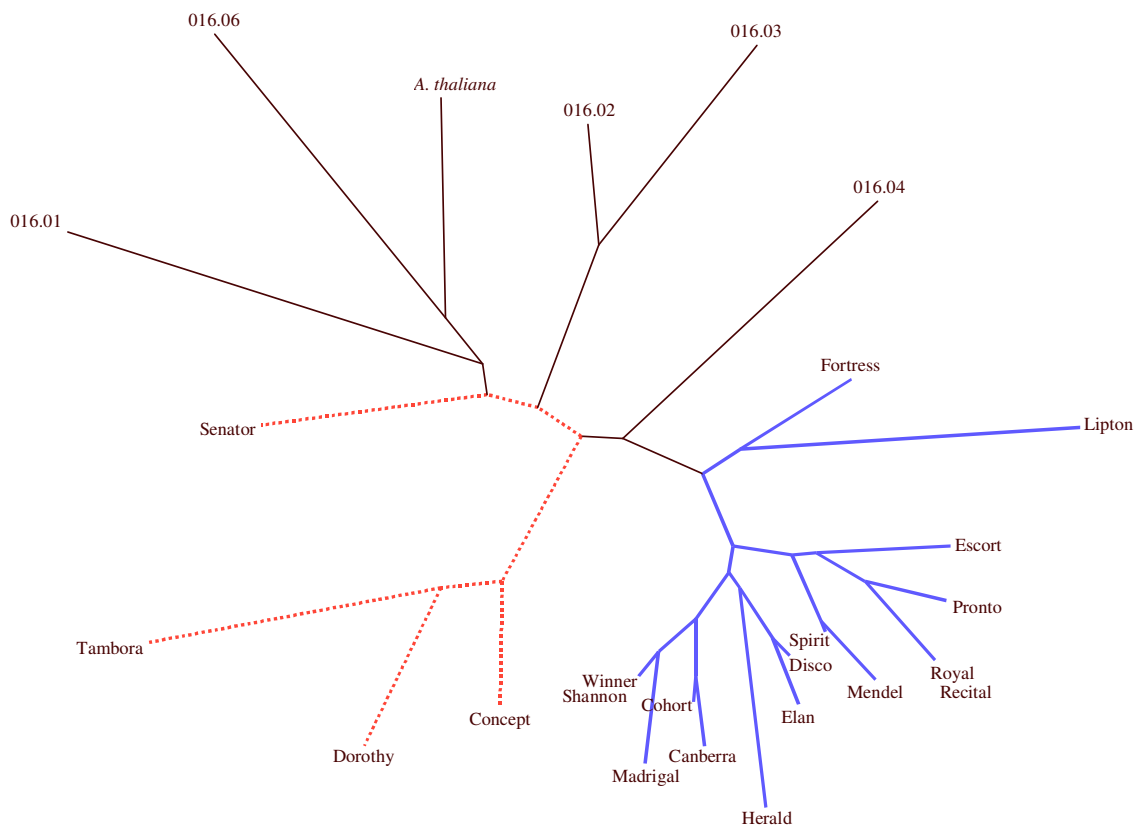


Figure 12. Neighbour-joining tree of pair-wise genetic distances among all OSR samples. Dashed (red) branches are the spring OSR clade and bold (blue) branches are the winter OSR clade. 016.01-016.06 are *B. oleracea* and *B. rapa* as shown in Table 6. Units are simple distances and branch lengths are all on the same relative scale.

## Conclusions

The work described in this report has served to consolidate and expand the capability of our research to support the CSL GM Inspectorate for England and Wales. In addition it has shown where unforeseen problems and complexities exist.

The development of new primer sets and primer matrices of PCRs was limited to a relatively small set of GM plants which were available to CSL. It has proved difficult to obtain many more GM plant samples due to legal constraints of their provision for scientific research. Under recent EU legislation (Regulation (EC) 1829/2003 (April 2004)) the Central Reference Laboratory (CRL), based at the Joint Research Centre (JRC), Ispra, Italy, is required to receive reference samples of GMOs from companies seeking marketing consent in the EU in order to validate GM identification methods. This has meant that companies now refuse to supply such GM materials to any labs other than the CRL (and JRC). We hope that this issue can be resolved in the near future to enable a wider development and validation of PCR methods such as those described here.

Many (>200) DNA sequences were examined for primer design purposes in this project. This gave us a good understanding of the number and variety of sequences currently released in GM events worldwide. This information is summarised more specifically elsewhere (Bruderer & Leitner, 2004), but an interesting conclusion from the alignment of many of these elements is that in the vast majority of cases single primer pairs, designed from the conserved regions among all variants, can be used for their identification. This could not be done in only two GM element 'families': *bar / pat*; and *Cry* for which specific primer pairs were required for each variant. In most cases, therefore, 'universal' primers can be designed for particular GM elements, e.g. P-35S, T-*nos* etc. but this work has highlighted the need for constant monitoring of new variants of these sequences in order to ensure their detection with PCR tests.

In GM screening tests there remains the challenge to identify all possible events in a single test. Several researchers have worked towards this with multiplexed methods (Hernandez *et al.*, 2005; Heo *et al.*, 2004; Rudi *et al.*, 2003; Xu *et al.*, 2005). Most promising of these is the combination of PCR amplification combined with microarray detection of products. This has the potential to screen for many hundreds of GM elements simultaneously with a test equivalent in time and cost to perhaps a few dozen PCR tests. Preliminary work at CSL (results not shown in this report) show that the GM primers ('outward' and 'inward') reported here work well in such PCR microarray screens and it is hoped that further work will provide complete GM screening microarrays which will be at the disposal of the GM inspectorate.

The SPCR technique, successfully applied in this work to the rapid cloning and sequencing of GM event flanks, has the capability to detect and isolate flanking DNA of any known target sequence. It will therefore be an excellent method for the rapid identification of the origin of 'single markers' in GM detection (Kerins, 2005). 'Single markers' are anomalous PCR test results in which only a single, apparently GM origin, sequence is amplified and no accompanying sequences expected from known GM events are amplified. Such results could arise from three possible causes: (i) bacterial or fungal contamination of samples; (ii) previously unreported fragments of GM events genetically segregating from the characterised event; and (iii) GM events

becoming fragmented by recombination or some other mechanism. Since causes (ii) and (iii) would constitute an unauthorised GM, it is important that detection methods can make this distinction. SPCR could be used to do this as the flanking sequence would indicate either bacterial or fungal DNA (therefore cause (i)) or plant / GM DNA (cause (ii) or (iii)). However, so far SPCR has not been validated with very low concentrations of target which are likely in these situations. This further work would be needed before its application to single marker situations.

Real-time PCR analysis of maize kernels has demonstrated that in circumstances where the genotype of the GM presence is known a direct conversion between different measures of GM content is possible. Even where the genotype is not known, a worst-case conversion of 2.52 could be applied ( $\% \text{ GM seed} = 2.52 \times \% \text{ GM DNA}$ ). However, there remains a need to measure these values for other maize crop types (e.g. sweetcorn) and the effect of including the whole plant on  $\% \text{ GM DNA}$  (i.e. fodder maize). In the latter case, total amplifiable DNA could be estimated from various maize plant tissues using the same real-time PCR methods described here and the mass of each used to estimate expected GM values in the resulting fodder.

The potential of SINE markers in OSR genetics has been demonstrated by this research work. It is expected that they can be developed further to increase their resolving power for OSR variety identification and be made into single locus markers, which will further simplify their application. In addition to OSR identification, SINEs could be used to identify related *Brassica* species and putative hybrids with OSR. Application of similar methods to other species will serve as a valuable tool to study coexistence of GM with non-GM crops and GM environmental containment.

## References

- Anonymous, (2004). Plant Varieties & Seeds Gazette Special Edition, No. 47. HMSO.
- Arumuganathan, K. & Earle, E. D. (1991). Nuclear DNA Content of Some Important Plant Species. *Plant Molecular Biology Reporter* 9(3): 208-218.
- Bruderer, S. & Leitner, E. (2003). Genetically Modified (GM) Crops: molecular and regulatory details. Centre for Biosafety and Sustainability (BATS) <http://www.gmo-watch.org/GVO-report140703.pdf>
- Deragon, J-M., Landry, B. S., Pelissier, T., Tutois, S., Tourmente, S. & Picard, G. (1994) An analysis of retro-position in plants based on a family of SINEs from *Brassica napus*. *Journal of Molecular Evolution* 39:378-386.
- Deragon, J-M, Gilbert, N., Rouquet, L., Lenoir, A., Arnaud, P. & Picard, G. (1996) A transcriptional analysis of the S1Bn (*Brassica napus*) family of SINE retrotransposons. *Plant Molecular Biology* 32(5):869-78.
- Ellis, T. H. N., Poyser, S. J., Knox, M. R., Vershinin, A. V. & Ambrose, M. J. (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. *Molecular and General Genetics* 260 (1): 9-19.
- Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.
- Giuliani, C., Consonni, G., Gavazz, G., Colombo, M. & Dolfini, S. (2002). Programmed cell death during embryogenesis in maize. *Annals of Botany*. 90(2): 287-292.
- Heo, M.S., Kim, J.H., Park, S.H., Woo, G.J. & Kim, H.Y. (2004) Detection of genetically modified maize by multiplex PCR method. *J. Microbiol. Biotechnol.* 14 (6), 1150-1156.
- Hernandez, M., Rodriguez-Lazaro, 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 (9), 3333-3337.
- Hsu, S-C. & Tsen, H-Y. (2001). PCR primers designed from malic acid dehydrogenase gene and their use for detection of *Escherichia coli* in water and milk samples. *International Journal of Food Microbiology* 64:1-11.
- Kerins, G. (2005) A review of single marker issues in GM enforcement and inspection. GM Inspectorate, CSL, Report to Defra.
- Kumar, A., & Bennetzen, J., L. (1993) Plant retrotransposons. *Annual Reviews of Genetics* 33:479-532.

- Lenoir, A., Cournoyer, B., Warwick, S., Picard, G. & Deragon, J-M. (1997) Evolution of SINE SI Retroposons in Cruciferae Plant Species. *Molecular Biology & Evolution* 14(9):934-941.
- Liu, Z. W., Fu, T. D., Tu, J. X. & Chen, B. Y. (2005) Inheritance of seed colour and identification of RAPD and AFLP markers linked to the seed colour gene in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics* 110 (2): 303-310.
- Pearce, S. R., Knox, M., Ellis, T. H. N., Flavell, A. J. & Kumar, A. (2000) Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Molecular and General Genetics* 263 (6): 898-907.
- Rudi, K., Rud, I. & Holck, A. (2003) A novel multiplex quantitative DNA array based PCR (MQDA-PCR) for quantification of transgenic maize in food and feed. *Nucleic Acids Research*. 31 (11), art. no.-e62.
- Spertini, D., Beliveau, C. & Bellemare, G. (1999). Screening Transgenic Plants by Amplification of Unknown Genomic DNA Flanking T-DNA. *Biotechniques* 27:308-314.
- Suwabe, K., Iketani, H., Nunome, T., Ohyama, A., Hirai, M. & Fukuoka, H. (2004) Characteristics of Microsatellites in *Brassica rapa* genome and their potential utilization for comparative genomics in Cruciferae. *Breeding Science* 54 (2): 85-90.
- Thangavelu, M., James, A., B., Bankier, A., Bryan, G., J., Dear, P., H., & Waugh, R. (2003) Happy mapping in a plant genome: reconstruction and analysis of a high-resolution physical map of a 1.9 Mbp region of *Arabidopsis thaliana* chromosome 4. *Plant biotechnology Journal* 1: 23-31.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673-80.
- Varshney, A., Mohapatra & T., Sharma, R. P. (2004) Development and validation of CAPS and AFLP markers for white rust resistance gene in *Brassica juncea* *Theoretical and Applied Genetics* 109 (1): 153-159.
- Vershinin, A. V., Allnutt, T. R., Knox, M. R., Ambrose, M. J. & Ellis, T. H. N. (2003) Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. *Molecular Biology and Evolution* 20 (12): 2067-2075.
- Weekes, R., Allnutt, T. R., Boffey, C., Morgan, S., Bilton, M., Daniels, R. & Henry, C. (2005). A study of crop to crop gene flow using farm scale sites of fodder maize (*Zea mays* L.) in the UK. *Transgenic Research*. In Review.
- Weller, S. A., Simpkins, S. A., Stead, D. E., Kurdziel, A., Hird, H., & Weekes, R. J. (2002). Identification of *Agrobacterium* spp. within *Brassica napus* seed by TaqMan

PCR - implications for GM screening procedures. *Archives of Microbiology*. 178: 338-343.

Windels, P., Taverniers, I., Depicker, A., Van Bockstaele, E. & De Loose, M. (2001). Characterisation of the Roundup Ready soybean insert. *European Food Research Technology* 213 (2): 107-112.

Xu, X.D., Li, Y.C., Zhao, H., Wen, S.Y., Wang, S.Q., Huang, J., Huang, K.L. & Luo, Y.B. (2005) Rapid and reliable detection and identification of GM events using multiplex PCR coupled with oligonucleotide microarray. *Journal of Agricultural and Food Chemistry* 53 (10), 3789-3794.

Yu, C. Y., Hu, S. W., Zhao, H. X., Guo A. G. & Sun, G. L. (2005) Genetic distances revealed by morphological characters, isozymes, proteins and RAPD markers and their relationships with hybrid performance in oilseed rape (*Brassica napus* L.) *Theoretical and Applied Genetics* 110 (3): 511-518.

## Appendix 1. PCR Primer Sequences.

NAME	Target Element	Sequence 5' - 3'	Length	T <sub>m</sub> (°C)
35SP1in	P-35S	ATCGTTCAAGATGCTCTGCCGA	23	59.6
35SP1out	P-35S	TCGGCAGAGGCATCTTGAACGAT	23	59.6
35SP2in	P-35S	TATAGAGGAAGGGTCTTGCGAAGGA	25	59.6
35sP4in	P-35S	TGCGTCATCCCTTACGTCAGTG	22	59.5
35sP4out	P-35S	CACTGACGTAAGGGATGACGCA	22	59.5
RR35s1	P-35S	GATAGTGGGATTTGTGCGTCA	20	55.4
35sP5in	P-35S	CAAATGCCATCATTTGCGATAAAGG	24	56.2
35sP5out	P-35S	CCTTTATCGCAATGATGGCATTTG	24	56.2
AMP1in	B-lactamase	GCATCTTACGGATGGCATGACAGT	24	59.6
AMP1out	B-lactamase	ACTGTCATGCCATCCGTAAGATGC	24	59.6
AMP2in	B-lactamase	CCCGTCGTGTAGATAACTACGATA	24	57.9
AMP2out	B-lactamase	TATCGTAGTTATCTACACGACGGG	24	57.9
BARNASE1in	<i>barnase</i>	GCACAGGTTATCAACACGTTTGA	23	56.0
BARNASE1out	<i>barnase</i>	TCAAACGTGTTGATAACCTGTGTC	23	56.0
BARNASE2in	<i>barnase</i>	CGTTGTTTTGTAAATCAGCCAGTC	24	56.2
BARNASE2out	<i>barnase</i>	GACTGGCTGATTACAAAACAACG	24	56.2
barnase3in	<i>barnase</i>	TACCGGTTATCAACACGTTTGAC	23	56.0
barnase3out	<i>barnase</i>	GTCAAACGTGTTGATAACCGGTA	23	56.0
BARSTAR1in	<i>barstar</i>	AGTATCAGCGACCTCCACCAGAC	23	61.3
BARSTAR1out	<i>barstar</i>	GTCTGGTGGAGGTCGCTGATACT	23	61.3
BARSTAR2in	<i>barstar</i>	ATGATGGTGATGTCGCAGCCTTC	23	59.6
BARSTAR2out	<i>barstar</i>	GAAGGCTGCGACATCACCATCAT	23	59.6
EPSPS1in	<i>epsps</i>	TTCGACAGCACCTTCATCGGCGA	23	61.3
EPSPS1out	<i>epsps</i>	TCGCCGATGAAGGTGCTGTGCAA	23	61.3
EPSPS2in	<i>epsps</i>	GTTCTTCAGACCGTTTCATCACGGT	25	61.2
EPSPS2out	<i>epsps</i>	ACCGTGATGAACGGTCTGGAAGAAC	25	61.2
EPSPS3out	<i>epsps</i>	ATTGCCGACGCCATCGATGATCC	23	61.3
EPSPS4in	<i>epsps</i>	CTGCAGGGCTTTGGCGCCA	19	61.6
EPSPS4out	<i>epsps</i>	TGGCGCCAAGCCCTGCAG	19	61.6
EPSPS5in	<i>epsps</i>	CGACAGCGAGAATCGGATATTC	22	57.7
EPSPS5out	<i>epsps</i>	GAATATCCGATTTCTCGTGTCG	22	57.7
FMV1in	P-FMV35S	GCACATGCATCATGGTCAGTAAGT	24	57.9
FMV1out	P-FMV35S	ACTTACTGACCATGATGCATGTGC	24	57.9
FMV2in	P-FMV35S	ATTAGTGAGTGGGCTGTGAGGACA	24	59.6
FMV2out	P-FMV35S	TGTCCTGACAGCCACTACTAAT	24	59.6
GUS1in	<i>uidA</i>	AACGATCAGTTCCCGATGCAGAT	24	59.6
GUS1out	<i>uidA</i>	ATCTGCATCGCGAACTGATCGTT	24	59.6
GUS2in	<i>uidA</i>	CATCACGCAGTTCAACGCTGACAT	24	59.6
GUS2out	<i>uidA</i>	ATGTCAGCGTTGAACTGCGTGATG	24	59.6
NOSP1in	<i>P-nos</i>	ACGTCAGAAACCATATTGCGCGTT	25	57.9
NOSP1out	<i>P-nos</i>	AACGCGCAATAATGGTTTCTGACGT	25	57.9
NOSP2in	<i>P-nos</i>	TTGGATACCGAGGGGAATTTATGGA	25	57.9
NOSP2out	<i>P-nos</i>	TCCATAAATCCCTCGGTATCCAA	25	57.9
NOST1in	<i>T-nos</i>	TGTTGCCGGTCTTGCGATGATTA	23	57.8
NOST1out	<i>T-nos</i>	TAATCATCGCAAGACCGCAACA	23	57.8
NOST2in	<i>T-nos</i>	CGGGACTCTAATCATAAAAACCCAT	25	56.3
NOST2out	<i>T-nos</i>	ATGGGTTTTTATGATTAGAGTCCCG	25	56.3
NPT1in	<i>nptII</i>	ATGCTCTTCGTCAGATCATCCIGA	25	59.6
NPT1out	<i>nptII</i>	TCAGGATGATCTGGACGAAGAGCAT	25	59.6

NPT2in	<i>nptII</i>	AGCTGTGCTCGACGTTGTCAGTAA	25	61.2
NPT2out	<i>nptII</i>	TTCAGTGACAACGTCGAGCACAGCT	25	61.2
OCST1in	<i>T-ocs</i>	AGCGACATCTATGATAGAGCGCCA	24	59.6
OCST1out	<i>T-ocs</i>	TGGCGCTCTATCATAGATGTCGCT	24	59.6
OCST2in	<i>T-ocs</i>	TGTCAAGGTTTGACCTGCACTTCA	24	57.9
OCST2out	<i>T-ocs</i>	TGAAGTGCAGGTCAAACCTTGACA	24	57.9
PATBAR1in	<i>pat or bar</i>	AGACAAGCACGGTCAACTTCCGTA	24	59.6
PATBAR1out	<i>pat or bar</i>	TACGGAAGTTGACCGTGCTTGTCT	24	59.6
PATBAR2in	<i>pat or bar</i>	GGACTTCAGCAGGTGGGTGTAGA	23	61.3
PATBAR2out	<i>pat or bar</i>	TCTACACCCACCTGCTGAAGTCC	23	61.3
PATBAR5in	<i>pat or bar</i>	ATCGTTAACCATTTACATTGAGACGTC	26	56.4
PATBAR5out	<i>pat or bar</i>	GACGCTCAATGTAAATGGTTAACGAT	26	56.4
PATBAR6in	<i>pat or bar</i>	GCAAAATGTGTGTACAATGTGGATCC	25	57.9
PATBAR6out	<i>pat or bar</i>	GGATCCACATTGTACACACATTTGC	25	57.9
35sT1in	<i>T-35S</i>	GATCGACAAGCTCGACGATCTTCG	25	62.9
35sT1out	<i>T-35S</i>	CGAAGATCCGTCGAGCTTGTGATC	25	62.9
35sT2in	<i>T-35S</i>	GGGTTTCTTATATGCTCAACACATG	25	56.3
35sT2out	<i>T-35S</i>	CATGIGTTGAGCATATAAGAAACCC	25	56.3
35sT3in	<i>T-35S</i>	TAATAATGTGTGAGTAGTCCCAGA	25	54.7
35sT3out	<i>T-35S</i>	TCTGGGAATACTACACATTATTA	25	54.7
35sT4in	<i>T-35S</i>	AATAATGTGTGAGTAGTCCCAGA	24	54.4
35sT4out	<i>T-35S</i>	TCTGGGAATACTACACATTAT	24	54.4
35sT5in	<i>T-35S</i>	CCAAAATCCAGTACTAAAATCCAG	24	54.4
35sT5out	<i>T-35S</i>	CTGGATTTTAGTACTGGATTTTGG	24	54.4
TA29Ain	<i>P-TA29</i>	GCATAAGTGGGTGGCTGGACTAGA	25	62.9
TA29Aout	<i>P-TA29</i>	TCTAGTCCAGCCACCCACCTTATGC	25	62.9
TA29Bout	<i>P-TA29</i>	GCCCTTGTGGTGCAAGTAAACAGT	25	61.2
TA29Bin	<i>P-TA29</i>	ACTGTTACACTTGCACCACAAGGGC	25	61.2
T7t1in	<i>T-g7</i>	CAAGTCAGGTATTATAGTCCAAGC	24	56.2
T7t2in	<i>T-g7</i>	ATCTACGGCAATGTACCAGCTGA	23	57.8
T7t1out	<i>T-g7</i>	GCTTGGACTATAATACCTGACTTG	24	56.2
T7t2out	<i>T-g7</i>	TCAGCTGGTACATTGCCGTAGAT	23	57.8
LelF	soybean endogenous	GCCCTCTACTCCACCCCATCC	22	65.1
LelR	soybean endogenous	GCCCATCTGCAAGCCTTTTGTG	23	59.6
LE3	soybean endogenous	ATTGTGACCTCCTCGGAAAGTTAC	25	59.6
LE4	soybean endogenous	GGCTTAGTGTCAATTGGTGCGAGA	24	59.6
Puc18a	pUC plasmid backbone	GATCTTTTCTACGGGGTCTG	20	55.4
Puc18b	pUC plasmid backbone	GGTCTGACGCTCAGTGGA	18	57.2
OSR6F	OSR endogenous	AACAGAGGATTGATTCTGCTTCG	23	56.0
OSR6R	OSR endogenous	CAAAAGTCAAATCACAAGAGCCA	23	54.2
SSIIA	maize endogenous	TGGTACAAGCGGAAGCAGCAG	21	59.5
SSIIB	maize endogenous	TCAGCTTTGGGTCCGGACAC	20	59.5
ECOL1	<i>E. coli</i>	ACTGAAAGGCAAACAGCCAAG	21	55.6
ECOL2	<i>E. coli</i>	CGTCTGTTCAAATGGCCTCAGG	23	59.6
FLIG1	<i>A. tumefaciens</i>	CCGGCAAGCTGCTGAAAT	18	54.9
FLIG2	<i>A. tumefaciens</i>	CGTCTGAGCCGAGGAAATGA	20	57.5
CAMVRT1in	<i>RT</i>	CCATCAGCAAAGTCTGCAAAGACAT	25	57.9
CAMVRT1out	<i>RT</i>	ATGTCTTTGCAGACTTTGCTGATGG	25	57.9
CAMVRT2in	<i>RT</i>	GCTTTCATCCATTGCTTAGTCTTGT	25	56.3
CAMVRT2out	<i>RT</i>	ACAAGACTAAGCAATGGATGAAAGC	25	56.3
CAMVRT3	<i>RT</i>	CCTCGAATGAGTGTAAAGTAACTCG	24	57.9

## Appendix 2. Standard Operating Procedures

The following SOPs constitute detailed methods recorded for quality control purposes within the Central Science Laboratory in the context of the project, "Detection and Traceability Technologies to Underpin the GM Inspectorate". They should not be interpreted as procedures formally approved by Defra.

## SOP 1. General PCR Protocol

### Reagents:

RedTaq Readymix PCR mix (Sigma R-2523)

Molecular biology grade water

Filtered sterile pipette tips

Appropriate PCR well plate or tubes for thermal cycler

1. Prepare a master reaction mix for all PCR reactions as follows:

Total volume per PCR = 10  $\mu$ l

Reagent	Stock Concentration	Volume ( $\mu$ l)	Final Concentration
RedTaq Readymix	2X	5	1X
primer 1	5 $\mu$ M	1	0.5 $\mu$ M
primer 2	5 $\mu$ M	1	0.5 $\mu$ M
Water	-	1	-

2. Add 8 $\mu$ l master mix to each reaction well / tube.

3. Add 2 $\mu$ l sample DNA (approx. 50ng) to each reaction well / tube.

n.b. Where more than two primers are used, reduce water volume accordingly. Where primers are to be added separately, reduce volume of mix to be added to each reaction accordingly.

4. Cycle parameters:

i) Initial 95 °C, 1 min.

ii) 95 °C, 1 min.

iii) 55 °C, 1 min.

iv) 72 °C, 1 min.

Repeat (ii) to (iv) for 45 cycles.

## SOP 2. Suppression PCR.

### Equipment:

RedTaq Readymix PCR mix (Sigma R-2523)  
Appropriate restriction enzymes and supplied buffers  
Appropriate oligonucleotides (see tables below)  
Streptavidin Magnasphere paramagnetic particles (PMPs) (Promega #Z5481)  
T4 DNA ligase and supplied buffer (Promega)

### Protocol:

#### A. Preparation

1. Select restriction enzyme (RE) and adapter combination from Table 1. Adapter sequences are shown in Table 2.

Table 1. Adapter - RE combinations

Adapter	RE	Digest Temperature (°C)
CG	<i>TaqI</i>	65
	<i>Clal</i>	37
	<i>MspI</i>	37
GATC	<i>BstYI</i>	60
	<i>BclI</i>	50
	<i>BglII</i>	37
	<i>BamHI</i>	37
	<i>MboI</i>	37
CTAG	<i>SpeI</i>	37
	<i>NheI</i>	37
	<i>XbaI</i>	37

Table 2. Adapter oligo sequences

Name	Oligo 1 5'-3'	Oligo 2 5'-3'
CG	AGCGTCCAGCTGAGCTGACCTGGTTCATGCCGTCGTGCGCCGTGTC	CGGACACGGCAAA
GATC	AGCGTCCAGCTGAGCTGACCTGGTTCATGCCGTCGTGCGCCGTGTC	GATCGACACGGCAAA
CTAG	AGCGTCCAGCTGAGCTGACCTGGTTCATGCCGTCGTGCGCCGTGTC	CTAGGACACGGCTTT

2. To make up adapters, mix 50 µl of oligo 1 & 2 (both at 50µM in TE or water) together in a 500 µl eppendorf tube. Heat to 95°C for 5 mins. Allow to cool to room temperature and stand for 30 mins. Store at -20°C. This is a 50X stock solution (25 µM). When required make a 10X stock solution by 1:5 dilution of the 50X in water. The volume required should be calculated beforehand.

3. DNA samples should be of sufficient quality for restriction digestion (can be tested by digestion beforehand).

4. Two nested forward primers (Fwd1 and Fwd2) are required for the PCR step, these should be as close to each other as possible in the target sequence and as close as

possible to the sequence flank to be isolated, while providing sufficient target sequence for positive identification of the flank sequence.

## B. Restriction Digestion and Ligation

1. Make up restriction digests of each DNA sample as shown below. Incubate at appropriate temperature for at least 3 hours.

Total volume	20 $\mu$ l
RE	1 $\mu$ l of (1 Unit = 1:10 dilution of 10U/ $\mu$ l stock or equivalent)
RE Buffer	1 $\mu$ l (of 10X stock as supplied)
DNA	10 $\mu$ l (approx. 1 $\mu$ g DNA)
Water	8 $\mu$ l

2. To each digestion add the following ligation reagents. Incubated at room temperature overnight.

Total volume	40 $\mu$ l
Above Digestion solution	20 $\mu$ l
Adapter	4 $\mu$ l (of 10X solution)
T4 DNA ligase	1 $\mu$ l(1U= 1 $\mu$ l of 1:5 dilution of 5U stock or equivalent)
ligase buffer	4 $\mu$ l (10X suppliers buffer)

## C. Nested PCR

1. Adapter primer sequences are as follows:

P3 (1'PCR) 5'-3' = AGCGTCCAGCTGAGCTG;

P4 (2'PCR) = CCTGGTTCATGCCGTCG.

2. Set-up the primary PCR as follows using the restriction / ligations (RLs) prepared above:

Total volume	10 $\mu$ l
RL	1 $\mu$ l
5X PCR mix	5 $\mu$ l
Primer Fwd1	1 $\mu$ l (5 pmol)
Primer P3	1 $\mu$ l (5 pmol)

Cycling conditions

95°C 1min

95°C 1min; 60°C 1min; 72°C 1min, for 25 cycles

3. Following primary PCR, dilute reactions 1:100 with water (add 1  $\mu$ l of reaction to 99  $\mu$ l water in new tube) and mix well. If using PMPs follow alternative protocol below.

4. Set-up the secondary PCR as follows:

Total volume	10 $\mu$ l
1:100 1' PCR	1 $\mu$ l
5X PCR mix	5 $\mu$ l
Primer Fwd2	1 $\mu$ l (5pmol)
Primer P4	1 $\mu$ l (5pmol)

Cycling conditions

95°C 1min

95°C 1min; 60°C 1min; 72°C 1min, for 35 cycles

5. Note that annealing temperature for 1' and 2' PCRs may require optimisation for different primers.

6. Analyse 2' PCR products on appropriate agarose or acrylamide electrophoresis gels.

#### **D. Using PMPs**

1. PMPs can increase the sensitivity and specificity of the SPCR. If specific products are not obtained using ordinary primers then labelled primers can be used. The primary PCR forward primer of the sequence of interest is labelled at 5' end with biotin (e.g. see Sigma Genosys oligo synthesis service).

2. Following 1'PCR with labelled Fwd primer, add entire 10  $\mu$ l of PCR to 185  $\mu$ l water. Add 5  $\mu$ l of resuspended PMPs and mix well.

3. Place the suspension in a suitable magnetic separation stand and allow PMPs to collect against magnet side of tube.

4. Carefully remove the liquid and wash PMPs with 500  $\mu$ l water. Repeat step 3.

5. Resuspend PMPs in 5  $\mu$ l water. Add 1  $\mu$ l of the re-suspension to 2' PCR (in place of 100  $\mu$ l diluted 1'PCR).

### SOP 3. DNA Extraction.

#### Equipment:

Waterbath (preheated to 65°C)

1.5 ml Eppendorf tubes

Grinding equipment: mortars and pestles, liquid nitrogen; or screw cap 2 ml tubes and Zirconia beads (0.3 - 0.5mm)

Bench-top centrifuge

CTAB Extraction buffer:

2% CTAB (20 g)

100mM Tris-HCl, pH 8.0 (100 ms, 1M)

20mM EDTA (40 ml, 0.5M)

1.4M NaCl (82 g)

1.0% NaSO\*

2.0% Polyvinylpyrrolidone(PVP)-40

\*PVP and NaSO added fresh to an aliquot of stock buffer immediately prior to extraction. Stock buffer can be autoclaved and stored at room temperature. n.b. values in brackets are amounts to make up 1 litre of buffer.

TE Buffer: 10mM Tris-HCl, pH 8.0; 1mM EDTA

Isopropanol

70% Ethanol

TE Buffer

Chloroform: isoamyl alcohol 24:1

#### Protocol:

1. Weigh 0.2 - 0.5 g of sample plant tissue.
2. For tough tissues, e.g. maize leaves, freeze with liquid nitrogen in a mortar and grind to fine powder. for soft tissues, e.g. seed embryos, *Brassica* leaves, add sample and CTAB buffer as step 2 to 2 ml screw cap tube with approx. 10 zirconia grinding beads (0.3-0.5 mm) and grind in mixer mill for 1 min.
3. Transfer ground sample to 1.5 ml eppendorf, add 700 µl CTAB buffer (if not already added at grinding stage), mix and incubate 65°C for 30 min.
4. Add 500 µl chloroform: isoamyl alcohol, mix very thoroughly, centrifuge in bench top minifuge at full speed for 10 mins.
5. Remove the supernatant and place in a fresh 1.5 ml eppendorf and repeat step 4.
6. Carefully remove upper (aqueous) layer to avoid transfer of interface residues and transfer to a fresh 1.5 ml tube. Add 700 µl isopropanol and mix.
7. Pellet the DNA by placing the tube in a microfuge for 10 min at full speed.
8. Carefully remove supernatant from DNA pellet. The pellet may be gelatinous at this stage.
9. Wash the pellet with 1 ml 70% ethanol, centrifuge 2 mins full speed.
10. Remove ethanol and carefully dry pellet with a twist of tissue.
11. Dissolve pellet in 100 µl TE buffer.

## SOP 4. Cloning of PCR Products

### Equipment:

Competent *E. coli* cells, stored at -70°C (JM109 or DH5 $\alpha$ );  
pGEM-T vector kit (Promega)  
Machery-Nagel Nucleospin gel extraction kit (or equivalent)  
LB broth  
LB agar plates (50  $\mu$ g/ml ampicillin, 0.5 mM IPTG, 40  $\mu$ g/ml X-Gal)  
RedTaq Readymix PCR mix (Sigma R-2523)

### Protocol:

#### A. Extraction of PCR product

1. Excise agarose gel slice containing target PCR product and place in 1.5 ml eppendorf tube.
2. Follow Machery-Nagel Nucleospin gel extraction kit procedure to extract DNA from gel slice, elute in 60  $\mu$ l.
3. Add 100  $\mu$ l isopropanol and centrifuge in minifuge full speed for 10 mins.
4. Remove supernatant and dry pellet carefully using a twist of tissue

#### B. Ligation of PCR product

1. Defrost components of pGEM-T kit on ice.
4. Add 3.5  $\mu$ l water, 5  $\mu$ l 2Xligase buffer, 0.5  $\mu$ l pGEM-T vector and 1  $\mu$ l T4 DNA ligase, to DNA pellet from A mix well.
5. Incubate 4°C overnight.

#### C. Transformation

1. Ensure sufficient agar plates and LB broth are available (one plate per transformation).
2. Defrost 50  $\mu$ l aliquot of competent cells on ice, use as soon as defrosted, do not leave for a prolonged time.
3. Add entire ligation reaction (10  $\mu$ l) to competent cells and mix gently, incubate on ice for 20 mins.
4. Heat shock cells at exactly 41°C, 1 min 30 secs for JM109 or 37°C, 15 secs for DH5 $\alpha$ .
5. Place cells on ice 2mins.
6. Gently pipette cells into chilled 250 ml LB broth (no antibiotic). Mix gently and incubate 37°C, 45 mins.
7. Pipette gently 100  $\mu$ l of culture onto LB agar (+AMP,X-GAL,IPTG - IPTG not required for DH5 $\alpha$ ) and spread evenly. Dry plate in laminar flow for 5 mins, incubate 37°C overnight.

#### D. Colony PCR Screening

1. Set up PCR reaction mix, per tube, as follows:

5  $\mu$ l RedTaq Readymix  
1  $\mu$ l M13 forward primer (#195 Appendix 1) 5  $\mu$ M  
1  $\mu$ l M13 reverse primer (#196 Appendix 1) 5  $\mu$ M  
3  $\mu$ l molecular biology grade water

2. Add 10  $\mu$ l above mix to each colony PCR tube or well of 96 well PCR plate (typically 8 wells per LB plate).
3. Use a small pipette tip to pick colonies for screening. If required, touch the tip onto a replica agar plate (+ampicillin) marked with a grid for future culturing and reference. Place the tip into a reaction mix tube.
4. Once all colonies to be screened have been picked, remove tips from PCR tubes and seal.
5. Run PCR with following cycles: 95°C, 1'; 60°C, 1'; 72°C, 1' for 25 cycles. Separate products on 1.4% agarose gels.
6. Excise candidate bands from gel and extract according to (SOP4 A). For each band pipette 30  $\mu$ l to two separate tubes and submit for sequencing (forward and reverse primers).

## SOP 5. General Real-Time PCR (TaqMan) Protocol

### Equipment:

ABI 770 or 7900 sequence detection system (SDS).

ABgene QPCR readymix, with ROX.

96 or 384 well PCR plate.

Primer and probe sets

Molecular biology grade water

Filtered sterile pipette tips

### Protocol:

Primer and probe should be designed in accordance with Applied Biosystems guidance (see ABI SDS user manual). The following protocol was designed with 5'FAM and 3'TAMRA labelled probes.

1. Prepare reaction mastermix per reaction as follows:

Total volume per PCR = 25  $\mu$ l

Reagent	Stock Concentration	Volume ( $\mu$ l)	Final Concentration
ABgene QPCR readymix	2X	12.5	1X
primer 1	5 $\mu$ M	2.5	0.5 $\mu$ M
primer 2	5 $\mu$ M	2.5	0.5 $\mu$ M
probe	5 $\mu$ M	2.0	0.4 $\mu$ M
Water	-	0.5	-

2. Add 20  $\mu$ l of mastermix to each plate well.

3. Add DNA sample (approximately 50 - 100 ng) in 5  $\mu$ l. Seal the plate and place in SDS.

4. Use following general parameters:

i) Initial 95 °C, 15 min.

ii) 95 °C, 1 min.

iii) 60 °C, 1 min.

Repeat (ii) to (iii) for 35-45 cycles, depending on sensitivity of assay.