Final Technical Report for Project G03026

Development of a standard, validated procedure for the isolation of transgene flanking regions in GM crops and detailed analysis of transgene insertion.

Contractors:

JIC - John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH (Lead contractor)

SCRI - Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA

Project Leaders:

Dr Wendy Harwood, JIC

Dr Mark Taylor, SCRI

Project Staff:

Mark Smedley, JIC

Danny Cullen, SCRI

Contents:

1. **Lay Person's Executive Summary** **Page 3**
2. **Background / Introduction** **Page 5**
3. **Objectives** **Page 7**
4. **Results** **Page 9**
   - **4.1 Objective 1** **Page 9**
   - **4.2 Objective 2** **Page 31**
   - **4.3 Objective 3** **Page 39**
   - **4.4 Objective 4** **Page 50**
   - **4.5 Objective 5** **Page 53**
5. **Conclusions and Recommendations** **Page 59**
6. **References** **Page 60**
7. **Outputs** **Page 60**
8. **Appendices**

Standard Operating Procedures

8.1 SOP for the isolation of plant genomic DNA from potato leaf material.

8.2 SOP for the isolation of plant genomic DNA from barley leaf material.

8.3 SOP for the isolation of transgene flanking regions in single copy lines of GM potato.

8.4 SOP for the isolation of transgene flanking regions in multicopy lines of GM potato.

8.5 SOP for the isolation of transgene flanking regions in single copy lines of GM barley (incorporating SOP for the isolation of transgene flanking regions in biolistic-derived lines and in a range of different GM crop types).

1. Lay Person's Executive Summary

Background

Methods for the development of genetically modified (GM) crops are continually improving such that the process can be more precisely controlled. One aspect of the genetic modification process which it has not been possible to control to date is the exact location of the additional inserted DNA within the plant DNA. Therefore, as part of the analysis and safety assessment of GM crops it is important to find out exactly where the additional DNA (transgene) has inserted. This is important because it may highlight whether a plant gene has been disrupted or whether there are rearrangements of the inserted DNA that might be of concern. The site where the transgene inserts is analysed by isolating the DNA at the junction where the transgene and the plant DNA join. Although there are a number of methods available that allow this 'transgene junction region' to be isolated, in general these methods have been used only with the insertion of specific transgenes or only in specific plant species. The aim of this project was to develop standard and validated methods to allow the routine isolation of the transgene junction regions in a range of different GM crops with diverse transgene insertions of different complexity.

Objectives and Approach

The objectives of the project were firstly to develop suitable methods for the isolation of transgene junction regions in relatively simple GM lines of barley and potato. This was done by comparing a number of commercially available kits with a standard laboratory protocol developed in an earlier project. Secondly, the methods developed were extended and tested in more complex GM potato lines and also in a range of other GM crops. During the project, GM barley lines produced using two different methodologies for introducing the transgene (Biolistics and *Agrobacterium*) were analysed. Finally, the methods developed were validated. As two laboratories were involved in the project (the John Innes Centre (JIC) and the Scottish Crop Research Institute (SCRI)), the methods developed at each site were tested at the other site and the results compared. For particular GM lines additional analysis was carried out to more fully characterise the entire transgene insertion.

Key Results

During the project it emerged that two commercially available kits were appropriate for the isolation of transgene junction regions. The first kit, the DNA Walking SpeedUp™ Kit II [Seegene] was found to be suitable for use in all GM crop types tested where the number of transgene insertions was low. This kit produced reliable and rapid results from small amounts of DNA. However, it was necessary to make some modifications to the recommended protocol to ensure that no artifacts were produced. The second kit, the APAgene™ GOLD Genome Walking Kit [BIO S&T] was found to be appropriate for the analysis of potato lines with both simple and more complex transgene insertions and gave reliable results. In order to analyse the more complex lines a number of additional steps had to be introduced to the protocol. For both kits additional steps to confirm that the correct DNA regions had been isolated were included. Full standard operating procedures (SOPs) are included in the final report.

During the validation stage, the SOPs developed by both sites were proved to be robust and were easily followed by the other site and yielded identical results. During the isolation and

analysis of transgene junction regions throughout the project, a number of lines were found to have additional or rearranged sequences of DNA present. The most frequent rearrangement seen was the insertion of additional sequence at one end of the expected transgene insertion, often this was in reverse orientation. It is important to note that many of the GM lines analysed in this project were developed a number of years ago using DNA vectors that have subsequently been improved. This may account for the high levels of rearrangements observed. Analysis of more recent GM lines with only single copies of the transgene present found that 14% had rearrangements with additional sequence inserted.

Conclusions

From the work carried out in this project, it is now possible to recommend the SOP using the Seegene DNA Walking SpeedUp™ Kit II for the analysis of a wide range of GM plant material with low numbers of transgene insertions. For more complex material such as the potato lines analysed during the project, the SOP using the APAgene™ GOLD Genome Walking Kit together with the additional analysis steps is appropriate.

The results obtained during the project have identified frequent rearrangements at the transgene insertion site. Further analysis is required to identify the reason for such rearrangements and in some cases to confirm where the additional sequences originate from.

2. Background / Introduction

Methodologies for the production of GM plants are continually being improved such that there is greater control over the transformation process. However, one aspect of the process that has not been controlled to date is the exact location of the transgene insertion within the host genome. For this reason, analysis of the transgene insertion site is an important part of any safety assessment. By determining the genomic regions flanking the transgene insertion it is possible, by sequence alignment with published mapping, EST and sequence data, to determine the position of the transgene in the genome. It can also determine whether any functional or regulatory genes have been disrupted and whether any new open reading frames (ORFs) have been created, resulting in the possible synthesis of novel proteins. Knowledge of the flanking sequences also provides information on unique identification sequences that can be used to detect the transgenic event in question (traceability).

There are currently a number of PCR based protocols that have been used to isolate the genomic regions flanking transgenes. However, methodology for the isolation of transgene flanking regions needs to be applicable to a range of crop species, a range of insertion types and complexity and to be able to cope with transgenic material derived using different techniques. The primary aim of this project was therefore to develop standard, validated procedures for the isolation of transgene flanking regions in a range of GM crops, with a range of different transgene insertion patterns.

PCR-based methodologies that have been used for isolating transgene flanking regions include inverse PCR (IPCR), ligation mediated or adaptor mediated PCR (LM-PCR) and thermal asymmetric interlaced PCR (Tail PCR). Each of these methods has been used in the large scale isolation of transgene flanking regions, for example LM-PCR (Sallaud et al., 2004), IPCR (An et al., 2003) and TAIL-PCR (Sha et al., 2004). In the G02 programme these methodologies were tested in the cereal species, wheat and barley (Salvo-Garrido et al. 2004). It was found that the best technology for the routine isolation of transgene flanking regions from different insertion events was LM-PCR (Harwood et al., 2005). Although each of these methods have been employed effectively in particular crops containing particular types of transgene insertion, they can be difficult to optimize for a range of different insertions as they are hampered by steps in the protocols that are laborious and technically demanding. This makes the routine and reproducible analysis of transgene insertion sites challenging and places a greater burden on both the applicant and the regulator in the risk assessment procedure. Additionally quality control of the entire process is difficult as the protocols generally involve multiple steps, employing reagents and equipment from many different suppliers. Thus there is still a requirement for stream-lining protocols and determining the key quality control parameters that must be addressed. A direct comparison of results achieved in different laboratories with the same methodology is a fundamental requirement in the development of a standard validated protocol.

Recently a number of genome walking kits have become available that may be appropriate for the isolation of transgene flanking regions. These include the APAgene™ GOLD Genome Walking Kit [BIO S&T], the DNA Walking SpeedUp™ Kit II [Seegene] and the Universal Vectorette™ System [Sigma]). During this project, these kits were compared to the best methodology identified in G02 for the isolation of transgene flanking regions. All of the transgenic plants tested in this project have been developed and characterized previously.

Part of the project focused in detail on a range of potato and barley transgenic material demonstrated to contain low copy number inserts. In addition, potato lines with multiple copies of the same insert or lines that had been co-transformed or re-transformed were included together with transgenic lines of a range of other crop species. In addition to GM material derived using *Agrobacterium*-mediated techniques, material derived by biolistics was also included. Previous research has shown that GM lines derived by biolistics tend to have a higher number of transgene copies, more rearrangements of the inserted DNA and tend to be more susceptible to silencing of the transgenes in later generations (Travella et al 2005). The outputs of the project therefore include recommended and validated standard operating procedures for the isolation and analysis of transgene flanking regions in a range of GM crop plants with a range of insert complexity.

3. Objectives

Objective 01

Compare the adaptor-mediated PCR technique developed in G02 with the three kits currently available for isolation of transgene flanking regions. For each method determine the false positive rate, reproducibility, cost and throughput.

01.1 Compare the optimized adaptor-mediated PCR method for junction sequence analysis with the kits – develop a standard operating procedure for potato using transgenic lines with low copy number.
01.2 Compare the optimized adaptor-mediated PCR method for junction sequence analysis with the kits – develop a standard operating procedure for barley using transgenic lines with low copy number.

Objective 02

Apply the best methods from objective 1 to transgenic potato lines with multiple insertions. Optimise protocols to cope with multiple insertion lines or stacked transgenes including lines derived by co-transformation and re-transformation. Develop a standard operating procedure for multiple copy transgenics. If appropriate, investigate whether it is possible to multiplex the genome walking kit for the simultaneous determination of several transgene junction sequences.

02.1 Optimize the protocol for transgenic lines with multiple insertions of the same transgene and for stacked transgenes derived by co-transformation and re-transformation.
02.2 Develop a multiplex PCR method for the analysis of stacked transgenes.

Objective 03

Test the best method(s) from objective 1 in a range of different transgenic crop species including wheat, barley, potato, sugarbeet, oilseed rape, *Brassica oleracea*, maize and pea. Develop a standard operating procedure for this method considering factors such as the best protocol for genomic DNA preparation from leaf tissue of a range of crops.

03.1 Evaluate charge switch technology for DNA extraction in a range of crop plants.

Objective 04

Develop protocols to allow the analysis of GM material derived by biolistics rather than *Agrobacterium*. Develop a standard operating procedure for biolistic derived material

Objective 05

Validate the procedures developed in objectives 1-4 by isolating transgene junction regions from the same standard set of GM samples at two different sites and compare the results. Further validate by sequencing the entire transgene insertion in a subset of GM lines. Compare the pattern of transgene insertion seen in a range of transgenic crops.

05.1 Complete two-site validation of standard operating procedures for junction sequence analysis.
05.2 Sequence and analyse the entire transgene insertion in co-transformed potato lines and barley lines transformed by both biolistics and *Agrobacterium.*

4. Results

4.1 Objective 01

Compare the adaptor-mediated PCR technique developed in G02 with the three kits currently available for isolation of transgene flanking regions. For each method determine the false positive rate, reproducibility, cost and throughput.

01.1 Compare the optimized adaptor-mediated PCR method for junction sequence analysis with the kits – develop a standard operating procedure for potato using transgenic lines with low copy number.

Evaluation of DNA extraction methods for potato leaf material.

Six DNA extraction methods were initially compared to determine the optimal procedure for potato leaf material in three transgenic lines. These methods included the main principles of DNA purification: spin columns (DNeasy®, NucleoSpin®), magnetic resins (ChargeSwitch®), in-house manual methods (CTAB-based), and proprietary reagents (DNAzol®, AquaGenomic™). The methods were assessed for yield and purity using spectrophotometry and the PicoGreen fluorescence-based DNA quantification assay. Real-time amplification of the *npt*II gene (encodes kanamycin resistance) located on vector pBIN19 was also used to assess template DNA quality. Details of the procedures are outlined in the SOP (appendix 8.1).

Agarose gel electrophoresis indicated that all of the extraction methods enabled recovery of predominantly large molecular weight fragments of DNA (20-25 kb) from *S. tuberosum* or *S. phureja* leaf material (data not shown), suggesting that the DNA was not excessively sheared and therefore suitable for PCR-based analyses. There were considerable differences in the yields of DNA extracted between the commercial kits and the manual and proprietary reagent based procedures, and between spectrophotometry and PicoGreen measurements. The highest yields of DNA based on PicoGreen quantification were obtained with the CTAB method (16.9-32.7 µg/100mg) followed by AquaGenomic Solution (8.34-10.38 µg/100 mg), whereas the DNeasy, Nucleospin and ChargeSwitch kits yielded lower (0.16 to 7.48 ug/100mg) but similar DNA amounts **(Table 1).**

It was evident that spectrophotometric measurements at 260 nm were overestimating the concentration of DNA by up to 10x in relation to fluorometric quantification using the PicoGreen dye, particularly with the more manual based procedures. It is well known that the major disadvantage of absorbance methods are the large contribution to the signal and interference caused by contaminants co-extracted in DNA preparations (RNA, proteins, carbohydrates, phenolics), whereas the PicoGreen assay is selective for double-stranded DNA only and therefore regarded as a more accurate and sensitive quantification method.

Table 1. Performance Summary of DNA Extraction Methods

Method	Concentration Range (ng/µl): Spectrophotometry	PicoGreen	Mean Purity Ratios: $A_{260/A280}$	$A_{260/A230}$	Range of Yield (µg / 100 mg)[a]	Real-time PCR Mean Ct-values[b]
DNeasy® Plant Mini Kit	21.7 – 95.4	12.9 – 40.4	1.79	2.12	1.93 – 6.06	22.7 ± 0.13
NucleoSpin® Mini Plant	2.6 – 9.6	1.1 – 10.7	1.75	1.47	0.16 – 1.15	22.8 ± 0.18
ChargeSwitch® gDNA Plant Kit	17.0 – 107.4	7.7 – 49.9	1.88	2.00	1.15 – 7.48	22.51 ± 0.21
CTAB method	1083 – 1693	113 – 218	1.94	1.87	16.9 – 32.7	23.3 ± 0.10
Plant DNAzol® Reagent	77.4 – 212.2	15.6 – 35.6	2.00	0.59	2.34 – 5.34	23.3 ± 0.14
AquaGenomic™ Solution	165.6 – 241.6	55.6 – 69.2	1.93	1.92	8.34 – 10.38	23.2 ± 0.16

[a]Based on PicoGreen fluorometry measurements.
[b]Purity of DNA was evaluated by real-time PCR amplification of the *npt*II gene present on the vector pBIN19.
A_{260}/A_{280} <1.8 indicates protein and phenolic contamination;
A_{260}/A_{230} <2.0 indicates phenolic/humic and carbohydrate contamination.

In terms of the quality of DNA, comparable Ct-values were obtained for all extracts standardised to 10 ng/µl during real-time amplification of the *npt*II gene indicating that there was no difference in the PCR efficiency and hence amplifiability. Any DNA extraction method must suit its intended end purpose and in the case of potato leaf material, any of the tested methods would be suitable for PCR-based techniques even though some yields recovered were low **(Table 1).** However, although commercial kits offer speed and considerable convenience they do lack the scalability that in-house methods can provide and often a choice has to be made between expenditure, yield and purity. For the scaled-up extraction of potato leaf, only the CTAB method, AquaGenomic and Plant DNAzol reagents were easily amenable to scaling, whereas the ChargeSwitch kit would be more suitable for semi-automatation. Although the CTAB method extracted the highest DNA yields and was the most economical in terms of cost per sample, it was also time-consuming and labor intensive. In this study, the AquaGenomic solution was the preferred high-yielding method due to its speed and scalability.

Evaluation of DNA Walking Methods for the Isolation of Transgene Flanking Regions in Single Copy Lines of Potato.

A summary of the transgenic potato lines selected for obtaining genomic DNA flanking the transgene insertion site by various DNA walking methods is shown in **Table 2.** A novel pre-screen method based on real-time PCR for determination of the length of vector backbone was developed due to the common occurrence of backbone integrated in transgenic potato lines. A series of 19 assays were designed using the Universal Probe Library (Roche) to cover the pBin19 vector backbone from both the Right (RB) and Left borders (LB). In the single copy lines, Dxs-19 (*S. tuberosum*) had LB vector backbone up to ~2.7kb in length integrated, but no RB backbone was present, whereas Bkt1-2 (*S. phureja*) and CrtB-9 (*S. tuberosum*) appeared to have the total cloning vector integrated, and the orientation of this integration using an anchor primer was from the RB. This pre-screening for vector backbone was used to identify a starting point for a DNA Walking procedure. It can be hypothesised that the transformation events leading to integration in lines Bkt1-2 and CrtB-9 involved T-strand processing that started at the RB but failed to terminate at the LB, and continued round the vector and back to the T-DNA for a second time. The real-time qPCR assay was also developed for quantifying the copy number of transgenes based on detection of the *Npt*II gene of pBIN19. Future work is required to determine the copy number of vector backbone sequence integrated in transgenic lines. Transgene-specific primers were subsequently designed for each protocol to determine flanking regions from both the LB and RB in three lines of potato, Bkt1-2 (*S. phureja*), CrtB-9 and Dxs-19 (both *S. tuberosum*), and three commercial DNA Walking Kits (APAgene™ GOLD Genome Walking Kit [BIO S&T], DNA Walking SpeedUp™ Kit II [Seegene], Universal Vectorette™ System [Sigma]) were compared with the Adaptor-Mediated PCR technique in terms of success and reliability. PCR products generated by each method were sequenced either directly after clean-up, following gel excision and purification, or *via* cloning in *E. coli*. Details of all the procedures are outlined in the appropriate SOP.

Table 2. Summary of Transgenic Potato Lines Selected for Obtaining Genomic DNA Flanking the Transgene Insertion Site by Various DNA Walking Methods

Potato transgenic line[a]	Species	Cultivar	Construct(s) present	Copy number[b]
Single-transformation				
Bkt1-1	*Solanum phureja*	Mayan Gold	pBIN19	2
Bkt1-2	*Solanum phureja*	Mayan Gold	pBIN19	1
Dxs-19	*Solanum tuberosum*	Desiree	pBIN19	1
Dxs-36	*Solanum tuberosum*	Desiree	pBIN19	2
TPS4	*Solanum tuberosum*	Desiree	pBIN19	3
TPS6	*Solanum tuberosum*	Desiree	pBIN19	1
TPS8	*Solanum tuberosum*	Desiree	pBIN19	2
TPS9	*Solanum tuberosum*	Desiree	pBIN19	5
TPS42	*Solanum tuberosum*	Desiree	pBIN19	2
TPS51	*Solanum tuberosum*	Desiree	pBIN19	3
Co-transformed				
CT-3	*Solanum tuberosum*	Desiree	pBIN19	3
CT-88	*Solanum tuberosum*	Desiree	pBIN19	3
CT-184	*Solanum tuberosum*	Desiree	pBIN19	2
Re-transformed				
B9Zep-4	*Solanum tuberosum*	Desiree	pBIN19 pGPTV	2
B9Zep-13	*Solanum tuberosum*	Desiree	pBIN19 pGPTV	1

[a] Refer to the following publications for details of the transgenic potato lines:

Morris WL, Ducreux LJM, Hedden P, Millam S and **Taylor MA**. (2006) Over-expression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: implications for the control of the tuber life-cycle. J Exp Bot. 57, 3007-3018.

Morris WL, Ducreux LJM, Fraser PD, Millam S and **Taylor MA**. (2006) Engineering ketocarotenoid biosynthesis in potato tubers Metabolic Engineering 8, 253-263.

Ducreux LJM, Morris WL, Hedley PE, Shepherd T, Davies HV, Millam S, **Taylor MA** (2005). Metabolic engineering of high carotenoid potato tubers containing enhanced levels of β-carotene and lutein. J Exp Bot 56, 81-89

[b] Copy number determined by Southern analysis and real-time PCR.

In terms of success rate, both the APAgene™ and SpeedUp™ kits were superior and comparable in that 5 out of 6 flanking regions were determined **(Table 3).** In one example for the LB of Bkt1-2, four and two independent bands tested produced the same flanking regions using the different APAgene™ and SpeedUp™ kit primers, respectively **(Figures 1 and 2),** and indicated a truncation to all of the T-DNA left border and *lac*I sequences, and part of the M13 *ori* sequence of pBIN19. The same flanking regions determined for all three potato lines were in agreement using both commercial kits. All junctions were subsequently confirmed following PCR analysis and sequencing. Both the Adaptor-Mediated PCR and Vectorette System are similar methods in principle but were less successful and more time consuming in comparison due to the need for DNA digestion with restriction enzymes (eg., *Bam*HI or *Hind*III) and ligation steps in addition to two rounds of PCR **(Table 3).** Only one PCR product was generated using Adaptor-Mediated PCR for the LB of line Bkt1-2, but the junction was determined by sequencing and this was in agreement with the APAgene™ and SpeedUp™ kits **(Table 3).** Although many products were amplified with the Vectorette System, vector backbone sequence was only reached for one product of the LB for line CrtB-9, while the remaining bands isolated were identified as artefacts generated by the kit primer. It is clear that both of these methods would require further optimization for determining the flanking regions in transgenic potato lines. The main disadvantages of such methods are: the presence of suitable restriction enzyme sites and successful ligation of the specific adaptors to restriction fragments obtained from genomic DNA, and the recommendation of an initial concentration of 500 to 1000 ng genomic DNA to perform the digestion step. This is in contrast to only ~100 ng DNA required for the first PCR step with the APAgene™ and SpeedUp™ kits. In order become useful for large-scale screening of transgenic plants, these procedures would have to be optimized and streamlined to allow fast and cheap analysis of many samples.

Table 3. Results Summary of DNA Walking Methods for Single Copy Potato Transgenic Lines

	Junctions Detected in Potato Transgenic Lines (single copy):					
	Bkt1-2 (*Solanum phureja*)		CrtB-9 (*Solanum tuberosum*)		Dxs-19 (*Solanum tuberosum*)	
Methods	LB	RB	LB	RB	LB	RB
APAgene™ GOLD Genome Walking Kit (BIO S&T)	+ (4)[a]	1 (1)	+ (1)	+ (1)	+ (2)	V (2)
DNA Walking SpeedUp™ Kit II (Seegene)	+ (2)	V (1)	+ (1)	+ (1)	+ (2)	+ (2)
Universal Vectorette™ System (Sigma)	-	-	V (1)	-	-	-
Adaptor-Mediated PCR	+ (1)	-	-	-	-	-

[a] Numbers in parenthesis indicate the number of different size products in which sequence information was obtained using the method.
KEY: LB = Left Border; RB = Right Border.
+ = Junction identified; - = No products or PCR artefacts; V = Vector backbone detected, further walking required.

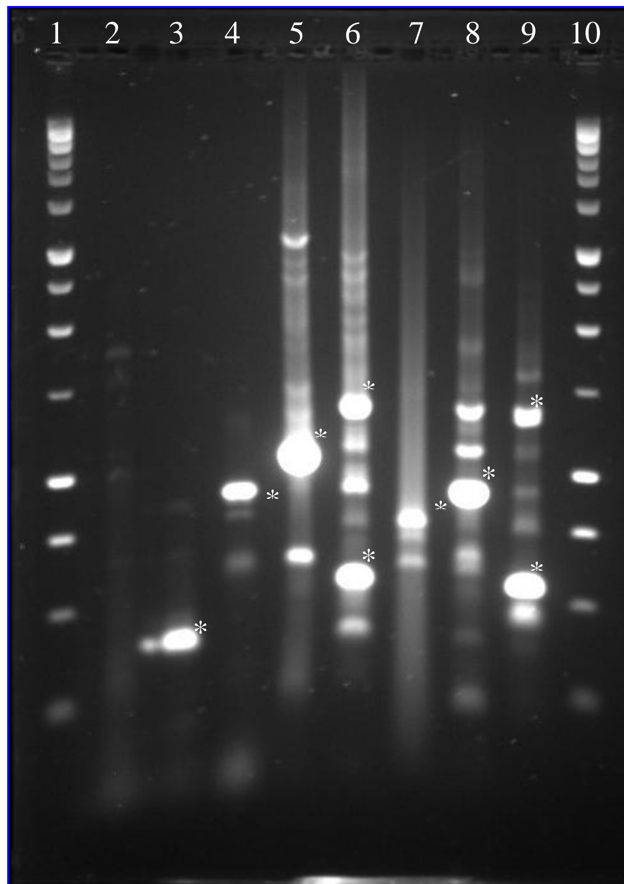


Figure 1. Amplification of T-DNA Left Border flanking regions from a single copy line of transgenic potato (Bkt1-2) using the DNA Walking SpeedUp™ Kit II (Seegene) and APAgene™ GOLD Genome Walking Kit (BIO S&T); 5 µl of PCR products loaded per lane. Lane 1, 1 kb DNA ladder, 250 – 10,000 bp; Final round nested PCR products: lanes 2-5, ACP1 to ACP4 SpeedUp™ primers, respectively; lanes 6-9, DRTA to DRTD APAgene™ primers, respectively; lane 10, 1 kb DNA ladder.
*All bands excised and purified using Wizard® Gel Clean-Up System (Promega) for direct sequencing.

```
4-61.2ANTI      -AGCGGGCAATCAGCTTCCACCGCCGCGGTGAAATCCG--CGACCGTTCCGCCGAAGGCC 633
4-65.1ANTI      -AGCGGGCAATCAGCTTCCACCGCCGCGGTGAAATCCG--CGACCGTTCCGCCGAAGGCC 483
4-66.1ANTI      -AGCGGGCAATCAGCTTCCACCGCCGCGGTGAAATCCG--CGACCGTTCCGCCGAAGGCC 123
4-62.1ANTI      -AGCGGGCAATCAGCTTCCACCGCCGCGGTGAAATCCG--CGACCGTTCCGCCGAAGGCC 127
pBIN19          TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACTTAATAACACATTGCG 6172
                  ** **  **    *        *    ***        * **       *  *  **

4-61.2ANTI      AGCATTACCGCATCGCCGGCATGAATCTGCTCGCCGCCAT---------CATCATCTTCT 684
4-65.1ANTI      AGCATTACCGCATCGCCGGCATGAATCTGCTCGCCGCCAT---------CATCATCTTCT 534
4-66.1ANTI      AGCATTACCGCATCGCCGGCATGAATCTGCTCGCCGCCAT---------CATCATCTTCT 174
4-62.1ANTI      AGCATTACCGCATCGCCGGCATGAATCTGCTCGCCGCCAT---------CATCATCTTCT 178
pBIN19          GACGTTTTTAATGTACTGGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGAT 6232
                  * **         * **  **  * * **   * ***          ** ** **  *

4-61.2ANTI      GGAAC---ACCATGAAGCTCGGCGAG-GTCGTTGCAAACCAGAAACGCGATGGAAAGCTG 740
4-65.1ANTI      GGAAC---ACCATGAAGCTCGGCGAG-GTCGTTGCAAACCAGAAACGCGATGGAAAGCTG 590
4-66.1ANTI      GGAAC---ACCATGAAGCTCGGCGAG-GTCGTTGCAAACCAGAAACGCGATGGAAAGCTG 230
4-62.1ANTI      GGAAC---ACCATGAAGCTCGGCGAG-GTCGTTGCAAACCAGAAACGCGATGGAAAGCTG 234
pBIN19          TGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGC--GGTCCACGCTGGTTTGCCC 6290
                 *  *   ***     ** * * *** ** *  **** *  *   * ** ***   **

4-61.2ANTI      CTATCGCCCGATCTCTTGGCCCATGTTTCGCCGCTCGGATGGGAACACATCAATCTCACC 800
4-65.1ANTI      CTATCGCCCGATCTCTTGGCCCATGTTTCGCCGCTCGGATGGGAACACATCAATCTCACC 650
4-66.1ANTI      CTATCGCCCGATCTCTTGGCCCATGTTTCGCCGCTCGGATGGGAACACATCAATCTCACC 290
4-62.1ANTI      CTATCGCCCGATCTCTTGGCCCATGTTTCGCCGCTCGGATGGGAACACATCAATCTCACC 294
pBIN19          CAGCAGGCGAAAATCCTGTTTGATGGT--GGTTCCGAAATCGGCAAAATCCCTTATAAAT 6348
                *    * *  *  ** **    *** *  *   *    ** ** * *   *  * * *

4-61.2ANTI      GGAGAATATCGCTGGCCAAAGCCTTAGCGTAGGATTCCGCCCCCTCCCGCAAACGACCCC 860
4-65.1ANTI      GGAGAATATCGCTGGCCAAAGCCTTAGCGTAGGATTCCGCCCCCTCCCGCAAACGACCCC 710
4-66.1ANTI      GGAGAATATCGCTGGCCAAAGCCTTAGCGTAGGATTCCGCCCCCTCCCGCAAACGACCCC 350
4-62.1ANTI      GGAGAATATCGCTGGCCAAAGCCTTAGCGTAGGATTCCGCCCCCTCCCGCAAACGACCCC 354
pBIN19          CAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTAT 6408
                  * *  ** **  *  * **  **       * * * *       *   *  * **

4-61.2ANTI      --AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATTCAGGGCGATGGCCCA 918
4-65.1ANTI      --AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACC-TCTAT-CAGGGCGATGGCCCA 766
4-66.1ANTI      --AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT-CAGGGCGATGGCCCA 407
4-62.1ANTI      --AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT-CAGGGCGATGGCCCA 411
pBIN19          TAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT-CAGGGCGATGGCCCA 6467
                  ************************************ ***** **************
```

Figure 2. Nucleotide sequence of four independent products generated from the Left Border flanking region of line Bkt1-2 using the APAgene™ GOLD Genome Walking Kit (BIO S&T). The arrow indicates the location of the junction in Bkt1-2 after alignment of sequences with vector pBIN19. Letters in bold indicate position of the Left Border of pBIN19.

A BLASTn program search of the nonredundant nucleotide (nr/nt) databases at the NCBI revealed that the most probable related genomic regions flanking the transgene insertion sites of *S. tuberosum* lines CrtB-9 and Dxs-19 were of *Solanum* origin **(Table 4).** However, in the case of *S. phureja* line Bkt1-2, the most significant alignments revealed that the transgene had inserted between sequences related to bacterial antibiotic resistance plasmid/transposon-like sequences; PCR verification confirmed all these genomic sequences as non-host genomic DNA nor vector-derived material. Interestingly, the same bacterial-like sequences were also identified by the BLASTn search at one of the left and right borders of the two copy line, *S. phureja* Bkt1-1 **(Table 5).** The origin of these bacterial-like sequences is unknown and we can only speculate without further DNA walking and sequencing from these flanking regions. A possible explanation for the presence of prokaryotic sequences in transgenic potato lines could be due to the incidental transposition of mobile genetic elements accompanying cloning steps (Kovřík et al., 2001). It was also reported by

the same authors that many published eukaryotic sequences contained the flanking inverted repeats of transposon Tn*10*, IS*10* and other prokaryotic IS elements.

Table 4. Most Significant Alignments for the Transgene Flanking Regions of Single Copy Potato Transgenic Lines Based on BLASTn searches of the NCBI databases

Line[a]		BLASTn Results of NCBI Nucleotide Databases		
		Accession number	*E* value	Alignment
Bkt1-2	LB	AF313472.2	0.0	*Pseudomonas aeruginosa* plasmid RPL11 transposon *Tn*1403.
	RB	AM992204.1	4e-37	*Psychrobacter psychrophilus* transposon *Tn*5080.
CrtB-9	LB	AC151957.1	6e-89	*Solanum tuberosum* chromosome 5 clone PGEC13.
	RB	EU124732.1	6.3	*Solanum lycopersicum* chromosome 3 clone C03HBa0030O03.
Dxs-19	LB	AC232052.1	5e-71	*Solanum tuberosum* chromosome 5 clone RH072I03.
	RB	AK325770.1	3e-125	*Solanum lycopersicum* cDNA, clone LEFL2025I06.

[a] Data for the most significant alignment was selected from those sequences generated from the largest PCR products recovered by DNA Walking using the BIO S&T and Seegene kits.
KEY: LB = Left Border; RB = Right Border.

Objective 01.2 Compare the optimized adaptor-mediated PCR method for junction sequence analysis with the kits – develop a standard operating procedure for barley using transgenic lines with low copy number.

Evaluation of DNA extraction methods for barley leaf material

In order to compare the three commercial DNA Walking Kits (SpeedUp™ Kit II [Seegene], APAgene™ GOLD [BIO S&T], Universal Vectorette™ [Sigma]) with the Adaptor-Mediated PCR in barley, it was necessary to firstly evaluate different DNA extraction methods and develop a standard operating procedure (SOP) for plant genomic DNA extraction appropriate for use with the DNA walking kits (Appendix 8.2).

Plant genomic DNA extraction methods were evaluated for DNA yield, quality, ease of use and scalability. Four standard lab protocols and five commercially available kits were assessed using leaf material from two barley cultivars 'Golden Promise' and 'Optic'. Leaf material was harvested, flash frozen with liquid nitrogen, ground to a fine powder in bulk and stored at -80 °C in 75mg aliquots prior to use. The majority of DNA extraction protocols stated a maximum of 100mg of ground leaf material. Of the five commercially available kits tested, Qiagen, ENZA and MoBi were silicon column based, the Charge Switch system used magnetic beads to bind DNA, and the Aqua Genomic was a liquid based extraction system. The standard lab protocols for DNA extraction all included a single phenol cleanup step. Three protocols, Edwards, Ellis and buffer "S" used similar alkaline extraction buffer. Buffer "S" and CTAB protocols required 1 hour incubation at 65 °C, the Edwards and Ellis protocols were performed at room temperature. The Ellis protocol differed from the Edwards, in that SDS was not included in the extraction buffer; it was added separately to the extraction. With all methods tested DNA was eluted with analytical grade water, except with the Charge Switch kit as elution with water was not recommended. The Qiagen DNeasy kit method was repeated using two different elution volumes. DNA yield and quality was assessed by spectrophotometry and gel electrophoresis.

All methods tested yielded DNA of reasonable quality, but the kits all produced lower yields than the standard lab protocols. Some of the DNA extraction kits did not yield sufficient DNA for use with transgene flanking region isolation methods which require an initial DNA digestion, such as the Adaptor-Mediated PCR technique or Vectorette kit. The Charge Switch method yielded good quality DNA, however, it scored badly for ease of use by the operators. All of the standard lab protocols were scalable, however, of the kits tested, only the liquid based Aqua Genomic was scalable.

A spectrophotometer was used to determine the concentration of double-stranded DNA (dsDNA) in the sample solutions. DNA absorbs ultraviolet light, with an absorption peak at 260 nm wavelength, Beer-Lambert law states that "the amount of light absorbed can be related to the concentration of the absorbing molecule". At a wavelength of 260 nm, the extinction coefficient for dsDNA is 0.020, and single-

stranded nucleotides, such as RNA, it is 0.027 $(\mu g/ml)^{-1}$ cm^{-1}. Therefore, an optical density (OD) reading of 1, at a wavelength of 260 nm, corresponds to a dsDNA concentration of 50 µg/ml. A summary of DNA yields obtained from the different extraction methods is presented in **Table 5**.

Table 5. Summary of DNA extraction yields.

	DNA yields µg/ml							
Method or kit used	**GP1**	**GP2**	**Optic 1**	**Optic 2**	**Mean**	**SD**	**Total yield mean µg**	**Volume µl**
Edwards et al.	919	1089	1073	1030	1027.75	76.66	103	100
Ellis 1994	1120	1346	1860	1434	1440	309.66	144	100
Buffer "S"	3170	1770	2980	2110	2507.5	674.31	250	100
CTAB	1312	1413	936	960	1155.25	243.04	116	100
Qiagen DNeasy	26	22	19	25	23	3.16	4.6	200
Qiagen DNeasy (2)	28	23	31	92	43.5	32.51	4.35	100
ENZA	385	415	442	428	417.5	24.31	83.5	200
MoBio	270		276		273	4.25	27.3	100
Charge Switch	50	40	33	21	36	12.2	3.6	100
Aqua Genomic	131			230	180.5	70.04	18.05	100

The purity of dsDNA preparations was estimated by determination of the ratio of absorbance at 260/280 nm and at 260/230 nm. Pure dsDNA has a 260/280 absorbance ratio of 1.8, pure RNA of approximately 2 and protein <1. Contamination by phenol can contribute to the overestimation of DNA concentration. Phenol absorbs with a peak at 270nm and has a 260/280 ratio of 2 (**Table 6**).

Table 6. Ratio of absorbance 260/280

DNA Quality 260/280						
	GP 1	**GP 2**	**Optic 1**	**Optic 2**	**Mean**	**SD**
Edwards et al	1.82	1.84	1.86	1.81	1.8325	0.0222
Ellis 1994	1.9	1.84	1.74	1.74	1.805	0.0789
Buffer "S"	1.45	1.57	1.49	1.46	1.4925	0.0544
CTAB	1.83	1.76	1.86	1.85	1.83	0.0356
Qiagen DNeasy	1.52	1.47	1.25	1.55	1.4475	0.1357
Qiagen DNeasy (2)	1.62	1.49	1.66	1.64	1.6025	0.0768

ENZA	1.85	1.83	1.87	1.83	1.845	0.0191
MoBio	1.93		1.96		1.945	0.0212
Charge Switch	1.37	1.61	1.67	1.67	1.58	0.1428
Aqua Genomic	1.78		1.61		1.695	0.1202

Absorption at 230 nm can be caused by contaminates such as thiocyanates, which are often present in extraction buffers, and other organic compounds. This characteristic can be used to determine the level of contamination by examining absorbance ratios at 260/230 nm. A high 260/230 ratio (>2) is indicative of pure DNA, while a low ratio would suggest contamination.

Table 7. Ratio of absorbance at 260/230

DNA Quality 260/230						
	GP 1	**GP 2**	**Optic 1**	**Optic 2**	**Mean**	**SD**
Edwards et al	2.2	2.18	2.26	2.05	2.1725	0.0885
Ellis 1994	2.02	1.94	2.15	1.98	2.0225	0.0911
Buffer "S"	1.75	1.9	1.95	1.9	1.875	0.0866
CTAB	2.05	2.22	2.38	2.42	2.2675	0.1688
Qiagen DNeasy	1.85	1.4	1.24	2.04	1.6325	0.3748
Qiagen DNeasy (2)	1.42	1.29	1.81	1.87	1.5975	0.2861
ENZA	2.38	2.44	2.34	2.42	2.395	0.0443
MoBio	2.09		2.05		2.07	0.0283
Charge Switch	1.23	1.26	1.58	2.54	1.6525	0.6125
Aqua Genomic	1.91		1.58		1.745	0.2333

For their ability to deliver high yield, good quality DNA with a quick and simple protocol that could be scaled, two methods stood out, the Aqua Genomic liquid based kit and the standard lab protocol based on Ellis 1994. The standard lab DNA isolation protocol based on Ellis was adopted for further work at JIC and a SOP prepared (Appendix 8.2).

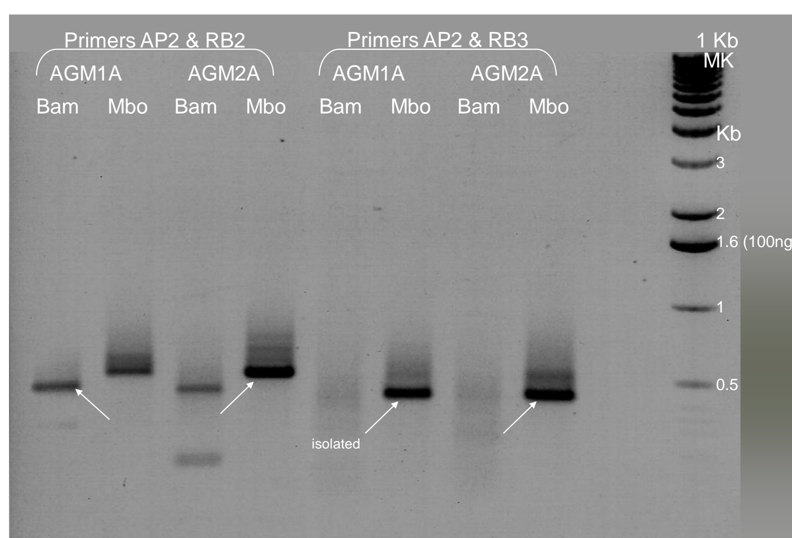
Evaluation of DNA Walking Methods for the Isolation of Transgene Flanking Regions in Single Copy Lines of barley.

Three commercial DNA Walking Kits (SpeedUp™ Kit II [Seegene], APAgene™ GOLD [BIO S&T], Universal Vectorette™ [Sigma]) were compared with the Adaptor-Mediated PCR technique previously used successfully in barley. Transgene specific primers were designed according to the kit specifications and each kit was tested on the same two transgenic barley lines AGM1 and AGM2, the results were compared to those obtained with the adaptor-mediated PCR method.

Adaptor-Mediated PCR, previously optimised in FSA project GO2, requires the genomic DNA containing the target sequence to be digested with restriction enzymes. This is followed by the ligation of an adaptor to the end of the target sequence. Nested PCR is then performed using T-DNA specific primers and primers complementary to the adaptor, thus amplifying the DNA sequence between the T-DNA and the adaptor. The PCR products are then separated by gel electrophoresis, isolated, purified and either sequenced directly or cloned into a vector and sequenced.

Adaptor-Mediated PCR was performed on barley lines AGM1 and 2 containing the T-DNA from the transformation vector pAL135. Two combinations of right border (RB) specific primers and two restriction enzymes were used for the analysis. The amplified DNA was separated on an agarose gel, isolated, purified and sequenced (**Figure 3**). Right border transgene flanking sequences were successfully obtained for both lines.

Figure 3. Adaptor-mediated PCR of Barley lines AGM1 and 2.



The Universal Vectorette™ system (Sigma) uses three steps to obtain the DNA junction sequence. It is similar to Adaptor-Mediated PCR, in that genomic DNA containing the target sequence is digested with a restriction enzyme and then goes through a ligation reaction. The ligation attaches a Vectorette unit to each end of the target sequence. Nested PCR is then performed using T-DNA specific primers and primers complementary to the Vectorette unit. The PCR amplifies the DNA sequence between the known T-DNA sequence and the Vectorette unit.

The Vectorette system was found not to be suitable for barley as one of the primers supplied with the kit was able to amplify products from the barley genome when tested without the presence of a transgene specific primer (**Figure 4**). Thus, this method would give false positives when used for obtaining junction sequence in barley (**Figure 5**)

Figure 4. First Vectorette PCR showing primer Vec 1 mispriming in barley controls.

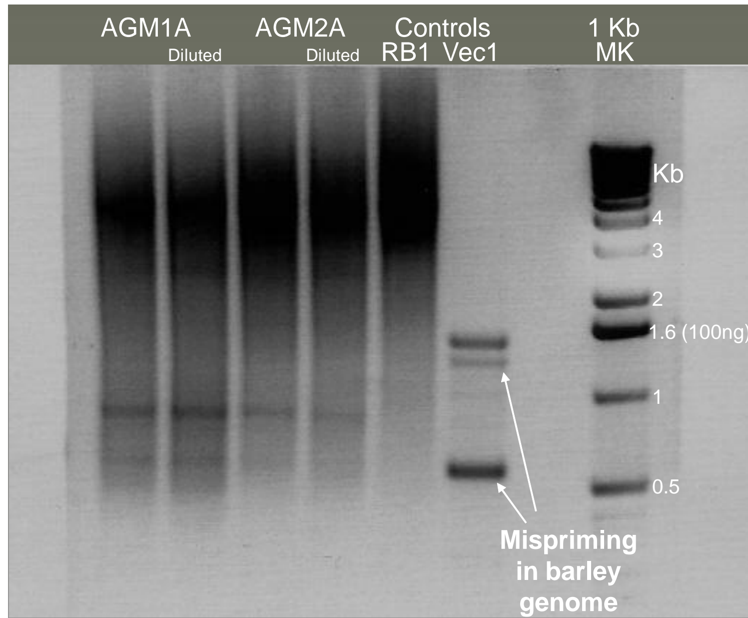
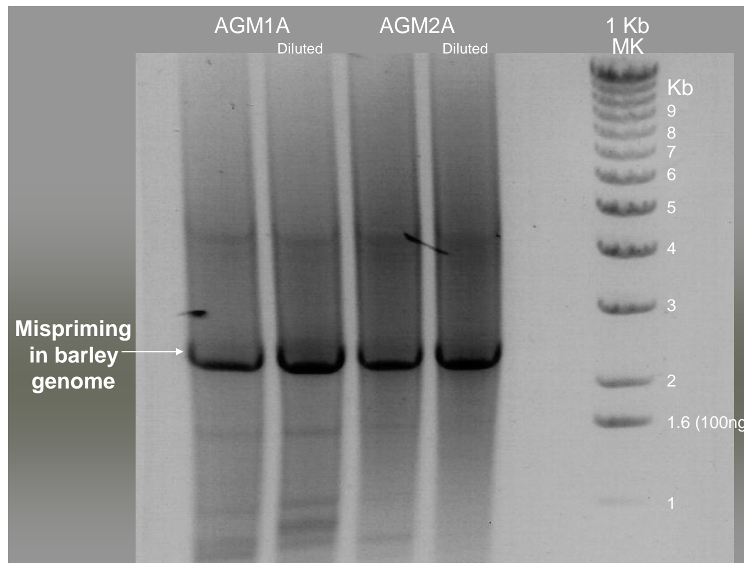


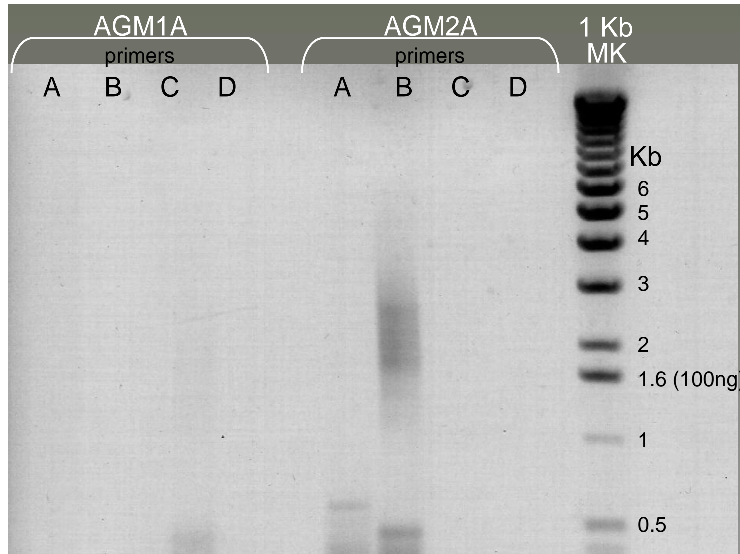
Figure 5. Final Vectorette PCR showing amplification of non-target sequence in barley



The APAgene™ GOLD (BIO S&T) is similar to Tail-PCR, in that it uses degenerate primers to randomly bind to the unknown sequence. The first PCR reaction is performed in two parts, the first using only T-DNA specific primers to amplify the target template, in the second part, four degenerate primers are added which randomly bind to the target DNA and amplify back to the T-DNA. A digestion step is preformed to digest excess primers remaining in the first PCR and this avoids these being carried over to the final nested-PCR.

The APAgene kit gave poor results for barley, in the initial tests, some amplification was achieved but only after increasing the concentration of the primers supplied and no clear bands were obtained (**Figure 6**.)

Figure 6. The non-specific amplification from the APAgene kit when used in barley.



The Seegene SpeedUp™ Kit II (Seegene) uses four Annealing Control Primers (ACP) to bind randomly to the unknown DNA sequence in conjunction with T-DNA target specific primers (TSP). In this system, three nested PCR's are performed and a cleanup step included after the initial PCR.

The Seegene kit gave very good results in barley yielding clear bands that could be sequenced directly after isolation from the gel (**Figure 7**). It was found that weaker bands could be re-amplified to increase product concentration by repeating the final PCR (**Figure 8**), however, at a later date, this was found to exacerbate the production of PCR artefacts which were generated by the Seegene Universal primer (UniP) (see section on PCR artefacts and false positives below).

Figure 7. An example of the Seegene method performed on single copy lines of barley.

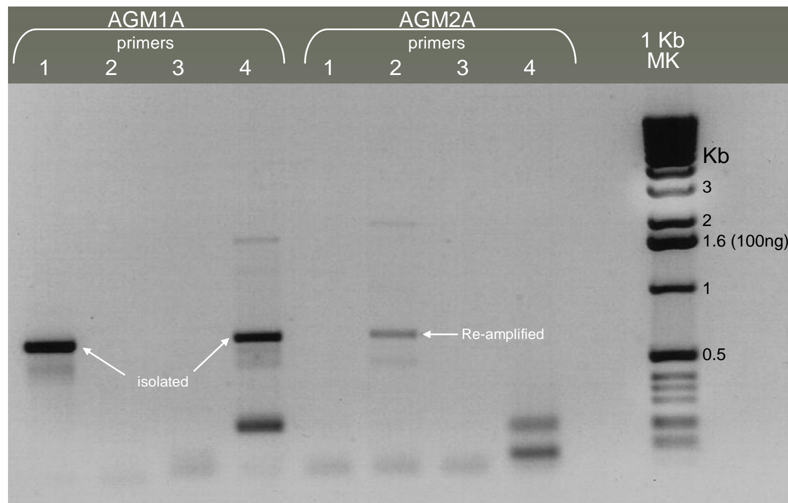
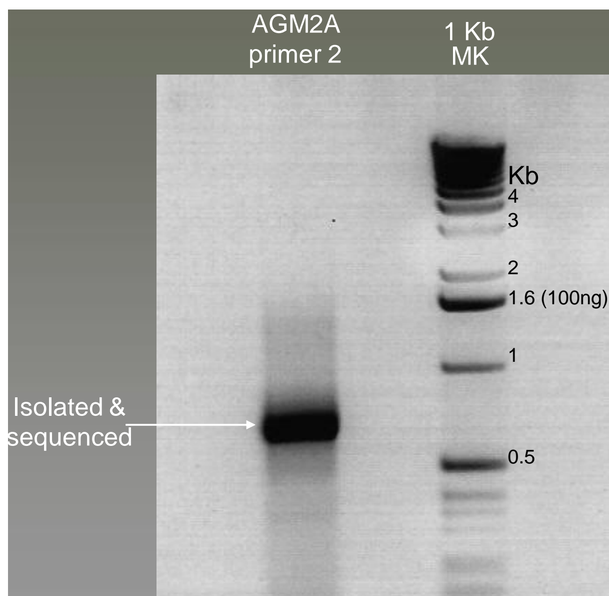
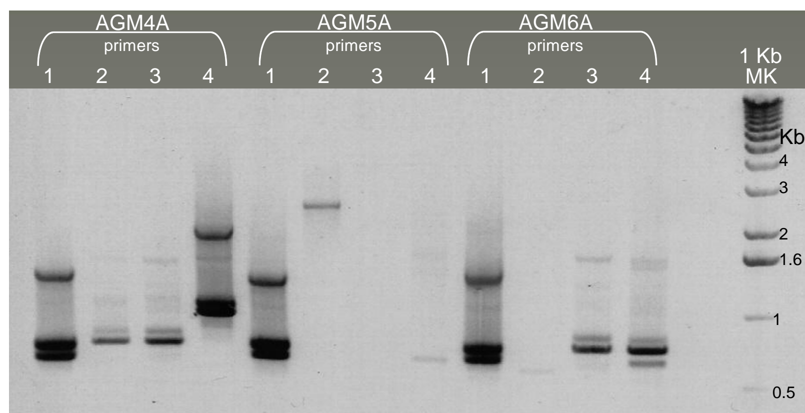


Figure 8. Re-amplification of the Seegene PCR product from line AGM2A



The Seegene kit was used on more complex transgenic lines, known to contain multiple transgene copies. The Seegene kit can be used to isolate two independent transgene junctions in one reaction; this is usually not possible with the adaptor-mediated PCR method (**Figure 9**).

Figure 9. Seegene kit used on multiple transgene copy barley lines.



The Seegene kit was considered the best method for junction sequence analysis in barley and was adopted as the method of choice. The original adaptor-mediated PCR technique also continues to work well for barley but is more time consuming than the Seegene kit. Further analysis was performed on a number of transgenic barley lines to optimise the process. This further analysis revealed some unexpected rearrangement at the T-DNA border region and has also led us to incorporate some additional checks into the recommended barley SOP to ensure that false positives are not a problem.

Twenty five independent transgenic barley lines, transformed with five different constructs, were analysed. Six out of the twenty five transgenic lines, which equates to 24 %, had some kind of rearrangement at the T-DNA border region, these rearrangements ranged from simple T-DNA truncations or read though, where part or all of the vector backbone was present, to more complex insertions. A subset consisting of twenty one of these lines were analysed in more detail. One line 84-14 was found to have read-through at the RB and contained the whole plasmid backbone. This was confirmed by a PCR-based walking strategy from the RB through the incorporated backbone sequence to the LB region. Of the remaining twenty lines, junction sequence validation was performed by PCR amplification and sequencing through the junction site as in the SOP section 7.5 (Appendix 8.5). Barley cultivar 'Golden Promise' DNA was run along side each line as a negative control. The sequences obtain from the validation PCR were aligned to the junction sequences obtained by the Seegene method. Primers were designed to anneal to the putative barley sequence and PCR amplification performed to confirm that the sequences originated from the barley genome. All but one of the twenty junction sequences could be validated. The validation process was repeated twice for this negative line, the second time using an alternative set of primers. This one unsubstantiated junction sequence, in twenty, equates to an accuracy of 95 % with the Seegene methodology.

One unexpected and commonly found rearrangement was the incorporation of additional sequence, in reverse complement to the T-DNA, close to and outside of the right border (RB) repeat. These additional sequences ranged in size from small insertions of 44 base pairs (bp) as in line AGM4A (**Figure 10**) to large insertions of up to 1209 bp as found in line 55-07-01. These inserts can originate from anywhere within the plasmid DNA and appear to be independent insertions not originating from

the inserted T-DNA. Examples of these complex types of rearrangements are given in **Figures 10 and 12**.

Figure 10. T-DNA RB junction of line AGM4A showing an insert from the luciferase gene.

pAL135 T-DNA

GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTT

GCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGC

GGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA

RB Sequence Luciferase (blastn 100%) reverse complement

ACGCAGGAAAGAACATGAAGGCCTTGACACTCTGATTGACAAATACGATTTATCTAATTT

ACACGAAATTGCT AAAATGAAATATGGACTAAAATGAGAACTACACAGTTGCCCTGTTTG

Barley Genomic DNA

AGAACAGGCCAGGGATCAAAATTCCTGCACTTTTCTCAAAAGAGCAGGGACTAGAAAGA

AAATTAAGCAAGGGCACTTCGCTCTGTTTCAGAACATGCCAGGGATCCAAATTTAATTTAA

AGGGATTGGCAGGGGCCGCGCGAACGCGTACCCCCTCTATCACAAATTATGACGTCCACT

CTCCCACTTGGGGAGATGGTAGGCCAGCGCACCCACCAAGTGCAATGTGGGCCACTGATT

TAGGCCACGGGCCTTCTTGATCGCCCGTCGACCCCGTCCAGGTAAGTATGTAAAAACGGA

GCACCACGCCCTCCTCATATAGCCAGCTCGGCCGCCCGACGCCGCTCGCTAGGCGAGGTG

GGACTAAAAC

Line AGM4A revealed a 44 bp insertion of part of the luciferase gene, inserted in reverse complementation to the T-DNA. The insert corresponds to pAL135 plasmid map coordinates 7061 to 7104 bp, some 1567 bp away from the RB (**Figure 11**). BLASTn analysis on the barley DNA gave top 10 BLASTn hits to rice genomic DNA (Oryza sativa japonica cultivar-group) ranging from 89 to 97% homologies.

Figure 11. Plasmid Map of transformation vector pAL135

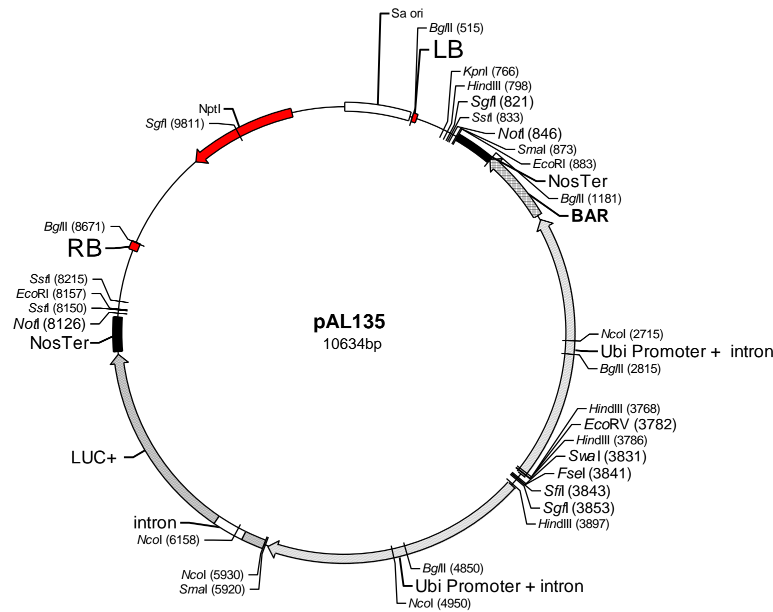


Figure 12. Line 55-14-01T-DNA RB junction showing an insert from close to the left border then a second insertion from the Hygromycin B resistance gene *Aph IV*.

pBRACT204 T-DNA

AAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGA

ATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTT

TATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC

AAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTCATCGAT

GATATCAGATCCGGGCCCTCTAGATGCGGCCATCTAGACCCAGCTTTCTTGTACAAAGTGG

RB Sequence Maps 569 to 600 bp on plasmid

TTGATCCTTGAACAAGTTTGTACAAAAAAGCAGGCTCCGCGGCCATCGGTCCAGACGGCC

Reverse complement Hyg fragment maps 2219 bp to 2005 bp on plasmid map

GCGCTTCTGCGGGCGATTTGTGTACGCCCGACAGTCCCGGCTCCGGATCGGACGATTGCGT

CGCATCGACCCTGCGCCCAAGCTGCATCATCGAAATTGCCGTCAACCAAGCTCTGATAGA

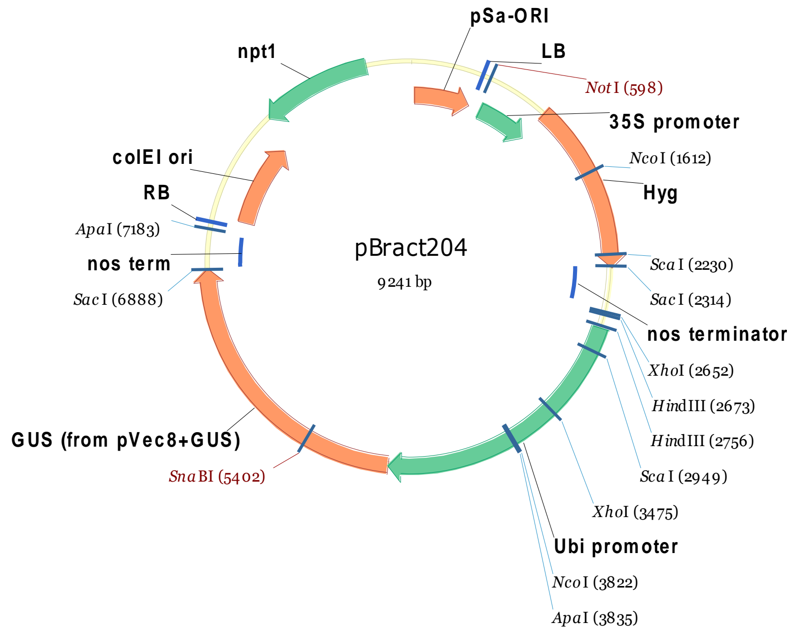
GTTGGTCAAGACCAATGCGGAGCATATACGCCCGGAGCCGCGGCGATCCTGCAAGCTCCG

GATGCCTCCGCTC

Line 55-14-01 Right border analysis revealed a rearrangement consisting of two additional insertions, one inserted in the forward direction which originates from 5 bp before the Left border (LB) repeat, 569 to 600 bp on the pBRACT 204 plasmid map

(**Figure 13**), and the second insert 214 bp of the Hygromycin B resistance gene *Aph IV* in reverse complementation.

Figure 13. Plasmid map of barley transformation vector pBRACT 204.



The material analysed in this study may be broadly divided into two groups, firstly GM lines containing 'old' constructs that were developed a number of years ago, for example those from the pAL series based on the original pGreen vectors. The second group comprises lines containing 'new' constructs, for example the pBract series based on the newer improved pGreen vectors. When we only consider GM barley lines containing pBract constructs with single copies of the transgene inserted and stable transgene expression then rearrangements of the type described above were found in 3 out of 21 (14.3%) of lines.

PCR artefacts and false positives.

It became apparent that PCR artefacts were being amplified in barley when using the standard Seegene kit protocol. Also, it became apparent that re-amplifying the products produced by the final Seegene PCR exacerbated the production of artefacts. These PCR artefacts consisted of barley genomic DNA sequence with Seegene Universal primer sequence at either end (**Figure 14**).

Figure 14. An example of a PCR artefact isolated from barley.

GAGTTTAGGTCCAGCGTGTGGAGTTTAGGTCCAGCGTGTGGGGGGGAACGGT
GTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTA
TACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTAT
CGCGCGCGGTGTCATCTATGTTACTAGGATCGAAGGGTGGGCGCGCCGACTG
GGCGCGCCCCAGCTTTCTTGTACAAAGTGGTTGATCCAGTCGCTGTATGTGTT
TGTTTGAGATCTCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAA
AAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATC
ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATCCGT
CCCCCCCACACGCTGGACCTAAACTC

A bilateral approach was taken to resolve the problem of PCR artefacts causing false positives. Firstly, a strategy was implemented to screen gel extracts or colonies by standard PCR for the region just inside the T-DNA borders. The ability to distinguish between junction sequences and artefacts early, significantly reduced the time spent needlessly pursuing false positives. A reverse primer was designed just inside the border region, to be used in conjunction with the target specific primer 3 for screening. More details can be found in the SOP section 7.3 (Appendix 8.5).

Secondly, artefacts were found to occur when there was not a target specific bias within reactions. Therefore, this bias was created by doubling the concentration of Target Specific Primers and also performing a pre-amplification step before employing the first Seegene reaction. The pre-amplification step consists of an asymmetric PCR, using just a target specific primer to increase the target site prior to the first Seegene reaction. This pre-amplification step has proved particularly useful when analysing lines with multiple transgene copies (Appendix 8.5 SOP section 7.1)

The frequency and type of rearrangement found during these experiments highlights the importance of a full analysis of transgene flanking regions in transgenic crops. The methodology developed during this project and presented here for the analysis of transgene junction sequences in barley is now routine, reproducible, with no false positives.

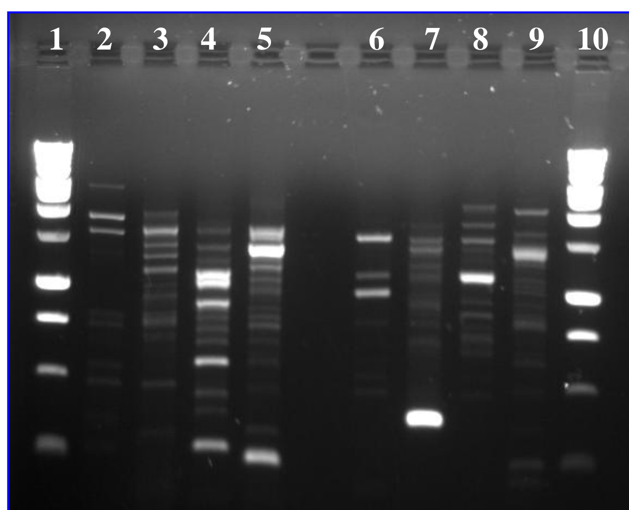
4.2 Objective 02

Apply the best methods from objective 1 to transgenic potato lines with multiple insertions. Optimise protocols to cope with multiple insertion lines or stacked transgenes including lines derived by co-transformation and re-transformation. Develop a standard operating procedure for multiple copy transgenics. If appropriate, investigate whether it is possible to multiplex the genome walking kit for the simultaneous determination of several transgene junction sequences.

02.1 Optimize the protocol for transgenic lines with multiple insertions of the same transgene and for stacked transgenes derived by co-transformation and re-transformation.

The next stage of this study was to use the APAgene™ and SpeedUp™ kits in locating the flanking regions in more challenging transgenic lines containing multiple T-DNA inserts, including co- and re-transformed lines. Indeed, it is more likely that those transgenic lines commercialized in the future will contain stacked transgenes. In an initial test with potato lines containing two transgene copies (e.g., Bkt1-1, Dxs-36, TPS6, TPS 8, and TPS 16), the APAgene™ kit was used more successfully in determining junctions as the SpeedUp™ kit suffered from the problem of producing artefact bands in a similar way to that described for the barley analysis above. This is highlighted in one example for determining the junctions from the LB of lines TPS6 and TPS8 **(Figure 15A & B).**

A. DNA Walking SpeedUp™ Kit



B. APAgene™ GOLD Kit

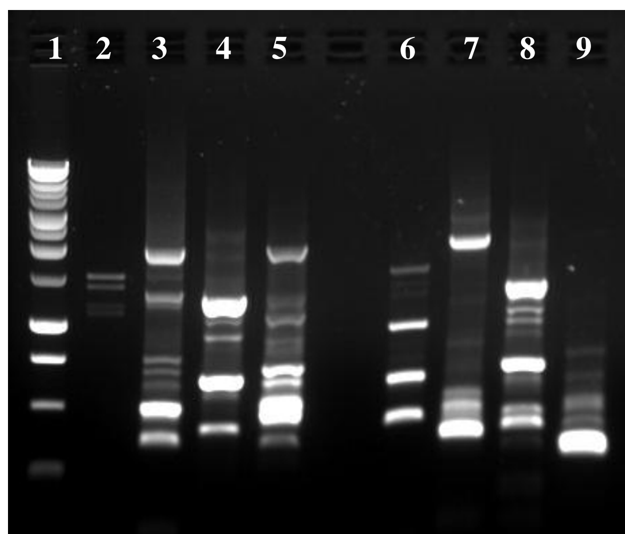
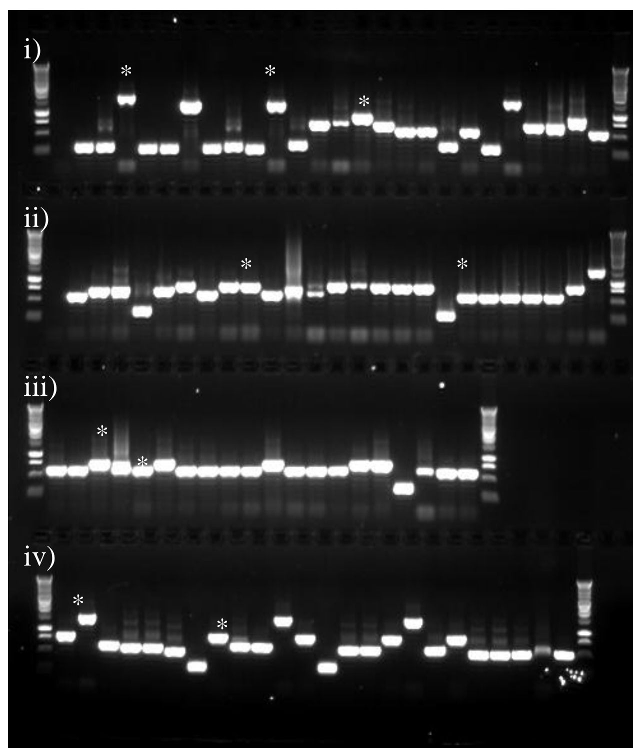


Figure 15. Final-round PCR amplification products generated from the Left Border of two copy transgenic lines of potato (TPS6 and TPS8) using the **(A)** DNA Walking SpeedUp™ Kit (Seegene) and **(B)** APAgene™ GOLD kits (BIO S&T); 5 µl of PCR products loaded per lane. **(A)** Lanes 1 and 10, 1 kb DNA ladder, 250 – 10,000 bp; Final round nested PCR products: lanes 2-5, ACP1 to ACP4 SpeedUp™ primers with TPS6, respectively; lanes 6-9, ACP1 to ACP4 SpeedUp™ primers with TPS8, respectively. **(B)** Lane 1, 1 kb DNA ladder, 250 – 10,000 bp; lanes 2-5, DRTA to DRTD APAgene™ primers with TPS6, respectively; lanes 6-9, DRTA to DRTD APAgene™ primers with TPS8, respectively.

Although many PCR products were amplified using both kits, the use of a PCR confirmation test to exclude artefacts and subsequent sequencing demonstrated that all bands isolated by gel extraction or cloning with the SpeedUp™ kit were infact artefacts generated by the UniP2 primer. In contrast, all bands isolated and sequenced from these five transgenic lines with the APAgene™ kit resulted in determining the transgene flanking regions, and this was confirmed by several independent bands in each case **(Table 8).** The successful application of the SOP for multicopy lines involving cloning and PCR verification of insert size and confirmation to exclude artefact bands for lines TPS6 and TPS8 is shown in **Figure 16A & B.**

A. Colony PCR with M13F/R



B. Nested PCR screen of above products*

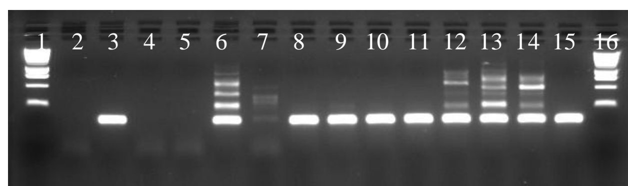


Figure 16. Example of the procedures used for downstream processing of PCR products using the APAgene™ GOLD kit (BIO S&T) for the Left Border of two copy transgenic lines of potato (TPS6 and TPS8); 5 µl of PCR products loaded per lane. **(A)** Colony PCR confirmation of cloned inserts obtained by electroporation using M13 forward and reverse primers from the LB of TPS6 (i and ii) and TPS8 (iii and iv). **(B)** Nested PCR screen of some of the above M13 products* to exclude artefacts using the confirmation test primer LBT1; a 1µl volume of the M13 PCR product was re-amplified. Lanes 1 and 16, 1 kb DNA ladder, 250 – 10,000 bp; lanes 2-3, negative (1 µl SDW) and positive (1 µl pBIN19 at 1ng/µl) controls for confirmation PCR; lanes 4-15 show results using nested PCR following the amplification of (1 µl) first-round M13 products; lanes 4-5, negative controls (SDW); lane 6, positive control (pBIN19); lanes 7-15, a range of different cloned inserts from lines TPS6 and TPS8.
*Indicates M13 PCR products used in nested PCR confirmation step.

The APAgene™ kit was used successfully in locating both of the flanking regions at the LB and RB of TPS8, both junctions at the LB of TPS42, all three junctions at the LB of lines TPS4 and TPS51, and three out of the five flanking regions at the LB of TPS9 **(Table 8)**; the screening of additional transformant colonies may have resulted in the isolation of the remaining junctions in the latter transgenic line. However, it was only possible to isolate one of the two flanking regions at the LB of Bkt1-1, one each at the LB and RB of TPS6, and one at the RB of TPS16 **(Table 8)** with the APAgene™ kit. This failure to isolate both flanking regions could be explained by rearrangements in some of the T-DNA LB and RB regions selected for the design of gene-specific primers due to the integration process, leading to the loss of the sequences specific to the primers. Although the SpeedUp™ kit was used successfully in locating one of the flanking regions from both the LB and RB of Bkt1-1, and one junction each from the RB of lines Dxs-36 and TPS6, the majority of the bands analyzed were artefacts. For this reason, the APAgene™ kit was subsequently used in testing other potato transgenic lines with multiple insert copies **(Table 8).**

Table 8. Results Summary for Transgene Flanking Regions of Multicopy Potato Transgenic Lines using the APAgene™ GOLD Genome Walking Kit (BIO S&T)

Transgenic Potato Lines	Copy Number[a]	Junctions isolated (number of products analyzed per junction)[b]	
		Left Border	Right Border
Single Transformation:			
Bkt1-1	2	1 (2[c])	2 (6[c] : 1)
Dxs-36	2	ND	2 (7[c] : 1)
TPS6	2	1 (13)	1 (5[c])
TPS8	2	2 (8 : 3)	2 (2 : 1)
TPS16	2	ND	1 (11)
TPS42	2	2 (11 : 3)	ND
TPS4	3	3 (5v : 4 : 2)	ND
TPS51	3	3 (6 : 1 : 1)	ND
TPS9	5	3 (1 : 3 : 5)	ND
Co-Transformed:			
CT-184	2	4 (4 : 6 : 3 : 1v)	2 (1 : 10)
CT-3	3	4 (7 : 1 : 3 : 1v)	2 (1 : 5)
CT-88	3	3 (5 : 3 : 2v)	3 (11 : 3 : 1v)
Re-Transformed:			
B9ZEP-13	1	4 (2 : 1v : 2)	1 (1)
B9ZEP-4	2	2 (2 : 5)	3 (2v : 1 : 3)

[a] Copy number determined by Southern analysis and real-time PCR.
[b]Number of independent PCR products isolated and sequenced that identified the same junction.
ND, Not Determined.
[c]The same flanking regions were obtained using the Seegene kit.
v = vector backbone detected, further DNA walking required.

In the case of the re- and co-transformed potato lines, there were several examples of the detection of additional flanking regions being isolated compared to the actual copy number estimation. An example of this is highlighted with the single copy (*zep* gene) re-transformed line B9ZEP-13, in which 3 flanking regions/or backbone were isolated from the LB whereas only one junction was obtained at the RB **(Table 8).** In the remaining re-transformed two copy (*zep* gene) line B9ZEP-4, three flanking regions/or backbone were detected at the RB and two junctions at the LB. The same pattern emerged with co-transformed potato lines tested. In the two copy line CT-184, four flanking sequences/or backbone emerged at the LB and two junctions were isolated at the RB **(Table 8).** In addition, when three copy lines were tested, four flanking sequences/or backbone were recorded at the LB and two junctions at the RB for CT-3, whereas three each were isolated at the LB and RB in line CT-88, in agreement with the copy number determination. Following the analysis of sequences, a PCR assay was designed to amplify a sequence crossing over the junction into the genomic DNA/or backbone to confirm the identification of flanking regions/or backbone, and this was positive in all cases described. Several explanations for the discrepancy between the copy number determination and the number of flanking regions detected in several transgenic lines can be offered. Firstly, the inherent inaccuracies of Southern analysis and the possibility of DNA rearrangements in the *npt*II gene sequence used for real-time PCR could account for an underestimation of copy number. Secondly, the integration of vector backbone without a complete T-DNA sequence (*npt*II gene deleted) could also account for the discrepancy. The presence of vector backbone DNA sequences in transgene loci was detected in several potato lines following DNA Walking and sequencing at both left and right borders **(Table 8).**

The application of the DNA walking methodology was not only successful in locating the transgene flanking regions but was also useful in the detection of additional DNA rearrangements in the vector backbone of some potato transgenic lines. In one example for a re-transformed line (B9ZEP-4), there was a truncation to the T-DNA right border sequence and immediately downstream from this, two separate rearrangements in the pBIN19 vector sequence **(Figure 17A-C);** a short piece of vector backbone sequence (36 bp) with 100 % homology to part of the *tet*A gene (nucleotides 9438 – 9473) was followed by 205 bp of backbone inserted in reverse complement and showing 100 % homology to part of the *kil*A gene of pBIN19 (nucleotides 634 – 838).

(A)

```
Zep-4RB1        ---------TGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCT 51
pBIN19          AGTTCCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCT 9240
                          **************************************************

Zep-4RB1        TAATTCTCCGCTCATGATCAGATTGTCGTTTCCCGCCTTCAGTTTAAACTATCAGTGATT 111
pBIN19          TAATTCTCCGCTCATGATCAGATTGTCGTTTCCCGCCTTCAGTTTAAACTATCAGTG--T 9298
                *********************************************************  *

Zep-4RB1        TTGTGCCGAGCTGCCGGCCAGCGAGACGAGCAAGATTGGCCGCCGCCCTCACTGCCCGG- 170
pBIN19          TTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAAGAGCGTTTATTAGAATAATCGGA 9358
                ***    **  *   ***  *  *  * *  *  *  *  **        * *   ***

Zep-4RB1        CACCT----GGTCGCTGAATG-TCGATGCCAGCACCTGCG-GCACGTCAATGCTTCCGGG 224
pBIN19          TATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCGTCCATTTGTATGTGCATGCCAACCAC 9418
                 *  *    ** **    ** * *  ** *   *  *     * * **  ****   *
```

```
Zep-4RB1          CGTCGCGCTCGGGCTGATCGCCCATCC-CGTTACTGCCCCGATC--CCGGCAATGGCAAG 281
pBIN19            AGGGTTCCCCAGATCTGGCGCCGGCCAGCGAGACGAGCAAGATTGGCCGCCGCCCGAAAC 9478
                   *     * * *      ****   *  **  **   *  ***   *** *    * **

Zep-4RB1          GA-CTGCCAGCGCTGCCATTTTTGGGGTG-AGGCCGTTCGCGGCCGAGGGGCGCAGCCCC 339
pBIN19            GATCCGACAGCGCGCCCAGCACAGGTGCGCAGGCAAATTGCA-CCAACGCATACAGCGCC 9537
                  ** * * ******  ***     ** * * ****   * **  ** * *    **** **

Zep-4RB1          TGGGGGGATGGGAG---------------------------------------------- 353
pBIN19            -AGCAGAATGCCATAGTGGGCGGTGACGTCGTTCGAGTGAACCAGATCGCGCAGGAGGCC 9596
                    *  * ***  *


(B)
Zep-4RB1          --ATTTTGTGCCGAGCTGCCGGCCAGCGAGACGAGCAAGATTGGCCGCCGCCCTCACTGC 58
pBIN19            GGTTCCCCAGATCTGGCGCCGGCCAGCGAGACGAGCAAGATTGGCCGCCGCCCGAAACGA 9480
                     *     *    *  ************************************  *  *

Zep-4RB1          CCGGCACCTGGTCGCTGAATGTCGATGCCAGCACCTGCGGCACGTCAATGCTTCCGGGCG 118
pBIN19            TC--CGACAGCGCGCCCAGCACAGGTGC--GCAGGCAAATTGCACCAACGCATACAGCGC 9536
                   *  *  * *  ***  *     * ***  ***         *  *** ** * * *

Zep-4RB1          TCGCGCTCGGGC-TGATCGCCCATCCCGTTACTGCCCCGATCCCGGCAATG-GCAAGGAC 176
pBIN19            CAGCAGAATGCCATAGTGGGCGGTGACGTCGTTCGAGTGAACCAGATCGCGCAGGAGGCC 9596
                    **     * * *  * * *  *  ***   *     ** ** *     *    *** *

Zep-4RB1          TGCCAGCGCTGCCATTTTTGGGGTGAGGCCGTTCGCGGCCGAGGGGCGCAGCCCCTGGGG 236
pBIN19            CGGCAGCACCGGCATAATCAGGCCGATGCCGACAGCG-TCGAGCGCGACAGTGCTCAGAA 9655
                   * **** * * ***  *  **  ** ****   ***  **** *   ***  *   *

Zep-4RB1          GGATGGGAG--------------------------------------------------- 245
pBIN19            TTACGATCAGGGGTATGTTGGGTTTCACGTCTGGCCTCCGGACCAGCCTCCGCTGGTCCG 9715
                    * *


(C)
Zep-4asRB1        ---------------------------------CTCCCATCCCCCCAGGGGCTGCGCCCC 27
pBIN19            CTTCTCGAACCCTCCCGGCCCGCTAACGCGGGCCTCCCATCCCCCCAGGGGCTGCGCCCC 660
                                                   ***************************

Zep-4asRB1        TCGGCCGCGAACGGCCTCACCCCAAAAATGGCAGCGCTGGCAGTCCTTGCCATTGCCGGG 87
pBIN19            TCGGCCGCGAACGGCCTCACCCCAAAAATGGCAGCGCTGGCAGTCCTTGCCATTGCCGGG 720
                  ************************************************************

Zep-4asRB1        ATCGGGGCAGTAACGGGATGGGCGATCAGCCCGAGCGCGACGCCCGGAAGCATTGACGTG 147
pBIN19            ATCGGGGCAGTAACGGGATGGGCGATCAGCCCGAGCGCGACGCCCGGAAGCATTGACGTG 780
                  ************************************************************

Zep-4asRB1        CCGCAGGTGCTGGCATCGACATTCAGCGACCAGGTGCCGGGCAGTGAGGGCGGCGGCCAA 207
pBIN19            CCGCAGGTGCTGGCATCGACATTCAGCGACCAGGTGCCGGGCAGTGAGGGCGGCGGCCTG 840
                  **********************************************************

Zep-4asRB1        T-------CTTGCTCGTCTCGCTGGCCGGCAGCTCGG-CACAAAAT-------------- 245
pBIN19            GGTGGCGGCCTGCCCTTCACTTCGGCCGTCGGGGCATTCACGGACTTCATGGCGGGGCCG 900
                          * *** * ** *   ***** * *  *   ***  * *
```

Figure 17. Nucleotide sequence isolated from the Right Border of line B9ZEP-4 showing DNA rearrangements in the vector backbone of pBIN19. **(A)** The alignment of the total 353 bp of sequence isolated using the APAgene™ GOLD kit (BIO S&T) with pBIN19; 100% homology (108 bp) to part of the Nos promoter and the T-DNA right border (nucleotides 9191-9297). **(B)** A short stretch of vector backbone sequence (36 bp) in light grey font showing 100 % homology to part of the *tet*A gene of pBIN19 (nucleotides 9438 – 9473). **(C)** A stretch of vector backbone of 205 bp (underlined font) inserted in reverse complement showing 100 % homology to part of the *kil*A gene of pBIN19 (nucleotides 634 – 838).

Overall, the APAgene™ kit was used with a high success rate with potato lines containing 1-3 transgene copies in the isolation of flanking regions/and or identification of vector backbone and it was possible to obtain products ranging in size from 300 to 2500 bp **(Figures 1** and **15).** It is likely that additional screening of PCR products can be employed to successfully isolate flanking regions in potato lines containing >3 transgene copies. Using this method, we isolated genomic flanking sequences of T-DNA insertions from transgenic potato lines derived from single transformation, co- and re-transformation events. The reasons for the greater success rate of this kit over the SpeedUp™ kit may be linked to the initial PCR step with a gene-specific primer to generate single strand DNA fragments, thus increasing the concentration of potential flanking sequences before the addition of degenerate random tagging (DRT) kit primers to form double strand DNA products. The addition of a DRT-primer digestion reaction to eliminate non-specific background caused by both free and incorporated DRT primers may have also prevented the generation of artefact products, which was an inherent problem using the SpeedUp™ kit with potato lines. Changing the annealing temperatures and length of extension at 72 ^{o}C, number of cycles during each PCR step, and altering the ratios of the kit primers DW2-ACPN and/or UniP2 to gene-specific primers did not reduce the number of artefact bands generated using the SpeedUp™ kit with lines TPS6 and TPS8 (data not shown). However, it was found from the barley work that introducing an additional pre-amplification step to the SpeedUp™ kit protocol, similar to that included in the APAgene™ kit protocol, did solve the problem of artefacts and this step has been included in the SOP (Appendix 8.5).

02.02 Develop a multiplex PCR method for the analysis of stacked transgenes.

Because the protocol of the APAgene™ GOLD Genome Walking Kit (BIO S&T) uses two different PCR buffers I and II for the four kit primers DRTA-B and DRTC-D, respectively, it was only possible to assess the kits reliability in isolating flanking regions in a duplex PCR format. Four potato transgenic lines, Bkt1-2, CT-88, B9ZEP-4, and B9ZEP-13 containing 1 to 3 transgene copies **(Table 2)** were tested using the APAgene™ kit during both a singleplex and duplex PCR format. In all cases, it was possible to isolate all flanking regions for each line in a duplex format as determined using the recommended singleplex protocol of the kit and the SOP developed (**Table 4** shows results for singleplex PCR). **Figure 18** shows the amplification results for lines Bkt1-2 and CT-88 in both singleplex and duplex PCR formats. In conclusion, the duplex PCR format with the APAgene™ kit was successful and had the advantage of being less time-consuming and more economical in terms of reagent use per sample.

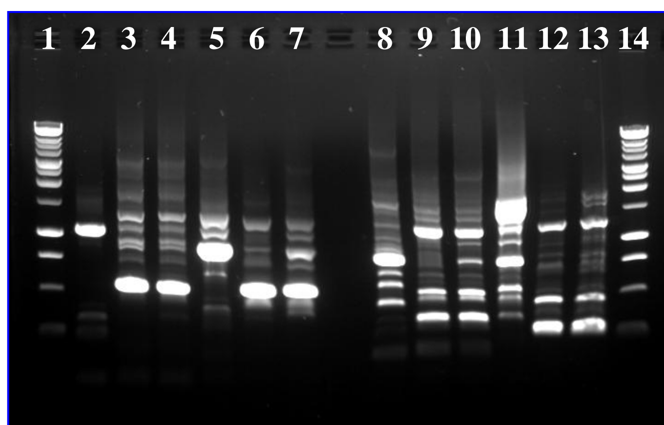


Figure 18. Final-round PCR amplification products generated using the APAgene™ GOLD kit (BIO S&T) in a singleplex and duplex PCR format from the Left Border of two transgenic lines of potato (Bkt1-2 and CT-88); 5 µl of PCR products loaded per lane. Lanes 1 and 14, 1 kb DNA ladder, 250 – 10,000 bp; Final round nested PCR products: lanes 2-3, DRTA and DRTB APAgene™ primers in singleplex with Bkt1-2, respectively; lane 4, DRTA+DRTB primers in duplex with Bkt1-2; lanes 5-6, DRTC and DRTD primers in singlepelx with Bkt1-2, respectively; lanes 7, DRTC+DRTD primers in duplex with Bkt1-2; lanes 8-9, DRTA and DRTB APAgene™ primers in singleplex with CT-88, respectively; lane 10, DRTA+DRTB primers in duplex with CT-88; lanes 11-12, DRTC and DRTD primers in singlepelx with Bkt1-2, respectively; lane 13, DRTC+DRTD primers in duplex with Bkt1-2.

4.3 Objective 03

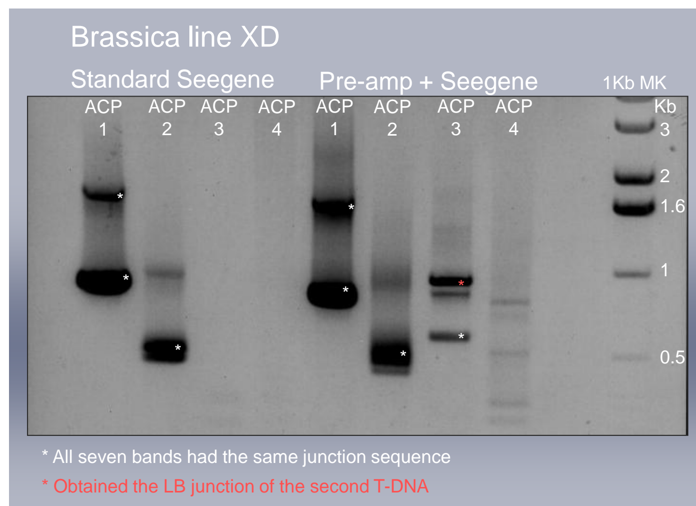
Test the best method(s) from objective 1 in a range of different transgenic crop species including wheat, barley, potato, sugarbeet, oilseed rape, *Brassica oleracea*, maize and pea. Develop a standard operating procedure for this method considering factors such as the best protocol for genomic DNA preparation from leaf tissue of a range of crops.

The Ellis DNA extraction method and the Seegene kit, reported as the best method for barley in section 4.1 and as SOPs in appendices 8.2 and 8.5, were successfully tested on a range of different transgenic crop plants. The different transgenic crops had all been produced using different plasmids and therefore each crop required its own set of target specific primers. On occasion, because of the makeup of the T-DNA, designing target specific primers could prove difficult as reported for maize. The Ellis DNA extraction method continued to perform well, giving a high yield of good quality DNA with an operator friendly protocol. Further investigation of the Charge Switch method of DNA extraction (objective 03.1) was not carried out as the method was found to be unsuitable during the work described in objective 1. Examples of the isolation and analysis of transgene flanking regions in different crops are described below.

Brassica line XD

The *Brassica* line XD contains two T-DNA insertions. In this line the standard Seegene protocol was compared to the modified protocol with an additional pre-amplification step prior to the Seegene reactions. It was found that including the pre-amplification step enabled the isolation of the second LB junction that it was not possible to obtain using just the standard protocol (**Figure 19**). In total eight bands were isolated from the gel; seven contained the same junction sequence. The gel bands were cloned into pGem-T and bacterial colonies were screened for the T-DNA junction prior to sequencing with M13 forward and reverse primers.

Figure 19. *Brassica* line XD gel showing additional LB junction obtained using the pre-amplification step.



The sequence obtained for both T-DNA junctions using a pre-amplification step and Seegene Anealing Control Primer 3 (ACP 3) are shown in **Figure 20.** Both T-DNA insertions were found to be truncated at the LB, the first by 61 bp, the second by 52 bp from the start of the LB repeat. NCBI databases were searched using BLASTn analysis on the putative plant DNA sequence; both T-DNA junctions had good alignments to *Arabidopsis* sequence. In both cases the top seven hits were *Arabidopsis* (**Figure 21**).

Figure 20. *Brassica* XD line sequences obtained for the two T-DNA junctions.

First T-DNA junction sequence.

**GTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCTACAGCCTCGGGAATTGCTACC
GAGCTCGGCCGGCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATC
CTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT
AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTC
CCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGAT
AAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTCACCGGTTAACTA
TAACGGTCCTAAGGTAGCGATGGCAAACAGCTATTATGGGTATTATGGGTGGTTCTT
TATGCGGACACTGACGGCTTTATGCCTGCAGGTCGCGAGCGATCGCGGTACCGCCC
GGGCGTCGACAGGCCTGAATTCAATTCGGCGACTTTTTTCCTCCTCATTATACTTTCT
CGTCTTATTTCAGTTTCATAACATTTGCTTGCTTACTATCATTTTAGGGACTGCGGGA
ATATGGAGAATATCGGCATTAAACGAAGCTGGTGGTTGGAAAGATAGAACAACCGTC
CCCCCCCACGGACGCTGGACCTAAACTCAATCACTAGTGCGGCCGCCTGCAGGTCGA
CCATATGGGAGAGCTCCC**

Blue pGemT
Red T-DNA truncated -61 bases from LB start
Green Plant DNA 137 bases
Pink UniP primer
Violet ACP linker
Plum ACP3 binding site

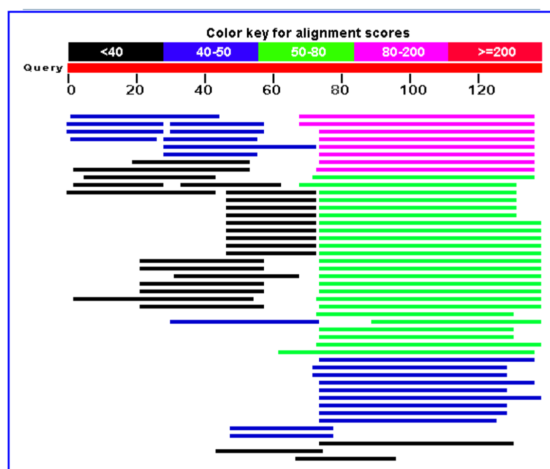
Second T-DNA junction sequence.

**CTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCAC
TAGTGATTGAGTTTAGGTCCAGCGTCCGTGGGGGGGGTACGGTTCTCGAGAGAGAC
GTCGGTTCCTGAGAGAGAAGTCAGTTCTCGAGAGAGACGTCGGTTTATCTCTTACCA
TCTTGGACAAGAGGTCGTTTGAGTTTGGTTGGGAAGTCTTAGGAACTCAGGGTTCAA
TATCGGTGAACAAGGACTCGTGTCTTGTTCATTGTGGTGGAGAGAATTCGACGTTGT
GGTGCTTTAAGCTATGTGTTTGTGGTTTGATGTTGAACAATCTAACAAAAAGAGACG
ATTAAATCCGACGAAGACGATCTCTCCTCTGCATCAGCCGTCGTAAAACCCCTTGTCT
CTGCTGAAAGCTACGAGGGAGTAAGGATACTCCGTCTCACTTAGGCTCGGCATGTCT
TGAGACTCTCATTGATCGTGTGCTCGTTCTTGTAAAAGTGAGCTAAACTATGACTAAT
TACTTTTAATATCTTCATTGAAAGAGTTACTTTTGTTGTGTATTTTTATCTCAGGTTAC
CTGATTTTGCTCTCTGTGAAACCGACTTGACATCTGAATTAACGCCGAATTGAATTCA
GGCCTGTCGACGCCCGGGCGGTACCGCGATCGCTCGCGACCTGCAGGCATAAAGCC
GTCAGTGTCCGCATAAAGAACCACCCATAATACCCATAATAGCTGTTTGCCATCGCT
ACCTTAGGACCGTTATAGTTAACCGGTGAATTCCCGATCTAGTAACATAGATGACAC
CGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTCTATCGCGTATTAA
ATGTATAATTGCGGGACTCTAATCATAAAAACCCATCTCATAAATAACGTCATGCATT
ACATGTTAATTATTACATGCTTAACGTAATTCAACAGAAATTATATGATAATCATCGC
AAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTGCCAAATGTTTGAACGATC
GGCCGGCCGAGCTCGGTAGCAATTCCCGAGGCTGTAGAATCCCGCGGCCATGGCGG
CCGGGAGCATGCGAC**

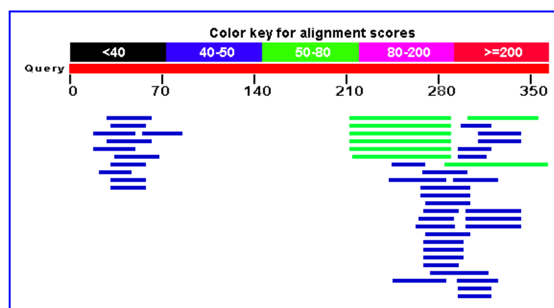
Blue pGemT
Red T-DNA truncated -52 bases from LB start
Green Plant DNA 508 bases
Pink UniP primer
Violet ACP linker
Plum ACP3 binding site

Figure 21. BLASTn analysis of *Brassica* XD line sequences.

First T-DNA junction. The 137 bases of putative *Brassica* DNA had a bit score of 114, 7e-23, with a maximum homology of 97 % to *Arabidopsis* Chromosome 5 BAC clone Accession AL162506.1



Second T-DNA junction. The 508 bases of putative *Brassica* DNA had a bit score of 77, 5e-11, with a maximum homology of 81 % to *Arabidopsis* IMK1 inflorescence meristem receptor Accession NM115033.3



Pea lines 8-3 and 8-5

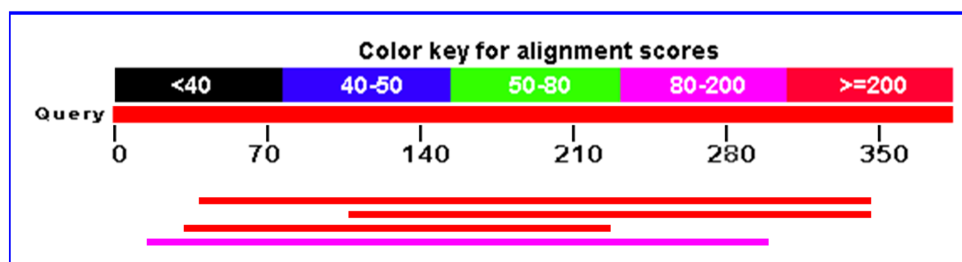
Pea lines 8-3 and 8-5 both contained single T-DNA insertions; the plasmid used to transform both lines was an early prototype plasmid of the popular pG series. It was found that the 25 bp right border repeat had been omitted from the plasmid when it was made. Not unsurprisingly, both lines had read-through at the RB incorporating large amounts of superfluous backbone sequences. For line 8-3 over 700 bp of backbone was sequenced. Line 8-5 contained 358 bp of superfluous backbone sequences from the point where the RB repeat had been omitted (**Figure 22a**). BLASTn analysis was performed on the 362 bases of putative plant DNA sequence, the top 4 hits were alignments to pea transposable element**s**. The top alignment was to accession AY299397.1 *Pisum sativum* clone Ps-cos14 Ogre retrotransposon, bit score of 872, E-value 4e-93, with maximum identities of 95 % homology (**Figure 22b**).

Figure 22(a). Pea line 8-5 sequence obtained using Seegene ACP3.

TGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCACAACATACGA
GCCGGAAGGAAACACAGATGAAAGACGGAAGAAACGGACTCGGCATGATATCGCGT
CCTGGGCCTACGTAGTCTGTCAGACACAAACATCAGAGTTGACGTAGTTTGGGGACA
GGGGGGAACGTGCTCGCTAGGATGTCGCATCCTATGCATACGTATCTTCTCTACCCA
GAGAAAGAATCAGAGCACTCGTAGCTCGGCTAACGCACGCCAAACAGAACACAAACA
GGAAACCGACTGCCAATCGCTGGACTTACGTCAGACTCTGAGCAAACAACACACACA
GGAAACCGACTGCCAATCGCTGGACTTATGTCAGACTCCAACAAACAAACAAGACAC
GCCCCCCCCACGGACGCTGGACCTAAACTCAAGCTTGGCGTAATCATGGTCATAGCT
GTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAG
CATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT
GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAAT
CGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGC
TCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
AGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGAAAGAACATGTGAG
CAAAAGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCATCACTAG
TGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAG
CTTGAGTATTCTATAGTGTCACCTAAAT

Green Plant DNA 362 bases
Red T-DNA and backbone missing RB position underlined
Blue pGemT

(b). BLASTn alignment results for pea line 8-5.



Wheat lines 22-5 and 44.

Wheat lines 22-5 and 44 were both transformed with an early version of the dual binary transformation system based on plasmids pAL154 and pAL156. The system utilises the additional virulence genes on pAL154 to enhance the transfer of T-DNA from pAL156 to the plant cell. Both lines contain single copy T-DNA insertions from pAL156. Analysis of the RB junction was performed using a pre-amplification step followed by the Seegene reactions. Four bands were isolated from the gel (**Figure 23**), cloned into the pGem-T vector and sequenced using M13 forward and reverse primers. Sequence data for all four bands gave a similar configuration of the rearrangement shown in **Figure 24a**.

Figure 23. Gel showing bands obtained for wheat lines 22-5 and 44.

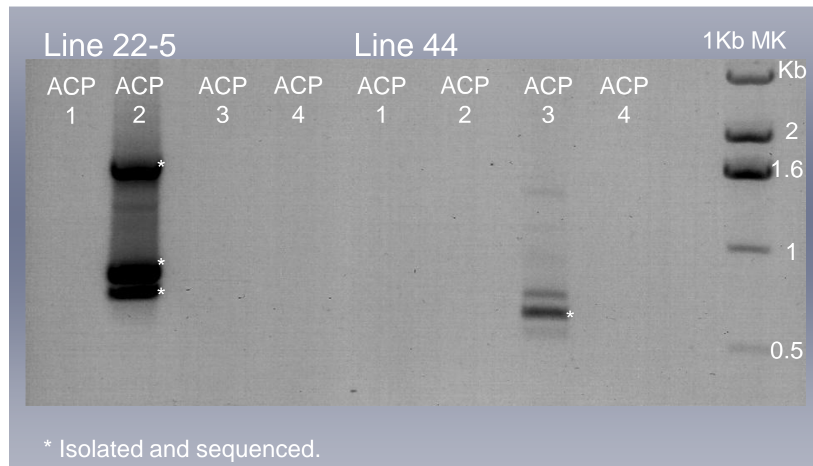
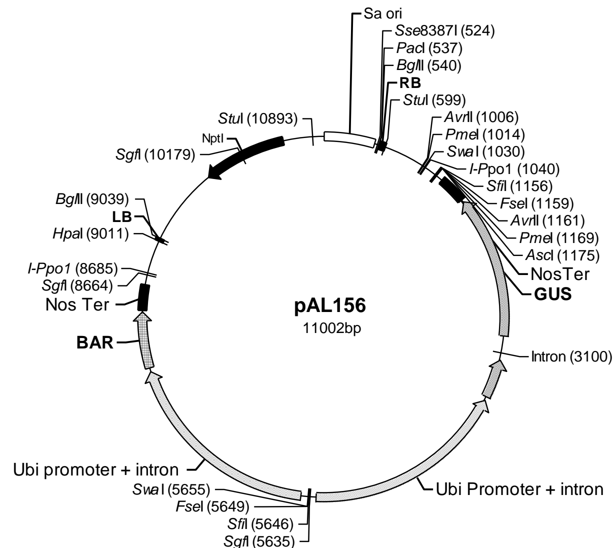


Figure 24(a). Wheat line 22-5 top band produced by Seegene ACP2, sequenced in the forward direction.

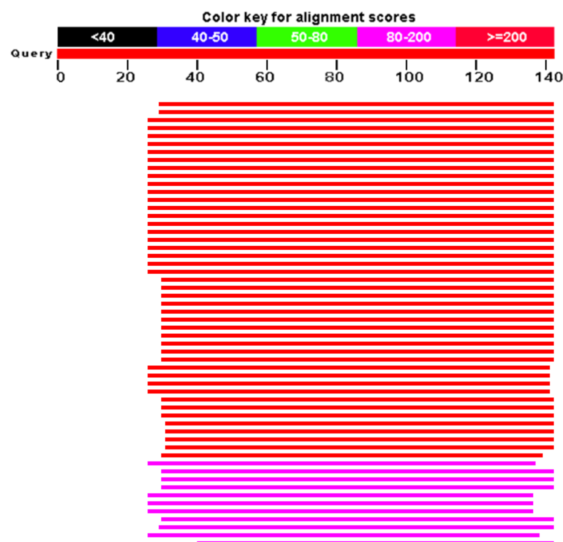
GTCGCATGCTCCCGGCCGCCATGGCCGCGGATTTAATCATGGTCATAGCTGTTTCCTGTG
TGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAA
GCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTT
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAG
GCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCA
GGGGATAACGCAGGAAAGAACATGAAGGCCTTGACAGGATATATTGGCGGGTAAACTAA
GTCGCTGTATGTGTTTGTTTGAGATCTCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
GTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTC
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTACCCCCCC
CACGGACGCTGGACCTAAACTCAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGG
AGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTG
GCGTAATCATGG

Blue pGemT
Red T-DNA
Red/yellow RB repeat
Red/Blue Micro homology
Violet Inserted plasmid Backbone DNA in reverse complement (Map 9039-9374 bp)
Black unknown
Dark yellow *Phytophthora cinnamomi* 112 bases

(**b**). Plasmid map of pAL156 wheat transformation vector.



(c). BLASTn alignment results for wheat line 22-5.



The sequence obtained revealed 55 bases of read-through from the start of and including the RB repeat. The rearrangement includes an insertion of backbone DNA, in reverse complement, from the area just outside the LB repeat which aligns to plasmid map coordinates 9039 to 9374 bp (**Figure 24b**). An area of micro-homology consisting of 6 bp was found between the T-DNA read-through and backbone insert. BLASTn analysis was performed on the remaining 141 bases, of which 29 were found to be of unknown origin and 112 bases had a strong alignment to the soil based plant pathogen *Phytophthora cinnamomi* (**Figure 24c**). The top 22 alignments were to soil

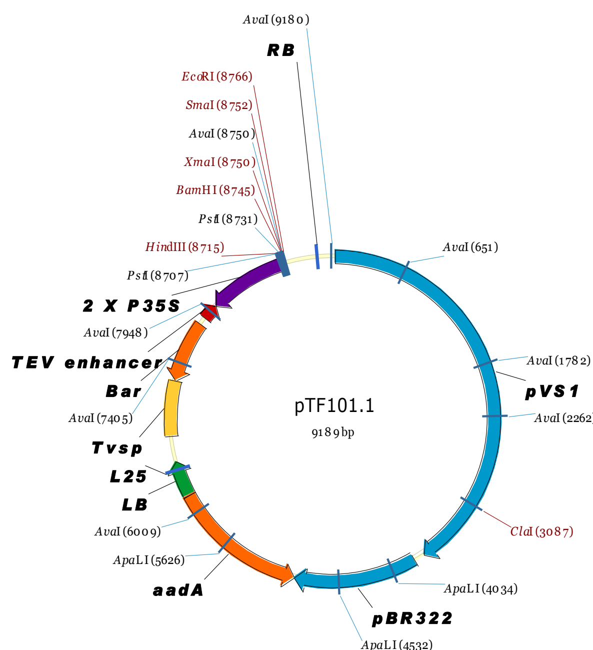
borne fungi from groups such as Ascomycota, Basidiomycota and Zygomycota. The top alignment was to accession AM412177.1 *Phytophthora cinnamomi Tub1* gene for alpha-tubulin, strain Pr120, bit score of 207, E-value 8e-51, with maximum identities of 100 % homology.

Both wheat lines contained the same rearrangement at the RB. This signifies that either the two lines arose from the same transformation event or the rearrangement occurred prior to transformation, perhaps in the bacterium. Instability in these small early dual binary vectors in bacteria is well documented, particularly insertions in the backbone. When this rearrangement arose could be ascertained by obtaining the junction between the *Phytophthora* and wheat DNA. However, designing target specific primers for this task may prove difficult because of the reverse compliment backbone insertion.

Maize line 70-20.

The maize 70-20 line contains the T-DNA from the plant transformation vector pTF101.1 (**Figure 25**), which contains the double or twinned CaMV 35S promoter close to the RB region. Many vectors contain the double 35S promoter, or similar versions with doubled upstream enhancer regions. The duplicated regions can pose a problem when designing target specific primers. Target specific primers should not be designed to bind within these duplicated regions in order to maintain their specificity. For the maize line 70-20 target specific primers were designed just outside the double 35S promoter, between the multiple cloning site and the RB.

Figure 25. Plant transformation vector pTF101.1 showing the double 35S promoter.



Junction sequence analysis, on the RB region, was performed on maize line 70-20 and three bands derived from TSP1, 3 and 4 isolated and sequenced. It was ascertained

from this and subsequent analysis that this line contains multiple T-DNA insertions. The first two junctions obtained had rearrangements where areas of truncated RB were butted against LB backbone sequence; this would suggest an insertion hotspot, within the maize DNA, where multiple T-DNAs have inserted in tandem. The first rearrangement, shown in **Figure 26a**, was confirmed by SCRI during the site validation. The second rearrangement, obtained and confirmed by PCR from right to left border, has an insertion of the *beta-D-glucuronidase* gene (*GUS*); it is unclear if this originates from the vector or from *Escherichia coli* chromosomal DNA. **Figure 26b**.

Figure 26(a). First T-DNA insertion in maize line 70-20.

CGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG
CGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGAT
ACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA
GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCACGC
CCCCCCCACGGACGCTGGACCTAA

Blue: Truncated RB sequence
Red: Vector backbone sequence from the pBR322 near LB
Orange: Seegene universal primer binding site

(b). Second T-DNA insertion found in maize line 70-20.

TGCAGCACATCCCCCTTTCGCCAGCTGGCGCTGAACCGTTATTACGGATGGTATGTCCAAA
GCGGCGATTTGGAAACGGCAGAGAAGGTACTGGAAAAAGAACTTCTGGCCTGGCAGGAG
AAACTGCATCAGCCGATTATCATCACCGAATACGGTGGAAAAACGCCAGCAACGCGGCCT
TTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTG
ATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAAC
GACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC CAATACGCAAACCG
CCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA

Red/Grey:3'end of LB3 primer binding sequence.
Red: Vector backbone sequence from the near LB
Violet: *beta-D-glucuronidase* gene (*GUS*)
Blue: Truncated RB sequence
Blue/Grey: 3'end of RB3 primer binding site
Violet: Seegene universal primer binding site

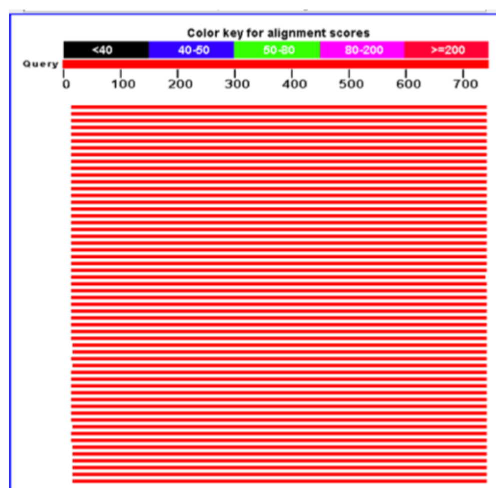
(c). Third T-DNA insertion showing junction between RB and maize 70-20 DNA.

ACGGCGCTGCTTATGATCGGCCACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT
CCGCTTCCTCGCTCACTGACTTGTTCCATGATTAACCAAAGTTGAACATAATAAAATAAGA
AATTGAAAGATTTCGTTGAAATTGTTCGTTTGGAAATTTCCCGAAAAGCTAATTATAGGTT
CTTCTCTCCATCGGAACAATAGGGCCGTTATGCTTATTACTAAACTTGTTGAAGAGATGAA
ATAGAACCAAGGTCTATCTTTTTGATCAGAGGTTAAATCGATCATCAGAAGAAGAATTAG
GCCAAAAATTAGGATACATTCTGGGAAAATGAAACTTCCATGGAAGAGAAGCAAATGAA
ACGCTTTCATAAAAATTCTCGTAGAATCGAGAATGAAGTTTTCATTCTGTACATGCCAGAT
CATGAATTAGTAACTGCAGCCAATCTCCGAAAAGTCCCGATTGTTTCGATTTTTGGAATGG
GATATTTACGGAATCCCCATGAATAGGATCAAACCTTATTCCATGCTATTTCCATAAGATT
CCTCTTTCTTATTCTTAAGCAAGCCCCCGAGAGGGCTTAGTTGATCATGATTTCTGTTTTCT
CTTTTTTTTCCTTTTTATTTGTTTCGAAAAAGATATCGTCCGATTCTCCTTCTATTGATTCTT

TTCCGATCGAGATGTATGGATCCATGTGTCTACATACCTAGATTCTGTTCATGGATTAACG
AAAATGTGCAAGAGCTCTATTTGCCTCTGCCATTCTATGAGTCGCTTCCTTTTTGCGTATGG
CACCCCCACTCCCTTGGCAGCATCA

Blue: truncated RB DNA – 56 bp before the start of the RB repeat.
Red: backbone insertion 4091 to 4120 on plasmid map
Black: unknown
Green: maize Zea perennis mitochondrion – NADH dehydrogenase subunit 1.

(d). BLASTn alignment results for maize line 70-20 third T-DNA insertion.



The third T-DNA insertion was truncated at the RB by 56 bases from the start of the RB repeat, had a 29 base insertion of vector backbone, and 721 bases of maize sequence was recovered (**Figure 26c**).

BLASTn analysis was performed on the 721 bases of maize sequence. The top 4 alignments were to maize species and the top 10 alignments gave hits to mitochondrion or chloroplast DNA (**Figure 26d**). The top alignment was to accession DQ645538.1 *Zea perennis* mitochondrion, complete genome, the feature in this part of the sequence was NADH dehydrogenase subunit 1, with a bit score of 1325, E-value 0.0, with maximum identities of 99% homology.

Potato line B-1 Seegene analysis with pre-amplification step.

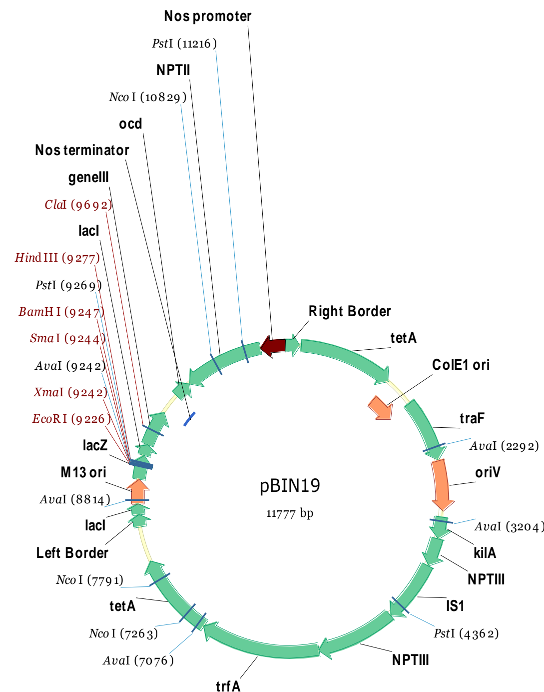
Potato line B-1, containing plasmid pBin19 was analysed using Seegene RB analysis. Both ACP1 and ACP4 gave identical junction/rearrangement and this was confirmed by PCR back to the T-DNA. The RB region of the T-DNA was truncated -37 bases from the RB repeat start (**Figure 27a**). A short piece of unknown DNA was present, then an insertion of Lac Z and Lac I (9284-9511 bp on NTI pBin19 map) from the LB region in the forward direction (**Figure 27b**). Still in a forward direction (3 bp micro-homology) a 377 bp piece of vector backbone was found from the ColE1 origin of replication from pBin backbone (1287-1664 bp on NTI pBin19 map). Finally, 65 bases were found of unknown origin but with blastn matches to expression or cloning vectors, however not to pBin 19.

Figure 27(a). Sequence obtained for potato line B-1 using Seegene analysis.

TCGGCGGGGGTCATAACGTGACTCCCTTAATTCTCCGCTCATGATCGATATCGATCAATTC
GTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACA
TACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACAT
TAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTA
ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAA
GGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGA
CCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAT
AGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTCCCCCCCCACG

Blue: T-DNA
Violet : rearrangement/insertion (Lac Z and I)
Violet/blue: Micro-homology
Red: vector backbone (ColE1 ori)
Black/green: Unknown seq blastn to cloning/expression vectors pARA and pOsAct not pBin19
Black: Seegene primer binding site

(b). Plasmid map of pBin 19 transformation vector.



Potato line TPS8 Seegene analysis.

In this line, Seegene primer 1 was used and read through at the right border going into the backbone was seen with no potato / junction found. Primers were designed to PCR back to the T-DNA for the previously found junctions 1 and 2 and for the backbone.

Junction 1 and the backbone read through could be validated; however, junction 2 obtained by the Apagene kit could not.

We have shown that the Seegene system can be used successfully in a range of different transgenic crop plants, containing a range of different plasmid T-DNA insertions. The pre-amplification step has proved useful when the amount of sample DNA has been limited and when increased Seegene product yield was needed. The increase in product yield has enabled direct sequencing in most cases, reducing time and expense spent on cloning procedures. In particular, the pre-amplification step has proven useful when obtaining transgene junction sequences in lines containing multiple T-DNA insertions and also artefacts have not been observed when using the pre-amplification. Complex rearrangements continued to be found during the analysis of GM crop species other than potato and barley.

4.4 Objective 04.

Develop protocols to allow the analysis of GM material derived by biolistics rather than *Agrobacterium*. Develop a standard operating procedure for biolistic derived material.

A protocol was developed to enable the isolation of transgene junction sequence from transgenic material derived by partical bombardment (biolistics). Biolistic transformation is achieved by coating particles of heavy metal, usually gold, with plasmid DNA; the particles are then bombarded at high velocity into plant material delivering the plasmid DNA to the cell. Transgenic plant material transformed by this method often contains multiple transgene copies, with many rearrangements, caused primarily by the physical nature of the transformation process.

In order, to perform junction sequence analysis on biolistics derived material it is necessary to first ascertain the position of the plasmid breakpoint. This is achieved by designing a number of PCR primer pairs that will amplify ~ 1 Kb regions throughout the plasmid, these regions or amplicons should overlap one another by ~ 100bp. By performing standard PCR analysis, on the biolistic derived plant material, areas of the plasmid that remain intact will produce an amplicon of known size. However, the area containing the plasmid breakpoint and therefore the junction will not. Once the 1 Kb area containing the breakpoint has been determined, it is prudent to “fine map” the area by designing another series of primer pairs that will produce amplicons of ~300 bp within this 1 Kb region. Once the position of the breakpoint is known, within ~300 bp, target specific primers can be designed to use with the Seegene kit following the protocol given in the SOP (Appendix 8.5).

Analysis of the biolistics derived barley line 8-1a.

Barley line 8-1a contains a single copy of the biolistic transformation plasmid pAL51; 32 primers (16 pairs) were designed to enable the amplification of the whole insert and identify the breakpoint (**Figure 28**). In the initial PCR reactions amplicons ranged in size from 761 to 1654 bp and the breakpoint was identified in a region between the

3’ end of the luciferase gene and the 5’ end of the ubiquitin promoter driving the *Bar* gene. By using combinations of primers to fine map in this area the breakpoint was believed to be within the 311 bp region between primers Luc LB4 and pAH nos term. Each amplicon was sequenced in forward and reverse directions. Target specific primers were designed within the 3’ end of the luciferase gene and junction sequence analysis performed with the Seegene protocol. Junction sequence was obtained for barley line 8.1a, the breakpoint and junction being at the 3’ end of the luciferase gene (**Figure 29a**). 365 bases of putative barley DNA were obtained (**Figure 29b**). BLASTn analysis was performed, giving a single alignment to barley accession DQ249273.1 *Hordeum vulgare* subsp. vulgare cultivar Morex BAC 631P8, with a bit score of 425, E-value 6e-116 and identities 325 / 370 = 87 % (**Figure 29c**). The BAC length was 101158 bp, looking at the BAC sequence in more detail, it was ascertained that the insert was in a retrotransposon called Viva-1p.

Figure 28. Biolistic transformation plasmid pAL51, primer combinations shown.

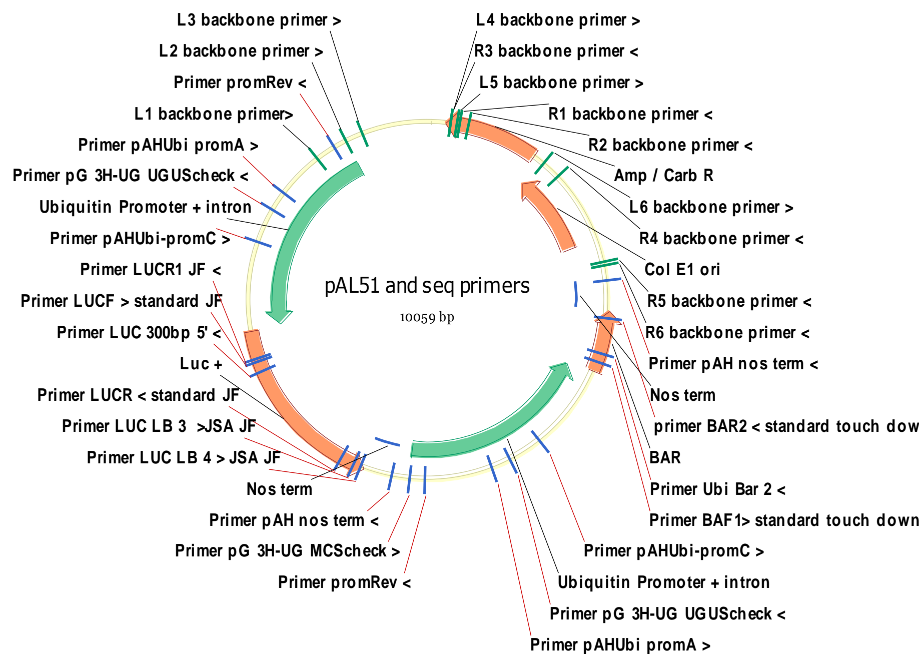
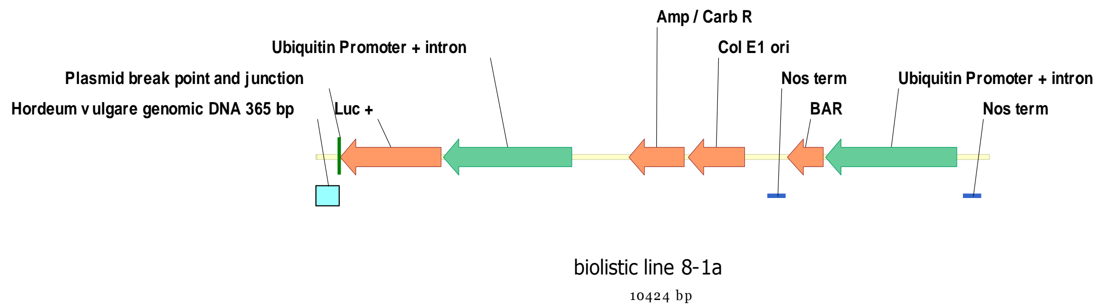


Figure 29(a). Diagram showing the plasmid breakpoint and insert junction.

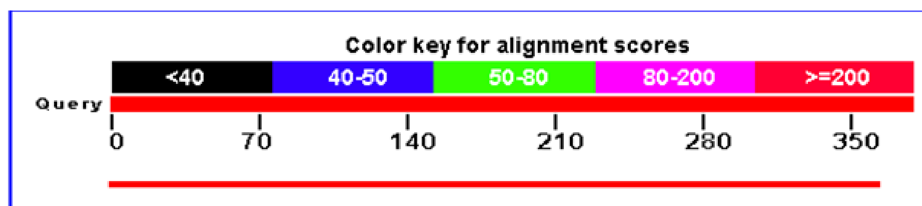


(b). Junction sequence obtained for line 8.1a, sequenced UniP in reverse direction.

TTCCGGAGTTGTGACACGTGGAATACGTTGTGCACACCCAAAAGTTCTGGGGGTAGCTCC
AACTGATAAGCAACTTCTCCACGCCTTGCCATGATCTTAAAAGGTCCAATGAACCTTGGAG
CGAGTTTCCCCTTAACATGGAAACGCTTGGTTCCCTTGAGAGGTATGACACGGAGGTACAC
CTGGTCTCCGGGAATGAACTCCACATCCCTGCGGCGAGAATCCGCATAGCTCTTTTGT
CGCGACTTGGCTGCCTTTAGATTTTCTCAAACAAGTTGCACCTTCTCTTCTGCCTTGTGCAC
AACTTCCGGCCCGAAAATCAGCCCTTCACCTGTTTCGGACCAGTTGAGCAGGGTGCGGTAC
TTCGTCCACAAACACAA

Green: barley DNA
Red: 3’ end of Luc

(c). BLASTn analysis of junction sequence obtained for barley line 8.1a.



We have demonstrated that it is possible to obtain junction sequence from biolistics derived transgenic lines using the same Seegene protocol successfully used for *Agrobacterium*-derived lines. The only difference is the requirement for the initial PCR analysis to determine the breakpoint within the plasmid. A description of this additional analysis step has been included in the SOP in Appendix 8.5.

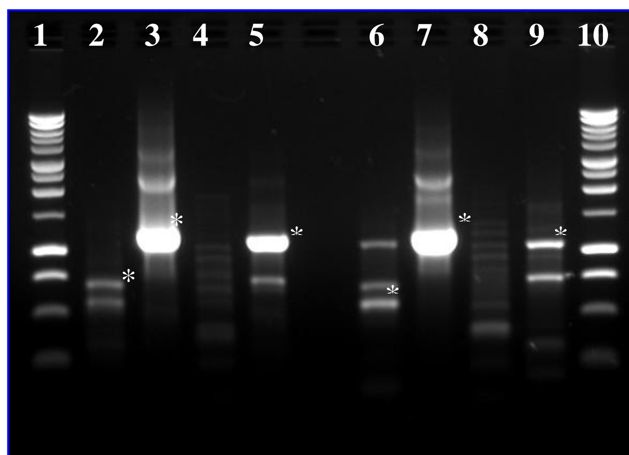
4.5 Objective 05

Validate the procedures developed in objectives 1-4 by isolating transgene junction regions from the same standard set of GM samples at two different sites and compare the results. Further validate by sequencing the entire transgene insertion in a subset of GM lines. Compare the pattern of transgene insertion seen in a range of transgenic crops.

05.1 Complete two-site validation of standard operating procedures for junction sequence analysis.

Leaf DNA extracts and primer reagents were exchanged between JIC and SCRI in order to carry out an inter-laboratory validation of SOPs. SCRI was supplied by JIC with two single copy transgenic lines of barley (87-10-1 and 92-02-01), a single copy line of maize (70-20), and a two copy line of *Brassica* species (XD). These were tested at SCRI using the DNA Walking SpeedUp™ Kit II (Seegene) following the SOP developed at the JIC. PCR products **(Figure 30A and B)** were gel extracted and purified before direct sequencing according to the SOP. The number and size range of PCR products detected correlated to those previously generated at the JIC. Sequence data obtained from each transgenic line was aligned with the appropriate cloning vector sequence using the ClustalW2 multiple sequence alignment program. Several examples of aligned sequences generated from independent PCR products for each line were forwarded to JIC for a validation check.

A. Barley



B. Maize and *Brassica*

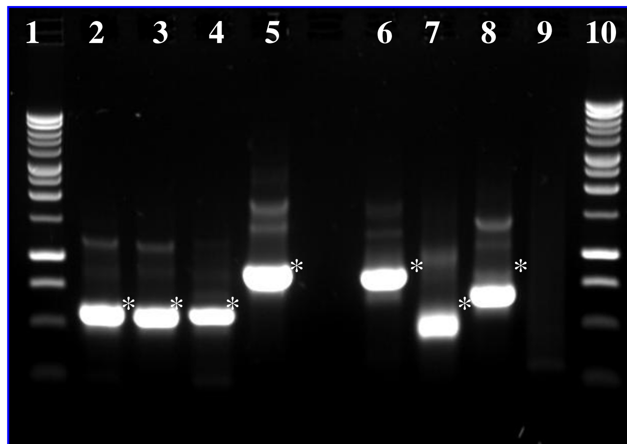


Figure 30. Final-round PCR products generated during the two-site validation of standard operating procedures for junction sequence analysis in Barley, *Maize*, and *Brassica* species using the DNA Walking SpeedUp™ Kit (Seegene); 5 µl of PCR products loaded per lane. **(A)** Lanes 1 and 10, 1 kb DNA ladder, 250 – 10,000 bp; Final round nested PCR products: lanes 2-5, ACP1 to ACP4 SpeedUp™ primers with Barley D87-10, respectively; lanes 6-9, ACP1 to ACP4 primers with Barley V92-2, respectively.

(B) Lanes 1 and 10, 1 kb DNA ladder, 250 – 10,000 bp; Final round nested PCR products: lanes 2-5, ACP1 to ACP4 primers with *Maize* 70-20, respectively; lanes 6-9, ACP1 to ACP4 primers with *Brassica* XD, respectively.
*All bands excised and purified using Wizard® Gel Clean-Up System (Promega) for direct sequencing.

After performing the junction sequence analysis, SCRI forwarded the sequence files they obtained to JIC for comparison. The sequence data obtained at SCRI using the Seegene SOP matched the junction sequences obtained at JIC exactly for the two barley lines 87-10-1 and 92-02-01, and the *Maize* line 70-20. The SCRI obtained one junction sequence for the two copy B*rassica* line XD; this matched the junction for the second T-DNA (see objective 03, section 4.3), however, an additional 188 bases of plant sequence was obtained in the initial experiment at JIC.

JIC was supplied by SCRI with two potato lines CT 88 and TPS4, the relevant primers and the SOP developed at SCRI for use with the Apagene kit. Junction sequence analysis was performed following the SOP. The junction sequences were successfully obtained and validated by PCR analysis, amplifying through the junction site. The sequences obtained were analysed and forwarded to SCRI.

For line CT 88 RB analysed using the Apagene primer B and sequenced directly using RF3 primer (**Figure 31**), three bands were isolated and gave the same rearrangement. Read-through was found at the RB + 174 bases from the RB 25 bp repeat start point. In addition, a major rearrangement, part of the OriV and kilA genes from the pBin19 plasmid backbone was found inserted in reverse complement (RC). An area of micro-

homology was also seen between the end of the backbone sequence and the insert start point.

Figure 31. Potato line CT88 RB junction sequence obtained using the Apagene SOP.

GTCATCGGCGGGGGTCATAACGTGACTCCCTTAATTCTCCGCTCATGATCAGATTGTCGTT
TCCCGCCTTCAGTTTAAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAAGAG
AAAAGAGCGTTTATTAGAATAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTC
GTCCATTTGTATGTGCATGCCAACCACAGGGTTCCCCAGATCTGGCGCCGGCCAACGAGA
CAAGCAAGATTGGACGCCGCCCTCACTGCCCGGCACCTGGTCGCTGAATGTCAATGCCAG
CACCTGCGGCACGTCAATGCTTCCGGGCGTCGCGCTCGGGCTGATCGCCCATCCCGTTACT
GCCCCGATCCCGGCAATGGCAAGGACTGCCAGCGCTGCCATTTTTGGGGTGAGGCCGTTC
GCGGCCGAGGGGCGCAGCCCCTGGGGGGATGGGAGG

Blue T-DNA
Red vector backbone
RB repeat
Micro-homology
Violet rearrangement/insertion

Line TPS4 RB was analysed using Apagene primer C and sequenced directly using RF3 primer (**Figure 32**). Five of the five bands isolated from this line gave the same rearrangement. Truncation was found at the RB -35 bases from the RB repeat start point. Following this a rearrangement of *Lac I* from the LB region of the T-DNA was found. It is possible that this represents two truncated T-DNA regions, both running in the forward direction. The junction sequence obtained was confirmed by PCR validation as described in the SOP.

Figure 32. Potato line TPS4 RB junction sequence obtained using the Apagene SOP.

GGCTTGTCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAATTCTCCGCTCATGATC
AGATTGCTGCAAATTGACGCTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTAC
TGGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGG
CCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGT
TTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAG
GGTTGAGTGTTGTTCCAGTTTGGAACACGAGTCCACTATTAAAGAACGTGGACTCC

Blue T-DNA
Violet rearrangement/insertion (Lac I)

To summarise, the standard operating procedures developed at both JIC and SCRI were validated by being successfully repeated by the other laboratory. In addition, in all cases where identical GM material was analysed at both sites, the same transgene flanking sequences were isolated.

05.2 Sequence and analyse the entire transgene insertion in co-transformed potato lines and barley lines transformed by both biolistics and *Agrobacterium*.

For the analysis through the entire transgene insertion in potato we designed a T-DNA, with left/right border flanking region, PCR-based walking approach to sequence the entire transgene insertion site in the *Solanum phureja* line Bkt1-1 (**Table 2).** A series of six primer sets were designed to PCR walk from the genomic DNA flanking region across the left border and into the T-DNA, and then crossing over to the right border and flanking sequence. Individual PCR products were purified using the Wizard® SV PCR Clean-Up System (Promega) and sequenced directly. The sequence data obtained for each PCR product was aligned with the pBIN19 cloning vector and flanking sequences at the LB and RB using the ClustalW2 multiple sequence alignment program. A contig map was generated using all PCR products and was 7041 bases in length, and this was found to be 100 % homologous to the pBIN19 cloning vector and flanking sequences.

In barley, entire transgene insertions were sequenced in lines transformed by biolistic and *Agrobacterium* mediated methods. For the biolistics transformed line 8-1a, containing the plasmid pAL51, 16 PCR primer pairs were designed to walk though the transgene insertion (see section 4.4 objective 04). For the single copy barley lines 86-3 and 89-2, containing a T-DNA insertion from pBRACT 217, a series of 10 primer pairs were designed to walk through the T-DNA from the LB to the RB junction (**Figure 33**). The individual amplicons were sequenced directly in both forward and reverse directions. The sequence data was aligned in ContigExpress with the relevant plasmid sequence and contig maps drawn, an example is given in **Figure 34**. In all cases, the sequence data alignment was as expected, no unexpected rearrangements were found within the transgene insertions for the biolistics or *Agrobacterium* derived barley lines.

To summarise the analysis of the entire transgene insertions in potato and barley, no additional rearrangements were found within the inserted DNA. Although the number of lines analysed in this way was small, it does indicate that most rearrangements are found at the transgene flanking region and not within the inserted DNA. Therefore, full analysis of the transgene flanking regions would normally be sufficient to detect most rearrangements.

Figure 33. Map of pBRACT 217 with primers used for sequencing though the T-DNA.

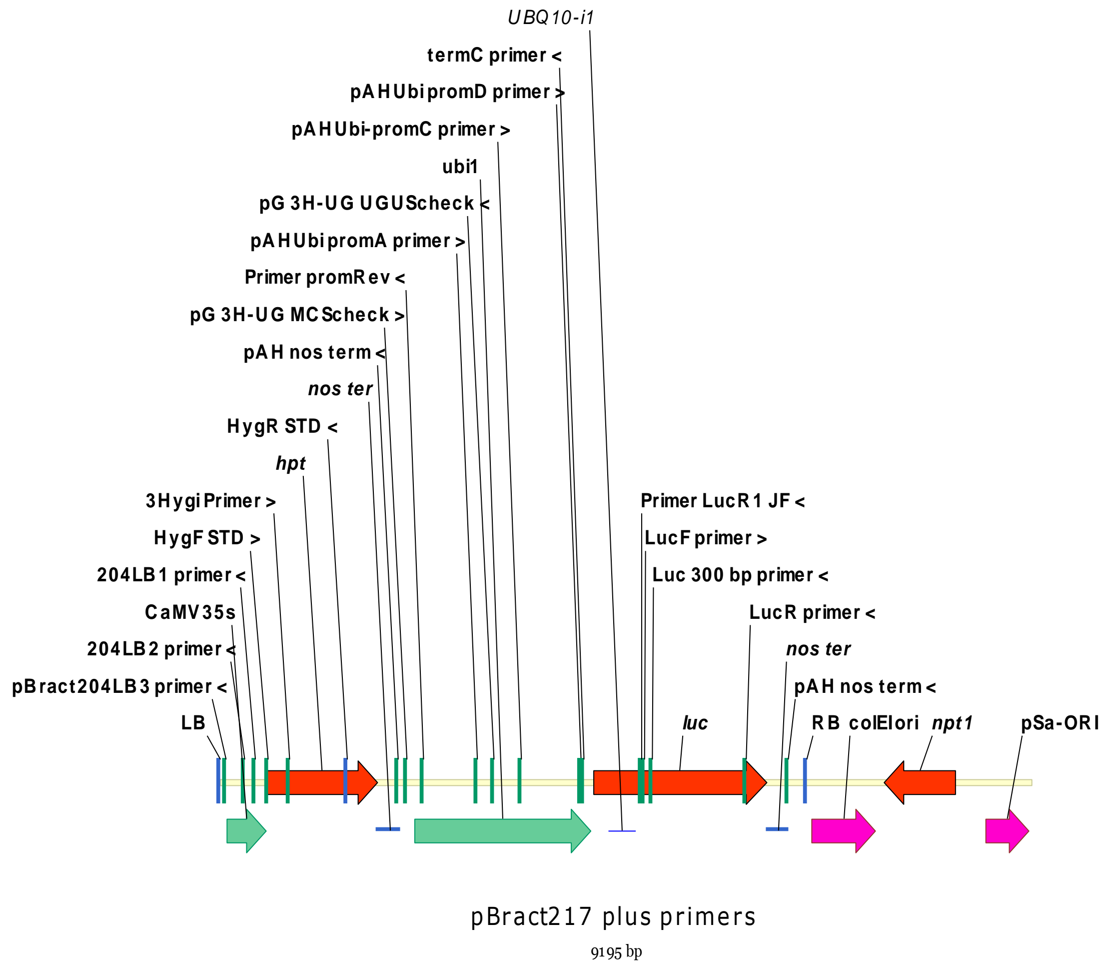
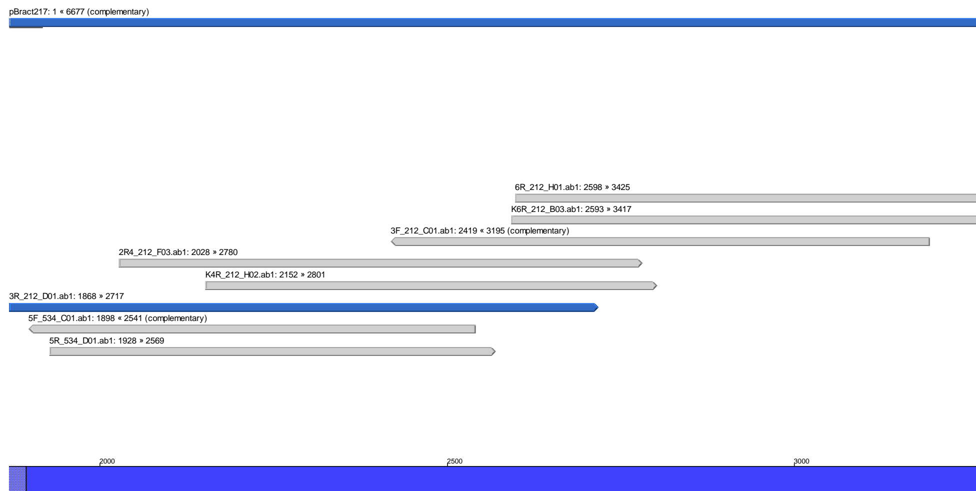


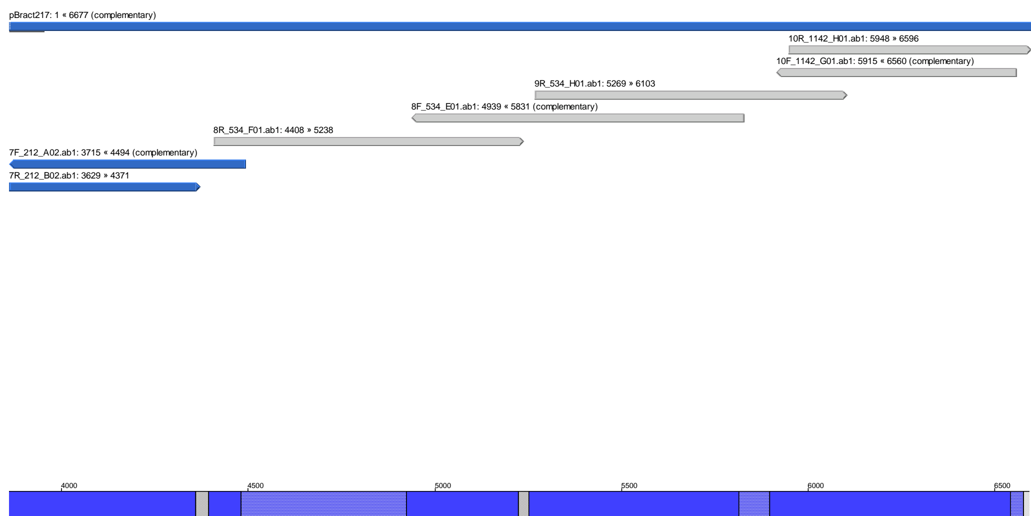
Figure 34. ContigExpress map for barley line 86-3 aligned with pBRACT 217. Parts a, b and c of the figure represent consecutive sections of the T-DNA. The top blue bar in each part of the figure shows the original pBract217 plasmid sequence. Below this the individual amplicons are shown aligned to the plasmid sequence. The final wider blue bar shows the ContigExpress map made up of the individual amplicons and shaded according to the number of overlapping amplicons contributing to that section of the map.

(a)

(b)



(c)



5. Conclusions and Recommendations

The current methodologies for the production of GM crops result in random insertions of transgenes. It is therefore possible that a transgene insertion will result in disruption of a plant gene or regulatory sequence leading to the synthesis of a novel protein. The isolation and analysis of the genomic regions flanking transgenes and of the junction between transgenic and plant DNA is therefore an essential part of the safety assessment of GM crops. In the future, this uncertainly about the location of the transgene insertion is likely to be removed by the availability of methodologies such as the use of zinc-finger nucleases that will allow the precise insertion of transgenes into plant genomes (Shukla et al. 2009, Townsend et al. 2009). However, the current GM crops, and most of those in the pipeline, do not benefit from this new technology. The availability of robust, reliable and validated procedures to isolate and analyse transgenic flanking regions is therefore of great importance.

From the work carried out in this project, it is now possible to recommend the SOP using the Seegene DNA Walking SpeedUp™ Kit II for the isolation of transgene flanking sequences from a wide range of GM plant material with low numbers of transgene insertions. For more complex material, the SOP using the APAgene™ GOLD Genome Walking Kit is appropriate. For both kits it was found necessary to make some modifications and to introduce some additional steps to ensure that artefacts were not amplified and that the transgene flanking sequences could be fully validated.

The results obtained during the project have identified frequent insertions / rearrangements at the transgene insertion site. The explanation for such insertions / rearrangements, and in some cases the origin of the additional inserted DNA, is not clear. For this reason we recommend additional investigation of transgene insertion including the entire process from plasmid vector preparation and introduction into *Agrobacterium* through to plant transformation. The insertion of DNA originating from *Agrobacterium* chromosomal DNA during plant transformation has already been documented (Ulker et al. 2008). The development of an efficient screening process to detect sequences originating from *Agrobacterium* chromosomal DNA or from bacterial cloning vectors might be one way to address this problem. It is important to note that many of the GM lines analysed in this project were developed a number of years ago using DNA vectors that have subsequently been improved. This may account for the high levels of rearrangements observed. Analysis of more recent GM lines with only single copies of the transgene present found that 14% had rearrangements with additional sequence inserted. However, the frequency and type of rearrangements found during these experiments highlights the importance of a full analysis of transgene flanking regions in transgenic crops.

All of the milestones for this project have been successfully met and SOPs for the isolation of transgene flanking regions from different GM crops with different complexities of transgene insertion and those developed using either *Agrobacterium* or biolistics have been prepared and are included in the Appendices of this report.

6. References

An S, Park S, Jeong D-H, Lee D-Y, Kang H-G, Yu J-H *et al.* (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiology 133: 2040-2047.

Ellis THN (1994) Approaches to the genetic mapping of pea. In: Modern methods of plant analysis, vol 16. Springer, Berlin Heidelberg New York, pp 117–160.

Harwood W A, *et al.* G02002 final technical report 2005, Food Standards Agency.

Kovřík, A.; Matzke, M. A.; Matzke, A. J. M.; Koukalová, B. (2001) Transposition of IS*10* from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. Mol Genet Genomics 266: 216-222.

Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, et al. (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. The Plant Journal 39: 450-464.

Salvo-Garrido H, Travella S, Bilham LJ, Harwood WA, Snape JW (2004) The distribution of transgene insertion sites in barley determined by physical and genetic mapping. Genetics 167: 1371-1379

Sha Y, Li S, Pei Z, Luo L, Tian Y, He C (2004) Generation and flanking sequence analysis of a rice T-DNA tagged population. Theor Appl Genet. 108: 306-314.

Shukla V K et al. (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature 459, 437-441

Townsend J A et al. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459, 442-445

Travella S, Ross S M, Harden J, Everett C, Snape J W, Harwood W A (2005) A comparison of transgenic barley lines produced by particle bombardment and *Agrobacterium*-mediated techniques. Plant Cell Reports, **23**: 780-789.

Ulker B, Li Y, Rosso M G, Logemann E, Somssich I E Weisshaar B, (2008) T-DNA-mediated transfer of *Agrobacterium tumefaciens* chromosomal DNA into plants. Nature Biotechnology 26: 1015-1017.

7. Outputs

Bartlett J G, Smedley M A, Harwood W A. 'Analysis of T-DNA/host-plant DNA junctions in single-copy transgenic barley lines'. Molecular Genetics and Genomics (in preparation)

Cullen D W, Harwood W A, Smedley M, Davies H V & Taylor, M A (2008). Development of a standard, validated procedure for the isolation of transgene flanking regions in GM crops and detailed analysis of transgene insertion. Abstracts of the 1st Global Conference on GMO Analysis, Villa Erba, Como, Italy, 24-27 June 2008, 112.

Cullen D W, Harwood W A, Smedley M, Davies H V & Taylor M A (2009). Comparison of DNA Walking Methods for Isolation of Transgene Flanking Regions in GM Potato. Journal of Agriculture & Food Chemistry (in preparation).

Harwood W A, Transgene insertion and implications for food safety, JIC seminar, March 2007.

Smedley M, Taylor M A, Cullen D W, Harwood W A (2009). Isolation and analysis of transgene flanking regions in a range of GM crops. (in preparation).

8. Appendices