VILNIUS UNIVERSITY

RAPOLAS SPALINSKAS

DEVELOPMENT, OPTIMIZATION AND APPLICATION OF PCR-BASED METHODS FOR MOLECULAR CHARACTERIZATION OF TRANSGENIC INSERTS IN GENETICALLY MODIFIED PLANTS

Doctoral Dissertation
Biomedical Sciences, biology (01 B)

Vilnius, 2014

This doctoral dissertation was prepared from 2008 to 2014 in European Commission Directorate General Joint Research Centre, Institute for Health and Consumer Protection, Molecular Biology and Genomics Unit; and Vilnius University, Faculty of Natural Sciences, Department of Botany and Genetics.

Scientific supervisor:

Prof., habil. dr. Juozas Rimantas Lazutka (Vilnius University, biomedical sciences, biology – 01 B)

Scientific consultant:

Dr. Anne Milcamps (European Commission Directorate General Joint Research Centre)

VILNIAUS UNIVERSITETAS

RAPOLAS SPALINSKAS

GENETIŠKAI MODIFIKUOTŲ AUGALŲ TRANSGENINIŲ INTARPŲ MOLEKULINIS APIBŪDINIMAS KURIANT, OPTIMIZUOJANT IR TAIKANT POLIMERAZINE GRANDININE REAKCIJA PAGRĮSTUS METODUS

Daktaro disertacija

Biomedicinos mokslai, biologija (01 B)

Vilnius, 2014

Disertacija rengta 2008 – 2014 metais Europos Komisijos Generalinio Direktorato Jungtinio tyrimų centro Sveikatos ir vartotojų apsaugos instituto Molekulinės biologijos ir genomikos padalinyje bei Vilniaus universiteto Gamtos mokslų fakulteto Botanikos ir genetikos katedroje.

Mokslinis vadovas:

Prof. habil. dr. Juozas Rimantas Lazutka (Vilniaus universitetas, biomedicinos mokslai, biologija – 01B).

Konsultantė:

Dr. Anne Milcamps (Europos Komisijos Generalinis Direktoratas Jungtinis tyrimų centras).

TABLE OF CONTENTS

ABBREVIATIONS

AAP - Abridged Anchor Primer

acc. No. - Access number

Amp - Ampicillin

AOCS - American Oil Chemists' Society

AUAP - Abridged Universal Amplification Primer

BAC - Bacterial artificial chromosome

BDA - Boomerang DNA amplification

BLAST - Basic Local Alignment Search Tool

BSA - Bovine serum albumin

CaMV - Cauliflower mosaic virus

cat.no. - Catalogue number

CDS - Coding DNA sequence

CERA - Center for Environmental Risk Assessment

CNT - Carbon Nanotubes

CPCR - Capture PCR

CRM - Certified reference material

dC7GTP - 7-deaza-2′-deoxyguanosine

dCTP - Deoxycytidine triphosphate

ddNTP - dideoxynucleoside triphosphate

delta-G - Gibbs free energy (free enthalpy)

dGDP - deoxyguanosine diphosphate

DLA - Digestion ligation amplification

DMSO - Dimethyl- sulfoxide

dsDNA - double stranded DNA

DTT - Dithiothreitol

DWCNT - Double walled carbon nanotubes

EC - European Commission

EDTA - Ethylenediaminetetraacetic acid

EFSA - European Food Safety Authority

E-GW - Extension-based genome walking

EU-RL GMFF - European Union Reference Laboratory for GM Food & Feed

FLEA-PCR - Flanking-sequence exponential anchored-PCR

gDNA - Genomic DNA

GM - Genetically modified; genetic modification

GMDD - GMO Detection Method Database

GMO - Genetically modified organism

GMP - Genetically modified plants

GSP - Gene specific primer

GW - Genome walking

I - Inosine

IHCP - Institute for Health and Consumer Protection

IPCR - Inverse PCR

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IRMM - Institute for Reference Materials and Measurements

JRC - Joint Research Centre

Kan - Kanamycin

LaNe PCR - Lariat-dependent nested PCR

LB media - Lysogeny broth media

LDV PCR - Long Distance Vectorette PCR

LF-PCR - Locus-Finding PCR

LM-PCR - Ligation-mediated PCR

LT-PCR - Long template polymerase chain reaction

LTR - Long tandem repeat

LT-RADE - Long template rapid amplification of genomic DNA ends

M (M1, M2...) - Marker

MWCNT - Multi walled carbon nanotubes

NCBI - National Center for Biotechnology Information

P-35S - Cauliflower mosaic virus 35S promoter

PCR - Polymerase chain reaction

P-E35S - Enhanced cauliflower mosaic virus 35S promoter

P-GW - Primer-based genome walking

PNK - Polynucleotide kinase

PVA - Polyvinyl alcohol

PVP - Polyvinyl pyrrolidone

RACE - Rapid amplification of cDNA ends

RADE - Rapid amplification of genomic DNA ends

RAGE - Rapid amplification of genomic ends

R-GW - Restriction-based genome walking

RSE-PCR - Restriction site extension PCR

RSO - Restriction site oligonucleotide

rxn - reaction

SDBS - Sodium dodecyl benzene sulfonate

SD-PCR - Shine–Dalgarno sequence based PCR

SEFA-PCR - Self-Formed Adaptor PCR

SEQ (or seq) - Sequence

SHP-PCR - Sequential Hybrid Primer PCR

SRPCW - Semi-Random PCR Chromosome Walking

ssDNA - Single-stranded DNA

SSP-PCR - Single-specific-primer PCR

SWCNT - Single walled carbon nanotubes

T35S - Cauliflower mosaic virus signal terminator

TAIL-PCR - Thermal asymmetric interlaced PCR

TBE - Tris/Borate/EDTA buffer system

T-DNA - Transfer DNA

TdT - Terminal deoxynuleotidil transferase

TE - Tris-EDTA buffer system

TMAC - Tetramethylammonium chloride

TRIS (or Tris)- tris(hydroxymethyl)aminomethane

TVL-PCR - TOPO Vector-ligation PCR

UFW - Universal fast walking

USA - United States of America

USPTO - United States Patent and Trademark Office

X-gal - Bromo-chloro-indolyl-galactopyranoside

YT broth - Yeast and tryptone broth

INTRODUCTION

One of the fastest developing fields of biotechnology in the past three decades is that of genetically modified (GM) plants for industrial applications, as well as food and feed. In 2013 alone, 1.5 billion hectares of crops were planted worldwide comprising 12% of biotech crops which signifies a 100-fold growth increase of such a subgroup since 1996 (James 2013). As this progress involves the consumer, being directly or indirectly in the food chain, a comprehensive and strict regulatory system must be implemented that is supported by the outcome of scientific studies. Therefore, a GM crop and everything that it is composed of has to be extensively studied at the molecular level, reliable analytical data has to be provided and the political decisions for cultivation, marketing, labeling, tolerance and other aspects have to be taken. The enormity of such a phenomenon is exemplified by new institutions appearing and a large number of life science specialists shifting towards GMO studies. The control of GM food, feed and other products derived from such material is still in constant development, thus what was not long ago considered sufficient as qualitative data on GM presence, is now shifting toward "how much" of it is present (quantitative data analysis). Subsequently, such investigations lead to more substantiated decisions that have positive outcomes for the consumer.

The European Union (EU) has probably the most rigorous regulations for GMOs in the world. The granting of authorizations for a novel GM event in the EU is governed by the verdict of the European Commission (EC) that relies on case study evaluations by European Food Safety Authority (EFSA). As of 2014, the EC Register lists a total of 48 authorized GMOs and 15 events of pending or expired authorization (EU Register of authorized GMOs 2014). Regarding GM crops, at the time of drafting, two were permitted for cultivation in separate EU member states: MON810 maize (trade name "YieldGard"; Bt δ-endotoxin expressing maize; a food crop) and EH92-527-1

potato (trade name "Amflora"; amylopectin producing trait; crop for industrial application where waxy starch is used). The sequence of the authorization events in the EU can only begin after an analysis of full-scale reliable scientific data that is provided during application for authorization of a new GM event. This thesis seeks to explain what constitutes the experimental part behind the GMO analysis and evaluation for authorization, as well as what else can be done to obtain additional knowledge on the GM crop.

This manuscript presents the strategies of molecular characterization of transgenic DNA as well as the plant genome sequences around it. The first technique to perform a case-study is a well established method called Long Template Polymerase Chain Reaction (LT-PCR) which has a potential to retrieve the transgenic insert in one single uninterrupted fragment. The possibility of obtaining the full fragment with plant flanking sequences bordering it and therefore confirming the fragments integrity may establish LT-PCR as a stand-alone preferential method to apply in GMO studies. As the integrity of the transgenic material is of key importance in GMO safety studies, the further manipulation of the retrieved fragment may be of service to this cause while the primer walking type of sequencing is applied. As in this case the DNA template is sequenced by a series of direct or indirect (plasmid) primer walking reactions, the highest possible accuracy is achieved on long fragments. As previous sequencing run results act as a primer design spot for all consequent reactions, the integrity of the DNA fragment may be further confirmed. The aim was to establish primer walking sequencing as a preferential method for long fragment nucleotide sequence decoding. After the method is set, the other crucial aspect would be the source of DNA bringing the quality sufficiency issue, sensitivity of testing and concentration. In the event of the traces of unknown GMO appearance on the market, the DNA template may be at very low GM quantity. On the contrary, some of the GM plant DNA can be purchased commercially as pure leaf nucleic acid of close to 100% GM quantity. Additionally, differently concentrated Certified Reference

Material (CRM) is also available on the market. All of these sources may or may not be suitable for complex applications in molecular characterization of the transgene. It is therefore more favorable to test the DNA sources in our study.

As LT-PCR is a sensitive reaction with many variables that may lead to road-blocks when mistakenly selected, such as premature polymerase drop from the template, secondary DNA structures blocking polymerization, ineffective primer annealing. In such occurrences the additives to the reactions are considered. One of the most common ones is brought from the material sciences and it is a hypothetically potent PCR additive called carbon nanotubes (CNT). It has a foundation to increase the LT-PCR specificity, reaction stability and to impact the reaction yield. The preparation of such a typically insoluble additive with very limited application in molecular biology data was established as a separate objective in this study for scientific novelty reasons. In addition, working with this additive would also bring merits of gaining information on handling, preparation and application of MWCNTs that would not only yield benefit in applied molecular biology but also in any field in which material science co-applies.

The other strategy in the molecular characterization of GMOs study deals with another fundamental issue that is the knowledge of plant genomic sequences that surround the transgenic insert. As such, the possibility to identify unknown DNA regions adjacent to known sequences as in general application of various Genome Walking (GW) methods needs to be tested on GMOs. To this day a stand-alone method for testing the unknown DNA around the insertion site of the GM cassette was not established. As it will be presented in the manuscript, the GW technique brings the most potential to obtain the required sequences in the least time-consuming way with an appropriately positive success rate. However there is no unique method that stands out as the preferred one and carries a lot of scientific potential and novelty in development. If properly

developed, the method should be robust, easily applicable and fit the needs of GMO studies. The optimization of the method should include the leveling up of the parameters to obtain positive results in the most cases where the method is applied, particularly in GMO testing. As statistically the most widespread element in the transgenic plants is a P-35S promoter (promoter 35S from Cauliflower Mosaic Virus), it is an advantage to have a set of primers that would fit any genome walking application on the material containing this element. This promoter being strong and effective it is also being used in the development of the new GM plants. Thus any further knowledge of this element is valuable.

Finally, the genome walking method should be tested out in the laboratory routine application study and enough variety of specimens in order to prove the robustness of the method. This study on five distinct GM events was established as a follow-up to the development and optimization of the method. The largest possible assortment of species that have a 100% GM DNA material on the market is selected to carry on with this objective. The expected proof that the method works on different GM plant species would expand the possibilities of the continuation of method's further optimization and application in real-life emergency situations involving detection, identification and characterization of unknown GMOs. The outcome of this application study was to be published as a separate scientific publication with nucleotide sequence data deposited into the worldwide databases as GenBank.

This thesis deals with a series of GMO research issues when trying to detect and identify, and most significantly characterize the transgenic event at a molecular level. In this manuscript, two approaches to obtain the nucleotide sequence of synthetic DNA insert and the plant genome sequences flanking the insert are presented. Applicability and/or arising difficulties of these techniques are thoroughly described. While carrying out the experiments and planning day-to-day strategies of approach, optimizations of all the relevant methods

were taken into consideration. The rationale behind that was the peak specificity and the highest yield of the expected product of any method that was applied; however the feasibility of any developed or optimized method was also considered.

THE AIM OF THE STUDY

Development, optimization and application of PCR-based methods for molecular characterization of transgenic inserts and plant flanking sequences in GM plants.

THE OBJECTIVES OF THE STUDY

(1) Demonstrate the probability of obtaining the full uninterrupted LT-PCR fragment containing the entire transgenic insert of the selected GM events and display the integrity via "primer walking" sequencing;

(2) To test the hypothetically potent hydrophobic PCR additives carbon nanotubes (CNT) as a foundation of increased LT-PCR specificity, reaction stability and impact on the reaction yield;

(3) To develop and optimize a new robust method which would allow the identification of unknown DNA regions adjacent to known sequences aiming at the needs of analysis of transgenic DNA present in GMOs;

(4) Based on statistical calculations (frequency of the most commonly used transgenic element) a set of universal primers for genome walking or other PCR-based application was aimed to be designed on the promoter 35S from Cauliflower Mosaic Virus);

(5) Application study of genome walking method on five distinct GM events was established as an objective of the study for proof of the method robustness and for obtaining publically unavailable sequences of GM plants.

SCIENTIFIC NOVELTY

- The application of the Multi Walled Carbon Nanotubes (MWCNT) for the improved LT-PCR specificity, reaction stability and yield increase has no or very limited published data. No application of such PCR additive on GMO is currently known.
- The development and optimization of a user-friendly restriction-independent adapter ligation-mediated genome walking method LT-RADE for mapping the unknown DNA adjacent to known sequences.
- Application of aforementioned genome walking method LT-RADE to obtain the nucleotide sequences of unpublished transgenic sequences of the selected GM events.

WORK APPROVAL

The results of this study were published in two scientific publications in the peer-reviewed journals Food Analytical Methods (IF_{2012} = 1.969) and European Food Research and Technology (IF_{2012} = 1.436). Twelve (12) nucleotide sequences were submitted to the GenBank database. The results were presented in the European Commission Directorate General Joint Research Centre, Institute for Health and Consumer Protection, Molecular Biology and Genomics Unit scientific seminar.

CHAPTER I: LITERATURE ANALYSIS

If we took a quick glance back in time, it would be hard to believe how swiftly genetic engineering became a field of immense knowledge. When Paul Berg performed the landmark gene-splicing experiment creating the first synthetic recombinant DNA in 1971, the world of science was hardly ready for an upcoming revolution that was about to happen.

1. SAFETY AND REGULATION OF GMO

1.1. Regulation of the GMO in the European Union in brief

To establish the control of a very important group in foodstuffs and feed *i.e.* GMO, European Union legislation provides with important guidelines to govern all aspects of GMO including the most important ones: direct and/or indirect effect on human health, and the environment. The main legislative acts that are in force in the EU are:

1. Directive 2001/18/EC of the European Parliament and of the Council (OJ L 106 of 17.4.2001) on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. This directive was later amended by directive 2008/27/EC (OJ L 81 of 20.3.2008);
2. Regulation (EC) No. 1829/2003 of the European Parliament and of the Council (OJ L 268 of 18.10.2003) on genetically modified food and feed;
3. Regulation (EC) No. 1830/2003 of the European Parliament and of the Council (OJ L 268 of 18.10.2003) concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC

The main highlight of the fundamental biotechnology directive 2001/18/EC is the transparency for the procedure of granting consent for the deliberate release and placing on the market of GMOs. This consent is limited time-wise (to 10 years; renewable) and obliges to perform the monitoring of the released GMO on market or for experimental purposes. Common objectives for such monitoring are also described in this directive. Additionally the 2001/18/EC directive covers the common methodology for case-by-case assessment of the environmental risks associated with the release of GMOs. Several articles and one annex of directive 2001/18/EC was later replaced or rephrased by directive 2008/27/EC.

Two regulations: 1829/2003 and 1830/2003 act as important adjustments to the aforementioned main directive. These regulations apply to GMOs for food and feed use, food and feed containing GMOs and food and feed produced from GMOs or containing ingredients produced from GMO. The regulation (EC) No. 1829/2003 describes the single authorization procedure, indicating that the inquirer submits one application for authorization of a GMO for food and feed uses and for cultivation. The assessment of this application for authorization is performed by the European Food Safety Authority (EFSA) which is responsible for the risk assessment. The subsequent risk management is the responsibility of the European Commission and the Standing Committee on the Food Chain and Animal Health. With respect to the Regulation 1930/2003 which sets out a framework for assuring the traceability of GMOs throughout the food chain (including processed foods) and crops and labeling particularities of the GMOs; the regulation 1829/2003 sets out some general labeling rules for GMOs and the exception of labeling of the products which contain a proportion of GMOs of less than 0.9 % of each ingredient as this threshold may be technically unavoidable in the traditional crop growth-to-processing cycle. This threshold in GMO legislation is referred to as the *GMO adventitious presence threshold*. The Regulation 1930/2003 further explains

the Member State's role in sampling and analyzing the food and feed via qualitative and quantitative methods as regarding to this threshold.

A set of additional regulations and Council and Commission decisions support the main legislative acts on GMOs, such as (the list is not exhaustive):

- Regulation (EC) No 1946/2003 of the European Parliament and of the Council on transboundary movement of genetically modified organisms;
- Commission Regulation (EU) No. 619/2011 laying down the methods of sampling and analysis for the official control of feed as regards presence of genetically modified material for which an authorisation procedure is pending or the authorisation of which has expired;
- Commission Regulation (EC) No 152/2009 laying down the methods of sampling and analysis for the official control of feed;
- Regulation (EC) No 882/2004 of the European Parliament and of the Council on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules.

All of the mentioned legislations and other Council and Commission decisions and reports act as the sets of guidelines on which the GMOs undergo a strict control. These sets of guidelines thoroughly describe such procedures as sampling of the material, screening for GMOs, quantification of the content of GM material, free goods movement rules between the Member States internally and import and export outside the EU, and other detailed actions and measures involving GMOs. Worth pointing out is the Cartagena Protocol which ensures that the transfer, handling and use of GMOs does not have any adverse effects on the environment and human health, specifically focusing on transboundary movements. The Cartagena Protocol was signed in year 2000 between the Member States of the Community and went into force in 2003 via the regulation 1946/2003.

The self-evident foundation of the legislative apparatus for GMO control in the EU is the probable direct and indirect effect of such material on human health and environment. Nonetheless at present the social discontent of the GM containing products on the market and of the GMO authorization overall in Europe is undeniable.

1.2. The safety profile: evaluation of the GMO effect on human health and the environment

The most apparent reason for GMO safety evaluation is the probability of undesirable and/or unknown interaction between the human and the GM material, or the environment and GMOs. As transgenic DNA enters the plant genome subsequently granting a novel trait to the host organism, thus the other than transgene-coded traits may be obtained by the plant. The occurrence of such unintended qualities may be responsible for a direct effect on human and environment, like the shift of the natural toxin concentrations and novel natural toxins in a GM plant and/or in the food and feed produced from the GMOs. This trait would not necessarily mean that the resulting GM containing product will be toxic *per se*, however the probable risk should not be overlooked. In such case the allergenicity profile in humans may also be altered as such becoming health threatening. Worth emphasizing that the transgene coded protein may not be the direct potential threat, but some metabolites that are not present in the traditional crops may appear and pose the same risks as described before. It is noteworthy that such potential risks of allergenicity and toxic effect does not have any scientific background (Nicolia *et al.* 2013) as lacking the real-life situations in which such occurrences have emerged as such to be considered only theoretical. At present, two cases are known of which one deals with the potential Brazil-Nut allergen presence in the transgenic soybean (Nordlee *et al.* 1996) and the other is an unverified case of maize event CBH-351 (StarLink) (Siruguri *et al.* 2004). The soybean in question has never been marketed and the maize event CBH-351 case was never verified

with this event being only approved in the United States for feed purposes (CERA 2014).

The other safety profiling example would be an impact of the GMO on the biodiversity *i.e.* the variety of life forms. This potential threat is environment-related with several possibilities to materialize such as (1) out-crossing, a happening when the GM plant passes the new trait to the traditional form of plant; (2) the impact on the weed population (as in herbicide tolerant crop fields), which may then impact the animal behaviors and populations; (3) the loss of traditional plant varieties when a GM cultivar becomes widely used instead of the conventional plants; or (4) passing the transgenes into the soil microorganisms. However the gene flow is not the single issue when dealing with coexistence *i.e.* organic, conventional and biotech crops existence as one system, because there is also the economic and management issue (Nicolia *et al.* 2013). As practically the isolated fields of GM crops are being planted, harvested and worked further, some wide information on coexistence became available (Devos *et al*. 2008; Colbach *et al*. 2008; Gryson *et al*. 2009; and other studies).

Seemingly the politicians, the public voice and the scientists unanimously agree that the threats may be potential at certain degrees. Therefore it becomes the concern of the GMO developer and producer to provide a safe GM plant as well as abundant information and descriptions required for authorization. Additionally the main concern of the scientists becomes the testing of the novel GM event and further scientific description of the plant and primarily the transgene. As such, the most crucial technical information about the GM plant as a whole and emphasizing the trasgenic cassette, can be broadly generalized as the molecular characterization. All in all, when assessing the safety of the GMO, it should be concluded that over three decades the biotech crops are being used and in parallel investigated with no major threats to the user and environment occurrence (Nicolia *et al.* 2013).

Finally, when assessing the safety profile of the GMOs, one must be cautious of the wrong or inaccurate interpretation of the threats to the consumer when

assessing the political decisions. For example, the continuous notice on the EU market of unauthorized GM rice in food products originating from China (see RASFF notifications) demanded severe measures (EU decision 2011/884/EU). Therein, it was decided that any product that tested positive for the presence of the P-35S or the T-nos or the cry1Ab genetic element should be withdrawn from the market. Such emergency measure was invoked by the lack of information available on cultivated or field-released GM rice in China. Self evidently, such decisions should not be extrapolated to the extent that the removal of products with unauthorized GMO immediately alerts about a direct health threat to the consumer, however such conclusions are sometimes drawn by the public voice. However, such decisions are rather the standard regulatory procedure where only the authorized and tolerated GMOs is allowed to appear on the European market.

1.3. GMO detection, identification and characterization: an experimental perspective of bioengineered plants

There are numerous occasions in which the presence of GMO has to be screened and/ or quantified; also the GM have to be identified on regular basis as well as characterized for expanding the knowledge of a biotech crop in molecular level. GMO being a socially sensitive subject *per se*, it is crucial to obtain reliable data for decision making. All these manipulations require different analytical experimental approaches serving the purpose (Emons 2010). To date GM testing has focused largely on the development of specific quantification methods (Van den Eede 2010). Lesser information is available on the stability of the transgenic insert in the plant (Papazova *et al.* 2006, 2008; Aguilera *et al.* 2008; La Paz *et al.* 2010a), the overall constitution of the insert in nucleotide sequence outline (GMO detection method database; Dong *et al.* 2008) and the insertion sites of the transgenic insert within the plant's genome (Windels *et al.* 2001; Hernández *et al.* 2003; Babekova *et al.* 2008; and other studies). When discussing the qualitative GM detection and identification

methods via PCR, the principal aims are: (1) the screening and detection of GMO presence, (2) the GM construct-specific detection, and (3) the GM event identification (see figure 1). For the purposes of screening and detection *i.e.* checking for the presence of GM content, a transgenic element is targeted via selected qualitative or quantitative PCR method. In this case, the expected result is a product specific to the transgene allowing to detect the GM content in the genetic material (Querci *et al.* 2006). In most cases the screening and detection target is situated in the commonly used in biotechnology promoter or terminator regions of the transgene (Mendoza *et al.* 2006). This may simplify the task due to common promoters and terminators ordinarily used in transgene construction. The more precise methodology is used for construct-specific detection. In this case, a border-junction of two transgenic elements present in GM cassette is targeted and typically it allows to determine the nature of the synthetic insert *i.e.* what kind of event is present in the genetic material (Mäde *et al.* 2006). Even though there is a possibility to detect multiple events with the same type of transgene, the construct-specific detection allows the researcher to know the type of the synthetic construct that is present in the material. Finally, the most specific detection and identification method is via event-specific PCR application (EU-RL GMFF 2014; Directive 2001/18/EC). In this case, the target is a junction between synthetic cassette and the flanking sequence of plant DNA. This type of method allows to describe the species and the particular event that is being screened precisely due to the unique DNA fragment obtained via PCR-based assay. There is no existing likelihood for the detection of two distinct GM events via event-specific method solely because of the presence of the unique insertion site covering the cross-link of plant genomic DNA and synthetic DNA. It is in a regulatory framework for the GM crop developer to provide with a functional, sensitive and reliable event specific method for the assessment of application for authorization (Directive 2001/18/EC).

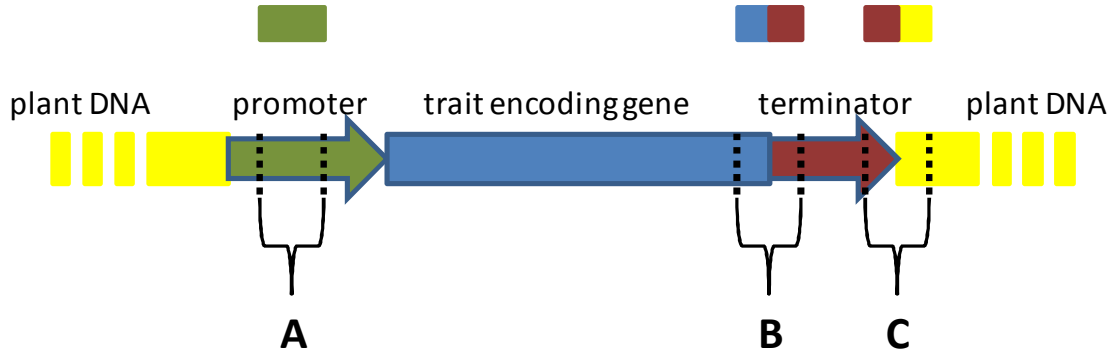

Figure 1: GM detection and identification via PCR. **(A)** detection of GM presence during screening will target a common GM construct element; **(B)** construct specific detection will target the junction of two GM construct elements which in turn lead to the knowledge of the nature of the GM construct; **(C)** GM event specific detection targets the junction between the synthetic DNA and plant genome to detect the exact species and GM event presence.

All the above mentioned PCR-based methods are related to detection, screening, identification and, if needed, quantification of GM contents. The use of these methods is functioning as the backbone of routine testing and analysis (EU-RL GMFF 2014). However, all the detection and/or identification related methods tend to require a lot of precise *a priori* information on the application site to carry out the experiment. Failure in having correct nucleotide sequence, having too little data on molecular structure of the transgene of plant DNA or such may lead in failure of obtaining reliable results (Querci *et al.* 2010). Moreover, the presence of GM content may even be overlooked *per se*. To avoid any failures it is therefore always a merit to have as much molecular-level data on the biotech crop as possible. Therefore, the other side of the transgenic insert investigation is the study of genetic element description, the expression profile, nucleotide sequence composition, transgenic insert localization within the genome, stability and copy number of the insert, and relative information retrieval which could be broadly generalized as the molecular characterization. This type of analysis is always aimed at the ways

of nucleotide sequence retrieval, the implementation of nucleotide sequence to shift the analysis towards needed procedures (as the aforementioned GMO screening), the genetic description of the transgene and its location on the genetic map using the nucleotide sequence. Self-evidently the common trait to all molecular characterization procedures is the detection and analysis of the nucleotide sequence.

As it is generally known, the nucleotide sequence information is commonly highly classified intellectual property protected by the GM event producer company; however it is not illicit neither to perform research on GM material nor to retrieve and examine the nucleotide sequence independently. As such, GM material for DNA extraction or readily prepared GM plant DNA can be purchased as Certified Reference Material (CRM) also known as standards for qualitative and quantitative check, and traceability in products (Regulation (EC) 1830/2003; IRMM 2014). In this doctorate thesis, I describe the experiments where the CRM was the preferential choice of DNA source due to the simplicity of purchase of such material as well as no restrictive requirements for using this DNA source for research purposes.

Apart from other existing ways of transgenic sequence assessment such as via Southern blotting, the most convenient and accepted approach is using PCR-based technologies. Furthermore, the specificity and accuracy of the PCR as an analytical method for GMO testing has not been surpassed by any other method (Holst-Jensen 2009). This fact is mostly based on the simplicity of the method selection, the ample knowledge of the technique, and the likelihood of carrying out such reactions in most of the molecular biology laboratories. As such, the molecular characterization methods described in this manuscript are solely PCR-based. Evidently, the PCR experiments can only be designed according to the initial data of the nucleotide sequence that is known. These experiments may either (1) be designed in the known region of the transgene (application of Conventional PCR, Real Time PCR); (2) cover the unknown

region between two known segments (Long Template PCR); or (3) directed to the unknown region of the nucleotide sequence (Genome Walking). Regardless of which is the area of research, all methods usually appear in the experimental project design at once like the control reactions for LT-PCR or Genome Walking are mostly Conventional amplification reactions.

2. LONG TEMPLATE PCR FOR CHARACTERIZATION OF RECOMBINANT DNA SEQUENCES IN GM PLANTS

2.1. Long template polymerase chain reaction (LT-PCR): a method to amplify long DNA strands

Since the year 1983 when Nobel Prize winner biochemist Dr. Kary Banks Mullis developed the technique to generate copies of the selected parts of DNA in several folds of magnitude, the so called polymerase chain reaction (PCR) became an essential method in the entire field of molecular biology. To this day the original or Conventional PCR with many variations of the method stands as the fundamental technique in research and diagnostics where DNA amplification is involved. Moreover, due to the constant development and alterations, it seems that the PCR is here to stay for a long time.

Since the first publication on PCR in year 1985 (Saiki *et al.* 1985), there were numerous studies concerning the optimization and application of the method. Only in year 1989 the PCR was included in laboratory manuals of standard molecular biology methods (Sambrook *et al.*, 1989) and over 800 research and review articles were published on PCR (Bloch 1991). It was not long before scientists questioned whether it was possible to improve the reaction for obtaining long DNA fragments (>3 Kb). In 1991, first research results for Long Template PCR (LT-PCR; also known as Long Range PCR) were published (Maga & Richardson 1991; Ponce & Nicol 1991).

While attempts to polymerize long strands of DNA were made to retrieve up to 15.6 Kb of lambda DNA (Kainz *et al.* 1992), the lack of proofreading 3'- 5' exonuclease activity in *Thermus aquaticus* (*Taq*) polymerase was the biggest disadvantage of the reaction. It was reported that a pre-mature drop of *Taq* was occurring in cases where the misincorporation of nucleotides appeared during the strand synthesis (Innis *et al.* 1988). This occurrence may be explained by *Taq* polymerase "error rate" increase after synthesizing up to 3 Kb DNA fragment which leads to misincorporated nucleotides and most likely end in dissociation of the enzyme molecule from the template (also known as chain termination). The limits of *Taq* polymerase have therefore shown that an additional element was required for the synthesis of long and accurate DNA fragments. It was not long before a proofreading enzyme from *Pyrococcus furiosus* (*Pfu*) was blended with modified N-terminal deletion mutant *Taq* DNA polymerase (Klentaq1) to obtain up to 35 Kb long lambda DNA fragment (Barnes 1994). In his publication, Barnes addresses several problems of the PCR methodology as the dependency of heat-liable components from denaturation time and the enzyme drop effect at the depurination site of the amplification sequence. The same problems were also studied and well described in the "Long-distance PCR" manual written by Foord and Rose (1994) thus the latter as well as Barnes, could be rightfully called pioneers of LT-PCR.

Apart from the cleavage of misincorporated nucleotides using a proofreading enzyme in the reaction, the LT-PCR method usually has a different buffer constitution. Ohler and Rose (1992) made an early observation that synthesis of long DNA fragments might require higher concentration of tris(hydroxymethyl)aminomethane (Tris) to maintain a correct reaction pH during the prolonged heating periods. Lately there is a tendency to use non-ionic detergents in the buffer systems provided in commercial LT-PCR kits (see Additives section below). This additive is expected to stabilize *Taq*

polymerase and may also suppress the formation of secondary DNA structures (Gelfand 1989).

Although the well known manual "Molecular Cloning: A Laboratory Manual" by Sambrook and Russell (2001; third edition) and Barnes (1994) suggests that the oligonucleotide primers used in LT-PCR to be slightly longer (25 to 30 nt) to those used in conventional PCR, however this modification of the reaction protocol might not be strictly required when using the novel highly developed and optimized kits for LT-PCR available commercially. The concentration of primers (Foord and Rose, 1994) as well as the quantity of unique deoxynucleoside triphosphates (dNTP) and the metal ion (Mg^{2+} for magnesium dependent polymerases) do not differ from a standard PCR and have the same values if varied during the method optimization.

2.2. Additives

Differently from the simple amplification by Conventional PCR, the application of LT-PCR might require additional elements for the composition of the master mix. These elements generally called additives can be useful for many different processes such as a stabilization of the polymerase on the template DNA or protection of the enzyme in prolonged periods of high temperature. As the additives may act as stabilizing agents or enhancers of the reaction they must be used at very specific concentrations to avoid the inhibition of the reactions. However the effect of additives is very case specific, thus it has to be tested as a variable and always parallel to control reactions. Most commonly known additives are listed in the table1.

Table1: PCR additives that are known to aid the reaction process and/ or outcome

Additive	General functionality	Reference
Dimethyl- sulfoxide (DMSO)	Reduces secondary structures, particularly in GC rich templates; reduces primer melting temperature and primer dimmer formation	Winship 1989; Hung *et al.* 1990; Chakrabarti & Schutt 2001
Betaine monohydrate	Improves DNA template melting by removing the base constitution dependency; increases yield through increased PCR efficiency	Rees *et al.* 1993; Henke *et al.* 1997; Zhang *et al.* 2009
Non- ionic detergents (Triton X, Tween 20 or Nonidet P-40)	Stabilize *Taq* polymerase and may suppress the formation of secondary structures	Bachmann *et al.* 1990; Katzman 1993.
Bovine Serum Albumin (BSA)	Used to amplify strongly damaged DNA or template that contains inhibitors	Pääbo *et al.* 1988; Kreader 1996; Giambernardi *et al.* 1998; Nagai *et al.* 1998
Trehalose	Lowers the DNA melting temperature and supports thermal stabilization of *Taq* polymerase	Spiess *et al.* 2004
Formamide	Reduces non-specific priming and increases amplification specificity and efficiency	Sarkar *et al.* 1990
7-deaza-2′-deoxyguanosine	When used at a particular ratio with dGDP, aids the amplification	Mizusawa *et al.* 1986

(dC7GTP)	of templates with very stable secondary structures (85% GC content)	
Glycerol	Reduces secondary structures; increases PCR efficiency, especially in high temperatures	Varadaraj & Skinner 1994; Nagai *et al.* 1998
Dithiothreitol (DTT)	Increases PCR specificity and product yield	Nagai *et al.* 1998; Ralser *et al.* 2006
Tetramethylammonium chloride (TMAC)	Reduces or eliminates non-specific priming and DNA-RNA mismatches; improves stringency in hybridization reactions	Wood *et al.* 1985; Jacobs *et al.* 1988; Hung *et al.* 1990
Carbon Nanotubes (CNT)	Specificity enhancement; potential yield increase	Zhang *et al.* 2008

Even though additives as trehalose, 7-deaza-2'-deoxyguanosine (dC7GTP), formamide and carbon nananotubes (CNT) have not been well investigated, published results of the studies have shown that they offer a possibility to be useful in the PCR reactions. Most of the enhancers listed in table1 were initially tested on Conventional PCR and some on Real-Time PCR, however due to the prolonged thermocycling regime, the need to use additives in LT-PCR increases. In any case, when used, the additives should be first carefully empirically tested with the amplification protocol and particular template and primers. This caution must be taken for avoiding the inhibitory effect from the additive.

2.3. The use of LT-PCR

In comparison to the Conventional PCR covering the large variety of fields of application, the LT-PCR would more likely to become the method of choice in a smaller range of molecular biology manipulations such as the trials of the integrity of selected nucleotide sequence in the genome, cloning and analysis of full genes, characterization of cloned sequences, construction of large contiguous sequences and other. Historically the LT-PCR was applied in diagnostics of the mutation that causes hereditary Gaucher disease (Sidransky *et al.* 1996); affirmation of gene *qnrA*1 carrying plasmid detection in infectious *Enterobacter cloacae* outbreak (Paauw *et al.* 2006); lately it was also used for typing the infectious *Enterococcus faecium* (Weisser *et al.* 2012). In addition, this method was described as a strategy for detecting the chromosomal modifications as inversions (Lombardi *et al.* 2005), translocations (Waller *et al.* 1999) or deletion (Grant & Blondal 2001); full allele mutations were also the target of LT-PCR (Saluto *et al.* 2005). This method is also known to be useful in identification of chromosomal diversity (Wandeler & Camenisch 2011) and even cytochome P450 genotyping using modified version Real-Time Long PCR (Müller *et al.* 2003).

The existing possibility to clone large fragments of DNA (Sambrook and Russell 2001) and the availability of commercial vectors that accommodate very large target DNA inserts (such as pJAZZ and pSmart vectors by Lucigen) enables the analysis of LT-PCR products even at low concentrations of purified fragment. The size of the DNA fragments that can be cloned into modern vectors is not a limiting factor anymore as these vectors are known to be capacious.

LT-PCR has also found applications within the research field of GMOs. Due to the fact that LT-PCR is expected to produce a long uninterrupted DNA fragment, it can also serve as a method to characterize the DNA stability of transgenes and the contiguous state of a synthetic cassette *i.e.* in GMO

transgenic insert studies. Depending on the quantity of obtained product, the expected long fragment after application of LT-PCR allows the researcher to submit the amplicon for direct sequencing or continue with molecular cloning reactions. One of the better known applications of LT-PCR for GM molecular characterization in GMO analysis was described for unauthorized Bt10 maize event (Milcamps *et al.* 2009).

2.4. Carbon nanomaterial as an enhancer to LT-PCR

The increase of the demand in material science resulted that in the first decade of XXI century, many novel materials were introduced to molecular biology techniques which promptly became additives for enhancement or inhibition of molecular processes. In the last two decades, the increasing interest in material science has uncovered a number of novel substances to be used in different scientific fields. As such, molecular biology techniques faced yet another potential upgrade possibility following a large amount of research and testing to be done whether it would be the enhancement or inhibition of molecular processes with the addition of the nanomaterials of which particularly the graphene allotropes. As a novel prospective, the use of a nanomaterial called carbon nanotubes (CNTs) in PCR has emerged.

A CNT is a cylindrical formation of graphitic carbon molecules to create a tube like nanoparticle structure. Together with fullerenes (round) and graphite (sheet), the CNTs (cylinder) are one of three known graphene family allotropes (Kroto *et al.* 1985; Iijima 1991; Novoselov *et al.* 2004). Nanotubes are divided into 3 separate groups depending on the layer quantity:

- Single Walled Carbon Nanotubes (SWCNT) have one hollow tube like (cylinder) structure;
- Double Walled Carbon Nanotubes (DWCNT) have 2 layers: an inner and outer hollow cylinders (Endo *et al.* 2005);

- Multi Walled Carbon Nanotubes (MWCNT) have the *Russian doll* model structure of hollow concentric cylinders inside another (Zhou *et al.* 1994).

As for the chemical properties of CNTs they may be non-functionalized and functionalized to add other properties to the structure. In the case of multiple layers, the outer layer is selectively functionalized (Endo *et al.* 2010) Most common functionalization contain the addition of carboxyl (-COOH), hydroxyl (-OH) and aminic ($-NH_2$) group to the nanotube (Sun *et al.* 2002).

As being very hydrophobic (Li *et al.* 2005; Manivannan *et al.* 2009), CNTs cause a significant problem in putting them into aqueous solutions, therefore 2 major drawbacks occur:

1. Dispersion of nanotubes in aqueous solutions as well as deagglomeration (immediately after the treatment the nanotubes tend to form new structures *i.e.* agglomerates in between);
2. Obtaining stable working solutions with no agglomerates of CNTs.

Numerous organic solvents (Dalton *et al.* 2000; Ortiz-Acevedo *et al.* 2005; Li *et al.* 2005; Manivannan *et al.* 2009) and acids (Ramesh *et al.* 2004) are known to be good solvents for nanotubes but cannot be used for molecular biology applications. As informed by Cheap Tubes Inc. (USA) (personal communication with Dr. Mike Foley), the more common approach for having well dispersed nanotubes in a solution is an addition of a dispersing agent (surfactant) and using water as a solvent. Known surfactants include polyvinyl pyrrolidone (PVP), sodium dodecyl benzene sulfonate (SDBS), polyvinyl alcohol (PVA) or Triton X100. It is also known that good dispersion with no nanotube agglomerates may be achieved using dimethyl sulfoxide (DMSO) as a solvent. Any of the above described chemical components however have to be used through addition of a required ultrasonication step to disperse the agglomerates of the tubes. Upon the request, the company Cheap Tubes Inc

(USA) has also provided us the information on ultrasonication protocols necessary to obtain working solutions of CNTs. Dispersion of MWCNTs require a 2-8 min sonication preferably with probe style ultrasonics machine with an interruption of 10 s every 30 s in a high ultrasound amplitude. However SWCNTs require a constant sonication for 20 min at 40% amplitude to break Van der Waals bonds that are responsible for agglomerate formation.

In the early study of Cui *et al.* (2003), applications of Single Walled Carbon Nanotubes (SWCNT; 2 nm diameter) to a Conventional PCR reaction were described. The study thoroughly describes the significant yield increase of a short *BRCAA*1 gene fragment while applying different concentrations of nanoparticles to the reactions. Moreover, the study indicated that abovementioned nanoparticles SWCNT can completely substitute magnesium ions of the PCR reaction (Cui *et al.* 2003). A later study by Yi *et al.* (2006) analyzed the effect of the polymerase pre-treatment with SWCNT and Multi Walled Carbon Nanotubes (MWCNT). The experiments yet again suggest to an increased product yield when the reactions were altered by nanoparticles however this finding was not related to the objectives of the study. In 2007, Zhang *et al.* raised an objective to improve the results of multiple re-amplification reactions using carbon nanopowder suspension (CNP) *i.e.* the mixture of CNTs. The research team has proven that the reactions with the same primers on previously obtained product, gain unspecific products after the 4^{th} re-amplification round and result in agarose-gel smearing after 5-6^{th} rounds, however the product appeared to be easily re-amplified and a well-defined band on gel could be obtained when using CNPs in the reactions. The aid of these nanoparticles to stabilize the reaction for as many as six re-amplification rounds showed the potential of nanomaterial in molecular biology methods (Zhang *et al.* 2007).

However it was not before the year 2008 that nanoparticles were tested on LT-PCR. After the experiments with CNPs, Zhang *et al.* (2008) published the

results of an LT-PCR efficiency study when applying SWCNT and MWCNT to the reactions. With this study Zhang *et al.* showed how a familiar LT-PCR problem as the PCR product smearing on agarose gel, can be avoided when applying various concentrations of CNTs. These investigators successfully obtained a 14 Kb fragment from lambda DNA (Zhang *et al.* 2008).

3. GENOME WALKING STRATEGIES FOR OBTAINING UNKNOWN DNA OF TRANSGENIC INSERT/ FLANKING DNA SEQUENCES IN PLANTS

3.1. Genome Walking

Genome Walking is a molecular biology technique for the identification of genomic segments adjacent to known sequences. Genome Walking is a general term to refer to a number of methods that have been developed over the last 20 years (Leoni *et al.* 2011). Isolation and analysis techniques however may entail labor-intensive library constructions and subsequent steps of DNA sequencing. In order to simplify this process, PCR dependent techniques have been considered to allow the retrieval of unknown DNA at a smaller scale by focusing on a specific gene of interest. Several review publications listing and describing genome walking methods are available in literature, such as Hui *et al.* (1998), Tanooka & Fujishima (2009), Kotik (2009), and Leoni *et al.* (2011).

Various techniques are available which differ in terms of isolation and selection of the desired fragment such as genomic digestion (either enzymatic or chemically induced) followed by adherence of adapters (cassettes) which serve as anchors for designed primers (anchor PCR, vectorette PCR, splinkerette PCR and other methods) (Theuns *et al.* 2002; Arnold & Hodgson 1991; Devon *et al.* 1995). The combination of gene specific primers with the "adapter" primers allows the amplification of specific DNA fragments flanking the gene of interest. The successful retrieval of the DNA fragments depends on

a number of factors. One of the major limiting steps in the aforementioned techniques is the need of digesting the genome prior to the PCR amplification. Successful isolation of the unknown DNA of interest is only possible when a restriction site is present at a suitable distance to allow a PCR reaction to proceed. Moreover an efficient ligation of an adapter to a protruding or blunt-end DNA strand is needed. Limitations of the method due to specificity issues are often reported. Another technique that involves genomic digestion but is not dependent on the ligation of adapters is Inverse PCR, which entails self-ligation (Ochman *et al.* 1988). Due to the fact that Inverse PCR was described as a method for obtaining unknown sequences as early as 1988, it may be identified as the pioneering genome walking method.

Literature reports on a few methods that do not require the step of genomic restriction, namely: amplification between an *Alu* consensus sequence and a known sequence (novel *Alu* PCR) (Puskás *et al.* 1994); an amplification based on a combination of sequence specific primers with nonspecific walking primers or (partially) degenerated primers (targeted gene-walking, (Parker *et al.* 1991); thermal asymmetric interlaced PCR, (Liu *et al.* 1993); SiteFinding PCR, (Tan *et al.* 2005)); introduction of unique walker primer binding sites through the rolling circle mode of DNA synthesis (Reddy *et al.* 2008). Rudi *et al.* (1999) described a restriction independent genome walking method that includes as a first and important step a single strand elongation starting with a gene specific primer, which is a methodology normally applied in RNA research. This constitutes the first step for the isolation of cDNA (5' Rapid Amplification of cDNA Ends, also called 5' RACE). It was applied for the isolation of both upstream and downstream genomic regions of a tRNA *Leu* gene from a cynobacterial genome (*Microcystis auruginosa*) (Rudi *et al.* 1999). It has recently been described for a plant genome (spinach) to retrieve flanking sequence of the cDNA coding region *Lhcb1.1* gene (Leoni *et al.* 2008).

Lastly, some molecular biology suppliers (Sigma, BIO S&T, Seegene) have genome walking kits placed on the market. However very few analyses are

published on the application of these commercially available kits (study of Cullen *et al.* 2011). Given the difficult nature of the application of the technology, where adjustments and optimizations of the reactions are needed in many cases, it is not improbable that the application of such kits may be limiting.

3.2. The classification of existing Genome Walking methods

The development and optimization of genome walking methods is rapid at the moment, thus the reader might get lost in a large amount of titles and data. Even though four aforementioned authors have tried to introduce a classification to the available genome walking methods, there are still several methods to be defined and classified and some might have names diversified by different authors. Although not comprising all the available methods and containing duplications, the categories of the methods by Leoni *et al.* (2011) are the best at the moment. The author separates three groups of genome walking methods: restriction-based (R-GW), primer-based (P-GW) and extension-based (E-GW) with subgroups as Inverted PCR, Cassette PCR and others. This scheme is principally correct, but the classification is based on the first (and sometimes conditioning) step of the method (Leoni *et al.* 2011).

At the onset of our study, and in order to define the best genome walking method for our analysis, we classified the genome walking methods in a distinct system, that contains two large groups and two large subgroups: restriction dependent/ independent methods and methods that require or not the adapter (anchor, linker, homopolymeric tail, oligonucleotide repeat) ligation throughout the protocol. The tables 2 and 3 below provide an overview of our classification (see also figure 2 for a schematic overview).

Figure 2: The variety of the Genome Walking methods

3.2.1. Adapter ligation-mediated Genome Walking

The first group of methods is the methods that require ligation or formation (synthesis, secondary structure creation or other) of a small single or double stranded adapter at the end of the unknown DNA sequence that is targeted to be recovered. Throughout the literature, adapters are known also by synonyms such as anchors or linkers, and some methods are subsequently named on the basis of atypical adapters such as slinkerettes and vectorettes. A distinct type of adapter is also a homopolymeric tail (usually poly(dA) or poly(dC)). Frequently the addition of such adapter is followed with the amplification using a deoxyoligothymidine or deoxyoligoguanidine primer. The first sub-group of the adapter-requiring methods includes enzymatic digestion as an initial step and/or throughout the genome walking application. The methods that have no restriction step throughout genome walking have also such step but rather only to aid the cloning (at the end of the genome walking). The latter ones are categorized as restriction independent methods.

Table 2: An overview of genome walking via adapter ligation method application

ADAPTER LIGATION METHODS	
Enzymatic digestion dependent methods	
Method	References
Anchor PCR (Adapter PCR; Cassette PCR)	Theuns *et al.* 2002; Zheng *et al.* 2001; Rishi *et al.* 2004; Perez-Hernandez *et al.* 2006; O'Malley *et al.* 2007; Rosati *et al.* 2008.
T-linker PCR	Yuanxin *et al.* 2003.
Vectorette PCR; Long Distance Vectorette PCR (LDV PCR)	Riley *et al.* 1990; Hengen 1995; Arnold & Hodgson 1991; Fenton *et al.* 2002 (LDV PCR); Laitinen *et al.*

	2004.
Splinkerette PCR (SplinkTA-PCR as modification)	Devon *et al.* 1995; Hengen 1995 (Splinkerette PCR); Yin & Largaespada 2007 (SplinkTA-PCR); Uren *et al.* 2009.
Oligo-cassette mediated PCR; Capture PCR (CPCR)	Rosenthal & Jones 1990; Lagerström *et al.* 1991; Nielsen *et al.* 2008.
Panhandle PCR	Jones & Winistorfer 1992; Felix *et al.* 1997; Raffini *et al.* 2002.
Boomerang DNA amplification (BDA)	Hengen 1995.
Single-specific-primer PCR (SSP-PCR)	Shyamala & Ames 1990, 1993.
Ligation-mediated single-sided PCR (LM-PCR)	Müller & Wold 1989, 1991, 2001; Tanguay *et al.* 1990; Fors *et al.* 1990; Garrity & Wold 1992.
Simplified oligo-cassette PCR; versatile cassette ligation-mediated PCR	Kilstrup & Kristiansen 2000; Nthangeni *et al.* 2005.
Adapter ligation-based PCR-mediated walking	Padegimas & Reichert 1998.
Step-down PCR	Zhang & Gurr 2000.
One-base excess adapter ligation PCR	Tanooka *et al.* 2008.
Straight-walk	Tsuchiya *et al.* 2009.
Blocked digestion-ligation-amplification (DLA)	Liu *et al.* 2009.
Template-blocking PCR	Bae & Sohn 2010.
TOPO® Vector-Ligation PCR (TVL-PCR)	Orcheski & Davis 2010.
Supported-PCR	Rudenko *et al.* 1993.
Rapid amplification of genomic ends (RAGE)	Cormack & Somssich 1997.

Restriction site extension PCR (RSE-PCR)	Ji & Braam 2010.
Enzymatic digestion independent methods	
Site Finding PCR	Tan *et al.* 2005
Rudy *et al.* method; 5'RACE based genome walking	Rudy *et al.* 1999; Leoni *et al.* 2008, 2010
High-throughput genome-walking	Reddy *et al.* 2008.
Long-distance genome walking via long and accurate PCR	Min & Powell 1998.
Long template rapid amplification of genomic DNA ends (LT-RADE)	Spalinskas *et al.* 2012, 2013
Flanking-sequence exponential anchored-PCR (FLEA-PCR)	Pule *et al.* 2008.

The first method listed in the table 2 is a conventional method of genome walking called anchor PCR. This method was thoroughly described by Theuns *et al.* (2001) and Zheng *et al.* (2001) and comprises of DNA isolation and digestion followed by adapter ligation and adapter-PCR with isotope labeled primer. O'Malley et al (2007) has introduced a selective step in this method: the modification discards the radioactive labeling step and instead features adapter with an adapter-primer sequence flipped reverse and complement. This guarantees the specificity of the reaction because the PCR may start only when the first cycle creates the primer annealing site. None of the other method variations by Rishi *et al.* (2004), Perez-Hernandez *et al.* (2006) and Rosati *et al.* (2008) features a labeling step but introduce simple steps for an increasing reaction specificity like the nested-PCR in many cases.

T-linker PCR by Yuanxin *et al.* (2003) features a step of protruding 3' poly(dT) overhang addition prior to enzymatic restriction after which a single PCR step is carried out to add an additional A at the restriction site. An adapter with protruding T then may be ligated at the restriction site and two nested PCRs are carried out to obtain a specific product. A successful amplification of the genomic sequences flanking multi *S8* genes in the S8-1 rice mutant by T-linker PCR is presented in the study.

The Vectorette PCR introduces a new type of adapter, having a mismatch between the double stranded oligonucleotides. One of the strands carries the reverse complement sequence of the vectorette primer and thus after the adapter ligation, the first step entails DNA synthesis starting from the vectorette primer binding site. This method might have served as the basis for aforementioned O'Malley *et al.* (2007) method. Splinkerette PCR introduced by Devon *et al.* (1995) and later modified by Yin (2007) has a hairpin-structure as an adapter. Otherwise, the procedure resembles the vectorette genome walking.

The method named Oligo-cassette mediated PCR by Rosenthal & Jones (1990) is very similar to the anchor PCR method by Theuns *et al.* and has introduced a biotin-tagged genome specific primer (GSP) instead of radioactive labeling. The method of Rosenthal & Jones was renamed to Capture PCR by Langerstrom *et al.* (1991) to emphasize the step where streptavidin beads catch biotin labeled amplified DNA fragments.

Panhandle PCR (Jones & Winistorfer 1992) is a method that has identical sequence adapters annealed to the both sides of a DNA fragment. In room temperature, the fragment creates a hairpin structure between the ends resembling the actual *pan handle*, however the denaturation and primer annealing steps allow the single primer to perform amplification from both sides (the single primer acts as forward and reverse).

The method of Boomerang DNA amplification was never tested and has no research data, however Hengen (1995) introduced this method in a benchmark article as a potential genome walking technique. This method involves an adapter with non-complimentary midsection annealing to make a circular structure so the amplification primer would return to its start position when the extension step is done.

Leoni *et al.* (2011) classifies Mueller & Wold (1989) and Fors *et al.* (1990) method LM-PCR in separate groups however the method of Fors *et al.* does not introduce any actual modification. LM-PCR is a method that contains the ligation of the unidirectional (staggered) linker of two different sized complementary strands. The shorter strand has the 5' end dephosphorylated and serves only to ligate the linker to the end of cleaved DNA. The short strand is discarded during the DNA denaturation and primer annealing initiates on the longer adapter strand.

Regardless of the similarity in method titles (Oligo-cassette mediated PCR and Simplified Oligo-Cassette mediated PCR) of Rosenthal & Jones (1990) and Kilstrup & Kristiansen (2000) the latter does not implement a biotin labeled product. Instead, the method Simplified oligo-cassette PCR uses two dephosphorylated primers that target two different overhang sequences of digested DNA. Two different cassettes carrying three digestion sites are then annealed and rounds of amplification are performed.

The method of Nthangeni *et al.* (2005) is also listed in this category for having a dephosphorylated cassette that prevents self-ligation. In fact, the author describes the creation of so called "universal cassette" that provides a 200 bp sequence, which is used to design a variety of cassette-specific primers. The method however is greatly based on the Single specific primer-PCR of Shyamala & Ames (1990) (also Shyamala & Ames (1993)). The ligation

cassette called versatile cassette and described by Nthangeni *et al.* carries five different restriction sites (*EcoR1*, *SacI*, *Kpn1*, *SmaI/XmaI*, and *BamHI*) at one side and five restriction sites (*HindIII*, *XhoI*, *PstI*, *EcoRV*, and *XbaI*) as well as reverse and forward primer sites at the other side.

In the year 1998 Padegimas & Reichert have introduced another modified adapter ligation based PCR using uniquely blocked adaptors (hairpin formation as in vectorette and splinkerette methods) along with removal of unligated genomic DNA by exonuclease III digestion. Authors claim that removing unligated fragments via digestion, the specificity of PCR-mediated walking increases greatly. By using a blend of LT-PCR polymerases (as *Tth* and VENT) up to 6 Kb fragments were obtained in the study.

Step-down PCR described by Zhang & Gurr (2000) deals with a blend of restriction enzymes (up to 6) digesting DNA simultaneously for obtaining short DNA fragments to ligate to the adapters. The latter also carry restriction sites to avoid self-ligation. Even though this method is described as efficient and fast, the materials required to perform the reaction of step-down PCR seem unusually numerous. Tanooka *et al.* (2008) brings up another simple modification to the adapter PCR, however the complexity of the method allows it to be listed separately from adapter PCR. One-base excess adapter ligation PCR involves an adapter with one extra base at the 5' overhang so that T4 DNA ligase cuts the 3' end base of restriction site present in the genomic DNA and incorporates the phosphorylated overhang base of the adapter. The method called Straight Walk by Tsuchiya *et al.* (2009) is not a new method, rather a simple adapter PCR (ligation-mediated PCR) with improved specificity and sensitivity. The author describes the usage of delta-G values (free energy) utilized as a criterion to design an adapter with high binding efficiency and carrying highly specific and sensitive adapter primers.

The method Blocked DLA by Liu *et al.* (2009) employs two single stranded adapters with further ddNTP extension of 3' to avoid strand filling in primary PCR and allowing the overhang to be amplified carrying the adapter primer site. A similar technique called Template-blocking PCR was described by Bae & Sohn (2010), however a major modification was introduced. Instead of adding ddNTP to the adapter sequence, the authors present polymerization limitation at the 3' end of restriction product via the same ddNTP addition.

A simple, yet well-thought method is the Single-specific-primer PCR (SSP-PCR) introduced by Shyamala & Ames (1990) where digested DNA is cloned into a generic vector and then primed with a known GSP and vector primer to obtain the unknown flanking sequence. This method is very dependent on capturing the right fragment and ligating it with the vector which may be seen as a limiting factor, however in the year 2010 Orcheski & Davis presented their version of the method called TVL-PCR in which pCR4-TOPO vector is employed as linker. Again the simplicity of the method lies in the chimeric molecule creation that has the backbone of a well-known vector and randomly captured digested DNA fragments.

The method by Rudenko *et al.* (1993) is called Supported-PCR. After enzymatic restriction of the genome, an initial step of single-cycle PCR with gene specific primer and biotin-labeled dUTP is performed. Then this product is retrieved while also concentrating it in several orders of magnitude. A linker is then ligated to a poly(dA) tail and this product serves as a template for conventional PCR. Up to 1.6 Kb fragments were isolated by using this method.

In the year 1997 Cormack & Somssich have introduced a method RAGE which involves enzymatic digestion followed by polyadenilation of 3' with terminal deoxynuleotidil transferase (TdT) and an adapter PCR using an anchored deoxyoligothymidine primer.

The last genome walking method in our enzymatic digestion and adapter dependent method list is RSE-PCR by Ji & Braam (2010). This method has PCR-mediated adapter incorporation on the digested DNA via a short primary extension step. This method is based on polymerase 5-3' polymerization activity to effectively fill the adapter sequence brought by the primer that anneals on the sticky end of the DNA fragment left by restriction enzyme.

The second sub-group of methods of genome walking are the methods that do not require enzymatic digestion throughout the entire procedure. These methods still involve some kind of adapter ligation or formation, however the starting template DNA is typically high molecular weight non-digested gDNA.

The first method not requiring the enzymatic digestion is Site Finding PCR (Tan *et al.* 2005). This method is based on a long adapter (over 50 bp) containing two primer sequences, a restriction site (to aid the cloning after the genome walking) linked to four random nucleotide sequence at the 3' end. This adapter is called a Site Finder and it is randomly annealing to the DNA template via four unique bases at the end of adapter.

In 1998 and 1999, two methods were brought up by Min & Powell (1998) and Rudy *et al.* (1999). These methods were based on poly(dC) tail addition on the firstly obtained ssDNA product via single primer extension. In the case of Rudy *et al.*, a deoxyoligoguanidine primer was used; Min & Powell had a restriction site of *EcoRI* attached to poly(dG) stretch of a primer. These reactions were unfinished and highly dependent on homopolymeric tail priming and semi-nested PCR (if available). Leoni *et al.* has further modified this method in 2008 by using a primer from the 5'RACE kit by Invitrogen. The complexity of that primer, gave the possibility to prime the homopolymeric single nucleotide stretch much more effectively. However, from our observations, the full potential of the 5'RACE based method was not reached. An odd application of the same AAP primer in two reaction steps (when

adapter is already created) and the usage of several product cleaning steps as well as employing relatively inefficient PCR enzymes, exposed the method as prone to more advantageous improvements. In this doctoral dissertation we have presented a modified version of this method merging it with the LT-PCR, optimizing the resulting method and giving it the title of LT-RADE (Spalinskas *et al.* 2012, 2013; see Publications section at the end of the manuscript for these references).

Another method of the current category is high-throughput genome walking by Reddy *et al.* (2008) which employs unique walker primer binding sites set up through the rolling circle mode of DNA synthesis after annealing the partially degenerate primers to the denatured genomic DNA.

The last method on our list is by Pule *et al.* (2008) called FLEA-PCR where linear PCR is performed on gDNA with a single biotinylated primer that anneals close to the 5' end of the U3 region of the long tandem repeats (LTRs; the retroviral promoter sites). This PCR generates random length ssDNA fragments which are constituted of 5' part of sequence complementary to LTR and a fragment of flanking genomic sequence.

3.2.2. Genome Walking without adapter ligation

The second large group of the methods is those not requiring any adapter ligation. Most of these methods involve an element called degenerate primers i.e. a mixture of similar but not identical primers. As such, most of adapter ligation/ synthesis independent methods also do not require any enzymatic digestion. However, leaving the degenerate primers aside, there is one method that is enzymatic digestion dependent and not requiring an adapter. This method essentially was pioneering the early genome walking and it is called Inverse-PCR (Inverted PCR or IPCR) (Triglia *et al.* 1988; Ochman *et al.* 1988). Apart from the modified versions, the IPCR is a stand-alone method in

aforementioned sub-group. The lack of the adapters in this method is substituted by a step of circularization of the digested product. Thus the known DNA fragment flanking the unknown regions is circularized and reopened at the known DNA sequence making two arms that eventually act as priming sites for obtaining the unknown sequence. Many applications of this method exist (see table 3 for some references) including modified versions of IPCR as potential improvements.

Table 3: An overview of genome walking via adapter ligation independent method application

METHODS WITHOUT ADAPTER LIGATION	
Enzymatic digestion dependent methods	
Method	References
Inverse PCR (Inverted PCR; IPCR)	Triglia *et al.* 1988; Ochman *et al.* 1988; Tsuei *et al.* 1994; Arendse *et al.* 1999; Behura *et al.* 2001; *and other sources.*
Modified IPCR (Long range IPCR; Bridged IPCR; Rolling circle IPCR)	Benkel & Fong 1996; Kohda & Taira 2000; Tsaftaris *et al.* 2010.
Enzymatic digestion independent methods	
Targeted Gene-Walking PCR (also referred as PCR with random primers)	Parker *et al.* 1991.
Thermal asymmetric interlaced PCR (TAIL-PCR); high-efficiency TAIL-PCR	Liu *et al.* 1993, 1995a, 1995b, 2007; Singer & Burke 2003; Huang *et al.* 2010; *and other sources.*
Two-step Walking	Pilhofer *et al.* 2007.
Self-Formed Adaptor PCR (SEFA PCR)	Wang *et al.* 2007.
Universal Fast Walking (UFW);	Myrick *et al.* 2002; Walser *et al.*

Lariat-dependent nested PCR (LaNe PCR)	2006; Park 2011.
Uneven PCR	Chen & Wu 1997.
Restriction-site PCR	Sarkar *et al.* 1993.
Alu-PCR; Novel *Alu* PCR	Nelson *et al.* 1989; Puskás *et al.* 1994.
Sequential Hybrid Primer PCR (SHP-PCR)	Martin-Harris *et al.* 2010.
Nested-PCR based walking	Guo & Xiong 2006.
Single Primer Amplification	Hermann *et al.* 2000.
Semi-Random PCR Chromosome Walking (SRPCW)	Ge & Charon 1997, Ge *et al.* 1997.
Touchdown-PCR Based Walking	Levano-Garcia *et al.* 2005.
SD-PCR (Shine–Dalgarno sequence based PCR)	Ping *et al.* 2008.
Locus-Finding PCR (LF-PCR)	Thirulogachandar *et al.* 2011.

The table 3 above lists methods where the adapter is redundant yet it is still required to perform the enzymatic restriction. The first method is a simple Targeted Gene-Walking PCR (Parker *et al.* 1991) where a gene specific primer (GSP) is used in a mixture of random primers that anneal randomly, similar to studies in polymorphism research. Following this simple method is a popular genome walking protocol in applied science: TAIL-PCR. This method was described by Liu *et al.* (1993, 1995a, 1995b and 2007) and it involves three GSPs primed in pairs with the same short arbitrary primer (lower Tm) annealing through a special low stringency cycle. The method TAIL-PCR is thus largely based on the ability of primers to bind unspecifically in the unknown fragment at low stringency conditions. A simple yet very basic method for genome walking is Two-Step Walking (Pilhofer *et al.* 2007) that resembles TAIL-PCR procedures. This method is based on the usage of a

single GSP to produce a set of ssDNA fragments that subsequently serve as a template for unspecific priming with the same GSP using lowered annealing temperature and creating dsDNA product. These different length fragments are then directly sequenced using a nested primer in the reactions with stringent conditions.

Self-Formed Adaptor PCR (SEFA-PCR) presented by Wang *et al.* (2007) is another TAIL-PCR related method which uses a partially degenerate GSP in a sequence of four GSPs. The degenerate primer is first used with a low annealing temperature to bind close-by the sequence of four GSPs at an unknown DNA region. After this manipulation, the first GSP is used at the increased annealing temperature to produce a set of ssDNA fragments containing all GSP sequence at one side and a reversed complement degenerate GSP sequence at the other end. Self priming is then performed making a loop structure between the reverse complement degenerate GSP extension and the actual priming site of the same primer in the sequence of four GSPs. This way an adapter is created using only the original sequence.

Universal Fast Walk (UFW) is a different restriction and ligation independent method developed by Myrick & Gelbart (2002). This method starts with an extension of the first GSP, then, random annealing of the second GSP which has a complex structure: ten random bases at its 3' end but having a specific 5' part based on a motif from the known end of the first strand. When this primer is annealed, it creates a branched structure which is then trimmed on the expense of a template and repaired using the random bases of the primer as a correct template. This allows a first strand lariat with aforementioned motif to form. The so called lariat structure is a loop-stem-like structure between one end of DNA strand and its middle part. It is very much alike to a loop structure that forms in SEFA-PCR as a final point. Some authors compare the loop-like structure of SEFA-PCR and UFW to the "*pan handle*" concept (see above) (Park 2011). In fact, a book chapter on Lariat-dependent nested PCR (LaNe

PCR) by Park (2011) presents a possibility to construct this loop like structure in the process of other genome walking methods (as it is in method entitled RAGE).

The principle of the method Uneven PCR (Chen & Wu 1997) is to pair a GSP with a short (10 nt) arbitrary oligonucleotide that binds to many sites of the template. If the distance between a GSP and an arbitrary primer is within the radius of the polymerase's processivity, the fragment is effectively amplified. In favor of the desired PCR product production, two different annealing temperatures are used in consecutive PCR steps. The specificity of the method is further increased by a subsequent nested PCR step.

In 1993 Sarkar *et al.* used restriction sites in a close proximity to known DNA as priming sites for PCR. The research team named the method Restriction-Site PCR and the simplicity of the protocol lies in the possibility to find a digestion site and prime it with a so called restriction site oligonucleotide (RSO), coupling it with a GSP. A subsequent nested PCR step is also applied in the end for specificity increase.

A stand-out method for walking on human DNA is *Alu*-PCR developed by Nelson *et al.* (1989) and modified by Puskás *et al.* (1994). This method is based on a highly repetitive *Alu* sequence which is primed in a pair with a GSP. In an early version of the method, *Alu-Alu* nonspecific priming was unavoidable thus a modification was introduced by adding a chemically modified analogue of the *Alu* primer which is not extended by a polymerase.

Martin-Harris *et al.* (2010) presented a simple and robust method for genomic walking entitled Sequential Hybrid Primer PCR (SHP-PCR). The protocol of this method includes three PCR steps with three GSPs at the known DNA sequence. The primary PCR is performed with GSP1 paired with a degenerate primer which carries a tag. The resulting product is then primed with a GSP2

paired with a reverse complement tag sequence. The latter tag primer carries tag2 sequence adjacent which again serves as a new priming site in the next round of PCR (with GSP3). The tag2 primer carries a tag3 sequence which afterward may be used for sequencing. The produced fragment in this method is getting shorter at one side (as expected using nested primers), however is growing on the other side where tags are added (atypical in genome walking applications). The method has a very high specificity due to the repeated nested PCR steps. Due to the usage of a degenerate primer in the primary PCR, the tag sequences are not to be called adapters. The method by Martin-Harris *et al.* is a follow-up of a simpler protocol called Nested-PCR based walking (Guo & Xiong 2006) which has the same three GSPs paired with one single degenerate primer without tags. The same degenerate primer is used in all three rounds of PCR.

A method Single Primer Amplification (Hermann *et al.* 2000) employs two primers: a biothynilated one (B-primer) and a nested one (N-primer) in a close proximity of B-primer. The primary extension is performed with a B-primer with the following capturing of the product on streptavidin beads. After washing a mispriming PCR round at low stringency condititions is performed with N-primer thus creating another priming site for N-primer. The protocol then utilizes the same N-primer to prime from both sides of the template and to create a dsDNA product.

In 1997, Ge & Charon developed a method called Semi-Random PCR chromosome walking (SRPCW). This method is a modification of Targeted Gene-Walking PCR (Parker *et al.* 1991) to increase the specificity of the walking. The primary PCR is performed at low stringency conditions with a GSP and a random primer. All resulting products are immediately cloned into vector pGEM-T. Having a plasmid with a variety of non-specific products and, expectantly, the target fragment, the second GSP is used in a couple with a generic vector primer. This assures the specific selection of a target fragment.

The method Touchdown-PCR Based Walking by Levano-Garcia *et al.* (2005) is a simple fusion of the Touchdown PCR method with the usage of the degenerate primer in a couple with a GSP. The resulting product is afterwards directly sequenced using solely the specific primer. Another method SD-PCR (Ping *et al.* 2008) employs a long degenerate primer designed according the Shine–Dalgarno sequence of prokaryotic transcripts and paired with a gene specific primer. This method is based on the natural presence of 5' unique ribosome-binding sites called Shine–Dalgarno in prokaryotic high-level expression genes in a close proximity of the AUG codon.

Very recently, Thirulogachandar *et al.* (2011) presented another affinity-based genome walking method. It is a type of random-priming PCR where a set of different fragments is produced. The novelty of the method is an enriching technique to eliminate all the non-specific products via incorporation of capture primer binding region and biothynilated primer. After the randomly-primed PCR the biothynilated capture primer is used to purify and select only the desired product. The method is based on the principle that the primer will bind to its complementary strand and the resulting complex can be captured if it is marked with biotin.

CHAPTER II: OBJECT OF THE STUDY, MATERIALS AND METHODS

1. OBJECT OF THE STUDY

The following commercial trait containing GM events were studied throughout the preparation of this thesis: H7-1 sugar beet, Bt11 maize, MON810 maize, LLRICE62 rice, A2704-12 soybean, T45 rapeseed and LLCOTTON25 cotton.

2. MATERIALS

- H7-1 sugar beet Certified Reference Material (level 1), № ERM-BF419b was obtained from the Institute for Reference Materials and Measurements (IRMM), JRC, Belgium.
- H7-1 sugar beet seeds, dossier № CRL-VL-28/04 were received from KWS SAAT AG company.
- H7-1 Sugar beet leaf material, grown from the seeds described above.
- Bt11 maize Certified Reference Material, 5% GM content, № IRMM-412R-5 was obtained from the Institute for Reference Materials and Measurements (IRMM), JRC, Belgium.
- Bt11 maize leaf material, grown from the seeds (dossier № CRL-VL-12/05) provided by Syngenta company.
- MON810 maize Certified Reference Material, 5% GM content, № IRMM-412R-5 was obtained from the Institute for Reference Materials and Measurements (IRMM), JRC, Belgium.
- Lambda DNA *c*I857 *Sam7* isolated from infected *E. coli* strain W3350 was purchased from Promega, USA (cat.no. D150A).
- Multi Walled Carbon Nanotube (MWCNT) aqueous solution in 100% dimethyl sulfoxide (provided by Nanobiosciences Unit, IHCP, JRC, Italy; produced by Nanocyl s.a., Belgium).

- MON810 maize *var. Helen BT* seeds were provided by Instituto Nacional de Investigaciones Agrarias - Spanish Ministry of Science.
- Leaf tissue DNA of LLRICE62 rice (CRM 0306-I3^{+}) was purchased from American Oil Chemists' Society (AOCS).
- Leaf tissue DNA of T45 rapeseed (CRM 0208-A3^{+}) was purchased from AOCS.
- Leaf tissue DNA of A2704-12 soybean (CRM 0707-B4^{+}) was purchased from AOCS.
- Leaf tissue DNA of LLCOTTON25 cotton (CRM 0306-E) was purchased from AOCS.

3. METHODS

3.1. DNA extraction and quality control

DNA from Certified Reference Materials (H7-1 sugar beet, Bt11 maize and MON810 maize) and seeds (H7-1 sugar beet only) was extracted using the NucleoSpin Food DNA extraction kit by Macherey- Nagel (Düren, Germany). DNA from GM plants (H7-1 sugar beet and Bt11 maize) was extracted using DNeasy Plant Mini Kit by Qiagen (Hilden, Germany). DNA from 1,000 mg of MON810 maize leaf material was extracted using NucleoSpin Plant II Maxi kit (Macherey-Nagel, Germany) following the manufacturers protocol.
The quality of all extracted and purchased DNA was checked using gel-electrophoresis technique on agarose gels. The concentration of DNA was measured using PicoGreen dsDNA Assay kit by Invitrogen (USA) and Lambda DNA molecular size standards by Promega (USA).

3.2. Conventional PCR

A Conventional PCR method was used for primer and template check and/or control reactions in the experiments. Reactions were carried out using the

FastStart High Fidelity PCR System kit by Roche Applied Science (Mannheim, Germany). A standard PCR procedure described in the protocol of the kit was followed. A regular reaction is carried out in 50 µL total volume which contains 1X buffer with pre-titrated final concentration of 2 mM of $MgCl_2$, 0.2 mM of every dNTP in a mix, 2 U of *Taq* polymerase and 50 ng of DNA template. The concentration of primers was up to 100 pmoles per reaction. The thermal cycling protocol is presented in table 4. All thermal cycling was performed in Applied Biosystems GeneAmp PCR System 9700 with gold sample block module.

3.3. Long Template PCR

LT-PCR reactions were carried out using 2 different commercially available kits: Expand Long Template PCR system by Roche Applied Science (Mannheim, Germany) and Phusion™ Hot Start High-Fidelity DNA Polymerase (Finnzymes, Finland). The PCR procedure described in the protocol of Roche Applied Science kit was followed: amplification of 0.5-9 Kb fragments with System 1 and amplification of >12 Kb fragments with System 3. A typical reaction with Roche kit was performed in 25 µL and contained 1X PCR buffer, 0.2 mM of each dNTP, 1.5 mM of $MgCl_2$, 7.5 pmoles of each primer and 2.5 U of polymerase blend present in the kit: *Taq* and *Tgo*. In several cases, when needed, a reaction of a total volume of 50 µL was performed doubling-up all the reagent concentrations. The reactions with Finnzymes LT-PCR kit were performed in 25 µL total volumes containing 1X PCR buffer, 0.2 mM of each dNTP, 1.5 mM of $MgCl_2$, 12.5 pmoles of each primer and 0.5 U of Phusion polymerase. All LT-PCR experiments contained 200 ng of template DNA per reaction. The thermal cycling conditions are presented in table 4.

Table 4: thermal cycling protocols for conventional and long template PCR

Thermal cycling conditions						
	FastStart *Taq* (Roche)		*Taq*+*Tgo* (Roche)		Phusion (Finnzymes)	
Initial denaturation	94°C- 4 min		92°C- 2 min		98°C- 30 s	
Denaturation	94°C- 30 s	30 cycles	92°C- 30 s	30 cycles	98°C- 10 s	30 cycles
Primer annealing	60°C- 30 s		60-67°C- 30 s		60-74°C- 30 s	
Elongation	72°C- 1 min		68°C up to 12 min		72°C up to 7 min	
Final elongation	72°C- 7 min		68°C- 7 min		72°C- 5 min	

3.4. Rapid Amplification of Genomic DNA Ends (RADE and LT-RADE)

RADE includes several steps of the 5'RACE kit (Invitrogen) fused with conventional and nested PCR techniques and applied on gDNA templates as opposed to the original 5'RACE application on the cDNA. We introduced modifications to 5'RACE as pointed out below. The method consists of five main steps.

Single primer extension. 0.5 µg of genomic DNA was mixed with 60 pmoles of the first gene specific primer (GSP1) to obtain the single primer extension product. The reaction was carried out in a total volume of 100 µL in the presence of 0.4 mM of each dNTP and 7.5 U of *Taq* polymerase (FastStart *Taq* DNA polymerase, Roche, Germany). The thermal cycling conditions for the reaction are displayed in table 5.

Product purification. Product was purified using a S.N.A.P column purification kit included in the kit 5'RACE (Invitrogen).

Homopolymeric tailing. 10 µL of S.N.A.P purified DNA was used in a poly-dC tailing reaction catalyzed by template independent polymerase terminal deoxynucleotidyltransferase (TdT). A total volume of 25 µL contained 1X tailing buffer, 0.2 mM dCTP and TdT according to the protocol of the kit 5'RACE (Invitrogen).

Adapter PCR. 5 µL of the product with a poly-dC tail was used for PCR using the primers AAP (Invitrogen) and GSP2, each at a concentration of 20 pmoles. The reaction of a total volume of 50 µL contained 1X PCR buffer, 0.2 mM of each dNTP, 1.5 mM of $MgCl_2$, 2.5 U of *Taq* polymerase. The thermal cycling conditions for the reaction are summarized in table 5.

Nested PCR. A 100-fold dilution of the product was prepared in Tris-EDTA (TE) buffer and 5 µL of this dilution were used in a nested PCR reaction with AUAP (Invitrogen) and GSP3 primers. The thermocycling conditions and constitution of the reactions were the same as in the previous step except the concentrations of both primers were reduced to 10 pmoles.

LT-RADE: Application of a long template PCR within the RADE method was carried out as described above but with the following modifications. In all amplification steps, including the first step of single primer extension, *Taq* polymerase was substituted with the polymerase blend *Taq* + *Tgo* and the buffer replaced with buffers No.1 or No.3 (Expand Long Template PCR system (Roche, Germany)). In the single primer extension step, the elongation time of the reaction was set to 2 min (see table 5).

Table 5: thermal cycling protocols for PCR, LT-PCR or PCR-related (single primer extension) experiments as parts of RADE and LT-RADE methods

Thermal cycling conditions							
	RADE single primer extension			LT-RADE single primer extension			PCR and nested-PCR steps (RADE and LT-RADE)
Initial denaturation	94°C- 5 min			94°C- 4 min			94°C- 2 min
Denaturation	94°C- 1 min	35 cycles		94°C- 1 min	35 cycles		94°C- 1 min	35 cycles
Primer annealing	62°C- 1 min			62°C- 1 min			55°C- 1 min	
Elongation	72°C- 1 min			72°C- 2 min			72°C- 2 min	
Final elongation	72°C- 7 min			72°C- 7 min			72°C- 7 min	

3.5. Gel-electrophoresis

PCR products were analyzed on a 1% agarose gel containing the stain ethidium bromide which intercalates into DNA and allows monitoring DNA fragments under the UV light. The concentration of the dye in the ready-to-use agarose was 0.2 µg/mL (2 µL of 10 mg/mL stock solution to 100 mL of liquid agarose before running the electrophoresis). Visualization of the gels was performed with BioRad (USA) gel analyzing and documenting system. Four molecular size and/or mass standard ladders: GeneRuler^{TM} 100 bp Plus DNA Ladder; GeneRuler^{TM} 1 Kb DNA Ladder; GeneRuler^{TM} DNA Ladder Mix and MassRuler^{TM} Express Forward DNA Ladder Mix (Fermentas, Lithuania) were used to determine the size and/or concentration of the retrieved fragments.

3.6. Purification of the DNA

The products of the LT-PCR, intended for DNA sequencing, were analyzed on 1% agarose gels (TBE buffer, 5 V/cm) containing SYBR Safe DNA Gel Stain (Invitrogen, USA), as recommended in the protocol. This stain was chosen as it has very low or no mutagenic activity on DNA. The Safe Imager Blue-Light Transilluminator by Molecular Probes (Invitrogen Detection Technologies, USA) was used for visualization and isolation of the DNA fragments from gels. An amber filter was used with the transilluminator to make the DNA bands visible under blue light of approx 470 nm wave length.

Zymoclean Gel DNA Recovery Kit by Zymo Research (USA) was used to recover and purify DNA fragments from the agarose gel. According to the protocol, the kit is suitable for recovering DNA fragments up to 23 Kb with a relatively high column binding capacity of 5 μg of nucleic acids per membrane. The first elution from the membrane was performed with 10 μL of molecular biology grade water. One microliter of each retrieved fragment was run on an agarose gel for quality and quantity assessment.

3.7. DNA sequence analysis

All DNA sequencing was carried out by Microsynth GMBH in Balgach, Switzerland. The fragments for LT-PCR experiments were either sequenced in a direct mode (3.4 Kb fragment of H7-1 sugar beet and 1.7 Kb fragment of Bt11 maize) or cloned with BigEasy Long PCR Cloning Kit from Lucigen (Wisconsin, USA) (5.9 Kb fragment of Bt11 maize). All sequencing reactions were performed as bi-directional non-assembled primer walking reactions. Assembly of double stranded sequencing data and alignment analysis was performed with Lasergene 7 software suite (DNAstar, Madison, WI, USA), using in particular the tools EditSeq, SeqMan and MegAlign.

The fragments for RADE and LT-RADE reactions were either sequenced in a direct mode or cloned with a TOPO TA Cloning kit (Invitrogen) following the manufacturer's instructions.

3.8. Molecular cloning of the long DNA fragments

Molecular cloning of the long PCR fragments was performed with the BigEasy Long PCR Cloning Kit from Lucigen (Wisconsin, USA). Vector pJAZZ-OK Blunt was used for cloning a DNA fragment with blunt phosphorylated ends. LT-PCR reaction buffer was removed with DNA Clean and Concentrator kit from Zymo Research (USA) as the salty solution was not appropriate for the following steps. Kinase reaction was carried out prior to ligation a fragment with a vector to phosphorilize the ends of the fragment. Polynucleotide kinase (PNK) by Roche Applied Science (Mannheim, Germany) was used for the reaction. Additional gel-electrophoresis with SYBR Safe DNA Gel Stain (Invitrogen, USA) purification was performed to clean and concentrate only the phosphorilized target fragment. The purification from the gel step was performed by a DNA Recovery Kit from Zymo Research (USA). Ligation reaction was carried out as described in the protocol of the cloning kit. The resulting solution was introduced to BigEasy TSA electrocompetent cells with the use of Gene Pulser Xcell Electroporation System from BioRad (USA). The cells were recovered and plated on YT+Kan+X-gal+IPTG Agar medium with the concentrations of 30 mg/L kanamycin as the vector pJAZZ-OK has a resistance gene to kan; X-gal concentration of 20 mg/L and IPTG concentration to 1 mM. Additional selection with ampicillin was introduced adding 30 mg/L Amp solution to the agar medium as the bacteria strain has a resistance to Amp. Transformants were grown overnight and selection was made for bacteria colonies to work further. Additional re-plating of the selected colonies was performed to have a high yield of bacteria for colony PCR reaction and preparing the overnight culture. Colony PCR was performed with 2 primer pairs: 11GSFan2+pUCR3 and cry2F+cry2R in the Bt11 maize GM

insert. Regular PCR method described above was used with addition of 1% on SDS solution into the gel-electrophoresis loading dye. PCR results were analysed and overnight cultures were grown in liquid LB medium with 30 mg/L kanamycin. Plasmids were extracted the following day using NucleoSpin Plasmid kit from Macherey-Nagel (Düren, Germany) and prepared for sequencing determining the concentration and making dilutions. Prior to sending the plasmid preparations to the sequencing team, another verification analysis was carried out with restriction enzyme. As it is stated in the protocol, plasmid pJAZZ-OK is linear and releases the cloned fragment cutting with *NotI* restriction endonuclease producing two arm fragments and the target sequence. *NotI* was purchased from New England Biolabs (USA) and the reaction was carried out by a standard 1h incubation restriction protocol.

The BigEasy-TSA electrocompetent cells used for the molecular cloning of long PCR fragments has the following genotype descript: F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K *rps*L *nup*G λ- *ton*A *bla* (Amp^R) *sop*AB *tel*N *ant*A.

3.9. Molecular cloning of standard size fragments

TOPO TA Cloning kits contain pCR2.1-TOPO vector which was used to insert DNA fragments. One Shot TOP10 Electrocompetent *E.coli* cells (Invitrogen) were used for bacterial transformation. Electroporation was performed in 1 mm cuvettes in a BioRad Gene Pulser Xcell™ electroporator preset to following conditions: 10 μF, 600 Ohms, 1800 V. Colonies were grown in LB agar media containing 50 μg/mL kanamycin and 20 μg/mL bromo-chloro-indolyl-galactopyranoside (X-gal). Plasmids were purified following the protocol of NucleoSpin Plasmid (Macherey-Nagel, Germany). DNA sequencing reactions were carried out by Microsynth GMBH (Switzerland). Assembly of double stranded sequencing data and alignment analysis was performed with Lasergene 7 software suite (DNAstar, Madison, WI, USA).

The *E.coli* TOP10 electrocompetent_cells used for the molecular cloning of long PCR fragments has the following genotype descript: F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G.

3.10. Nanoparticle dispersion into aqueous solution in water

A probe style ultrasonic system Sonics VCX-750 (Sonics & Materials, Inc, USA) was used for ultrasonic dispersion of carbon nanotubes in water. The machine uses 750W uninterrupted power supply and a stepped micro tip probe for dispersion of volumes from 250 μL to 10 mL. The required sonication time was 5 minutes with an interruption of 10 seconds every 30 seconds at 35% amplitude. For avoiding the evaporation of water due to preheated solution during ultrasound treatment, the tubes were kept on ice.

3.11. Oligonucleotide primers

Standard primers of 20-25 nt length were used. Primers were designed with DNAstar Lasergene 7 (USA) PrimerSelect tool or by evaluating the sequence by eye. The primers were designed on the basis of the available DNA sequence of a particular fragment, taking into consideration the rules of primer stability and melting temperatures. The primers were checked for minimal or no possibilities to form "hairpin" or primer dimmer structures. All oligonucleotides used in the LT-PCR experiments are listed in table 6. BLAST similarity search was also performed for genomic walking primers as the specificity of this element is crucial for successful reactions. All oligonucleotides used in the RADE and LT-RADE development, optimization and application are listed in table 7.

Table 6: Oligonucleotide sequences of primers used in PCR

Template	Name	Sequence	Description
Lambda DNA	P3	TGGTTTATTGGAGTAG ATGCTTG	Within restriction enzyme-digested λ DNA (48,502bp)
	P4	GAGAGTTGTTCCGTTG TGGG	Within restriction enzyme-digested λ DNA (48,502bp)
Bt11	11GSFan1	TGATCATCGACCAGAG GCTC	Flanking DNA region 5’
	11GSFan2	CGTAGGTGTAGCCTCT AGTA	Flanking DNA region 5’
	11GSFanR 1	CTCTGGTCGATGATCA ATGG	Flanking DNA region 3’
	11GSRan2	GGTCGATGATAAATGG CCACA	Flanking DNA region 3’
	11GSR1	GAGCCTCTGGTCGATG ATCAATGG	Flanking DNA region 3’
	11GSF4	TCTGGTGGAGATCATT TCTTGGTC	Flanking DNA region 5’
	patintR2	GACTCAGATCTGGGTA ACTG	*pat* gene - *nos* terminator junction
	PATF	CACGAGGAGCATCGTG GAA	P-35S region
	PATR/ pUCF4	TCTTAGACGTCAGGTG GCACT	In the plasmid DNA of the insert
	pUCR3	CGTAATCATGTCATAG CTGT	In the plasmid DNA of the insert
	cry1F	TAGGAAGTTTGGTTGC ACTGG	*adh1* element of Bt11 insert

	cry1R	CATAGAGAGGAAAGG TAAACTC	*cry1Ab* gene
	cry2F	GAGTATTACTGGTCTG GACAC	*cry1Ab* gene
	cry2R	ATCACCTCCTGTGAAG CCTG	*cry1Ab* gene
	cry3R	CTTAACTATGCGGCAT CAGAG	In the plasmid DNA of the insert
H7-1	SBHG1	CAGGTGATGGTGGCTG TTAT	Flanking DNA region 5’
	SBHG2	TGATGTTTCTTTCTTCT CATTAC	Flanking DNA region 5’
	SBHTR1	TCTCCATATTGACCAT CATACT	T-DNA region 3’
	SBHTR2	TAACGCTGCGGACATC TACA	T-DNA region 3’
	SBHGR1	AACCTCATAGGTTTTA CGTAT	Flanking DNA region 3’
	SBHrbsSF 1	CACACACCAGAATCCT ACTGA	Located inside E9 3’ fragment
	SBHrbcSF 2	TGGAAATGGATGGAG AAGAGT	Located inside E9 3’ fragment
	SBHGSR3	GAAATACGTAAAACCT ATGAGGTT	Flanking DNA region 3’
	SBHIntF	GTCGTGGGGATTGAAG AAGAGTGG	*ctp2* and *cp4 epsps* junction
	SBHIntR	CTGTAGGTGATTGGCG TTGGAGTC	*ctp2* and *cp4 epsps* junction
	SBcox2F	TCAATGGACGAGGTAG TAGTAGAT	*cox2* gene in mtDNA

	SBcox2R	ATACGTATATGAGTTT TGGCTGGTA	*cox2* gene in mtDNA
	SBrrnF	TCCCGCTACTGACACT TGAAAA	*rrn* gene in mtDNA
	SBrrnR	AAAATACGGGGATGA CTTGTGG	*rrn* gene in mtDNA
MON810	5PL1-F	TGTCAGTGACTGTCGC TATTCTA	Flanking DNA region 5’
	5PL2-F	GTGAAGAAGCACAAA AGGAATAC	Flanking DNA region 5’
	5PL3-F	GTGCCTATAAATAGAT GAACAGTG	Flanking DNA region 5’
	5PL4-F	CCTTCTTCATTATTATA TCTTGTGC	Flanking DNA region 5’
	HSP-F	GTGTGGTGATTGCTGA GAGTG	*Hsp70* gene
	3PL1-R	GCCCAACAACAAGATC AAAGTCA	Flanking DNA region 3’
	3PL3-R	CCTTCATATCTGTTGC CCTTTCT	Flanking DNA region 3’
	35S-R	TGTCAGCGTGTCCTCT CCAAAT	P-35S promoter region
	Cry1-F	CTCTCCCTCACGCAGT TCCT	*cry1Ab* gene
	Cry4-F	AACCTGCAATTCCACA CCTCCA	*cry1Ab* gene
	CRY1-R	GGTGTAGATAGTGATG CTGTTGA	*cry1Ab* gene
	CRY2-R	ACTTGGTCGATGTGGT AGTCAG	*cry1Ab* gene

Table 7: Oligonucleotide sequences of primers used in RADE and LT-RADE steps of PCR

Target	Name	Sequence
P-35S	P35SRA	5'-TGTCGGCAGAGGCATCTTCAAC-3'
P-35S	P35SRB	5'-CTTTCCTTTATCGCAATGATGG-3'
P-35S	P35SRC	5'-GCAATGATGGCATTTGTAGGAG-3'
P-35S	P35SRC_rev	5'-CTCCTACAAATGCCATCATTGC-3'
P-35S	P35SRB_rev	5'-CCATCATTGCGATAAAGGAAAG-3'
P-35S	P35SRA_rev	5'-GTTGAAGATGCCTCTGCCGACA-3'
Cry1Ab	MON810F1	5'-TACATCGAAGACAGCCAAGAC-3'
Cry1Ab	MON810F2	5'-ACCTGATCCGCTACAACGCCAA-3'
Cry1Ab	MON810F3	5'-CACAGCCACCACTTCTCCTTG-3'
Poly(dC)	AAP	5'-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3'*
Adapter	AUAP	5'-GGCCACGCGTCGACTAGTAC-3'
T35S-*CaMV*	LL62_TF1	5'-CGCTGAAATCACCAGTCTCTCT-3'
T35S-*CaMV*	LL62_TF2	5'-TGTGTGAGTAGTTCCCAGATAAG-3'
T35S-*CaMV*	LL62_TF3	5'-GGTTCTTATAGGGTTTCGCTCAT-3'
Maize DNA	5PL4-F	5'-CCTTCTTCATTATTATATCTTGTGC-3'
Maize DNA	3PL2-R	5'-CAGTCTAGTAGGATTGGTATTGG-3'

* “I” stands for purine-type nucleoside inosine which indiscriminately forms pairs with adenine, thymine, or cytosine. In our case “I” forms a so called wobble base pair (non Watson-Crick base pair) between inosine-cytosine.

3.12. Submitting DNA sequences to genetic sequences database

The nucleotide sequences of the LT-RADE obtained fragments were accurately described using Sequin stand alone software by the NCBI for submitting and updating entries to the GenBank sequence database. All the LT-RADE fragments were then deposited to the NCBI database. Accession numbers are listed in table 8.

Table 8: GenBank accession number listing of the fragments obtained by LT-RADE

Organism	**Genome walking direction**	**Accession number**
MON810 maize	5’ insertion site, upstream (outside)	JQ406879
MON810 maize	3’ insertion site, downstream (outside)	JQ406878
MON810 maize	P-35S promoter, downstream (inside)	JX139718
LLRICE62 rice	5’ insertion site, upstream (outside)	JQ406881
LLRICE62 rice	3’ insertion site, downstream (outside)	JQ406880
LLRICE62 rice	P-35S promoter, downstream (inside)	JX139719
T45 rapeseed	5’ insertion site, upstream (outside)	JX139721
T45 rapeseed	P-35S promoter, downstream (inside)	JX139720
A2704-12 soybean	5’ insertion site, upstream (outside)	JX139723

A2704-12 soybean	P-35S promoter, downstream (inside)	JX139722
LLCOTTON25 cotton	5’ insertion site, upstream (outside)	JX139725
LLCOTTON25 cotton	P-35S promoter, downstream (inside)	JX139724

CHAPTER III: RESULTS AND DISCUSSION

1. STRATEGIES TO OBTAIN LONG SYNTHETIC DNA NUCLEOTIDE SEQUENCES

1.1. Characterization of the transgenic inserts in GM plants

The term “genetic characterization” is commonly used as a broad expression summarizing many forms of the genetic data collection for a specific description of an organism or its individual quality (Vicente *et al.* 2006). Genetic characterization can entail information from localization of the particular DNA locus (loci) in the genome to nucleotide sequence analysis and/or variation of these sequences between individual genomes. Gene functional analyses, classification, open reading frame detection, protein functions may as well be described while performing genetic characterization (Okamoto *et al.* 1986, Manseau *et al.* 1988). One of the significant factors in genetic material description is also the stability of a particular sequence in the genome. In the introduction of this manuscript, we have presented the indisputable necessity of the biotech plant research due to a strict regulatory system that controls the authorization and production of GM crops worldwide. At the moment, the application of authorization to introduce GM-derived food in the EU entails that an event-specific method is submitted and validated for every new biotech plant (Directive 2001/18/EC). However this requires knowledge of the nucleotide sequence between and overlapping the transgenic insert and the site of insertion in the plant genome. A "full" characterization of the entire genome of a biotech plant is as such not necessary.

One of the objectives in this study was the acquirement of the transgenic insert in a single uninterrupted fragment with pieces of flanking plant genomic sequences. Such results, if successfully obtained, not only would ensure the integrity of the transgenic insert thus already providing us with some data on

the stability of the insert; but also provide with much needed data via nucleotide sequencing. At the moment, the publically available data on the nucleotide sequences of GM inserts are very limited, thus experiment design may suffer from the lack of knowledge of the particularities: (1) not enough or no plant flanking sequence is available as nucleotide sequence; or (2) only the constitution of transgenic elements is known about the GM insert, but not the nucleotide sequence. Obtaining the sequences may give the possibility of continuation of the molecular characterization studies in a way that the nucleotide sequence obtained in one step of the experiments then becomes a basis for the next study design.

As the transgenic cassette is usually longer than the fragment obtainable by Conventional PCR, we have introduced the LT-PCR as a potential technique to achieve our aim. By successful application of LT-PCR to obtain full transgenic inserts we would obtain the information about the contiguous state (not-fragmented) of the synthetic DNA. We would then be able to continue with a molecular level characterization of these inserts by obtaining the nucleotide sequence, localizing the fragment in plants' genome and describing it in the light of publically available data.

1.2. Choosing GM events for LT-PCR application

Three GM events were chosen for experiments of LT-PCR application on transgenic plant DNA. All three events were chosen based on the knowledge of nucleotide sequence and characteristics of the transgene. Furthermore the available material for DNA extraction was considered as we wanted to investigate different DNA matrices and experimental limitations on them. The three GM events were as follows: H7-1 sugar beet; Bt11 maize; MON810 maize.

1.3. Bio- informatics analysis of chosen GM events

H7-1 sugar beet

The sequence of H7-1 sugar beet synthetic construct was retrieved from the publically available source GenBank. The nucleotide sequence is listed as "Glyphosate Tolerant Sugar Beet" under the accession number DD417675. This sequence is submitted as a patented sequence under the Japan Patent Office number JP 2006518205-A/6 on 10/08/2006. We discovered that there was a submission for H7-1 sugar beet patent in the United States Patent and Trademark Office (USPTO) on 26/02/2008. H7-1 sugar beet sequence of 3779 bp consists of a 3355 bp transgenic cassette sequence framed by flanking regions of 356 bp and 68 bp at respectively 5 and 3' ends of the sequence.

Bt11 maize

The DNA sequence for the Bt11 maize was retrieved from Patent No. US6114608, the GMO Detection Method Database (GMDD, China; Dong *et al.* 2008) and Molecular Biology and Genomics Unit service database described as GMO Event Bt11 'Sweet Corn Maize' (SYN-BT011-1) with transgenic insert and flanking 5 and 3-prime flanking sequences based on the Dossier provided by Syngenta 28 July 2004. As features narrative informs Bt11 has a transgenic insert of 6202 bp respectively framed by flanking regions of 328 bp and 537 bp from both sides. The file also describes the positions of Bt11 insert elements which are ColE1 ori, 35S promoter, IVS 6, *cry1Ab* gene, nosT terminator, repeat of 35S promoter, IVS 2, *pat* gene and again nosT terminator.

MON810

The DNA sequence for MON810 was retrieved from GMO Detection Method Database (GMDD, China; Dong *et al.* 2008). The 4983 bp fragment contains 803 bp of the plant 5' flanking sequence followed by 3591 bp of the transgenic insert and 588 bp of the plant 3' flanking sequence.

Schematics of the structural organisation of the 3 transgenic inserts are depicted in the figure 3 below. These schematic diagrams represent the summary of all data on the structures of the GM cassettes as retrieved from Application of Authorization documents (EFSA and GMO Compass), CERA database knowledge, BLAST analysis and GenBank entries as well as information from other publically available sources with relevant documentation as EU-RL GMFF laboratory website.

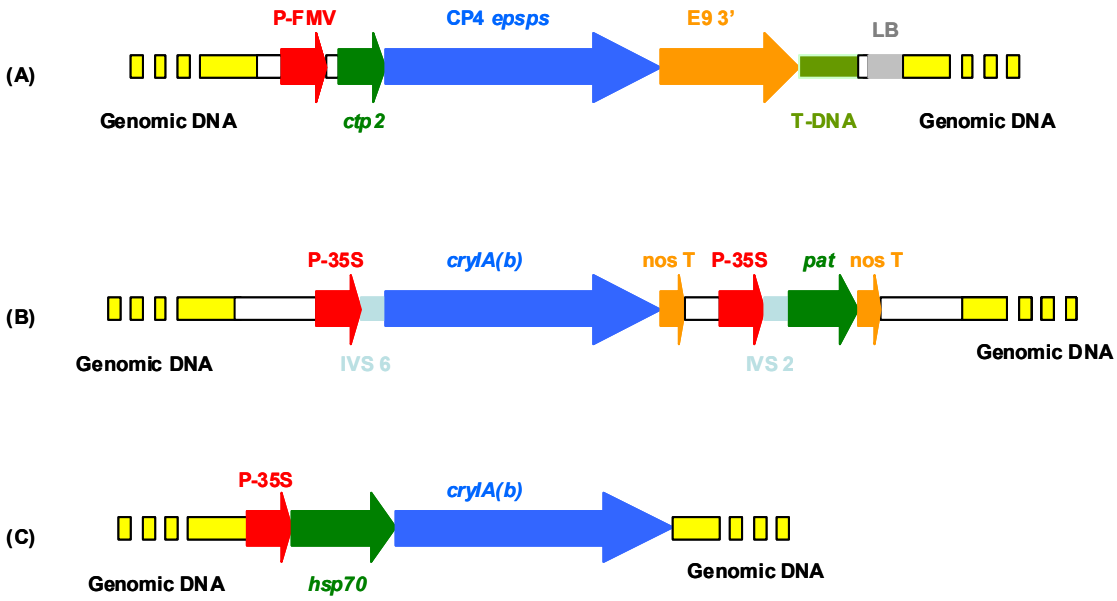

Figure 3: Mapping the genetic inserts of selected GM events H7-1 **(A)**, Bt11 **(B)** and MON810 **(C). P-FMV**- Figwort Mosaic Virus 35S promoter; ***ctp2***-chloroplast transit peptide (ctp2) from the *Arabidopsis thaliana* EPSPS gene; **CP4 *epsps***- 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene from *A.tumefaciens* strain CP4; **E9 3'**- transcription termination and polyadenylation signal sequences from the ribulose-1,5-bisphosphate carboxylase small subunit (*rbcS*) E9 encoding gene from pea (*Pisum sativum*); **LB**- Left Border sequence from *A. tumefaciens;* **P-35S**- Cauliflower Mosaic Virus 35S promoter; **IVS2** and **IVS 6**- introns 2 and 6 from the maize alcohol dehydrogenase gene; ***cryIA(b)***- gene encoding Cry1Ab δ-endotoxin (Btk HD-1) (*B. thuringiensis*

subsp. *kurstaki* (Btk)); **nosT**- *A. tumefaciens* nopaline synthase (*nos*) 3'-untranslated region acting as a terminator; ***pat***- phosphinothricin N-acetyltransferase (*S.viridochromogenes*); ***hsp70-*** heat-inducible enhancer fragment from heat shock protein 70 gene.

1.4. Template preparation for the experiments: challenging material

The DNA from plant leaves, seeds and/or certified reference material (CRM) was extracted as described in the section 3.1 of Methods subchapter. Different types of DNA were tested with the assumption that lower levels of DNA degradation in the preparations would give better results in the LT-PCR experiments. DNA for our experiments was extracted from the following:

- H7-1 sugar beet CRM of 100% GM content, plant leaves and ground seeds;
- Bt11 maize CRM of 5% GM content, plant leaves and ground seeds;
- MON810 maize CRM of 5% GM content.

One of our aims was to examine whether it is possible to amplify and visualize long DNA fragments from low GM content material. Figure 4 shows extracted DNA visualized on agarose gel (1μL per well).

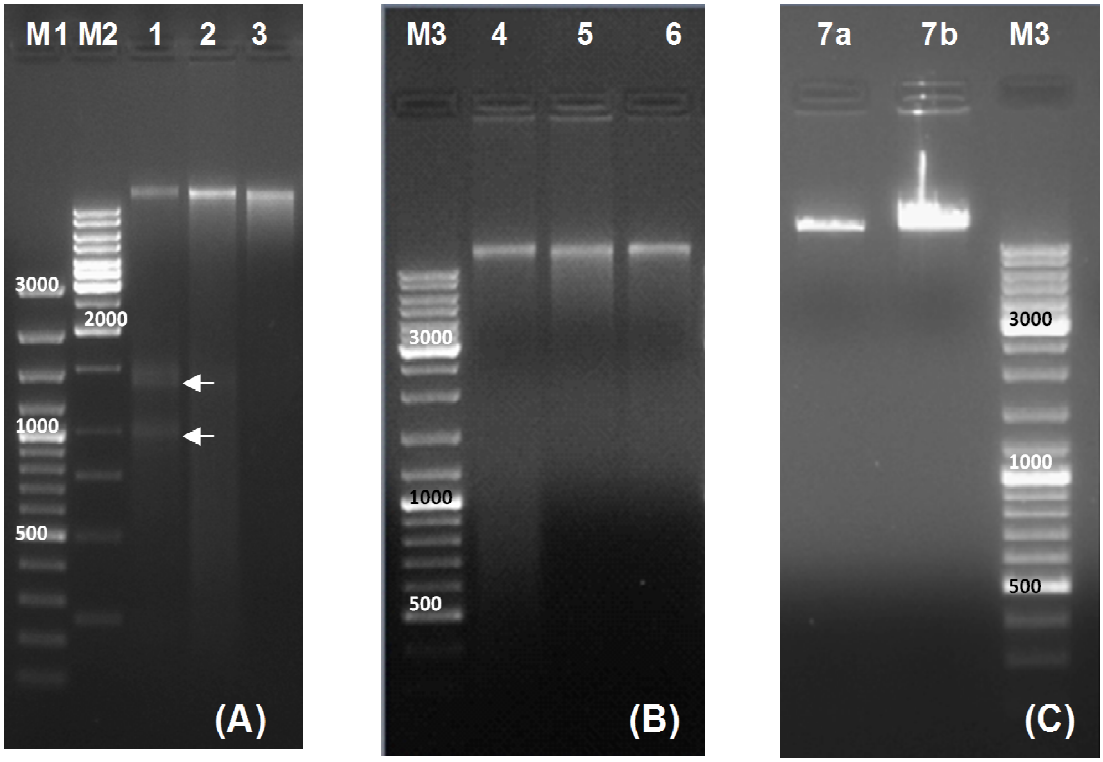

Figure 4: Assessment of different types of extracted DNA templates. (A) H7-1 sugar beet DNA: lane 1 contains DNA extracted from 100% GM content CRM; lane 2- plant leave DNA; lane 3- ground seeds. (B) Bt11 maize DNA: lane 4 displays DNA from 5% GM content CRM; lane 5- ground seeds; lane 6- plant leave DNA. (C) MON810 maize DNA: 7a and 7b show samples of 5% GM content CRM DNA. M1- GeneRuler™ 100 bp Plus DNA Ladder; M2- GeneRuler™ 1 Kb DNA Ladder; M3- GeneRuler™ DNA Ladder Mix (Fermentas, Lithuania). Selected size standards in bp are displayed in the lanes with DNA Ladders.

As it can be deduced from the figure 4(A), the quality of H7-1 sugar beet DNA varied amongst the matrices. The best quality DNA was extracted from plant leaves; the DNA from ground seed material showed some degradation but still had a large quantity of high molecular weight DNA; the DNA prep from CRM had the worse quality of the 3 investigated DNA sources: it showed a specific banding pattern marked with arrows in figure 4(A), with some high molecular weight DNA present. This pattern was observed in repeated DNA extractions.

This pattern (of 3 separate bands) suggests that the preparation of the CRM introduced a degradation of the DNA. Moreover, it seems that there are particular places in the high molecular weight genomic DNA of sugar beet which are more sensitive to degradation. As it was noted previously, this CRM of H7-1 sugar beet contains 100% GM material.

During the application of extraction protocol provided with the commercial kit (see Methods subchapter), we noticed that during the step of lysis of the cells, the lysate changed color to dark violet. This remarkable observation was only noticed using sugar beet CRM and it is most probable that the sugar levels of the preparation have determined the tinge of the solution.

Bt11 maize DNA from various sources is displayed in figure 4(B). A low level of degradation was observed in preparations of DNA extracted from seed and CRM but the ample amount of high molecular weight DNA suggests that the material is suitable for usage in experiments. The best quality DNA, as expected, was extracted from plant leaves. No particularities were encountered applying the standard protocols for DNA extraction as suggested by commercial kit manufacturers. It is very important to mention, however, that the percentage of GM content for Bt11 CRM was 5% (while the rest available CRM for Bt11 is 0.01, 0.1, 0.5, 1 and 2%) thus leading to much reduced amounts of GM DNA in the entire extraction. This has to be taken into consideration when using this particular DNA preparation in the experiments.

For MON810, only CRM was used for DNA extraction. The highest percentage of this CRM on the market is 5%. Figure 4(C) displays two different concentrations from two different samples of DNA extracted from MON810 5% CRM. The DNA quality is considered very good with no visible degradation patterns and well concentrated as a high molecular weight band on agarose gel.

1.5. Conventional PCR as DNA controls

Before the application of LT-PCR, the selected DNA templates were tested via Conventional PCR to investigate the quality and authenticity of the DNA. This testing was primarily aimed at the DNA extracted from CRM and ground seeds. In the case when CRM and/or ground seed DNA did not show the expected results, the plant DNA was used in the reactions. All the reactions were carried out according to instructions of the section 3.2. of the Methods subchapter.

3 pairs of primers were designed for H7-1 sugar beet DNA tests: SBHintF and SBHintR located on the junction of synthetic *ctp2* and *cp4 epsps* genes inside the transgenic insert (expected product: 591 bp, see figure 5(A1)) and two pairs SBcox2F and SBcox2R with SBrrnF and SBrrnR located in mtDNA (expected products: 216 and 459 bp respectively; see figure 5 (A2 and A3)). Two DNA matrices were tested with abovementioned primers: 100% GM content sugar beet CRM (figure 5(A1,2,3 lanes “a”)) and ground seed DNA (figure 5(A1,2,3 lanes “b”)). As it can be deduced from the agarose gel run, the experiments on ground seed DNA have concluded with a bigger yield of product, however the reactions containing CRM had very sufficient yield as well. We chose for the CRM DNA preparation for further experimental steps.

A primer pair cry1F and cry1R was designed for obtaining 1188 bp fragment from Bt11 maize DNA preps. These primers are situated between *adh1* and *cry1Ab* genes in transgenic insert of Bt11 maize and amplify the junctions of these elements. Figure 5(B4) part shows the amplification of this fragment using 5% GM content CRM (lane “a”) and ground seed DNA (lane “b”). Only traces of product were detected in the reaction containing CRM DNA, and therefore the GM content percentage is too low to obtain sufficient DNA product in Conventional PCR. Ground seed DNA has shown a clear well defined band on agarose gel and suggested that LT-PCR experiments on Bt11

maize should be carried out on ground seed DNA (see figures 7, 9 and 13 for primer locations on a transgene elements map for all objects of study).

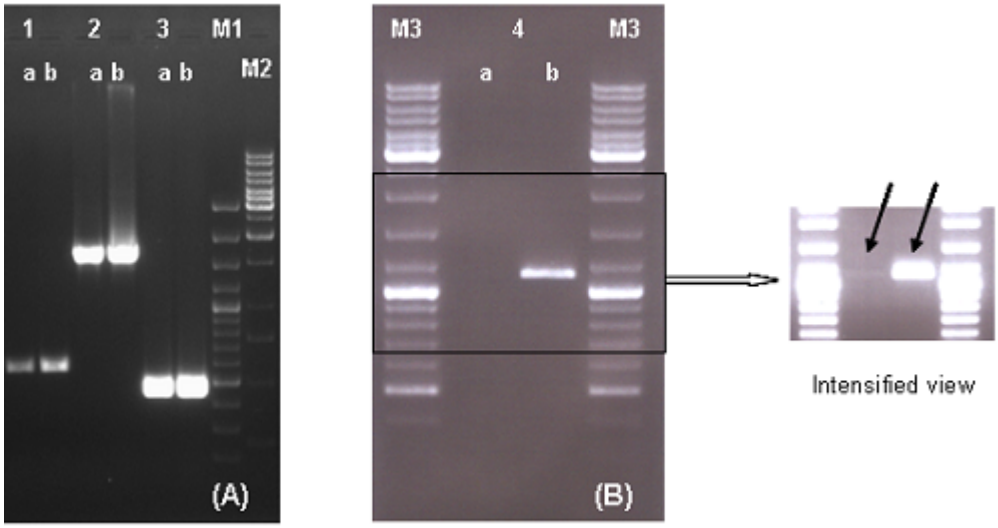

Figure 5: DNA matrices control: (A) H7-1 sugar beet 1st primer pair SBHintF and SBHintR; 2nd primer pair SBcox2F and SBcox2R; 3rd primer pair SBrrnF and SBrrnR. (B) 4th primer pair cry1F and cry1R. Experiments on DNA extracted from CRM are displayed in lanes "a"; experiments on ground seed DNA are shown in lanes "b". M1- GeneRuler™ 100 bp Plus DNA Ladder; M2- GeneRuler™ 1 kb DNA Ladder; M3- GeneRuler™ DNA Ladder Mix (Fermentas, Lithuania).

For MON810, the DNA preparation obtained from the CRM was tested with several primers. For this particular GM, besides testing the quality of the DNA, we wanted to investigate the integrity of the insert (as one large fragment). Therefore, every primer located in plant flanking DNA sequence (5 or 3') was paired with a primer inside the transgenic insert of MON810 or both amplification primers were located inside the construct. Forward primers 5PL1-F, 5PL2-F and 5PL4-F were paired with a primer 35S-R to amplify 1046, 817 and 434 bp fragments respectively (Figure 6 lanes 1, 2 and 3).

Internal primers located in *cry1Ab* gene of synthetic construct were paired as following: cry1-F and cry1-R also cry4-F and cry2-R to amplify solely synthetic DNA fragments of 803 and 343 bp in length (Figure 6 lanes 4 and 5). The last pair of primers tested covered 3' junction between plant genome and transgenic insert: cry4-F and 3PL1-R to retrieve a fragment of 1029 bp (Figure 6 lane 6). All expected fragments were obtained with satisfactory or good quantity, suggesting that 5% GM content DNA may be sufficient to perform LT-PCR experiments. Since the CRM is prepared with the highest standards and easily purchasable on the market, the successful application of the methods on these DNA matrices suggests that it can be a DNA source-of-choice when planning the experiments.

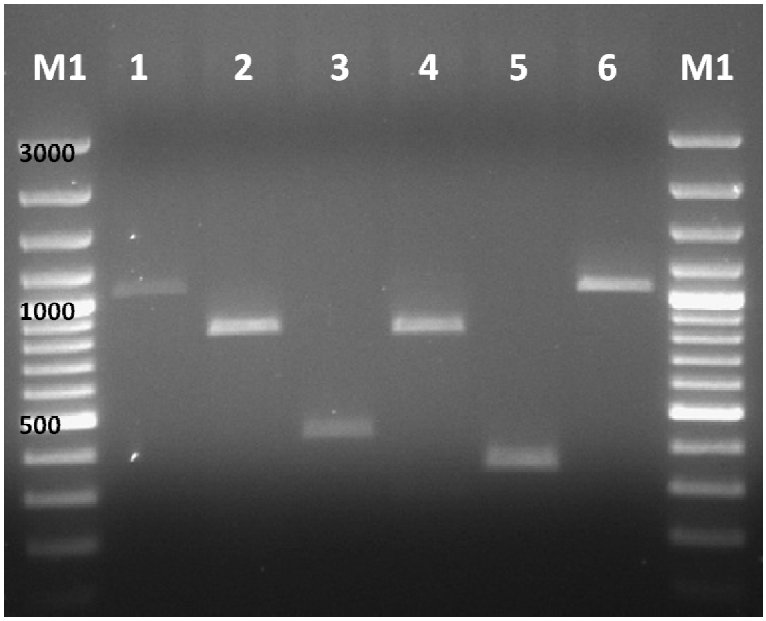

Figure 6: MON810 5% GM content CRM control reactions: lane 1- amplification with primer pair 5PL1-F and 35S-R; lane 2- 5PL2-F and 35S-R; lane 3- 5PL4-F and 35S-R; lane 4- cry1-F and cry1-R; lane 5- cry4-F and cry2-R; lane 6- cry4-F and 3PL1-R. M1- GeneRuler™ 100 bp Plus DNA Ladder. Selected size standards in bp are displayed in the lane with the DNA Ladder.

In conclusion, the selected DNA preparations to carry out LT-PCR experiments were as follows: 100% GM content CRM from H7-1 sugar beet;

ground seed DNA from Bt11 maize; 5% GM content CRM from MON810 maize.

1.6. LT-PCR on H7-1 sugar beet transgenic insert

Primers for LT-PCR were designed at the 5' and 3' plant flanking sequences of the insert to retrieve a long fragment carrying all structural elements of the insert. The two forward primers were designed in the 5' plant flanking region (SBHG1 and SBHG2). For the primers to be designed at the 3' end, the choice of a good primer was very limited, and almost impossible. Two primers were designed with sub-optimal characteristics (SBHGR1 and SBHGR3). The plant genomic region was very small (40 bp) and A/T-rich. Therefore, two other primers were designed within the very end of the insert (SBHTR1 and SBHTR2) in a sequence described as T-DNA (backbone).

In addition, primers SBHrbsSF1 and SBHrbsSF2 internal to the GM insert were designed in such manner that a small fragment could be obtained in combination with the primers in the 3' flanking region. This feature was included to check, if necessary, the suitability of the primers located outside of the insert. The primer locations on the transgenic insert map are demonstrated in figure 7.

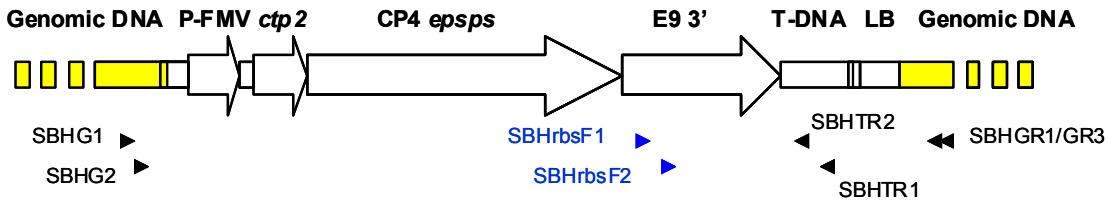

Figure 7: LT-PCR primer locations on structural map of H7-1 sugar beet synthetic construct

LT-PCR was carried out on DNA isolated from CRM. The primers were used in all possible combinations in order to maximise the rate of success. The combination of primers to obtain the largest possible fragment was not successful, in none of the cases (primer combinations between SBHG1/G2 and SBHGR1/GR3). This was very likely due to the failure of the primers designed in the small 3' flanking sugar beet DNA sequence (40 nt). This failure was expected as the possibility to design a suitable primer in that particular region was almost negligible. Moreover, a conventional PCR with SBHGR1/GR3 primers in combination with the 2 internal primers (SBHrbsF1/F2) was unsuccessful. This provided additional evidence for the non-suitability of the primers.

Successful amplification was obtained using forward primer SBHG1 in the 5' flanking region of the sugar beet DNA and 2 reverse primers in T-DNA region of the insert: SBHTR1 and SBHTR2. The fragments retrieved were respectively 3408 bp for the primer pair SBHG1 & SBHTR1 and 3347 bp for the primer pair SBHG1 & SBHTR2 (see figure 8). The specificity of the PCR was rather good, reflected by the high yield of the desired product. Some other bands can be noticed on the agarose gel though they are very weak. The presence of these bands nevertheless made it necessary to extract the desired PCR product from gel. For further experiments the longer fragment of 3408 bp was chosen.

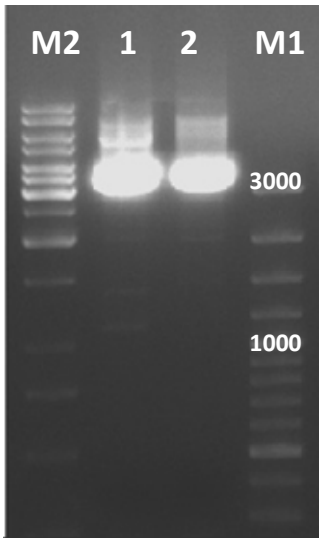

Figure 8: Products obtained via LT-PCR on H7-1 sugar beet DNA. Lane 1: 3347 bp fragment with SBHG1 & SBHTR2 primer pair; lane 2: 3408 bp fragment with SBHG1 & SBHTR1 primers. M1- GeneRuler™ 100 bp Plus DNA Ladder; M2- GeneRuler™ 1 kb DNA Ladder. Selected size standards in bp are displayed in the lane with the DNA Ladder.

In order to increase the yield of the obtained product, LT-PCR was carried out using the former PCR product (1 μL) as a template and the same primer pair. This step was necessary due to the requirement of a high concentration of pure DNA product for direct bi-directional sequencing reactions. It was observed that the DNA fragment was successfully amplified, with a high yield, and with no by-products. The fragment was purified and concentrated to obtain a sufficiently concentrated solution for sequencing reactions.

Sequencing reactions were performed by Microsynth Laboratory (Balgach, Switzerland). Primer-walking sequencing reactions were performed on the PCR fragment. The first 2 reactions were carried out with the primers that were used for the PCR amplification. Additional primers (10) were designed in the company, based on the sequence produced from the previous runs. As such, the 3.4 Kb fragment was sequenced as one uninterrupted fragment. Data analysis and sequence assembly were performed by us. The quality of the sequencing

was satisfactory. Good quality data were obtained with an average of 600 bases per run.

The consensus sequence was aligned with GenBank accession number DD417675 sequence to detect if there were any conflicts between the two sequences. The alignment of the two sequences was almost perfect, with no major differences. Only at the very end of the insert, we noted an additional base in our data (one more A in a string of 7 As). We have agreed that an extra adenine in the string of 7 of the same base would most likely reflect an error in the sequencing reaction and/or the amplification.

1.7. LT-PCR on Bt11 maize transgenic insert

Primers with different melting temperatures were designed at the 3' and 5' plant flanking DNA of Bt11 maize to retrieve a long fragment containing all the GM insert elements. Two forward (11GSF3 and 11GSF4) and 3 reverse primers (11GSR1, 11GSR2 and 11GSR3) with high melting temperatures were designed; another set of 2 forward (11GSFan1 and 11GSFan2) and 3 reverse primers (11GSR4, 11GSFanR1 and 11GSRan2) were designed with lowered melting temperatures to test different LT-PCR conditions and optimize the reactions. Primers with lowered melting temperatures were also fitted to pair up with the construct specific primers inside the transgenic insert. For the full primer list the reader is referred to materials and methods section. The primers essential for LT-PCR experiments (located in flanking plant DNA) and the other oligonucleotides necessary for retrieving the fragments of Bt11 maize transgenic insert are displayed in figure 9.

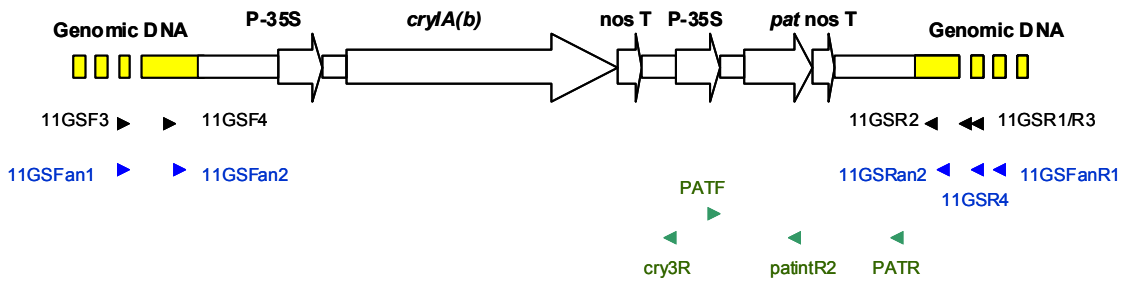

Figure 9: LT-PCR primer locations on structural map of Bt11 maize synthetic construct

A large set of LT-PCR reactions were designed of which pairs with correlating melting temperatures were used for LT-PCR runs. Transgenic insert length calculations have shown that the targeted amplification fragment was close to 6.5 Kb in length depending on the primer localization in the flanking region. However all the experiments with different primer combinations specified in table 9 repeatedly ended in producing high molecular weight smears in the gel-electrophoresis. All possible primer combinations did not result in a product from 5' to 3' flanking sequence covering all the structural elements of the insert. As the conditions and chemistry of a PCR reaction can be easily altered, the problem was approached by optimizing the experiments. However, the same high molecular weight smears also appeared applying lower enzyme and/or primer concentration, after shortening the extension time, reducing the number of reaction cycles, varying denaturation temperature and $MgCl_2$ concentration, also adding PCR additives such as betaine, trehalose and DMSO (data not shown). The origin of the smears was questioned due to the fact that the amplification product seemed to be stuck in the agarose gel wells; in rarer cases the amplification product was moving producing continuous smear on the gel. It is commonly known that the product remaining in the agarose gel wells yet allowing the DNA dye as ethidium bromide to stain it is due to the failure in normal dsDNA structure formation. High molecular smearing and smearing throughout the agarose gel lane are displayed in figure 10.

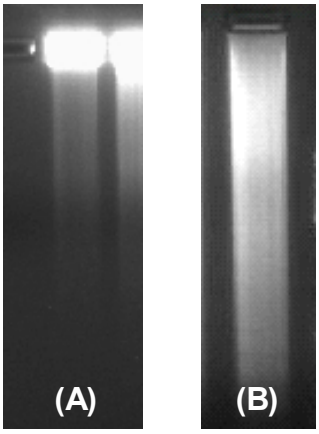

Figure 10: High molecular weight smearing or product stuck in agarose well (A); amplification product smearing (B).

Table 9: oligonucleotide pairs tested to obtain full Bt11 transgenic insert

Forward primer	Reverse primer
11GSF3	11GSR1
11GSF3	11GSR2
11GSF4	11GSR1
11GSF4	11GSR2
11GSFan1	11GSFanR1
11GSFan1	11GSRan2
11GSFan2	11GSFanR1
11GSFan2	11GSRan2
11GSFan1	PATR

For investigation of the smearing on agarose gel another set of LT-PCR reactions was planned to test all the primers localized either in 5' or 3' flanking plant sequences. These reactions had construct specific primers located inside the transgenic DNA paired with the ones in the flanking regions. All expected control products length was set to ~4 to 6 Kb to keep the LT-PCR conditions in the application. Reactions with primers in 5' plant flanking sequence have shown good results producing the longest fragment of 5902 bp (figure 11)

which is 77 bp short to reach the 3' plant flanking sequence. This fragment was produced using primer pair 11GSFan2 + PATR. However all reactions with primers in 3' flanking sequence were unsuccessful. This drew our attention to the short sequence of 77 bp (cloning vector pUC19 backbone sequence) in which presumably the amplification stops. The sequence is full of single type nucleotide stretches (several G nucleotides are closely situated to several C nucleotide stretch; many 3-5 nt repeats of T's) therefore is very likely to form secondary structures such as a stem-loop. Since the large fragment of 5902 bp was covering the transgenic insert as one uninterrupted sequence short by only 77 bp to reach the opposite end flank, we have decided to cover that short sequence with a conventional PCR that well overlaps the previously obtained large fragment. For such Primer pair PATF+11GSR4 was used to amplify a 1722 bp fragment that was purified and sent for direct primer-walking sequencing application.

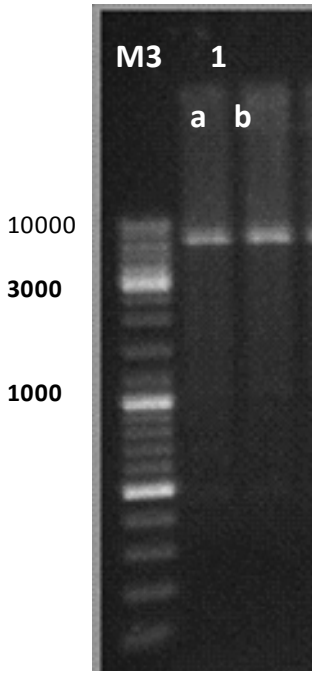

Figure 11: 5902 bp fragment obtained via LT-PCR on Bt11 maize DNA. Lines 1 "a" and "b" show 5902 bp fragment on agarose gel. M3- GeneRuler™ DNA Ladder Mix. Selected size standards in bp are displayed next to the lane with the DNA Ladder.

The 5902 bp fragment was purified from the agarose gel and cloned into a linear cloning vector. Cloning to a pJAZZ-OK Blunt Vector (figure 12) was selected as LT-PCR reaction has been carried out with a proofreading DNA polymerase Phusion that does not leave any overhangs at the end of the product. The vector pJAZZ-OK has two dephosphorylated arms which re-creates a linear plasmid after insert DNA is ligated in between. Vector has a protelomerase gene, replication factor and origin of replication, regulator of copy number and a terminator in the left arm of 10 Kb and a kanamycin resistance gene framed by two terminators in the right arm of 2.2 Kb. For particularities on cloning procedures see the 3.8. section in the Methods subchapter.

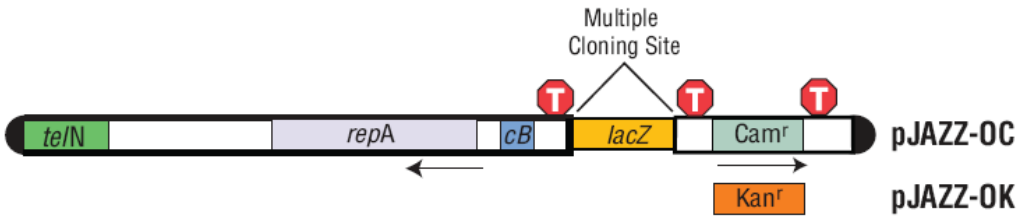

Figure 12. pJAZZ-OK Blunt vector map as provided by Lucigen Corp. The vector is linear with two arms stretching both sides of the cloning site. Figure taken from the manufacturers protocol provided with the commercial kit.

The set of plasmids resulting from molecular cloning was tested by enzymatic digestion. Restriction analysis with *NotI* released the insert of approximately 6 Kb and 2 vector arms. Three plasmids were sent to Microsynth Laboratory for primer walking sequencing reactions which were bi-directionally started with amplification primers and continued with newly designed oligonucleotides (seven in total) based on obtained sequence (so called primer walking sequencing). The quality of the sequencing was satisfactory: good quality data were obtained with an average of 750 bases per run. The

consensus sequence resulting from the retrieved data was assembled with Lasergene DNA sequence manipulation suite. Alignment was made between the consensus and patent sequence as well as consensus and the service database sequence. The similarity study showed 8 single nucleotide mismatches scattered throughout the sequence. The origin of these mismatches was not examined as the trustworthiness of the results was obvious due to nearly perfect alignment between sequences as such the mismatched nucleotides could be treated as polymerisation or sequencing errors.

1.8. LT-PCR on MON810 maize transgenic insert

The CRM containing 5% of GM content was used as a template though it could be potentially difficult to a retrieve full GM insert by LT-PCR technique given the "diluted" GM content in the matrix. Initially 4 primer pairs were designed in 5' and 3' flanking plant DNA sequences (except primer HSP-F targeting *hsp* gene) to retrieve a fragment of 3.8 to 4.4 Kb (figure 13). As can be observed in figure 14, Roche LT-PCR system was not efficient to retrieve a full GM insert. Partial sequences were obtained with the primers HSP-F and 3PL3-R. However, the Finnzymes LT-PCR system was efficient to retrieve a full GM insert with all 3 primer combinations: 5PL3-F and 3PL3-R, 5PL4-F and 3PL3-R, 5PL4-F and 3PL1-R. As seen in figure 14, all primer combinations resulted in successful retrieval of expected DNA fragment.

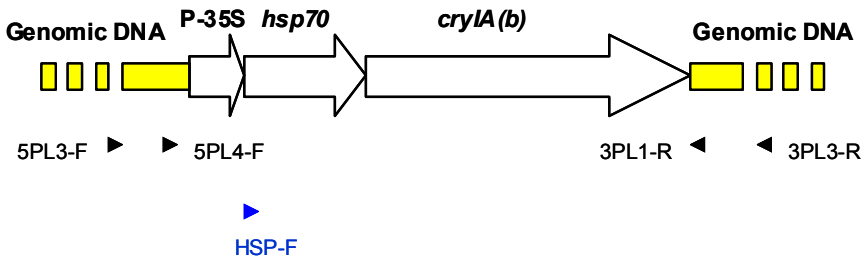

Figure 13: LT-PCR primer locations on structural map of MON810 maize synthetic construct

Due to the fact that 5% GM content DNA was used, the efficiency and specificity of the method LT-PCR was tested. As it can be deduced from the agarose gel run, the yield of the product is altering between different primer combinations: two primer pairs showed well defined bands, however the other two resulted in only the traces of the product. We therefore think that accurate primer design is essential when working with low target percentage templates of DNA.

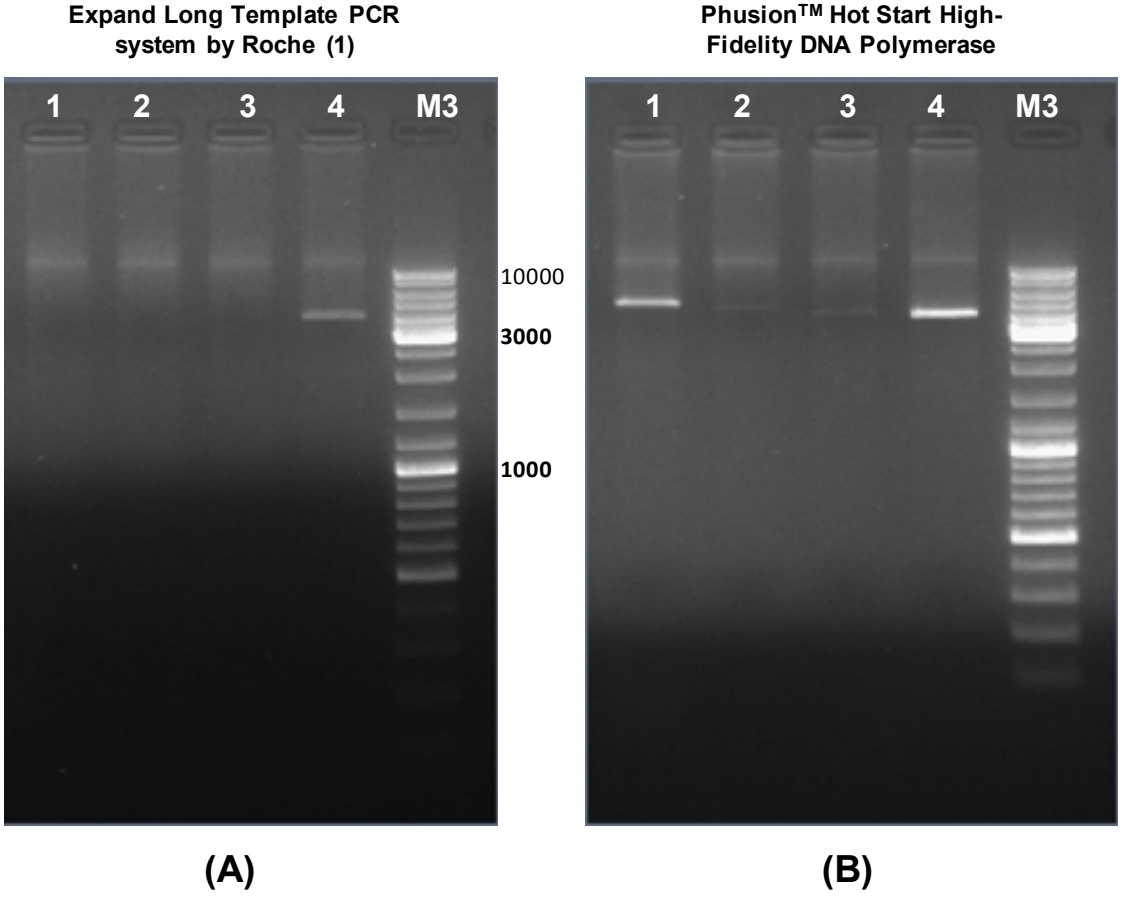

Figure 14: Products obtained via LT-PCR on MON810 maize DNA. (A) with Expand Long Template PCR system 1; (B) with Phusion Hot Start High-Fidelity DNA polymerase. Lanes 1- 4394 bp fragment with primers 5PL3-F and 3PL3-R; lanes 2- 4315 bp fragment with primers 5PL4-F and 3PL3-R; lanes 3- 3892 bp fragment with primers 5PL4-F and 3PL1-R; lanes 4- 3792 bp fragment with primers HSP-F and 3PL3-R. M3- GeneRuler™ DNA Ladder

Mix. Selected size standards in bp are displayed next to the lane with the DNA Ladder.

1.9. Dispersion of carbon nanoparticles, the novel LT-PCR additive

The initial purpose for CNTs application on molecular biology methods was to improve the specificity and efficiency in PCR when working with difficult templates as well as enhancement of the product yield. The hypothesis of carbon nanotube effect on PCR was based on one of the initial studies by Cui *et al.* (2003) which displays the interaction between SWCNT and DNA molecule and more precisely, the DNA being wrapped around the nanotube. This interaction as well as the one of *Taq* polymerase and the SWCNT was displayed via SEM and HRTEM imaging. One of the conclusions Cui *et al.* study presented was that strong interactions among SWCNTs, DNA templates and *Taq* polymerase took place during the PCR and as such these interactions affect the PCR efficiency and the amount of resulting product. The later study of Zhang *et al.* (2008) aimed to prove the benefits of the nanotubes on the LT-PCR (see below).

As the increase of the yield of reactions that actually produce a standard quantity of product do not require the PCR additives, our objective for the application of the nanotubes was to obtain the well concentrated long Bt11 transgenic insert fragment from 5' to 3' flanking sequence. This was not achieved with standard conditions or regular PCR additives (see above). Therefore, we opted for the following:

- Repetition of published experiments, *i.e.* 14 kb lambda DNA fragment amplification using the same CNT aqueous solutions with the exact same conditions as described by Zhang *et al.* (2008);
- Application of the CNTs on GM templates proven to be difficult in this study, *i.e.* amplification of Bt11 full GM insert; and to increase the low yield of the MON810 GM insert.

In 2008, Zhang *et al.* has published a study which is reporting the effects of SWCNTs and MWCNTs on LT-PCR amplification efficiency. The study showed positive results when amplifying a 14 Kb lambda DNA fragment with additions of MWCNTs in concentrations ranging from 0.4 to 1.6 μg/μL and from 0.6 to 1.2 μg/μL for SWCNTs. The standard application of LT-PCR to obtain the aforementioned fragment was unsuccessful producing a high molecular weight smear throughout the gel-electrophoresis. Displayed in figures of the experiments of Zhang *et al.* are the gradual improvements when carbon nanotubes are added to the reaction mixture when compared to the control reactions. It is clearly visible that the specificity of the reactions with CNTs increases greatly and well-concentrated DNA fragment is visible on agarose gel. Author also describes a visible problem of product stuck in gel wells and hypothetically presumes that this problem appears due to DNA that is wrongly folded and therefore cannot move in the gel lanes. This DNA cannot be refolded correctly nor digested with restriction enzymes by any applied DNA manipulations technique.

Aqueous CNTs solutions, used in this study, were prepared in water, first as described by Zhang *et al.* However, the application of 100W sonication for 1-2 hours did not result in nearly any dispersion of the nanotubes in water (the nanotubes stayed on the bottom of the clear thin-walled glass bottle with water over this sedimentation layer). We have received the personal confirmation that these were exactly the conditions the study team was using for dispersions, however the physical properties (CNTs being highly hydrophobic) determined that the nanomaterial did not even move to blend with water in 100W bath. It is therefore that we followed the advice by Dr. Mike Foley from Cheap Tubes Inc. (USA) (personal communication) to prepare the working dispersion preparations of MWCNTs in water. For the direct penetration of the solution a Micro-tip probe working on 35% amplitude was used for dispersion of MWCNTs in water. The ultrasound treatment was 5 minutes with an

interruption of 10 seconds every 30 seconds. After treating MWCNTs with ultrasound we applied them to PCR reactions to re-create the experiment done by Zhang *et al.* (2008).

1.10. Repetition of the study performed with *Taq* and *Pfu* polymerase blend and CNTs

The first step of our study was to repeat the experiments carried out by Zhang *et al.* study group in China. In his paper, the author thoroughly described the limitations of working with standard amplification enzymes as *Taq* or *Pfu* and the approach of blending two polymerases to overcome the limitations of processivity and stability on the template DNA. In long PCR experiments, Zhang *et al.* used 2:1 ratio of *Taq*:*Pfu* polymerase mix for amplification of 14 Kb Lambda DNA fragment and described the molecular smearing on agarose gel, therefore CNTs were applied to reactions for specificity tests (*i.e.* concentrating amplification product in one distinctive band on agarose gel) and amplification efficiency screening (*i.e.* product yield alteration).

The experiment described by Zhang *et al.* was repeated under exactly the same conditions. The enzyme for the LT-PCR constituted a mix of *Taq* and *Pfu* from the company Bio Basic Inc (Canada); the buffers were freshly prepared; the template for the reaction was lambda DNA and the primers were designed according to the sequence reported in the paper; the thermal cycling conditions as reported in the paper. Zhang *et al.* described an unsuccessful experiment when trying to amplify a 14 Kb DNA fragment from lambda DNA: a smear was observed along the gel lane. The application of specific CNTs aided significantly in retrieving a clear 14 Kb band.

When repeating the experiment of Zhang, surprisingly, we came to the opposite results: instead of a smear in the LT-PCR agarose gel run, we obtained a clear 14 Kb band. When adding CNTs in the reactions (dispersed in

water and according to concentrations described in the Zhang publication; see above) the yield of the LT-PCR reaction was actually reduced. As such, we were not able to repeat the described experiments by Zhang *et al.* Moreover, when performing this reaction with the polymerase that we were applying in our studies (Phusion High-Fidelity DNA Polymerase), we obtained clear bands of 14 Kb. Given that we used identical analytes and identical conditions, the result is rather remarkable.

Figure 15 displays a typical result while repeating the Zhang *et al.* study and is given as an example. Lane "C" (control) is an exact replication of a lambda DNA 14 Kb fragment amplification as described above. The fragment is very well defined and of high concentration on agarose gel opposed to the intensive smear obtained by Zhang *et al.* A non-gradual but very strong inhibition is visible in the other lanes (1-4) when applying 0.2 to 0.8 µg/µL of non-functionalized MWCNTs in 0.2 µg/µL increments per lane opposed to smears disappearing and yield growing in Zhang *et al.* experiments. The application of the same concentrations of functionalized MWCNT (-OH; -COOH; $-NH_2$) showed an even stronger inhibition of the reactions (figures not shown).

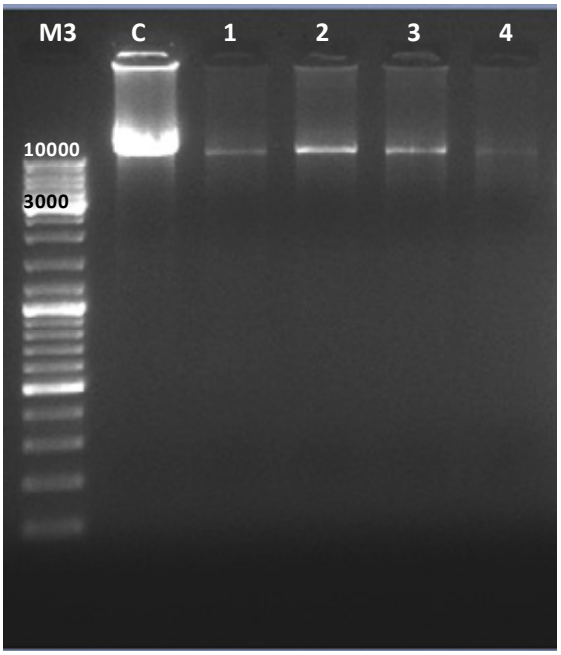

Figure 15: Repetition of Zhang et al (2008) experiment: non-functionalized MWCNT aqueous solution in water application to retrieve 14 Kb lambda DNA fragment. Lane C shows results obtained without nanoparticles; lanes 1,2,3 and 4 display results of the reactions containing 0.2, 0.4, 0.6 and 0.8 μg/μL of MWCNT respectively. M3- GeneRuler™ DNA Ladder Mix. Selected size standards in bp are displayed in the lane with the DNA Ladder.

1.11. Application of CNTs in our study to improve specificity and yield

Despite the inability of repeating the improved outcome as reported by Zhang *et al.* (2008) when applying nanotubes in the PCR reactions, but nevertheless having excellent results in our laboratory using only the polymerase, we decided to test the reported beneficial effect of CNTs in our study of Bt11 maize and to test whether switching the solvent for nanotubes may alter the results of lambda DNA amplification.

In our previous LT-PCR experiments, we have described a series of experiments on Bt11 maize which resulted in a display of high molecular smears and no defined bands of products on agarose gels. As described by Zhang *et al.* (2008), the addition of MWCNTs in the reaction helped to reduce the molecular smearing and concentrate the product in a well defined band. An experiment for screening the impact of MWCNTs on our PCR reactions (previously carried out reactions resulting in high molecular weight smearing) was designed.

The primer pair of 11GSFan2 + 11GSRan2 was used with Expand Long Template PCR system 3 from Roche to amplify a fragment of 6.5 Kb from Bt11 maize, containing the transgenic insert. Concentrations of 0.2 to 0.8 μg/μL of MWCNT in water were used as additives to the reaction. Three separate reactions with each MWCNT were performed along the control reaction. We have selected the concentrations of 0.2; 0.4 and 0.8 μg/μL of nude

MWCNTs, and MWCNTs with functional hydroxyl and amino groups, all nanomaterial was dispersed in water. The MWCNTs with functional carboxyl group was not used due to the complete inhibition of the reactions in several pre-experiment runs (data not shown).

Given the excellent results described by Zhang *et al.*, we expected that these concentrations of nanoparticles should show the disappearance of gel smears in LT-PCR reactions that were previously unsuccessful. However, though the experiment demonstrated that there was a pattern change and/or decrease in smear intensity, a clear visible PCR product of large size could not be obtained, regardless the MWCNTs applied or the concentration tested (figure 16). Several other primer pairs were tested yet the results were all the same with no large PCR products produced (data not shown). Every experiment was repeated up to three times. None of the reactions displayed the aid of MWCNTs to reduce gel smearing while concentrating the expected amplification product.

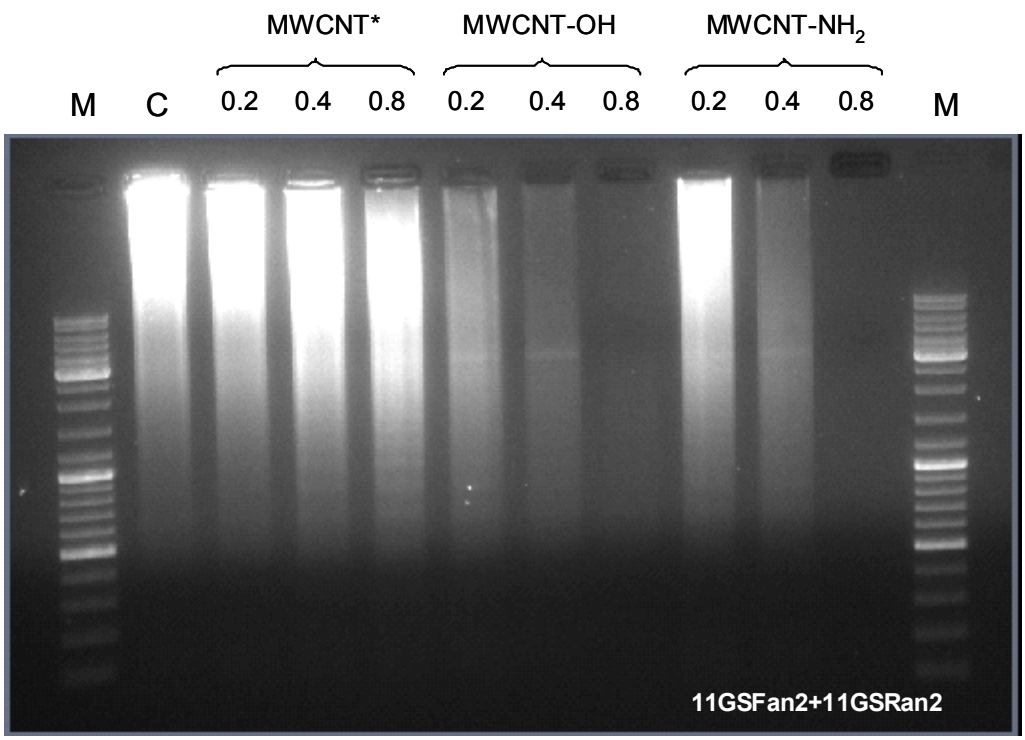

Figure 16: Application of 3 different MWCNTs to the LT-PCR reactions with Bt11 maize DNA as a template. Lane "C" displays the control reaction, being Bt11 maize amplified without nanoparticles. The concentrations applied (μg/μL) are indicated below the bracket. M- GeneRuler™ DNA Ladder Mix.

Next, we investigated the reported increased yield when applying MWCNTs. A set of reactions with either Phusion polymerase or *Taq+Tgo* polymerase blend were carried out, using lambda DNA as template and the primer pair P3+P4. These 2 polymerases (or blend of polymerases) are novel and highly proficient, and were expected to be much more progressive than the blend of enzymes used by Zhang (*Taq+Pfu*).

For this step, an additional set of the same MWCNTs dispersed in DMSO was prepared. These dispersions were of much greater stability as the DMSO is reportedly an optimal solvent for the nanotubes. The final concentration of the DMSO in the Control reactions was also adjusted to the same values of the final concentration of DMSO added with the MWCNTs. It is also important

mentioning that despite of DMSO being a well-known enhancer of PCR especially with the difficult templates, however it begins inhibiting the reaction at high concentrations.

Regardless of non-obtaining proof that MWCNT can increase reaction specificity, we have raised a hypothesis about the case-dependent product yield increase *i.e.* in some LT-PCR reactions MWCNT may facilitate better outcome of the reactions (only applicable to the reactions with Phusion polymerase and *Taq*+*Tgo* polymerase blend).

Indeed, after repeatedly carrying out the reactions, in several cases the concentration of the resulting product was visibly higher on agarose gel than in the control reactions. Due to this unusual observation, we have designed several experiments to see whether the increase in the yield is a constant and repeatable result. The results of the reactions are listed in table 10 and displayed in figure 17. Note, that three repetitions of these reactions were carried out with slight deviations to the results (see text below).

Table 10: Yield alterations obtained with MWCNT application on LT-PCR

Phusion DNA polymerase + MWCNTs solubilized in water	
MWCNT-Nude	Increase with Cf 0.2-0.4 μg/μL
MWCNT-OH	Inconsistent inhibition
MWCNT-NH_2	Gradual inhibition

Taq and *Tgo* DNA polymerases + MWCNTs solubilized in DMSO	
MWCNT-Nude	Increase with Cf 0.2-0.4 μg/μL
MWCNT-OH	Increase with Cf 0.2 μg/μL
MWCNT-NH_2	Increase with Cf 0.2-0.4 μg/μL

All experiments with DMSO had a final concentration of 2% of this reaction enhancer. The control reactions had the same concentration of DMSO for a

maximum precision. As seen from the table the results obtained were ambiguous. In most cases a gradual increase of product was observed when the polymerase *Taq* and *Tgo* was used with increasing concentrations of the CNTs (from 0.2 to 0.8 μg/μL). The highest applied concentration however showed a clear decrease of product. The observed increase did not seem to be different from the type of CNT used. All experiments were repeated three times and the positive results were obtained at least twice, sometimes 3 times. The observed inconsistency however of the obtained results show that agglomeration and/or other features of CNTs may affect the reactions.

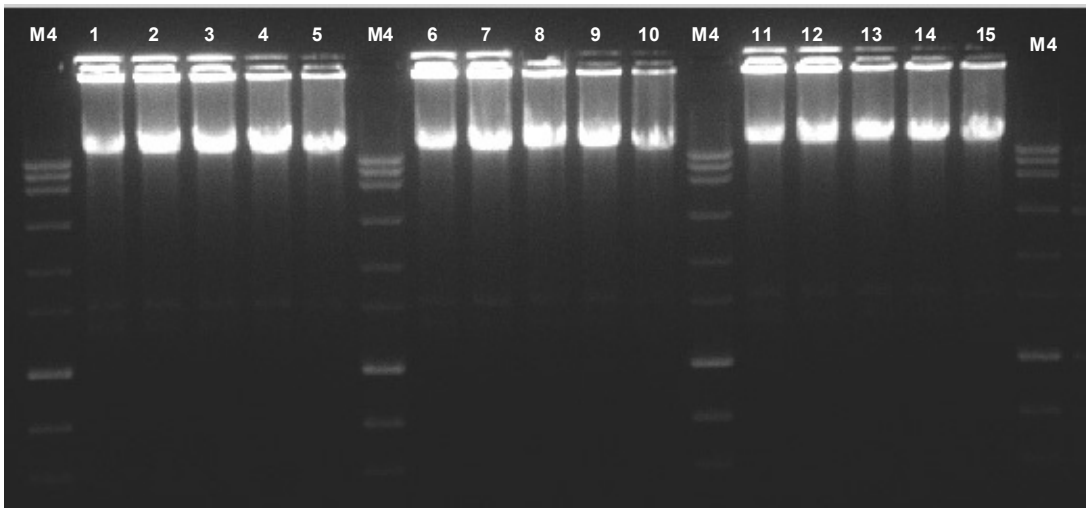

Figure 17. MWCNT in 100% DMSO solution application on LT-PCR for retrieving a 14 Kb lambda DNA fragment with *Taq*+*Tgo* polymerase blend. Lanes 1, 6 and 11 are controls without nanoparticles nor DMSO; lanes 2, 7 and 12 contain 2% of DMSO and no nanoparticles. Lanes 3, 4 and 5 contain 0.2, 0.4 and 0.8 μg/μL of non-functionalized MWCNT; lanes 8, 9 and 10 contain 0.2, 0.4 and 0.8 μg/μL of MWCNT-OH; lanes 13, 14 and 15 contain 0.2, 0.4 and 0.8 μg/μL of MWCNT-NH_2. M4- MassRuler™ Express Forward DNA Ladder Mix.

At the very end of the investigation of carbon nanotubes as a potential enhancer to PCR we have concluded that the benefit ratio against the

difficulties of the preparation and application of the material is very unfavorable. The mostly negative outcome of this investigation also raises questions as to which extent are some data published in peer-reviewed journals reliable. The failure in repeating the unsuccessful results of Zhang *et al.* might raise doubts if the results published have been trustworthy and genuine. Our observation about the yield increase in some PCR reactions when adding the MWCTs however shows that there is still work to be done with these applications of nanomaterial to PCR.

1.12. LT-PCR application on GMO templates

LT-PCR was tested on material of the event H7-1 sugar beet. As we did not have any previous experience with material of this crop, nor the genetic make-up of this plant, we firstly looked for reference genes. Two genes were chosen from the mitochondrial genome as this is the only source so far for public available DNA sequences: *cox2* and *rrn26*. PCRs targeting these genes were used to test the quality of the extracted DNA.

Three types of material were considered: CRM, ground seed and fresh leaves. DNA was extracted from all three without any major problems. We did observe a strange banding pattern for DNA extracted from CRM, but subsequent conventional PCRs showed that the DNA was suitable for further experiments. Therefore, final LT-PCR experiments were carried out with DNA extracted from CRM. A 3.4 kb PCR fragment was isolated, covering most of the H7-1 insert. It was successfully re-amplified to a suitable quantity, using the same primer pair of the first PCR reaction and purified. The fragment was submitted for a DNA sequence analysis via primer walking. It was the first time to use this type of sequencing, performing reactions on an isolated large PCR fragment, and, carrying out primer walking (sequencing via designing primers on the basis of previously obtained sequence) to cover the entire fragment. This procedure proved to be very efficient: rather quick and good quality DNA sequences.

The obtained DNA sequence for the insert matches almost perfectly the reported DNA sequence of H7-1 in the GenBank (acc. No. DD417675, Glyphosate tolerant sugar beet). We observed one extra nucleotide in the very end of the fragment. This difference however is most likely due to a sequencing mistake as it concerns one additional nucleotide, A, in a stretch of 7 As. The DNA sequence obtained covers the entire fragment except the very 3' end. Although our aim was to use primers directed towards both flanking regions of the insert, it was impossible to design a suitable primer in the limited DNA sequence that is available for the 3' end (40 bp). Few primers were tested but failed what was expected given that the region was AT rich. Therefore, as the best next alternative, a primer was designed in the T-DNA region, immediately upstream of the flanking region. The primers, directed to this T-DNA region, in combination with the 5' directed primers, provided successful LT-PCRs.

Analyzing the insert of the H7-1 sugar beet as one complete fragment, applying the technique of Long Template PCR, followed by a DNA sequence analysis via primer walking (performing sequencing on the basis of obtained sequences) proved to be an efficient procedure: not difficult, quick, good quality of DNA sequences. With this procedure, the obtained sequence result will not be affected by potential contaminations of the material with other and similar events nor by the presence of additional copies or additional inserts. This procedure could be recommended for analyzing GM-inserts. It needs to be noted that the technique of LT-PCR may not always work to retrieve the entire insert, and needs to be assessed on a case by case study. LT-PCR may not be successful due to secondary structures of the flanking regions (preliminary data). The size of the large PCR fragment to be produced may also pose limits on what can be achieved.

Next we have successfully applied a method long template PCR to isolate a large PCR fragment covering nearly full Bt11 maize transgenic insert. The technique allowed us to retrieve trustworthy results and optimize the

experiment conditions for a relatively difficult template DNA, moreover, a proof-reading DNA polymerase Phusion (Finnzymes) was used for the experiments (as opposed to the *Taq+Tgo* polymerase blend by Roche, used in H7-1 sugar beet study), assuring the accuracy of the data retrieved after sequencing analysis. All maters to avoid mutagenicity and damage to the DNA were applied, preventing UV-light damage and dye intercalation in the DNA.

A long DNA fragment has been successfully cloned into a linear cloning vector and productive bacterial transformation resulted in a fair copy plasmid DNA.

Three types of material for obtaining DNA matrices were tested: CRM, grounded seed and fresh leaves. DNA was extracted from 5% CRM material and Bt11 plant leaves, laboratory stock of previously extracted ground seed DNA was also tested. Good quality of all preparations was monitored although experiments on 5% CRM resulted in very low yield of the product therefore fresh leaves DNA was used for further conventional PCR and LT-PCR experiments. Two overlapping fragments of 1722 and 5902 bp were isolated. PCR purified fragment of 1722 bp and LT-PCR retrieved 5902 bp fragment in pJAZZ-OK linear plasmid were sent to perform primer walking reaction to cover the entire insert. Sequencing procedure was successful resulting in only 8 nucleotide mismatches aligning the JRC service sequence and the patent sequence with our sequence results after assembly. Second strand sequencing reactions for verifying the disambiguation were not performed.

One forward primer in 5' flanking DNA and one reverse primer in 3' flanking DNA were paired up with 2 primers inside the GM insert to cover all the sequence and 2 overlapping fragments carrying the full Bt11 maize transgenic insert were retrieved. All primer combinations to isolate an uninterrupted fragment from 5' to 3' flanking DNA regions were unsuccessful therefore sequence analysis was performed on 2 overlapping DNA fragments.

All the public available information on event Bt11 maize was collected prior to experimental data and analysis was performed. The experimental data confirmed that the data published on publicly available Chinese GMO Method Detection Database (GMDD) was incorrect.

For the final run of LT-PCR testing on the GMO matrices we have chosen a well-described GM event MON810 maize. As being one of the cultivation approved GMOs in EU, this MON810 went under a very strict, precise and random investigation providing a lot of information that was gathered and overviewed before our experiments. The DNA was extracted from 5% GM content CRM and four pairs of primers were designed to obtain the full uninterrupted DNA fragment carrying the transgenic insert. In this part of the LT-PCR study, the significance of correct primer design was tested as a well-characterized MON810 did not need a re-verification of the nucleotide sequence. The LT-PCR experiments with all four primer pairs resulted in a successful retrieval of the full transgenic insert as a DNA fragment, however the product yield was notably different throughout the reactions even though the primers that were used were designed within a very close proximity from each other. The difference of obtained product yield has proven that the primer design is crucial in the successful LT-PCR application and that the success rate depends on case by case basis.

It is common to consider PCR additives whenever a failure to obtain well-concentrated bands on agarose gel after running the amplification product. Many commercial producers add some standard additives (commonly DMSO) or altered ingredients (GC-rich buffer system) when selling the PCR kits, however this may be insufficient in some cases. In this study we also approached the concerns of PCR additives testing the up-most novelty, the nanomaterial Carbon Nanotubes. This extremely hydrophobic and reactive in air hollow carbon grid-tube-like particles had a surprising presentation of interaction with the PCR chemistry (particularly *Taq* DNA polymerase and DNA template) and proven alteration of the reaction specificity and increased yield of product. It is therefore that we designed and ran the experiments of MWCNTs as the LT-PCR additive. Even though we have overcome the extremities in preparation of the dispersed MWCNTs in water to obtain a fairly stable working solution for application, we have noticed some convincing

inconsistencies in other authors work *i.e.* Zhang *et al.* (2008) study. Little believable was the preparation of working solutions step of the study group and even less informative personal communication was achieved after we have performed the exact copy of the experiment using the identical materials as in original study. In fact, our study has shown that the 14 Kb lambda DNA fragment is easily obtainable via a simple LT-PCR application and not requiring any additives. We however have noticed that there is a pattern of product yield increase when using the LT-PCR enzymes that were commonly used in our laboratory *i.e.* the Phusion enzyme and *Taq+Tgo* blend. The testing of this yield increase was repeated several times due to inconsistency in results. These discrepancies appeared due to stability issues of the nanoparticle working solutions however we have obtained some proofs that the carbon nanotubes may increase the yield of the LT-PCR product. This result was obtained with the MWCNTs dispersed in water and DMSO as a chemical solvent (however also a stand-alone PCR additive).

We continued our study to improve the quality of the amplification of Bt11 6.5 Kb sequence which previously ended up in a high molecular smearing due to a particularly difficult secondary structure-prone short sequence at the end of the target sequence. MWCNTs were added to the reaction mixtures to reduce the smearing and obtain the well concentrated bands on agarose gels. This application as well proved to be unsuccessful. The high molecular smearing indeed was reduced, however we have not obtained the visible DNA fragment as such we could not explain if the smearing was happening due to reaction inhibition or it was the DNA folding effect and why was the amplification product not concentrating on the agarose gel.

2. GENOME WALKING STRATEGIES TO OBTAIN UNKNOWN DNA

2.1. RADE and LT-RADE, a 5'RACE-based genome walking procedure

The complexity of the work with GMO molecular characterization calls for a need to have an easy-applicable method for obtaining unknown DNA sequences. Even though a large variety of genome walking methods is available (see above), most of them are dependent on the enzymatic digestion or degenerate primers. This dependency means that the overall success of the selected genome walking method will only rely on the restriction sites appearing inside or in a close proximity to the transgenic insert; or the primary product retrieval will entirely rely on the success of annealing of degenerate primers. While performing the literature search, we have discovered that a very small selection of genome walking methods do not require enzymatic digestion or degenerate primers throughout the procedures. In this study we modified and optimized the method described by Rudi *et al.* (1999) and Leoni *et al.* (2008). In short, their approach is based on a 5' Rapid Amplification of cDNA Ends (5'RACE) protocol which is normally used to obtain full length sequence of RNA transcripts. Due to the method's relation to the 5'RACE method, we name it Rapid Amplification of gDNA Ends (RADE). For the convenience we have divided the method into five different steps (requiring distinct molecular biology manipulations), including three amplification steps, column purification and a homopolymeric tailing. The schematics of the method are displayed in Figure 18. After successful development of the modified version of Rudi/Leoni method we have further optimized it for retrieval of longer fragments merging the RADE method with the LT-PCR method.

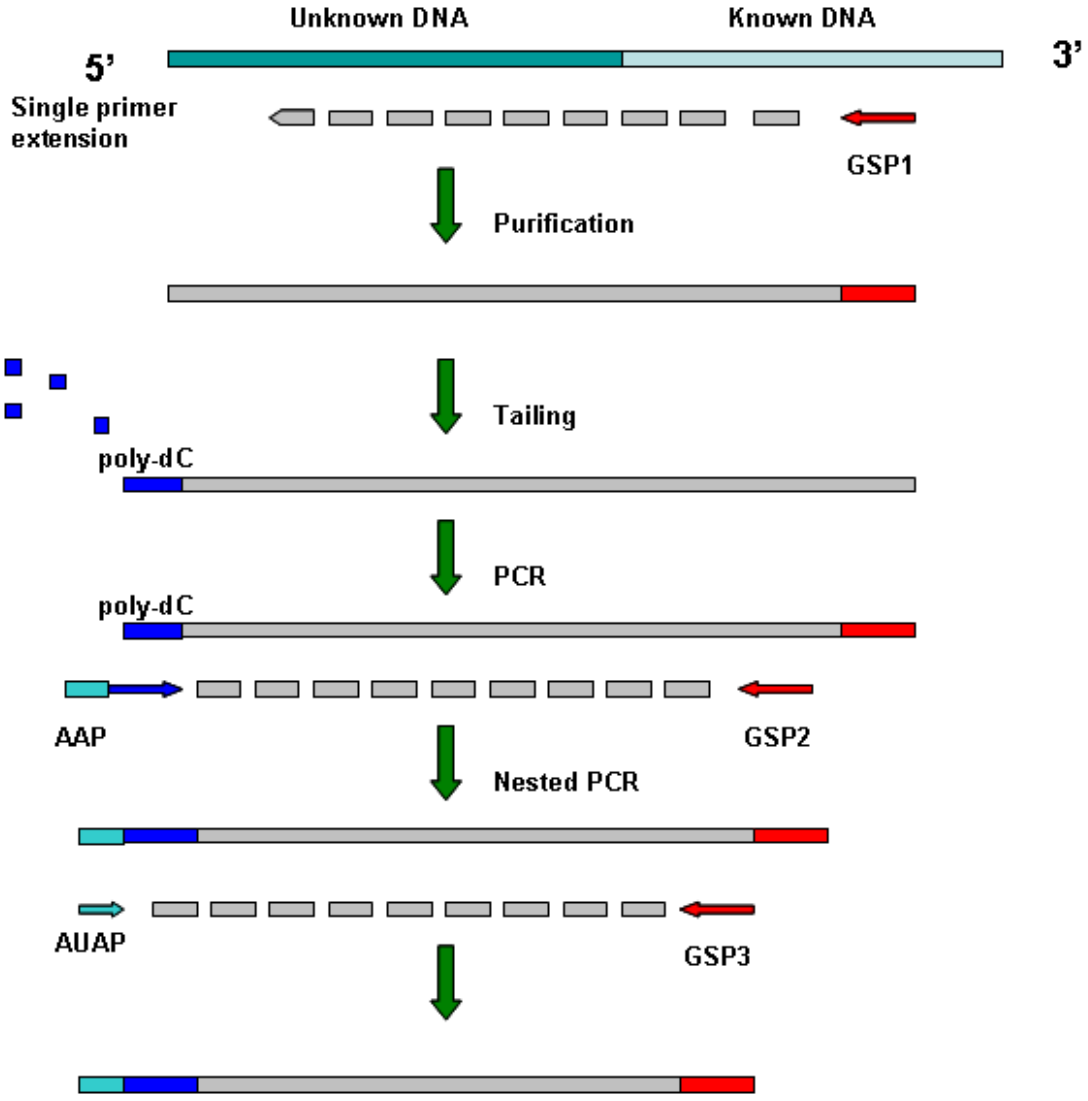

Figure 18: Schematic diagram of RADE and LT-RADE methods. Procedure steps displayed starting with the single strand extension and directed to the unknown DNA, followed by the column purification of the ssDNA fragment, homopolymeric tailing to create an adapter priming site and two subsequent nested adapter PCRs. **GSP1,2,3** – Gene Specific Primer; **poly-dC** – polydeoxycytidylic "tail"; **AAP** – Abridged Anchor Primer; **AUAP** - Abridged Universal Amplification Primer.

The first step is the elongation of a single stranded sequence, starting from one single primer. The GSP1 primer is used to amplify the ssDNA fragment of

interest starting from a known DNA sequence in the gDNA template. This step is crucial as the concentration of the product has a decisive role of the methods' further success *i.e.* the adapter anneals and the nested PCR is performed on purified product obtained via single primer extension. The study of Leoni *et al.* (2008) showed that the best product yield is obtained with 60 pmoles of primer. The next step is a column purification of the obtained product to remove the excess nucleotides and residual GSP1. The column purification of the product is needed for the excess GSP1 primer not to compete as a template in the addition of poly(dC) tail to the ends of target ssDNA molecules. The subsequent step of homopolymeric tailing with terminal deoxynucleotidyltransferase (TdT) adds a poly(dC) tail at the 3' end of the ssDNA fragment. This tail will serve as a primer landing site for the Abridged Anchor Primer (AAP) which carries a region of poly-dG (the constitution of primer used for adapter PRC and subsequent nested PCR are displayed in figure 19). The next two steps are amplification reactions resulting in dsDNA products. The first PCR constitutes a nested PCR and adapter PCR simultaneously. It is performed with a second GSP (GSP2) paired with the AAP which lands on the poly(dC) tail. Then a second nested PCR is carried out with a third GSP (GSP3) paired with the Abridged Universal Amplification Primer (AUAP; see figure 19). The latter primer has the same DNA sequence as AAP, missing only the poly(dG) stretch. The last two steps ensure that a specific product is obtained, and also with a sufficiently high yield allowing detection and isolation of the amplification product from an agarose gel.

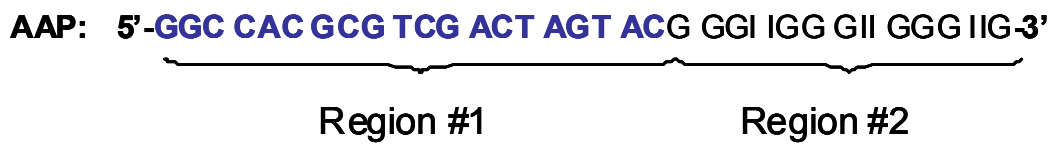

Figure 19: Constitution of the primers AAP and AUAP. Region #2 of AAP lands on poly(dC) tail and primes the reaction in a pair with GSP2; Region #1 of AAP carries an annealing site for next step nested PCR reaction primer AUAP that is exactly as AAP but without region #2. I stands for the nucleoside inosine which indiscriminately forms pairs with A, T or C.

To gather as much information as possible on the insertion site or the flanking regions of a specific target gene, it is desirable to obtain the longest possible fragments for a further DNA sequence analysis. Within RADE, the initial version of our genome walking method, a standard *Taq* polymerase is used which may not yield PCR products of long length. We therefore substituted the *Taq* polymerase of RADE with a polymerase mixture (a standard and a proofreading polymerase) used in PCR applications for obtaining long PCR products (see materials and methods). The main characteristic of a proofreading polymerase is the removal of misincorporated nucleotide thereby allowing the reaction to proceed at normal speed rather than to continue slowly or stop altogether as happens in the reactions with single non-proofreading enzyme. It is reported that adding a proofreading enzyme in a mix with *Taq* DNA polymerase greatly stabilizes the enzymes on the template DNA thus increasing processivity (Barnes 1994; Cheng *et al.* 1994).

We chose Expand Long Template PCR system (Roche) for LT-PCR applications. The system has a unique enzyme blend of thermostable *Taq* and *Tgo* DNA polymerases of which the latter one has a proofreading activity. Different buffers available in the kit were tested: No.1 with $MgCl_2$

concentration of 1.75 mM, and No.3 with 2.75 mM of $MgCl_2$ including a detergent. In all the steps that implied amplification, the polymerase and the buffers were accordingly substituted. In contrast to the RADE procedure, for LT-RADE in the first step (single primer extension) the elongation time is increased from 1 to 2 minutes.

2.2. Molecular characterization of the GM maize MON810 insert region

2.2.1. Isolation of MON810 5' and 3' flanking sequences

RADE and LT-RADE were applied to retrieve the plant flanking regions on both sides of the transgenic insert of maize event MON810. This insert consists of cauliflower mosaic virus promoter (P-35S) linked to a heat-inducible enhancer fragment (*hsp70*) and a δ-endotoxin coding sequence (*cry1Ab*) (Hernández *et al.* 2003) (Figure 20).

Limited availability of nucleotide sequence data for GMOs requires that an in-depth study of all available information is performed prior to primer design and PCR. The DNA sequence of the maize MON810 insert was assembled as a contiguous sequence from three overlapping sequences from GenBank: acc. No. AF434709 (Holck *et al.* 2002), acc. No. AY326434 (Hernández *et al.* 2003) and acc. No. AM749998 (Rosati *et al.* 2008). The 6401 bp fragment contains 803 bp of the plant 5' flanking sequence followed by 3591 bp of the transgenic insert and 2007 bp of the plant 3' flanking sequence (see figure 20). Another available online source for obtaining MON810 maize transgenic insert sequence was the Chinese GMO Detection Method Database (GMDD), however the database is not regularly updated thus contains a contiguous sequence of previously mentioned Holck *et al.* and Hernández *et al.* sequences lacking Rosati *et al.* sequence. The figure 20 shows the synthesis of the data of previously published MON810 maize studies, particularly the parts of transgenic insert that were aimed to retrieve and the methods that were used.

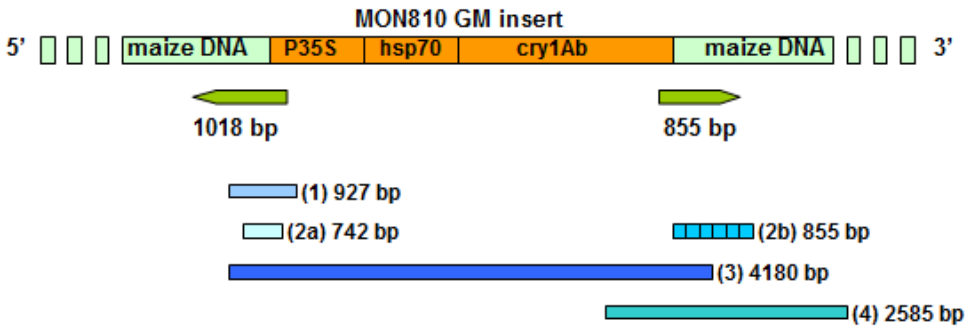

Figure 20: The longest fragments obtained by LT-RADE on 5' and 3' flanking sequence of MON810 GM insert are displayed in dark green arrows. Overview of previously published MON810 sequences and the techniques for obtaining them: (1) Holck *et al.* (2002) obtained by Ligation Mediated PCR; (2a) La Paz *et al.* (2010a) obtained by PCR as a part of mismatch endonuclease assay; (2b) La Paz *et al.* (2010b) 8 fragment consensus; fragments obtained by RACE 3'PCR; (3) Hernandez *et al.* (2003) obtained by Long Template PCR; (4) Rosati *et al.* (2008) obtained by PCR.

Applying RADE, two sets of primers were designed, both directed outward from the transgenic insert (table 7). Genome walking at the 3' end of the insert was carried out with a set of three nested gene specific primers, located at the very end of the cry1Ab gene and directed to the 3' plant flanking sequence of the MON810 insert (MON81F1, MON810F2 and MON810F3). The polymerase used in the different steps of the procedure was a standard *Taq* polymerase with the property to amplify gDNA fragments up to 3 kb in length (as described in the manufacturer's manual). Positive control reactions were included containing the purified ssDNA as template in a combination of the GSPs with a reverse primer (3PL2-R) located in the known 3' plant sequence in close proximity to the maize MON810 insert, yielding a fragment of known size (676 bp). This fragment was expected to decrease in length throughout the procedure as nested primers were used in the subsequent steps (not shown). At

step 4 (first nested PCR) agarose gel analysis did not yield a well-defined band due to the low concentration of obtained fragment, but at the end of the 5 step procedure, a fragment of approximately 550 bp was observed (see figure 21). This fragment was isolated by pCR2.1-TOPO cloning. Three independent clones were submitted for double strand sequencing of their insert, each harboring an insert of a slightly different size (456, 518 and 564 bp)..

For the analysis of the 5' flanking region of the maize MON810 insert, the gene specific primers P35SRA, P35SRB and P35SRC were used (see materials and methods). These primers are located at the very 5' end of the MON810 insert (in the P-35S) and directed to the 5' plant flanking sequence. The positive control included a reaction with a forward primer (5PL4-F) located in the 5' flanking plant sequence near maize MON810 insert, yielding an expected fragment of 434 bp in length (not shown). The five-step RADE procedure yielded two fragments of approximate sizes of 450 bp and 600 bp (see figure 21). Upon cloning of each fragment, three independent plasmids were submitted for double strand sequencing. DNA sequence analysis of the six fragments showed a slight difference in fragment sizes (342, 413 and 577 bp respectively) (data not shown).

For LT-RADE analysis of the flanking regions of the maize MON810, the same set of primers and the same controls as in RADE were applied. For the 3' flanking sequence analysis, we observed several fragments of which the largest length was approximately 900 bp. Better defined fragments were obtained when applying buffer No.1 instead of No.3. For the 5' flanking sequence analysis, a fragment of approximately 1050 bp was observed, with a better yield in buffer No.3 than with buffer No.1. Apparently the buffer compositions may help in a better yield of products (see figures 20 and 21). Sequence analysis of both fragments (for each 3 clones submitted for sequencing), showed that slightly different sizes were obtained (842, 849 and 855 bp for the 3' flanking region and 1018 bp for the 5' flanking region) (data not shown).

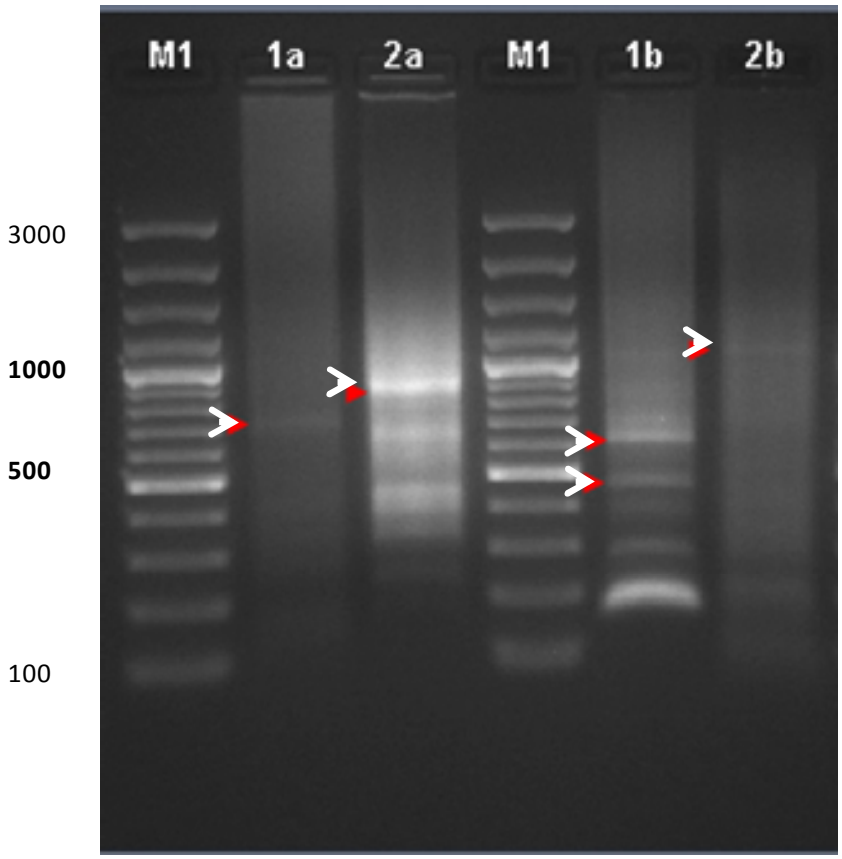

Figure 21: Gel electrophoresis of products obtained with RADE and LT-RADE. Fragments obtained in MON810 study. Arrows represent the fragments that were purified from 1% agarose gel, cloned and sequenced. 1a- pattern obtained by RADE on 3' flanking DNA; 1b- RADE on 5' flanking side; 2a- LT-RADE on 3' flanking side; 2b- LT-RADE on 5'flanking side. M1 stands for molecular standard GeneRuler™ 100 bp Plus DNA Ladder (Fermentas, Lithuania). Selected size standards in bp are displayed next to the lane with the DNA Ladder.

After the sequencing of the isolated bands a paradox was observed: the length of the insert in every plasmid was varying by several to the maximum of several dozens of nucleotides. The alignment study has shown that the fragments are a little longer or shorter on one side thus it was evident: the *Taq* polymerase seems to drop-off at various points from the template. This is happening in the first step of single primer elongation, these fragments are then

purified and the adapter is added and later on, the size difference is too small to be noticeable on the agarose gel when the first dsDNA product is obtained. After careful analysis of the nucleotide sequence, no particular constitution of the drop-off points was observed: no G/C or A/T rich regions or evident segments that are prone to form secondary structures in DNA strand were present. The largest fragment obtained by walking on the 3' end of the maize MON810 insert was of 564 bp in length; the fragment retrieved from the 5' end was 577 bp long. It is commonly known that *Taq* DNA polymerase can amplify up to 2000 bp per minute at its optimum temperature. Given that a one minute extension time had been applied, fragments of 1000 bp or longer were expected. However, the fact that smaller fragments were obtained may point to a premature drop of the *Taq* DNA polymerase from the template DNA. As the initial step of the RADE and LT-RADE deals with the single primer extension and is not limited at the other side with any stop signal, we found it normal to see minor length differences at one side of the amplified fragments.

The objective to optimize the developed genome walking method via merging it with LT-PCR was achieved. Considering the obtained product lengths, our data indicate clearly that LT-RADE performs better than RADE. Repeatedly, the application of a polymerase blend of *Taq* and *Tgo* enzymes (in LT-RADE) yields fragments which are about twice longer (see table 11). In this way, valuable additional DNA sequence information becomes available to find homologous sequences in databases.

Table 11: Size of DNA fragments obtained with RADE (using *Taq*) and LT-RADE (using *Taq* + *Tgo*) in MON810 maize

	Taq	***Taq* + *Tgo***
5' obtained flanking DNA	577 bp	1018 bp
3' obtained flanking DNA	564 bp	855 bp

2.2.2. Bioinformatics analysis of the MON810 5' and 3' flanking sequences

A BLAST analysis of all fragments showed that both flanking regions of GM event MON810 show high similarity to maize genomic DNA. Given that LT-RADE yielded longer fragments, we will discuss here only the analysis of those kind of DNA sequences (see table 12). The 855 bp maize fragment obtained at the 3' insertion site of the MON810 insert showed a 100% identity match with the sequence reported by Rosati *et al.* (2008), describing the 3' insertion site of the same event (acc. No. AM749998). The 1018 bp maize fragment obtained at the 5' insertion site of MON810 maize showed significant similarity to several maize clones (the highest similarity of 87% was found with BAC clone ZMMBBb0448F23, acc. No. AC160211).

Comparing the obtained flanking regions with sequences reported in the corresponding patent (SEQ 3 of 566 bp 5' flanking region and SEQ 4 of 879 bp 3' flanking region; Levine *et al.* 2004), a near perfect DNA sequence alignment could be demonstrated (two additional nucleotides were observed in our sequence but missing in SEQ 4; the obtained sequence and SEQ 3 were 100% identical). While for the 5' region a 100% identity was found, in the 3' region 2 additional bases were present in the sequence obtained in this study which were also found in the sequences by Rosati *et al.* The sequences found in our study at both the 3' and 5' junction of the insert were identical to the sequences described by Levine *et al.* (2004).

Sequence alignment of the obtained flanking regions with the contiguous MON810 sequence revealed an additional 139 bp at the 5' end that have not been reported so far. Similarity analysis showed significant matches with maize genomic DNA (see figure 22).

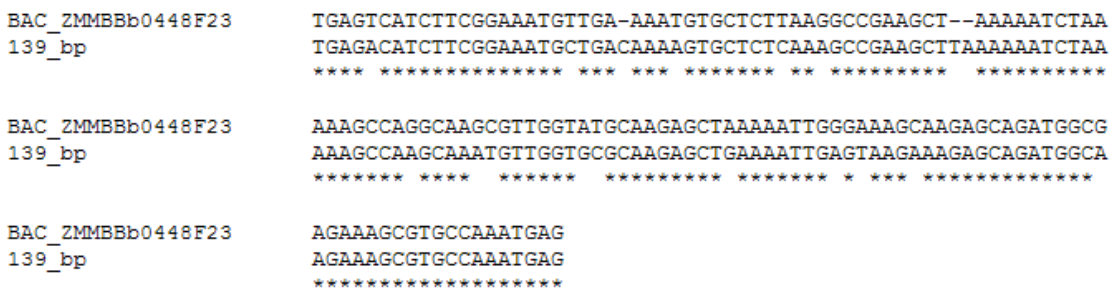

Figure 22: DNA sequence alignment of newly obtained flanking region of maize MON810. The alignment between the novel 139 bp DNA sequence obtained at the 5' flanking site of the maize MON810 insert and a part of BAC clone ZMMBBb0448F23 (acc. No. AC160211) sequence.

BLAST searches were performed to assign the identified MON810 insertion to a specific maize chromosome. Similarity searches for the obtained maize genome sequence at the 3' side of the MON810 insert point to chromosome 5. Also Rosati *et al.* assigned the matching BAC clone ZMMBBc0409B05 to chromosome 5. We have aligned our sequence of the 3' side with this reported BAC clone and found a 96% match (nucleotide locations from 93984 to 94796). On the contrary, BLAST analysis of the 5' genomic maize regions obtained in our study did not demonstrate any significant homology to that particular BAC clone. It has been suggested that the MON810 transformation event implied some rearrangements based in particular on the partial loss of the T-nos terminator of the transformation vector (Hernández *et al.* 2003). It seems not unlikely that also at the 5' end of the insert rearrangements have occurred. DNA sequence alignment may then not be feasible to find homologous maize genome sequences as such. Such rearrangements as T-DNA backbone insertion (Kohli *et al.* 2003) or chromosomal DNA deletions (Latham *et al.* 2006) are known and well described in literature. The fact that only one side of the MON810 maize flanking sequence aligns with an isolated BAC clone of

approximately 200 Kb may indicate that indeed severe rearrangements at the insertion site occurred during the transformation.

Table 12: Alignment analysis of obtained sequences using BLAST alignment tool

		Query coverage (nt)	Query coverage	Max ident	E value
MON810					
5'	Genomic sequence for *Zea mays* BAC clone ZMMBBb0448F23, complete sequence (acc. No. AC160211)	1-935	91%	87%	0.0
	Synthetic construct transgenic *Zea mays* with Cauliflower mosaic virus genomic sequence (acc. No. AF434709)	936-1018	9%	100%	2e-35
3'	*Zea mays* 3' insertion site between truncated cryI(A)b transgene and Zm-upI gene, isolate fragment 1 (acc. No. AM749998)	1-855	100%	100%	0.0

2.3. Characterization of the GM rice LLRICE62 insert region

The insert of the GM event LLRICE62 consists of cauliflower mosaic virus promoter (P-35S) linked to a glufosinate ammonium tolerance (*bar*) gene and a cauliflower mosaic virus signal terminator (T35S) sequence (Bayer Crop Science, 2003) (figure 23 (A)). The full length sequence of LLRICE62 rice

transgenic insert is not publically available, and the corresponding DNA sequence was deduced from available information of the genetic elements present in the insert. Given that this event contains a promoter P-35S, as also event MON810, the previously designed set of gene specific primers P35SRA, P35SRB and P35SRC were tested to obtain the region flanking the promoter at the 5' side. To analyze the 3' flanking region of the LLRICE62 insert, novel primers needed to be designed for the T35S sequence, a 3' located genetic element of the insert. Herefore, we used the available T35S sequence from GenBank for nested primer design (acc. No. GQ497217): LL62_TF1, LL62_TF2 and LL62_TF3.

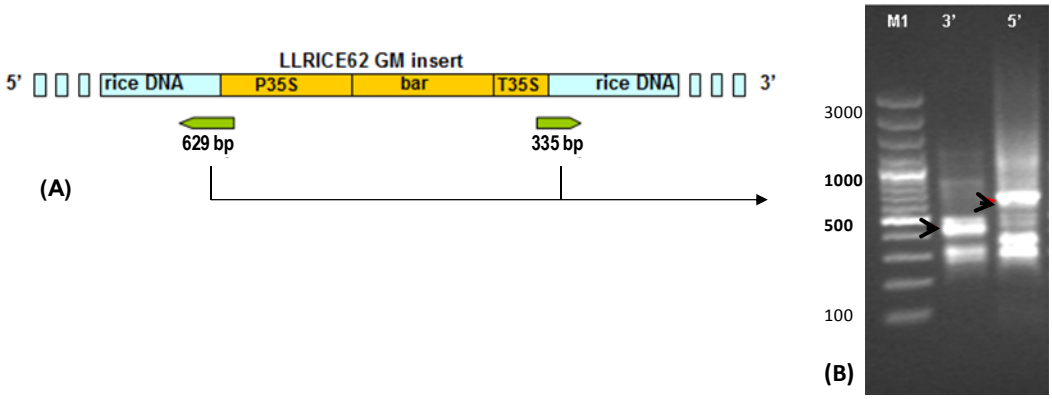

Figure 23: The longest fragments obtained by LT-RADE on 5' and 3' flanking sequence of LLRICE62 GM insert. (A) Schematic view displaying the transgenic insert elements; (B) gel electrophoresis of products obtained with LT-RADE on 3' and 5' plant flanking sequences. Arrows represent the fragments that were purified from 1% agarose gel, cloned and sequenced. M1 stands for molecular standard GeneRuler™ 100 bp Plus DNA Ladder (Fermentas, Lithuania). Selected size standards in bp are displayed next to the lane with the DNA Ladder.

Given that LT-RADE yields longer fragments than the RADE method, as observed in the MON810 insert analysis, we immediately applied LT-RADE for the investigation of LLRICE62. In contrast to the MON810 analysis, we

obtained several fragments of various lengths from both sides of the insert: the longest fragment at the 5' side was approximately 700 bp long and at the 3' side a fragment of approximately 500 bp (figure 23 (B)). Both the largest fragments were purified from agarose gel for direct sequencing reactions (a 629 bp fragment from the 5' flanking side, a 335 bp from the 3' flanking side, as deduced from the DNA sequence analysis).

Blast analysis of the 629 bp fragment (5' flanking side) showed that the most distal 5' part of the fragment (216 bp) had a 99% identity with rice genomic DNA (Oryza sativa Japonica Group genomic DNA, chromosome 6, BAC clone OSJNBa0029G06 (acc. No. AP004680), while the following 407 bp showed significant similarity to P-35S (acc. No. Z37515) (table 13, figure 24).

The 335 bp fragment obtained at the 3' side of the LLRICE62 insert showed for 109 bp 100% identity to T35S (acc. No. GQ497217), followed by 226 bp of rice genomic sequence (BAC clone: OSJNBa0029G06, acc. No. AP004680). By connecting the DNA sequences of both flanking regions (cleaned from the insert sequences), the insertion site could be analyzed. Similarity searches revealed a 96% match over the full length of the 487 bp fragment, confirming indeed that the insertion had taken place on chromosome 6 with a clear deletion of 18 bp at the site of integration (see figure 24).

Figure 24: Alignment between the sequences of rice event LLRICE62 insertion site. Sequence obtained with LT-RADE application is marked as 'study results' is aligned with BAC clone OSJNBa0029G06 (acc. No. AP004680) sequence retrieved from BLAST analysis.

Table 13: Alignment analysis of obtained sequences using BLAST alignment tool

		Query coverage (nt)	Query coverage	Max ident	E value
LLRICE62					
5'	*Oryza Sativa* Japonica Group genomic DNA, chromosome 6, BAC clone: OSJNBa0029G06 (acc. No. AP004680)	5-221	34%	99%	4e-106
	Binary vector BinHygTOp	222-629	64%	100%	0.0

	aph, tetA, tetR, traJ, insB, insA, aphA-3 and trfA genes (acc. No. Z37515)				
3'	*Glycine max* transgenic GMO cassette genomic sequence (acc. No. GQ497217)	1-109	29%	100%	1e-48
	Oryza Sativa Japonica Group genomic DNA, chromosome 6, BAC clone: OSJNBa0029G06 (acc. No. AP004680)	108-374	71%	100%	2e-136

2.4. Application of the LT-RADE genome walking method, based on CaMV P-35S, for 5 different genetically modified (GM) crops

2.4.1. Rationale for the usage of P-35S as a GMO target in LT-RADE

One of the critical features in molecular characterization studies of the biotech plants is the constitution of the transgenes *i.e.* the genetic elements comprising the full inserted sequence. One of such elements is the cauliflower mosaic virus promoter (P-35S). Since P-35S was initially described as a model plant nuclear promoter system (Odell *et al.* 1985), many thorough studies of this promoter's activity were carried out (Odell *et al.* 1988; Ow *et al.* 1986; Assaad and Signer 1990; and other studies). P-35S soon became one of the most commonly used promoters for chimeric gene constructs designed for expression in plant cells (Assaad and Signer 1990) Consequently it became the most common element in GMO constructs (Querci *et al.* 2010). To this day the P-35S is the most repeatedly used promoter in transgenic plants. From 46 events listed as approved or tolerated in the European Union, about one third

contains a P-35S or an enhanced version P-E35S as genetic element (Van den Eede 2010) (figure 25). Also, considering the information provided on field releases of GMOs in the EU, the so-called Summary Notification Information Form's, the P-35S promoter will still be applied in many future GMOs.

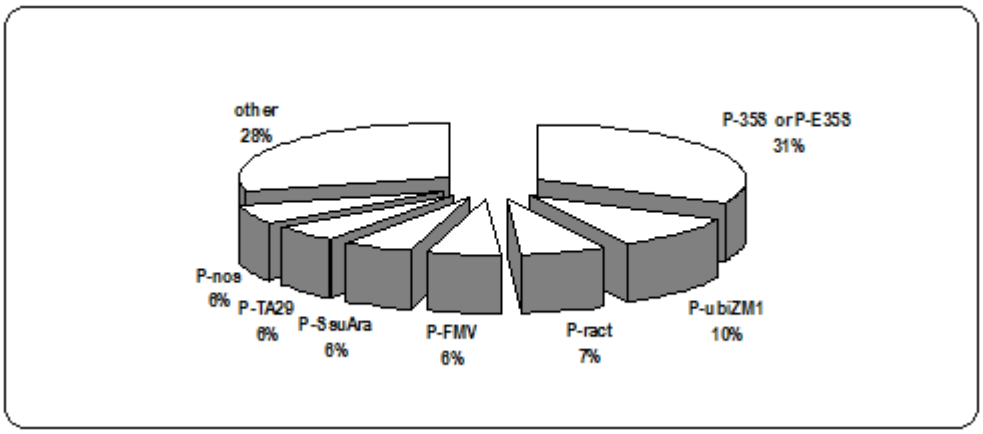

Figure 25: Promoters in EU approved or tolerated GM events. The statistical data for this graph was collected from Van den Eede (2010).

P-35S is of crucial importance when analyzing the unknown and/or unauthorized genetically modified plants (GMP) (Cankar *et al.* 2008). The potential of having a testing system that may allow a relatively easy retrieval and identification of any sequence adjacent to the P-35S brings much attention to this single element of many transgenic inserts; in fact this promoter is already the most characteristic target for GMO detection methods. Studies report on different methods for identification of unauthorized GMO events targeting the P-35S (Cankar *et al.* 2008; Ruttink *et al.*2009) as well as detection and identification of multiple GMO with one single technique (Raymond *et al.* 2009).

In this application study we present a genome walking method LT-RADE for detecting and identifying unknown GMOs through targeting the most commonly known transgenic element P-35S.

In the previous sections of the chapter I have presented an optimized genome walking method LT-RADE for obtaining comparatively long unknown DNA sequences adjacent to the known ones. The study was successfully carried out for MON810 maize and LLRICE62 rice GM events, targeting the transgenic insert insertion sites. This study extends the application for P-35S: (1) to a wider choice of plant species with varying genome sizes (from about 430 Mb (rice) to approx 2500 Mb (maize and cotton)) and (2) documents the LT-RADE genome walking as feasible in both directions (upstream and downstream) of the P-35S. As such, sequence information is obtained from the transgenic insert at the 3' end of P-35S but also from the 5' end. The schematic display of the LT-RADE directed outside and inside the GM insert is shown in figure 26.

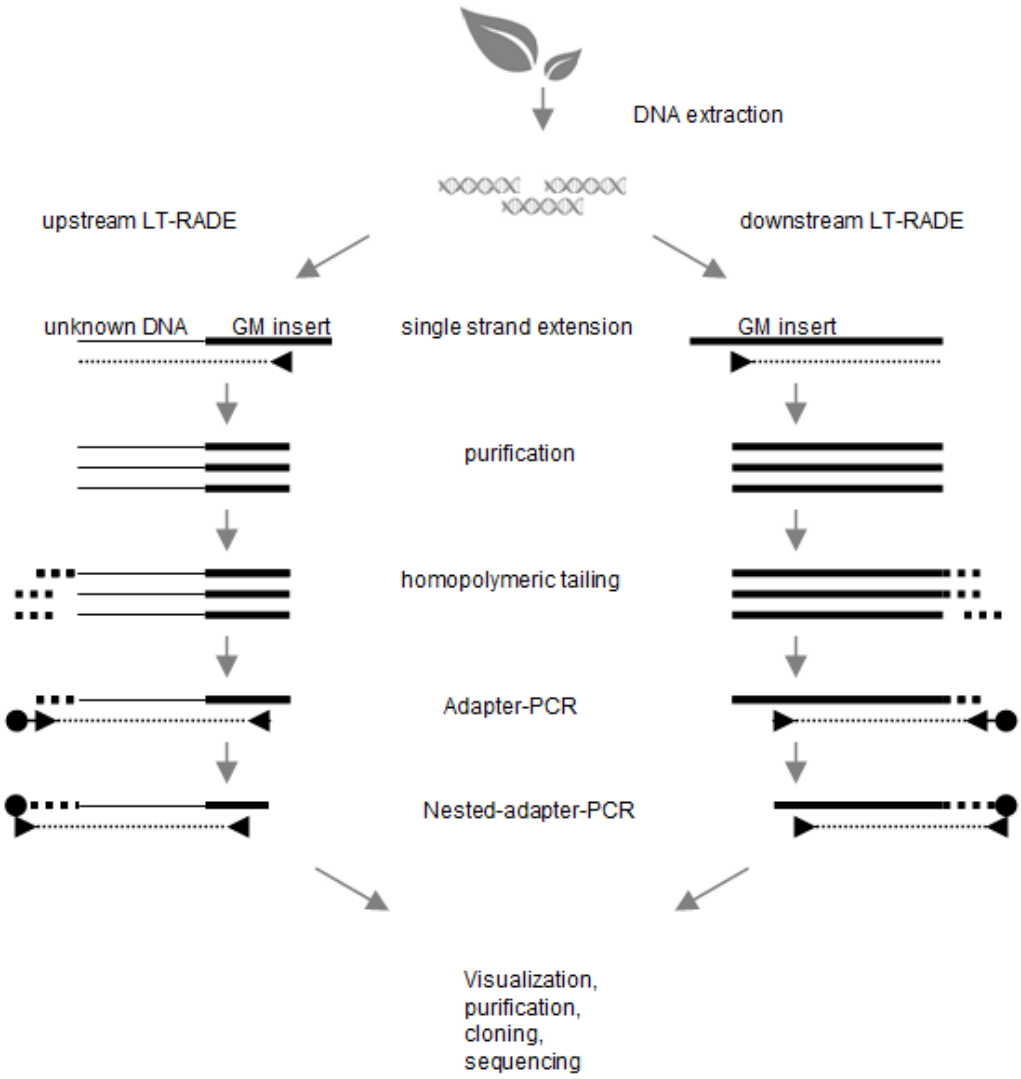

Figure 26: Bidirectional LT-RADE display where upstream and downstream "walking" is summarized. Two universal P-35S primer sets (gene specific primers) directed to the plant DNA or the transgenic insert are used. Initial step of single primer extension followed by the ssDNA fragment purification prepares for adapter synthesis. Adapter of poly-dC is then added using template independent TdT and two nested adapter PCRs conclude the procedure of obtaining the unknown DNA sequence.

This bidirectional genome walking was applied for the first time ever, and moreover it was applied for the retrieval of flanking sequences of P-35S from the following sources:

– T45 rapeseed,

- A2704-12 soybean and
- LLCOTTON25 cotton genomic DNA obtained from certified reference materials.

Additionally to the GM events above, the LT-RADE directed inside the transgenic insert was also planned for MON810 maize and LLRICE62 rice. These GM events were the original source of DNA for development and optimization of the method as it was described before. However all four previous genome walking applications on these events were successfully applied to map the DNA sequences outside the transgene, however it was never applied to read the GM inserts. The sequence of LLRICE62 rice transgenic insert was unpublished before this study.

2.4.2. Upstream genome walking

Upstream LT-RADE reactions directed to plant flanking sequences were performed using the primers P35SRA, P35SRB and P35SRC (figure 27). All three primers are situated in close proximity from each other (nested) and anneal to the P-35S promoter. This reaction will retrieve DNA sequences outside of the transgenic insert. These primers were designed in a previous study and successfully applied on events MON810 maize and LLRICE62 rice (yielding respectively 1018 bp and 631 bp fragments on the 5' insertion site of the events).

A similar upstream LT-RADE approach for obtaining unknown DNA flanking sequences confirmed the presence of P-35S and provided information on the preceding sequences in three additional GM events: T45 rapeseed, A2704-12 soybean and LLCOTTON25 cotton. As these species have different size genomes (table 14), different starting DNA concentrations ranging from 200 up to 500 ng of genomic DNA were tested to evaluate the impact of genome size on the outcome of LT-RADE application. The optimal concentrations for each selected event are displayed in table 14.

After carrying out the LT-RADE reactions the final product was analyzed by agarose gel. In all cases one or two well defined and concentrated bands were present (data not shown). Upstream genome walking displayed approximately 600 and 1000 bp long fragments in the reactions with T45 rapeseed; a 400 bp fragment with A2704-12 soybean DNA; and two bands of approximate length of 400 and 500 bp with LLCOTTON25 cotton DNA.

The fragments were isolated from the gel, purified and cloned. In all cases the largest possible and/or both obtained fragments were cloned for obtaining maximum nucleotide sequence data for an alignment study. From the DNA sequencing results, the following sizes of the inserts obtained by upstream walking could be calculated: a 847 bp fragment was obtained using T45 rapeseed DNA, a 493 bp fragment was revealed using A2704-12 soybean DNA and a fragment of 506 bp was sequenced in the reactions with LLCOTTON25 cotton DNA. The longest obtained fragments are listed in table 14 and displayed in figure 27.

2.4.3. Downstream genome walking

Next, it was aimed to extend the use of LT-RADE as a method permitting to obtain unknown regions of DNA regardless of which direction the reactions are performed on template DNA. The downstream genome walking would aid in a situation when the screening for GMOs shows the presence of the P-35S however other modes of identification of the GM event would not function. Considering that the binding of the three primer pairs was very successful in retrieving the upstream sequences, we first tested whether the reverse primers would allow collecting the downstream sequences. The primers P35SRC_rev, P35SRB_rev and P35SRA_rev represent as such the exact reverse complement to the original primers (see table 7 in M&M section). All reactions were performed using a range of 200 to 500 ng of pure genomic DNA and an

optimal concentration was then selected. All five previously used GM events were test subjects for downstream LT-RADE application (MON810 maize, LLRICE62 rice, T45 rapeseed, A2704-12 soybean and LLCOTTON25 cotton).

One to four distinct fragments were obtained in the downstream LT-RADE application. Fragments with an approximate length of 400, 450, 800 and 900 bp were retrieved in the reactions with MON810 maize DNA; around 1000 bp with LLRICE62 rice; two fragments of 400 and 600 bp with T45 rapeseed; approximately 400 bp and 450 bp fragments for A2704-12 soybean and LLCOTTON25 cotton respectively.

After purification, selected fragments (including the largest) were cloned and sequenced. The following fragments for downstream LT-RADE reactions were obtained: 856 bp for maize, 985 bp for rice, 550 bp for rapeseed, 493 bp for soybean and 428 bp for cotton. An overview of the results of these analyses is shown in table 14 (list of fragment lengths of the respective cloned inserts) and figure 27 (representation of the location of the retrieved fragments).

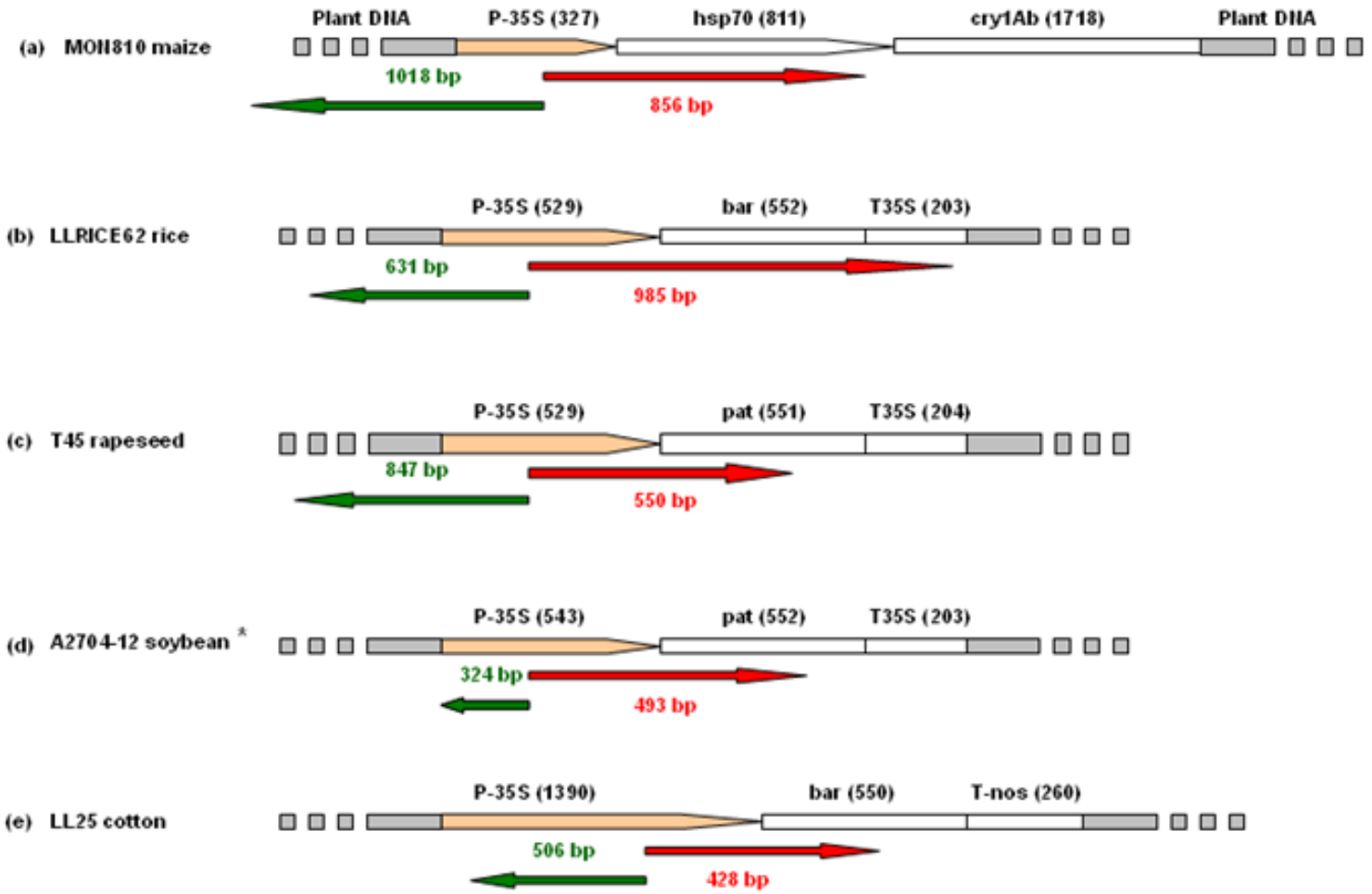
(a) MON810 maize
Plant DNA
P-35S (327)
hsp70 (811)
cry1Ab (1718)
Plant DNA
1018 bp
856 bp
(b) LLRICE62 rice
P-35S (529)
bar (552)
T35S (203)
631 bp
985 bp
(c) T45 rapeseed
P-35S (529)
pat (551)
T35S (204)
847 bp
550 bp
(d) A2704-12 soybean *
P-35S (543)
pat (552)
T35S (203)
324 bp
493 bp
(e) LL25 cotton
P-35S (1390)
bar (550)
T-nos (260)
506 bp
428 bp

Figure 27: Fragments obtained by LT-RADE application on five distinct GM events. Arrows demonstrate the genome walking direction. Schematic views of transgenic inserts of (a) MON810, (b) LLRICE62, (c) T45, (d) A2704-12, (e) LLCOTTON25 display the lengths of transgenic elements. * A2704-12 soybean contains two copies of identical transgenic cassette in a very close proximity.

Table 14: Species and GM event description and longest obtained genome walking results

	Transgenic elements	**Approximate genome size, Mb**	**Longest fragment obtained and sequenced, bp**		**Initial DNA conc, ng/rxn**	**Approximate genome copy number**
			Downstream*	**Upstream***		
MON810 maize	P35S-*hsp70*-*cry1Ab*	2500	856	1018	500	193.000
LLRICE62 rice	P35S-*bar*-T35S	430	985	631	200	400.000
T45 rapeseed	P35S-*pat*-T35S	1200	551	847	400	320.000
A2704-12 soybean	P35S-*pat*-T35S (2X)	1115	493	324	300	240.000
LLCOTTON25	P35S-*bar*-	2500	428	506	300	125.000

cotton	Tnos					

* as the GSP1 and GSP3 are interchangeable depending on the direction of the LT-RADE, there is 60 bp gap between obtained sequences via downstream and upstream genome walking.

2.4.4. Alignment study

The most significant matches in aligning the obtained sequences with GenBank are listed in table 15. The DNA sequences of most GM events analyzed in this study, but not all, have been reported and are available in GenBank. A significant partial or full similarity was observed in most cases when the sequences of the events were publically available in GenBank.

Two DNA sequences (partial CDS of LLRICE62 rice and T45 rapeseed transgenic inserts) were retrieved which are not reported yet in GenBank. The alignment study of these sequences showed the similarity to a binary vector sequence for the rice and the sequence of the rapeseed transgene has aligned with a transgenic maize sequence of the same constitution. Therefore, they do not show any alignment results when refining the similarity search to appropriate species from which DNA was extracted. The 985 bp fragment obtained from LLRICE62 rice with downstream genome walking covers most of the entire transgenic insert of this rice event, including part of the P-35S, the *bar* gene and the T-35S. The 551 bp fragment of T45 rapeseed, obtained from a downstream genome walking, encompasses the P-35S—*pat* gene junction of this event. The obtained fragment from T45 rapeseed however matched the other sequence when the alignment results were not refined to species level. This was a recombinant sequence in the GM maize having a similar transgenic insert as the rapeseed event we have selected (P-35S—*pat* gene junction aligned almost ideally).

Table 15: The outcome of the similarity analysis

LT-RADE direction	Fragment length	Organism and event/ alignment results	Query coverage	Max Ident
MON810 maize				
Downstream	856	Synthetic construct truncated CRYIA(b) (cryIA(b)) gene, partial CDS. Acc. No. AY326434	100%	99%
Upstream	1081	Synthetic construct transgenic *Zea mays* with Cauliflower mosaic virus genomic sequence. Acc. No. AF434709	86%	99%
LLRICE62 rice				
Downstream	985	Binary vector pLH7500, complete sequence. Acc. No. AY234331	100%	96%
Upstream	631	*Oryza Sativa* Japonica Group genomic DNA, chromosome 6, BAC clone: OSJNBa0029G06. Acc. No. AP004680 And Binary vector BinHygTOp aph, tetA, tetR, traJ, insB, insA, aphA-3 and trfA genes. Acc. No. Z37515	98%	100%
T45 rapeseed				
Downstream	551	*Zea mays* transgenic phosphinothricin acetyltransferase gene, partial CDS; and beta lactamase and phosphinothricin	100%	99%

		acetyltransferase genes, complete CDS. Acc. No. DQ156557		
Upstream	847	*Brassica napus* transgenic line T45 left border junction of transgenic event genomic sequence. Acc. No. FJ154954	92%	99%
A2704-12 soybean				
Downstream	493	*Glycine max* transgenic GMO cassette genomic sequence. Acc. No. GQ497217	100%	100%
Upstream	324	*Glycine max* transgenic GMO cassette genomic sequence. Acc. No. GQ497217	100%	99%
LLCOTTON25 cotton				
Downstream	428	*Gossypium hirsutum* transgenic clone LLcotton25 phosphinothricin N-acetyltransferase gene, complete CDS. Acc. No. HQ233646	100%	99%
Upstream	505	*Gossypium hirsutum* transgenic clone LLcotton25 phosphinothricin N-acetyltransferase gene, complete CDS. Acc. No. HQ233646	100%	100%

2.5. Development, optimization and application of RADE and LT-RADE

In the study designed for obtaining unknown DNA sequences we have developed and optimized an efficient enzymatic digestion independent Genome Walking technique to obtain unknown DNA sequences adjacent to a known region. It is demonstrated that these methods entitled RADE and LT-RADE can both be considered as useful, relatively easy and time-saving methods to analyze the insertion sites of GMOs. Both techniques rely on relative simple PCR-based manipulations with a high specificity outcome due to the nested PCR steps at the end of the procedure. The RADE method was optimized here by substituting the regular *Taq* polymerase with a polymerase blend of *Taq* and *Tgo* enzymes within all PCR steps of the procedure, which allowed to generate PCR products of about twice the size of the ones produced with a standard polymerase. All obtained fragments matched with high precision (>99%) reported GMO DNA sequences of the analyzed GM events, clearly documenting the robustness of these optimized methods.

RADE and LT-RADE can thus readily aid in the characterization of the transgenic inserts of unknown or poorly documented GMOs. Also, the three nested primers designed in this study for genome walking outward the P-35S promoter of cauliflower mosaic virus represent a particular interesting asset for rapid identification of traits expressed in transgenic plants. Indeed, to date the P-35S promoter still represents the most frequently used promoter in transgenic plants and will most probably remain very important considering the GMOs in the pipeline for commercialization reported in the EU field trial applications.

RADE and LT-RADE require relatively high amounts of initial target copies to obtain large fragments (about 10^5 copies). Babekova *et al.* (2008) demonstrated that fragments can be obtained with less starting material (about 40 ng), although with poor efficiency, using so-called SiteFinder walking. The size of

the obtained fragments however was limited to 300 bp. The precise number of targets in this particular GM event KMD1 rice was not determined. The application of genome walking techniques would in general greatly increase when lower amounts of targets could allow for obtaining information on unknown DNA sequences. One major bottleneck could be the relatively high background levels in the early amplification steps probably due to amplification starting from the large excess of unspecific targets present in the (plant) genome. Reducing this background by increasing the molar concentrations of target in the reaction may help resolving such inconvenience. Novel techniques such as single-cell analyses or digital PCR amplification could represent useful platforms for further improvements.

Further on the successful application of the P-35S target in a bidirectional LT-RADE for collecting flanking regions from different GM events present in maize, soybean, rapeseed, rice and cotton is presented. In all cases, sufficiently large fragments could be obtained yielding information to align and or to assign a function to the retrieved sequence. In this way, a further step in demonstrating the flexibility and robustness of LT-RADE in the easy retrieval of flanking fragments in GMO has been made.

A bidirectional LT-RADE was implemented by applying a previously designed nested set of primers to retrieve 5' flanking sequences of P-35S (upstream LT-RADE) and simply reversing the primer sequences to isolate fragments 3' adjacent of this promoter. The finding that this complementary set of primers successfully allowed the retrieval of sequences 3' flanking the P-35S (downstream LT-RADE), suggests that the T_m-value of the primers is a key parameter in the development of bidirectional LT-RADE approaches. Varying the DNA concentration used in the first reaction apparently had little influence on the outcome, with a smaller amount of starting DNA being enough for the reactions with rice, soybean and cotton. However we could not present the estimation that the size of genome impacts the positive outcome when different

size genome plant DNA preparations are used in different starting concentrations. The LT-RADE efficiency is presumably not hampered by genome complexity but presumably more limited by accessible number of target copies. Indeed, genomic DNA is largely fragmented upon extraction (average size of majority of fragments at about 20 Kb), what will disrupt any secondary of tertiary genome structures that could hinder primer annealing. On the contrary, LT-RADE as presented here still requires large numbers of targets (about 200.000 cps) what may indicate poor efficiency at any of the steps during the procedure.

The application of sets of complementary primers for both upstream and downstream retrieval of regions flanking the targeted P-35S sequence in simultaneously conducted LT-RADE reactions, provides an easy and efficient tool towards screening of GM events. Not only is a construct specific detection of a GM event achieved in the downstream LT-RADE (P-35S) but also the characterization of the GM event by analyzing the junction in the upstream LT-RADE. Moreover, it is shown that the same set of primers can be applied on different plant species showing the common applicability for a large number of GM events containing the P-35S, one of the most utilized promoters in GM constructs.

While this approach has been shown to be efficient for highly uniform material derived from GM plants (such a seeds), the application of LT-RADE in identifying flanking sequences of traces of the P-35S *e.g.* in food products, needs investigation. Given that the reaction efficiency seems to be low, improvements may be considered to increase: (1) the target ratio versus reagents (higher molarity) and (2) the target ratio versus background DNA (higher molarity). Such conditions could be reached by miniaturization the setup in digital PCR analysis.

The general positive outcome of the method's optimization and the sequencing results generated via the LT-RADE shows potential to adopt this method to any selective transgenic element. Recently some real-time screening assays for selective promoters and terminators were published (Debode *et al.* 2013). Similarly, the LT-RADE may be tested on a variety of targets, such as distinctive promoters and terminators used in the transgenic cassettes of the existing GMO as well as those in the development pipeline.

In conclusion, this study showed that LT-RADE can be a useful method for several plant species in obtaining small DNA fragments flanking known loci. For GM plants, the bidirectional P-35S method described here represents a good candidate for further optimization towards a reliable, sensitive retrieval system of non-registered genetically engineered DNA sequences detected in food and feed products.

2.6. Concluding remarks

Every day, many research teams are trying to optimize the existing methods to fit the needs or to create new robust, cost efficient and user-friendly techniques to apply them for the purpose of scientific knowledge of GMOs that are reaching the end-user: every consumer. With this doctoral thesis I have addressed several significant aspects of GMO analysis in particular, the molecular characterization of the transgenic insert and the plant DNA sequences surrounding the insert (so called insertion site). My work consisted of experiments planned to cover the transgenic insert via long fragment amplification using LT-PCR method, the optimization of the method in a matter of DNA template selection, parameter set and commercial polymerase choice to fit the needs of GMO testing; and further testing of the novel reaction additive, CNTs which bear a great potential yet pose extreme limitations for laboratory use. The testing of uninterrupted transgenic insert amplification has ended with an important observation that obtaining enough plant flanking

sequences is crucial for the successful LT-PCR of full transgenic insert amplification as the patent documents and/or publically available sequence data of the GM events are insufficient in most of the cases. This limiting factor has hinted to continue with the testing of the ways to obtain the insertion site sequences with the potential to follow through as much as needed to the unknown sequences of the plant DNA. The longer the plant flanking nucleotide sequences are obtained, the larger is the selection of primer annealing sites. As such, the second part of this doctoral thesis was the development and optimization of the genome walking method. Generally the aim of genome walking methods is obtaining the unknown DNA adjacent to known sequences, and there is a wide variety of ways to approach this. In this manuscript I have introduced a user-friendly robust genome walking method, constructed on the basis of other known molecular biology methods. For the success of the method, optimization of all possible variables was done and possibly the most advantageous conditions obtained. Thus the last part of the study dealt with the application of this restriction independent adapter ligation-mediated genome walking that we have called LT-RADE on the GMO matrices that have no publicly available nucleotide sequence and/or the bioinformatics data is limited.

CONCLUSIONS

1. Long fragments of 3408 and 5902 bp containing near entire transgenic insert were obtained and the integrity proven with the primer walking-type sequencing reactions.

2. The carbon nanotubes tested to have negative impact on the LT-PCR stability and specificity however the increase of reaction yield was observed in low concentrations of CNTs.

3. A genome walking method RADE and its advanced version LT-RADE were developed and optimized to retrieve unknown DNA sequences surrounding transgenic elements in the GMOs.

4. Two sets of universal primers (bidirectional) to test any GMO containing the P-35S element were created in the process of genome walking method development.

5. LT-RADE method on five distinct GM crop DNA was successfully applied to obtain the internal sequences of the transgenic insert as well as unknown plant DNA sequences surrounding the transgene.

PUBLICATIONS

1. Spalinskas R, Van den Bulcke M, Van den Eede G, Milcamps A. 2012. LT-RADE: An Efficient User-Friendly Genome Walking Method Applied to the Molecular Characterization of the Insertion Site of Genetically Modified Maize MON810 and Rice LLRICE62. Food Analytical Methods 6(2): 705-13. doi: 10.1007/s12161-012-9438-y

2. Spalinskas R, Van den Bulcke M, Milcamps A. 2013. Efficient retrieval of recombinant sequences of GM plants by Cauliflower Mosaic Virus 35S promoter-based bidirectional LT-RADE. European Food Research and Technology 237(6): 1025-31. doi: 10.1007/s00217-013-2078-7

ACKNOWLEDGEMENTS

I would like to sincerely thank my mentors: professor habil. dr. Juozas Rimantas Lazutka from Vilnius University and dr. Anne Milcamps from Joint Research Centre for all the ideas, discussions, tips, also trust, inspiration and overall most valuable input. My biggest respect and gratitute.

Dr. Marc Van den Bulcke from Scientific Institute of Public Health in Belgium for all the discussions we had and critical overview of my work.

All the staff of Joint Research Centre, Institute for Health and Consumer Protection, Molecular Biology and Genomic Unit; and all the staff of Vilnius University, Faculty of Natural Sciences, Department of Botanics and Genetics: my warmest gratitute for being with me throughout, being so friendly and nice, but also professional and supportive. It has helped me to grow in my professional and personal lives.

Finally, I would like to send my warmest regards and thank my family and my friends for the patience, endless motivation and believing in me. Your overall support kept me pushing forward.

To everyone who is going to read my manuscript, I hope you enjoy it!

REFFERENCES

Aguilera M, Querci M, Balla B, Prospero A, Ermolli M, Van den Eede G. 2008. A Qualitative Approach for the Assessment of the Genetic Stability of the MON 810 Trait in Commercial Seed Maize Varieties. Food Anal Methods 1(4):252–58.

Arendse MS, Dubery IA, Berger DK. 1999. Isolation by PCR-based methods of a plant antifungal polygalacturonase-inhibiting protein gene. Electron J Biotechnol 2 (3). doi: 10.2225/vol2-issue3-fulltext-4.

Arnold C, Hodgson IJ. 1991. Vectorette PCR: a novel approach to genomic walking. PCR Methods Appl 1(1):39-42.

Assaad FF, Signer ER. 1990. Cauliflower mosaic virus P35S promoter activity in Escherichia coli. Mol Gen Genet 223:517-520.

Babekova R, Funk T, Pecoraro S, Engel K-H, Busch U. 2008. Development of an event-specific Real-time PCR detection method for the transgenic Bt rice line KMD1. Eur Food Res Technol 228(5):707-16.

Bachmann B, Lüke W, Hunsmann G. 1990. Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucleic Acids Res 18(5):1309.

Bae JH, Sohn JH. 2010. Template-blocking PCR: an advanced PCR technique for genome walking. Anal Biochem 398(1):112-6. doi: 10.1016/j.ab.2009.11.003.

Barnes WM. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proc Natl Acad Sci U S A 91(6):2216-20.

Bayer Crop Science. 2003. http://www.gmocompass.org/pdf/regulation/rice/LL62_rice_1829_application_food_feed_efsa.pdf

Behura SK, Nair S, Mohan M. 2001. Polymorphisms flanking the mariner integration sites in the rice gall midge (Orseolia oryzae Wood-Mason) genome are biotype-specific. Genome 44(6):947-54.

Benkel BF, Fong Y. 1996. Long range-inverse PCR (LR-IPCR): extending the useful range of inverse PCR. Genet Anal 13(5):123-7.

Bloch W. 1991. A biochemical perspective of the polymerase chain reaction. Biochemistry 30 (11): 2735–47.

Cankar K, Chauvensy-Ancel V, Fortabat MN, Gruden K, Kobilinsky A, Žel J, Bertheau Y. 2008. Detection of nonauthorized genetically modified organisms using differential quantitative polymerase chain reaction: Application to 35S in maize. Anal Biochem 376:189-199.

CERA. 2014. GM Crop Database. Center for Environmental Risk Assessment (CERA), ILSI Research Foundation, Washington D.C. http://cera-gmc.org/GMCropDatabase.

Chakrabarti R, Schutt CE. 2001. The enhancement of PCR amplification by low molecular weight amides. Nucleic Acids Res 29(11):2377-81.

Chen X, Wu R. 1997. Direct amplification of unknown genes and fragments by Uneven polymerase chain reaction. Gene 185(2):195-9.

Cheng S, Fockler C, Barnes WM, Higuchi R. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Natl Acad Sci U S A 91:5695-99.

Colbach N, Dirr C, Gruber S, Pekrun C. 2008. Modelling the seed bank evolution and emergence of oilseed rape volunteers for managing co-existence of GM and non-GM varieties. Eur J of Agronomy 28(1):19-32.

Commission Implementing Decision. 2011. (EC) No. 2011/884/EU. OJ Eur Union 343:140-148.

Commission Regulation (EC) No. 152/2009. 2009. OJ Eur Union L 54.

Cormack RS, Somssich IE. 1997. Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. Gene 194(2):273-6.

Cui D, Tian F, Kong Y, Titushikin I, Gao H. 2003. Effects of single-walled carbon nanotubes on the polymerase chain reaction. Nanotechnology 15:154. doi:10.1088/0957-4484/15/1/030.

Cullen D, Harwood W, Smedley M, Davies H, Taylor M. 2011. Comparison of DNA walking methods for isolation of transgene-flanking regions in GM potato. Mol Biotechnol 49(1):19-31.

Dalton AB, Stephan C, Coleman JN, McCarthy B, Ajayan PM, Lefrant S, Bernier P, Blau WJ, Byrne HJ. 2000. Selective Interaction of a Semiconjugated Organic Polymer with Single-Wall Nanotubes. J Phys Chem B 104: 10012. doi: 10.1021/jp002857o.

Debode F, Janssen E, Berben G. 2013. Development of 10 new screening PCR assays for GMO detection targeting promoters (pFMV, pNOS, pSSuAra, pTA29, pUbi, pRice actin) and terminators (t35S, tE9, tOCS, tg7). Eur Food Res and Technol 236(4):659-669.

Devon RS, Porteous DJ, Brookes AJ. 1995. Splinkerettes-improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res 23(9):1644-5.

Devos Y, Demont M, Sanvido O. 2008. Coexistence in the EU- return of the moratorium on GM crops? Nature Biotech 26(11):1223–5.

Directive 2001/18/EC of the European Parliament and of the Council. 2001. OJ L 106.

Directive 2008/27/EC of the European Parliament and of the Council. 2008. OJ L 81.

Dong W, Yang LT, Shen KL, Kim B, Kleter GA, Marvin HJP, Guo R, Liang W, Zhang D. 2008. GMDD: a database of GMO detection methods. BMC Bioinformatics 9:260.

Emons H. 2010. GMO analysis – a complex and challenging undertaking. Anal Bioanal Chem 396(6): 1949–50. doi 10.1007/s00216-010-3453-x.

Endo M, Muramatsu H, Hayashi T, Kim YA, Terrones M, Dresselhaus MS. 2005. Buckypaper from coaxial structure. Nature 433(7025):476.

Endo M, Muramatsu H, Hayashi T, Kim YA. 2010. Coaxial carbon nanotubes: Double-Walled Carbon Nanotubes. Annual World Conference on Carbon 2010. 31-32.

EU Register of Authorized GMOs. 2014. http://ec.europa.eu/food/dyna/gm_register/index_en.cfm

EU-RL GMFF (European Union Reference Laboratory for Genetically Modified Food and Feed). 2014. http://gmo-crl.jrc.ec.europa.eu/

Felix CA, Kim CS, Megonigal MD, Slater DJ, Jones DH, Spinner NB, Stump T, Hosler MR, Nowell PC, Lange BJ, Rappaport EF. 1997. Panhandle polymerase chain reaction amplifies MLL genomic translocation breakpoint involving unknown partner gene. Blood 90(12):4679-86.

Fenton JA, Pratt G, Morgan GJ. 2002. Long distance vectorette PCR (LDV PCR). Methods Mol Biol. 192:275-84.

Foord OS, Rose EA. 1994. Long-distance PCR. PCR Methods Appl 3(6):S149-61.

Fors L, Saavedra RA, Hood L. 1990. Cloning of the shark Po promoter using a genomic walking technique based on the polymerase chain reaction. Nucleic Acids Res 18(9):2793-9.

Garrity PA, Wold BJ. 1992. Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting. Proc Natl Acad Sci U S A 89(3):1021-5.

Ge Y, Charon NW. 1997. Identification of a large motility operon in Borrelia burgdorferi by semi-random PCR chromosome walking. Gene 189(2):195-201.

Ge Y, Old IG, Saint Girons I, Charon NW. 1997. Molecular characterization of a large Borrelia burgdorferi motility operon which is initiated by a consensus sigma70 promoter. J Bacteriol 179(7):2289-99.

Gelfand DH. 1989. Taq DNA polymerase. In: PCR Techology: Principles and applications of DNA amplifications, Erlich, H.A., ed., Stockton Press, New York, 17–22.

Giambernardi TA, Rodeck U, Klebe RJ. 1998. Bovine serum albumin reverses inhibition of RT-PCR by melanin. Biotechniques 25(4):564-6.

Grant S, Blondal T. 2001. Patent US 6,225,093.

Gryson N, Eeckhout M, Trouillier A, Le Bail M, Soler LG. 2009. Strategies for coexistence of GM and non-GM soy from import to feed processing. Environ Biosafety Res 8(3):153-9. doi: 10.1051/ebr/2009008.

Guo H, Xiong J. 2006. A specific and versatile genome walking technique. Gene 381:18-23.

Hengen PN. 1995. Vectorette, splinkerette and boomerang DNA amplification. Trends Biochem Sci 20(9):372-3.

Henke W, Herdel K, Jung K, Schnorr D, Loening SA. 1997. Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res 25(19):3957-8.

Hermann SR, Miller JA, O'Neill S, Tsao TT, Harding RM, Dale JL. 2000. Single-primer amplification of flanking sequences. Biotechniques 29(6):1176-8, 1180.

Hernández M, Pla M, Esteve T, Prat S, Puigdomènech P, Ferrando A. 2003. A specific real-time quantitative PCR detection system for event MON810 in

maize YieldGard based on the 3'-transgene integration sequence. Transgenic Res 12(2):179-89.

Holck A, Va M, Didierjean L, Rudi K. 2002. 5'-Nuclease PCR for quantitative event-specific detection of the genetically modified Mon 810 MaisGard maize. Eur Food Res Technol 214:449-53.

Holst-Jensen A. 2009. Testing for genetically modified organisms (GMOs): Past, present and future perspectives. Biotech Adv 27(6):1071–82.

Huang H, Wang G, Zhao Y, Shi P, Luo H, Yao B. 2010. Direct and efficient cloning of full-length genes from environmental DNA by RT-qPCR and modified TAIL-PCR. Appl Microbiol Biotechnol 87(3):1141-9. doi: 10.1007/s00253-010-2613-4.

Hui EK, Wang PC, Lo SJ. 1998. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. Cell Mol Life Sci 54(12):1403-11. Review.

Hung T, Mak K, Fong K. 1990. A specificity enhancer for polymerase chain reaction. Nucleic Acids Res 18(16):4953.

Iijima S. Helical microtubules of graphitic carbon. 1991. Nature 354:56 – 8. doi 10.1038/354056a0.

Innis MA, Myambo KB, Gelfand DH, Brow MA. 1988. DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proc Natl Acad Sci U S A 85(24):9436-40.

IRMM: Institute for Reference Materials and Measurements. 2014. Online catalogue:

http://irmm.jrc.ec.europa.eu/reference_materials_catalogue/catalogue/Documents/rm_catalogue.pdf

Jacobs KA, Rudersdorf R, Neill SD, Dougherty JP, Brown EL, Fritsch EF. 1988. The thermal stability of oligonucleotide duplexes is sequence independent in tetraalkylammonium salt solutions: application to identifying recombinant DNA clones. Nucleic Acids Res 16(10):4637-50.

James, Clive. 2013. Global Status of Commercialized Biotech/GM Crops: 2013. ISAAA Brief No. 46. ISAAA: Ithaca, NY.

Ji J, Braam J. 2010. Restriction site extension PCR: a novel method for high-throughput characterization of tagged DNA fragments and genome walking. PLoS One 5(5):e10577. doi: 10.1371/journal.pone.0010577.

Jones DH, Winistorfer SC. 1992. Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. Nucleic Acids Res 20(3):595-600.

Kainz P, Schmiedlechner A, Strack HB. 1992. In vitro amplification of DNA fragments greater than 10 kb. Anal Biochem 202(1):46-9.

Katzman M. 1993. Use of oil overlays in "oil-free" PCR technology. Biotechniques 14(1):36-40.

Kilstrup M, Kristiansen KN. 2000. Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Res 28(11):E55.

Kohda T, Taira K. 2000. A simple and efficient method to determine the terminal sequences of restriction fragments containing known sequences. DNA Res 7(2):151-5.

Kohli A, Twyman RM, Abranches R, Wegel E, Stoger E, Christou P. 2003. Transgene integration, organization and interaction in plants. Plant Mol Biol 52:247-58.

Kotik M. 2009. Novel genes retrieved from environmental DNA by polymerase chain reaction: current genome-walking techniques for future metagenome applications. J Biotechnol 144(2):75-82. doi: 10.1016/j.jbiotec.2009.08.013.

Kreader CA. 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl Environ Microbiol 62(3):1102-6.

Kroto HW, Heath JR, O'Brien SC, Curl RF, Smalley RE. 1985. C(60): Buckminsterfullerene. Nature 318: 162-63. doi:10.1038/318162a0.

La Paz JL, Pla M, Papazova N, Puigdomènech P, Vicient CM. 2010a. Stability of the MON810 transgene in maize. Plant Mol Biol 74:563-71.

La Paz JL, Vicient C, Puigdomènech P, Pla M. 2010b. Characterization of polyadenylated cryIA(b) transcripts in maize MON810 commercial varieties. Anal Bioanal Chem 396:2125-33.

Lagerström M, Parik J, Malmgren H, Stewart J, Pettersson U, Landegren U. 1991. Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. PCR Methods Appl 1(2):111-9.

Laitinen J, Räkköläinen T, Hölttä E. 2004. Vectorettes for long chromosome walking in genomic DNA of the human p53 gene. Biotechniques 37(4):674-6, 678.

Latham JR, Wilson AK, Steinbrecher RA. 2006. The Mutational Consequences of Plant Transformation. J Biomed Biotechnol 2006:1–7.

Leoni C, Gallerani R, Ceci LR. 2008. A genome walking strategy for the identification of eukaryotic nucleotide sequences adjacent to known regions. Biotechniques 44:229–35.

Leoni C, Volpicella M, De Leo F, Gallerani R, Ceci LR. 2011. Genome walking in eukaryotes. FEBS J 278(21):3953-77. doi: 10.1111/j.1742-4658.2011.08307.x.

Levano-Garcia J, Verjovski-Almeida S, da Silva AC. 2005. Mapping transposon insertion sites by touchdown PCR and hybrid degenerate primers. Biotechniques 38(2):225-9.

Levine E. 2004. Patent US 6,713,259 B2.

Li Q, Kinlocha IA, Windle A. 2005. Discrete dispersion of single-walled carbon nanotubes. Chem Commun 3283-85. doi: 10.1039/B419039D.

Liu S, Dietrich CR, Schnable PS. 2009. DLA-based strategies for cloning insertion mutants: cloning the gl4 locus of maize using Mu transposon tagged alleles. Genetics 183(4):1215-25. doi: 10.1534/genetics.109.108936.

Liu YG, Mitsukawa N, Whittier RF. 1993. Rapid sequencing of unpurified PCR products by thermal asymmetric PCR cycle sequencing using unlabeled sequencing primers. Nucleic Acids Res 21(14):3333-4.

Liu YG, Whittier RF. 1995a. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25(3):674-81.

Liu YG, Mitsukawa N, Oosumi T, Whittier RF. 1995b. Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8(3):457-63.

Liu YG, Chen Y. 2007. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. Biotechniques 43(5):649-50, 652, 654 passim.

Lombardi AM, Cabrio L, Zanon E, Pellati D, Vettore S, Girolami A. 2005. Confirmation of the value of a modified long-distance polymerase chain reaction in the detection of inversion intron 22 in severe hemophilia a: a technical note. Clin Appl Thromb Hemost 11(4):493-6.

Maga EA, Richardson T. 1991. Amplification of a 9.0-kb fragment using PCR. Biotechniques 11(2):185-6.

Manivannan S, Jeong IO, Ryu JH, Lee CS, Kim KS, Jang J, Park KC. 2009. Dispersion of single-walled carbon nanotubes in aqueous and organic solvents through a polymer wrapping functionalization. J Mater Sci: Mater Electron 20:223-29. doi: 10.1007/s10854-008-9706-1.

Manseau LJ, Ganetzky B, Craigt EA. 1988. Molecular and Genetic Characterization of the Drosophila melanogaster 87E Actin Gene Region. Genetics 119:407-420.

Martin-Harris MH, Bartley PA, Morley AA. 2010. Gene walking using sequential hybrid primer polymerase chain reaction. Anal Biochem 399(2):308-10. doi: 10.1016/j.ab.2010.01.005.

Mäde D, Degner C, Grohmann L. 2006. Detection of genetically modified rice: a construct-specific real-time PCR method based on DNA sequences from transgenic Bt rice. Eur Food Res Technol 224:271–78. doi: 10.1007/s00217-006-0467-x.

Mendoza A, Fernández S, Cruz MA, Rodríguez Perez MA, Resendez Perez D, Barrera Saldaña HA. 2006. Detection of genetically modified maize food products by the polymerase chain reaction. Cienc Tecnol Aliment 5(3) 175-81.

Milcamps A, Rabe S, Cade R, De Framond AJ, Henriksson P, Kramer V, Lisboa D, Pastor-Benito S, Willits MG, Lawrence D, Van den Eede G. 2009. Validity assessment of the detection method of maize event Bt10 through investigation of its molecular structure. J Agric Food Chem 57(8):3156-63. doi: 10.1021/jf802627f.

Min GS, Powell JR. 1998. Long-distance genome walking using the long and accurate polymerase chain reaction. Biotechniques 24(3):398-400.

Mizusawa S, Nishimura S, Seela F. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res 14(3):1319-24.

Müller PR, Wold B. 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246(4931):780-6.

Müller PR, Wold B. 1991. Ligation-mediated PCR: Applications to genomic footprinting. Methods 2:20-31.

Müller PR, Wold B, Garrity PA. 2001. Current Protocols in Molecular Biology. 15.3.1-15.3.26.

Müller B, Zöpf K, Bachofer J, Steimer W. 2003. Optimized strategy for rapid cytochrome P450 2D6 genotyping by real-time long PCR. Clin Chem 49(10):1624-31.

Myrick KV, Gelbart WM. 2002. Universal Fast Walking for direct and versatile determination of flanking sequence. Gene 284(1-2):125-31.

Nagai M, Yoshida A, Sato N. 1998. Additive effects of bovine serum albumin, dithiothreitol, and glycerol on PCR. Biochem Mol Biol Int 44(1):157-63.

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramírez-Solis R, Webster TD, Ledbetter DH, Caskey CT. 1989. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci U S A 86(17):6686-90.

Nielsen CR, Berdal KG, Holst-Jensen A. 2008. Anchored PCR for possible detection and characterisation of foreign integrated DNA at near single molecule level. Eur Food Res Technol 226(5):949-56. doi: 10.1007/s00217-007-0616-x.

Nicolia A, Manzo A, Veronesi F, Rosellini D. 2013. An overview if the last 10 years of genetically engineered crop safety research. Crit Rev Biotechnol 34(1):77-88. doi 10.3109/07388551.2013.823595.

Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK. 1996. Identification of a Brazil-nut allergen in transgenic soybeans. N Engl J Med 334(11):688-92.

Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA. 2004. Electric Field Effect in Atomically Thin Carbon Films. Science 306(5696): 666-69.

Nthangeni MB, Ramagoma F, Tlou MG, Litthauer D. 2005. Development of a versatile cassette for directional genome walking using cassette ligation-mediated PCR and its application in the cloning of complete lipolytic genes from Bacillus species. J Microbiol Methods 61(2):225-34.

Ochman H, Gerber AS, Hartl DL. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120(3):621-3.

Odell JT, Nagy F, Chua NH. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313(6005):810-812.

Odell JT, Knowlton S, Lin W, Mauvais CJ. 1988. Properties of an isolated transcription stimulating sequence derived from the cauliflower mosaic 35S promoter. Plant Mol Biol 10:263-272.

Ohler LD, Rose EA. 1992. Optimization of long-distance PCR using a transposon-based model system. PCR Methods Appl 2(1):51-9.

Okamoto H, Hiromi Y, Ishikawa E, Yamada T, Isoda K, Maekawa H, Hotta Y. 1986. Molecular characterization of mutant actin genes which induce heat-shock proteins in Drosophila flight muscles. EMBO J 5(3):589-96.

O'Malley RC, Alonso JM, Kim CJ, Leisse TJ, Ecker JR. 2007. An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome. Nat Protoc 2(11):2910-7.

Orcheski BB, Davis TM. 2010. An enhanced method for sequence walking and paralog mining: TOPO(R) Vector-Ligation PCR. BMC Res Notes 3:61. doi: 10.1186/1756-0500-3-61.

Ortiz-Acevedo A, Xie H, Zorbas V, Sampson WM, Dalton AB, Baughman RH, Draper RK, Musselman IH, Dieckmann GR. 2005. Diameter-Selective Solubilization of Single-Walled Carbon Nanotubes by Reversible Cyclic Peptides. J Am Chem Soc 127: 9512. doi: 10.1021/ja050507f.

Ow DW, Wood KV, DeLuca M, De Wet JR, Helinski DR, Howell SH. 1986. Transient and Stable Expression of the Firefly Luciferase Gene in Plant Cells and Transgenic Plants. Science 234(4778):856-859.

Pääbo S, Wilson AC. 1988. Polymerase chain reaction reveals cloning artefacts. Nature 334(6181):387-8.

Paauw A, Fluit AC, Verhoef J, Leverstein-van Hall MA. 2006. Enterobacter cloacae outbreak and emergence of quinolone resistance gene in Dutch hospital. Emerg Infect Dis 12(5):807-12.

Padegimas LS, Reichert NA. 1998. Adaptor ligation-based polymerase chain reaction-mediated walking. Anal Biochem 260(2):149-53.

Papazova N, Windels P, Depicker A, Taverniers I, Roldan-Ruiz I, Milcamps A, Van Bockstaele E, Van Den Eede G, De Loose M. 2006. Sequence stability of the T-DNA - plant junctions in tissue culture in Arabidopsis transgenic lines. Plant Cell Rep 25:1362-68.

Papazova N, Ghedira R, Van Glabeke S, Bartegi A, Windels P, Taverniers I, Roldán-Ruiz I, Van Bockstaele E, Milcamps A, Van Den Eede G, Depicker A,

De Loose M. 2008. Stability of the T-DNA flanking regions in transgenic Arabidopsis thaliana plants under influence of abiotic stress and cultivation practices. Plant Cell Rep 27:749-57.

Park DJ. 2011. Lariat-Dependent Nested PCR for Flanking Sequence Determination. Methods in Mol Biol 687:43-55.

Parker JD, Rabinovitch PS, Burmer GC. 1991. Targeted gene walking polymerase chain reaction. Nucleic Acids Res 19(11):3055-60.

Pérez-Hernández JB, Swennen R, Sági L. 2006. Number and accuracy of T-DNA insertions in transgenic banana (Musa spp.) plants characterized by an improved anchored PCR technique. Transgenic Res 15(2):139-50.

Pilhofer M, Bauer AP, Schrallhammer M, Richter L, Ludwig W, Schleifer KH, Petroni G. 2007. Characterization of bacterial operons consisting of two tubulins and a kinesin-like gene by the novel Two-Step Gene Walking method. Nucleic Acids Res 35(20):e135.

Ping L, Vogel H, Boland W. 2008. Cloning of prokaryotic genes by a universal degenerate primer PCR. FEMS Microbiol Lett 287(2):192-8. doi: 10.1111/j.1574-6968.2008.01311.x.

Ponce MR, Nicol L. 1991. PCR amplification of long DNA fragments. Nucleic Acids Res 20:623.

Pule MA, Rousseau A, Vera J, Heslop HE, Brenner MK, Vanin EF. 2008. Flanking-sequence exponential anchored-polymerase chain reaction amplification: a sensitive and highly specific method for detecting retroviral integrant-host-junction sequences. Cytotherapy 10(5):526-39. doi: 10.1080/14653240802192636.

Puskás LG, Fartmann B, Bottka S. 1994. Restricted PCR: amplification of an individual sequence flanked by a highly repetitive element from total human DNA. Nucleic Acids Res 22(15): 3251-52.

Querci M, Jermini M, Van den Eede G. 2006. Training course on the analysis of food samples for the presence of genetically modified organisms. User Manual. http://gmo-crl.jrc.ec.europa.eu/capacitybuilding/docs-manual-EN.htm

Querci M, Van den Bulcke M, Zel J, Van den Eede G, Broll H. 2010. New approaches in GMO detection. Anal Bioanal Chem 396(6):1991-2002. doi: 10.1007/s00216-009-3237-3.

Raffini LJ, Slater DJ, Rappaport EF, Lo Nigro L, Cheung NK, Biegel JA, Nowell PC, Lange BJ, Felix CA. 2002. Panhandle and reverse-panhandle PCR enable cloning of der(11) and der(other) genomic breakpoint junctions of MLL translocations and identify complex translocation of MLL, AF-4, and CDK6. Proc Natl Acad Sci U S A 99(7):4568-73.

Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S. 2006. An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun 347(3):747-51.

Ramesh S, Ericson LM, Davis VA, Saini RK, Kittrell C, Pasquali M, Billups WE, Adams WW, Hauge RH, Smalley RE. 2004. Dissolution of Pristine Single Walled Carbon Nanotubes in Superacids by Direct Protonation. J Phys Chem B 108: 8794. doi: 10.1021/jp036971t.

RASFF portal (Rapid Alert System for Food and Feed), https://webgate.ec.europa.eu/rasff-window/portal/

Raymond P, Gendron L, Khalf M, Paul S, Dibley KL, Bhat S, Xie VR, Partis L, Moreau ME, Dollard C, Coté MJ, Laberge S, Emslie KR. 2009. Detection and identification of multiple genetically modified events using DNA insert fingerprinting. Anal Bioanal Chem 396:2091-2102.

Reddy PS, Mahanty S, Kaul T, Nair S, Sopory SK, Reddy MK. 2008. A high-throughput genome-walking method and its use for cloning unknown flanking sequences. Anal Biochem 381(2):248-53. doi: 10.1016/j.ab.2008.07.012.

Rees WA, Yager TD, Korte J, von Hippel PH. 1993. Betaine can eliminate the base pair composition dependence of DNA melting. Biochemistry 32(1):137-44.

Regulation (EC) No. 1829/2003 of the European Parliament and of the Council. 2003. OJ Eur Union L 268.

Regulation (EC) No. 1830/2003 of the European Parliament and of the Council. 2003. OJ Eur Union L 268.

Regulation (EC) No. 1846/2003 of the European Parliament and of the Council. 2003. OJ Eur Union L 287.

Regulation (EC) No. 882/2004 of the European Parliament and of the Council. 2004. OJ Eur Union L 165.

Regulation (EC) No. 619/2011 of the European Parliament and of the Council. 2011. OJ Eur Union 166:9.

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, Smith JC, Markham AF. 1990. A novel, rapid method for the isolation of terminal

sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res 18(10):2887-90.

Rishi AS, Nelson ND, Goyal A. 2004. Genome walking of large fragments: an improved method. J Biotechnol 111(1):9-15.

Rosati A, Bogani P, Santarlasci A, Buiatti M. 2008. Characterisation of 3' transgene insertion site and derived mRNAs in MON810 YieldGard maize. Plant Mol Biol 67(3):271-81. doi: 10.1007/s11103-008-9315-7.

Rosenthal A, Jones DS. 1990. Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. Nucleic Acids Res 18(10):3095-6.

Rudenko GN, Rommens CM, Nijkamp HJ, Hille J. 1993. Supported PCR: an efficient procedure to amplify sequences flanking a known DNA segment. Plant Mol Biol 21(4):723-8.

Rudi K, Fossheim T, Jakobsen KS. 1999. Restriction Cutting Independent Method for Cloning Genomic DNA Segments Outside the Boundaries of Known Sequences. Biotechniques 27:1170-77.

Ruttink T, Morisset D, Van Droogenbroeck B, Lavrač N, Van Den Eede G, Žel J, De Loose M. 2009. Knowledge-technology-based discovery of unauthorized genetically modified organisms. Anal Bioanal Chem. 396(6):1951-59.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic Amplification of β-globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. Science 230:1350-54.

Saluto A, Brussino A, Tassone F, Arduino C, Cagnoli C, Pappi P, Hagerman P, Migone N, Brusco A. 2005. An enhanced polymerase chain reaction assay to detect pre- and full mutation alleles of the fragile X mental retardation 1 gene. J Mol Diagn 7(5):605-12.

Sambrook J, Russell DW. 1989. Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sambrook J, Russell DW. 2001. Molecular Cloning. A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sarkar G, Kapelner S, Sommer SS. 1990. Formamide can dramatically improve the specificity of PCR. Nucleic Acids Res 18(24):7465.

Sarkar G, Turner RT, Bolander ME. 1993. Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. PCR Methods Appl 2(4):318-22.

Shyamala V, Ames GFL. 1990. Genome walking by single-specific-primer polymerase chain reaction; SSP-PCR. Gene 84:1-8.

Shyamala V, Ames GFL. 1993. Single Specific Primer-Polymerase Chain Reaction (SSP-PCR) and Genome Walking. Methods in Mol Biol 15:339-348.

Sidransky E, Tayebi N, Stubblefield BK, Eliason W, Klineburgess A, Pizzolato GP, Cox JN, Porta J, Bottani A, DeLozier-Blanchet CD. 1996. The clinical, molecular, and pathological characterisation of a family with two cases of lethal perinatal type 2 Gaucher disease. J Med Genet 33(2):132-6.

Singer T, Burke E. 2003. High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. Methods Mol Biol 236:241-72.

Siruguri V, Sesikeran B, Bhat RV. 2004. Starlink genetically modified corn and allergenicity in an individual. J Allergy Clin Immunol, 113, 1003–4; author reply 1004–5.

Spiess AN, Mueller N, Ivell R. 2004. Trehalose is a potent PCR enhancer: lowering of DNA melting temperature and thermal stabilization of taq polymerase by the disaccharide trehalose. Clin Chem 50(7):1256-9.

Sun YP, Fu K, Lin Y, Huang W. 2002. Functionalized Carbon Nanotubes: Properties and Applications. Accounts of Chem Res 35(12):1096–104. doi:10.1021/ar010160v.

Tan G, Gao Y, Shi M, Zhang X, He S, Chen Z, An C. 2005. SiteFinding-PCR: a simple and efficient PCR method for chromosome walking. Nucleic Acids Res 33(13):e122.

Tanguay RL, Pfeifer GP, Riggs AD. 1990. PCR-aided DNaseI footprinting of single copy gene sequences in permeabilized cells. Nucleic Acids Res 18(19):5902.

Theuns I, Windels P, De Buck S, Depicker A, Van Bockstaele E, De Loose M. 2002. Identification and characterization of T-DNA inserts by T-DNA fingerprinting. Euphytica 123: 75-84.

Thirulogachandar V, Pandey P, Vaishnavi CS, Reddy MK. 2011. An affinity-based genome walking method to find transgene integration loci in transgenic genome. Anal Biochem 416(2):196-201. doi: 10.1016/j.ab.2011.05.021.

Tonooka Y, Mizukami Y, Fujishima M. 2008. One-base excess adaptor ligation method for walking uncloned genomic DNA. Appl Microbiol Biotechnol 78(1):173-80.

Tonooka Y, Fujishima M. 2009. Comparison and critical evaluation of PCR-mediated methods to walk along the sequence of genomic DNA. Appl Microbiol Biotechnol 85(1):37-43. doi: 10.1007/s00253-009-2211-5.

Triglia T, Peterson MG, Kemp DJ. 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res 16(16):8186.

Tsaftaris A, Pasentzis K, Argiriou A. 2010. Rolling circle amplification of genomic templates for inverse PCR (RCA-GIP): a method for 5'- and 3'-genome walking without anchoring. Biotechnol Lett 32(1):157-61. doi: 10.1007/s10529-009-0128-9.

Tsuchiya T, Kameya N, Nakamura I. 2009. Straight walk: a modified method of ligation-mediated genome walking for plant species with large genomes. Anal Biochem 388(1):158-60. doi: 10.1016/j.ab.2009.02.002.

Tsuei DJ, Chen PJ, Lai MY, Chen DS, Yang CS, Chen JY, Hsu TY. 1994. Inverse polymerase chain reaction for cloning cellular sequences adjacent to integrated hepatitis B virus DNA in hepatocellular carcinomas. J Virol Methods 49(3):269-84.

Uren AG, Mikkers H, Kool J, van der Weyden L, Lund AH, Wilson CH, Rance R, Jonkers J, van Lohuizen M, Berns A, Adams DJ. 2009. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. Nat Protoc 4(5):789-98. doi: 10.1038/nprot.2009.64.

Van den Eede G. 2010. Compendium of reference methods for GMO analyses. Publications Office of the European Union. http://publications.jrc.ec.europa.eu/repository/handle/111111111/15068

Varadaraj K, Skinner DM. 1994. Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. Gene 140(1):1-5.

de Vicente MC, Guzmán FA, Engels J, Rao VR. 2005. Genetic characterization and its use in decision making for the conservation of crop germplasm. The role of biotechnology. Villa Gualino, Turin, Italy – 5-7 March, 2005.

Waller CF, Dennebaum G, Feldmann C, Lange W. 1999. Long-template DNA polymerase chain reaction for the detection of the bcr/abl translocation in patients with chronic myelogenous leukemia. Clin Cancer Res 5(12):4146-51.

Walser JC, Evgen'ev MB, Feder ME. 2006. A genomic walking method for screening sequence length polymorphism. Mol Ecol Notes 6(2): 563–567. doi: 10.1111/j.1471-8286.2006.01276.x.

Wandeler P, Camenisch G. 2011. Identifying Y-chromosomal diversity by long-template PCR. Mol Ecol Resour 11(5):835-41. doi: 10.1111/j.1755-0998.2011.03013.x.

Wang S, He J, Cui Z, Li S. 2007. Self-formed adaptor PCR: a simple and efficient method for chromosome walking. Appl Environ Microbiol 73(15):5048-51.

Weisser M, Oostdijk EA, Willems RJ, Bonten MJ, Frei R, Elzi L, Halter J, Widmer AF, Top J. 2012. Dynamics of ampicillin-resistant Enterococcus

faecium clones colonizing hospitalized patients: data from a prospective observational study. BMC Infect Dis 12:68. doi: 10.1186/1471-2334-12-68.

Windels P, Taverniers I, Depicker A, Bockstaele EV, Loose MD. 2001. Characterisation of the Roundup Ready soy insert. Eur Food Res Technol 213:107-12.

Winship PR. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res 17(3):1266.

Wood WI, Gitschier J, Lasky LA, Lawn RM. 1985. Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. Proc Natl Acad Sci U S A 82(6):1585-8.

Yi B, Rajagopalan R, Foley HC, Kim UJ, Liu X, Eklund PC. 2006. Catalytic polymerization and facile grafting of poly(furfuryl alcohol) to single-wall carbon nanotube: preparation of nanocomposite carbon. J Am Chem Soc 128(34):11307-13.

Yin B, Largaespada DA. 2007. PCR-based procedures to isolate insertion sites of DNA elements. Biotechniques 43(1):79-84.

Yuanxin Y, Chengcai A, Li L, Jiayu G, Guihong T, Zhangliang C. 2003. T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. Nucleic Acids Res 31(12):e68.

Zhang Z, Gurr SJ. 2000. Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. Gene 253(2):145-50.

Zhang Z, Wang M, An H. 2007. An aqueous suspension of carbon nanopowder enhances the efficiency of a polymerase chain reaction. Nanotechnology 18(35):355706.

Zhang Z, Shen C, Wang M, Han H, Cao X. 2008. Aqueous suspension of carbon nanotubes enhances the specificity of long PCR. Biotechniques. 44(4):537-8, 540, 542. doi: 10.2144/000112692.

Zhang Z, Yang X, Meng L, Liu F, Shen C, Yang W. 2009. Enhanced amplification of GC-rich DNA with two organic reagents. Biotechniques 47(3):775-9. doi: 10.2144/000113203.

Zheng SJ, Henken B, Sofiari E, Jacobsen E, Krens FA, Kik C. 2001. Molecular characterization of transgenic shallots (Allium cepa L.) by adaptor ligation PCR (AL-PCR) and sequencing of genomic DNA flanking T-DNA borders. Transgenic Res 10(3):237-45.

Zhou O, Fleming RM, Murphy DW, Chen CH, Haddon RC, Ramirez AP, Glarum SH. 1994. Defects in Carbon Nanostructures. Science 263(5154):1744-47. doi: 10.1126/science.263.5154.1744.