anses
French agency for food, environmental and occupational health & safety
Investigate, evaluate, protect

Request for scientific and technical support from the National Reference Laboratories on GMOs regarding methods for detecting NPBT products

Request No "2018-SA-0091"

Scientific and technical support REPORT

July 2018

French Agency for Food, Environmental and Occupational Health & Safety,
14 rue Pierre et Marie Curie, 94701 Maisons-Alfort Cedex
Telephone: +33 (0)1 49 77 13 50 - Fax: + 33 (0)1 49 77 26 26 - www.anses.fr

Keywords

GMO, mutation, selection, new breeding techniques, NBT, NGS

Presentation of the participants

PREAMBLE: The experts, members of the Expert Committees and Working Groups or designated rapporteurs are all appointed in a personal capacity, *intuitu personae*, and do not represent their parent organisation.

ANSES PARTICIPATION

Scientific coordination

– – Bacteriology, Virology and GMO unit – Plant Health Laboratory – ANSES

Scientific contribution

– – Bacteriology, Virology and GMO unit – Plant Health Laboratory – ANSES

– – Bacteriology, Virology and GMO unit – Plant Health Laboratory – ANSES

– – Bacteriology, Virology and GMO unit – Plant Health Laboratory – ANSES

CONTRIBUTIONS FROM OUTSIDE THE AGENCY

– – BioGEVES – GEVES

– – Strasbourg Laboratory – SCL

– – Strasbourg Laboratory – SCL

CONTENTS

Acronyms and abbreviations

DGAL: French Directorate General for Food

DNA: Deoxyribonucleic acid

EURL: European Union Reference Laboratory

GEVES: Study Group for the Control of Varieties and Seeds

GMO: Genetically modified organism

HCB: French High Council for Biotechnologies

NRL: National Reference Laboratory

NBT: New breeding technique

NGS: Next-generation sequencing

NPBT: New plant breeding technique

PCR: Polymerase chain reaction

RNA: Ribonucleic acid

SCL: Joint Laboratory Service

SDN: Site-directed nuclease

SNP: Single-nucleotide polymorphism

Definitions

In accordance with the request issued by the DGAL, this document directly relies on the opinion of the Scientific Committee of the French High Council for Biotechnologies published on 2 November 2017 and includes the definitions of the terms used therein.

However, certain terms have specific definitions in the context of this document:

- Detection: ability to detect traces of a modification event in a simple or complex product (e.g. batch of seeds, flour, foodstuff, etc.)
- Identification: ability to identify a modification event in a sample from only one individual (leaf fragments, individual seed, etc.)

1 Background, purpose and terms of implementation of the work

This document was written by the ANSES National Reference Laboratory (NRL) (Plant Health Laboratory), in collaboration with the two other NRLs mandated to undertake official analyses for the detection of GMOs (GEVES: BioGEVES Molecular Biology and Biochemistry Laboratory; SCL: Strasbourg Laboratory). It followed a request for scientific and technical support submitted by the Directorate General for Food (DGAL) of the Ministry of Agriculture, received by ANSES on 10 April 2018.

Based on the various situations examined by the Scientific Committee of the High Council for Biotechnologies (HCB) in its opinion on new breeding techniques (NBTs, or NPBTs for new plant breeding techniques) published on 2 November 2017 [1], the DGAL asked the NRLs to:

i) identify the techniques currently available for the detection, identification and quantification of NPBT products,

ii) take into account requirements in terms of the accreditation and performance of analytical methods as well as constraints such as analysis costs and times,

iii) indicate possibilities and prospects for the development of suitable methods and, where relevant, development or research needs.

2 Situations examined in the opinion of the HCB Scientific Committee

This document does not describe qualified NPBT technologies but focuses exclusively on analytical methods that can be used to detect, identify and/or quantify products derived from these technologies. In its opinion, the HCB Scientific Committee examined various situations described below.

2.1 Targeted genome modifications

There are various technologies for obtaining an organism whose genome has been modified at a specific pre-defined point via directed nucleases. The HCB opinion distinguishes between products derived from three main types of modifications.

An SDN-1 (site-directed nuclease 1) modification is an insertion, deletion or substitution of one or more nucleotides at a specific location within a genome.

An SDN-2 modification is an allelic conversion that involves modifying all or part of a gene sequence in a plant. The location and copy number of the gene are not modified.

An SDN-3 modification is a targeted integration of a DNA sequence regardless of its origin and length.

2.2 Epigenetics

The term "epigenetics" was used by Waddington in 1942 [2] to describe phenotype modifications that can be transmitted during cell division without any genotype changes. Epigenetic mutations can be classified into three types depending on their genotype dependency [3]. "Pure" epialleles are completely independent from the genome, "facilitated" epialleles are related to or even caused by a genome variation (but can persist after it disappears by excision or segregation), and "obligate" epialleles are directly related to a genome variation and co-segregate with it.

In eukaryotic cells, the DNA in the nucleus is arranged in the form of a compact and organised structure called chromatin. It is wound around protein complexes called histones. This organisation is extremely dynamic: the DNA/histone complexes can be assembled, stabilised, destabilised, disassembled, etc. These chromatin dynamics regulate DNA accessibility and are responsible for the regulation of gene expression during the lifetime of a cell or an individual. This regulation can also be transmitted during cell division or to offspring in which case it falls within the definition of epigenetics. The accessibility of the genetic information in chromatin relies on various mechanisms, the two most important being DNA methylation and covalent histone modifications [for review, see: 4, 5].

2.3 Integration of vectors and effectors

Vectorisation is the act of introducing and/or enabling the expression of one or more macromolecules (DNA, RNA, proteins) in a cell. These macromolecules can be transferred through physical methods (electroporation, biolistics, whiskers, etc.) or via a viral or bacterial vector (*Agrobacterium tumefaciens*, for example). The use of biological vectors can cause small DNA fragments, with lengths of only a few to around 100 base pairs, to be integrated into the genome of the target cell [6].

Effectors are proteins or nucleic acids (DNA or RNA) whose action within a cell leads to the occurrence of the expected modification. Directed nuclease techniques enabling specific genome sites to be modified repeatedly use a limited number of effectors. This frequent use can enable detection with tools that target preserved areas.

The presence of vectors and effectors in an organism is assumed to be temporary and unless a sequence has been intentionally integrated, no sequences derived from the vector or effector should persist in the transformed organism. If such a sequence is integrated, its detection can provide evidence of the technique used, although some of these sequences can also be naturally integrated in plant genomes [7].

3 Current methods for the detection, identification and quantification of SDN-3 GMOs

Currently, real-time PCR is the most widespread technique for the detection of genetically modified plants. A pair of oligonucleotides (primers) is used to amplify a specific region of the genome. The primer sequence is chosen to be specific to a region that is present in the genome of a genetically modified plant but absent from non-modified varieties. During PCR, amplification of the target DNA is monitored in each amplification cycle via a fluorescence signal. The amplification kinetics of samples of known and unknown concentrations are compared to determine both the status of the unknown sample (presence/absence of the target) and its concentration (amount of target DNA introduced into the reaction medium). A hydrolysis probe (TaqMan) can be used as needed to improve specificity.

The strategy recommended by the European Union Reference Laboratory (EURL), which is commonly implemented by the laboratories, starts with a step involving the detection of genetic elements common to several events (promoters, terminators, etc.). This step enables the sample's status to be determined (whether or not it contains a genetically modified organism (GMO)). It can also help guide the analyses to be undertaken to identify the GMO. Identification occurs through a battery of tests, each of which is specific to one or more GMOs. In parallel, the laboratories carry out a quantification analysis targeting a common region of all varieties of the target species as well as a quantification analysis specific to the identified GMO. The relationship between the two results is used to establish the level of the identified GMO within the species. This strategy is implemented for all types of matrices (seeds, processed products, etc.) and makes it possible to detect, identify and quantify GMOs in complex samples (mixtures of species, low level of GMOs, etc.).

Performance: The limit of detection for real-time PCR as currently used is around three to five copies of target DNA per well. It is commonly acknowledged that the minimum relative limit of detection reached is 0.01% per plant species. However, this constraint is primarily due to sample heterogeneity and the ability to collect a test sample representative of the entire sample, not to the performance of PCR itself. The absolute quantification range is between 20 and more than 100,000 copies, i.e. an order of magnitude of 10^5. In order to position samples in relation to the regulatory thresholds (0.1% and 0.9%), most of the methods that have been published have been validated between 0.1 and 5 to 10%, although it is technically possible to exceed this maximum threshold.

Accreditation: Real-time PCR has been a proven technique for more than 10 years. ISO standards are available [8-11], as are an implementation guide published by AFNOR [12] and several reference documents published by the EURL. The three NRLs routinely use the methods validated and published by the EURL and have been accredited for their implementation. A guidance document on flexible scope accreditation published by the EURL is also available [13]. Flexible scopes reduce accreditation costs in view of the rate of authorisation of genetically modified plants and the publication of detection and quantification methods.

Cost: Like any molecular biology method, real-time PCR requires suitable facilities enabling the risk of contamination to be controlled. This technology requires the acquisition of a compatible thermocycler, whose price ranges from €20,000 to €40,000 depending on the model and manufacturer. For information, the cost of consumables for the analysis of a negative maize sample is around €75 (extraction and detection PCR).

Duration: The analysis of a sample is broken down into several stages. Sample preparation can take one to two hours, depending on the nature of the sample and the quantity to be analysed. Most EU laboratories process runs of around eight samples in parallel. The extraction of a run lasts

between half a day and a whole day. In order to reduce handling times, automated extraction machines are available on the market (€60,000 to €75,000 depending on the type of machine), and some European laboratories are equipped with such machines. Identification and quantification PCR techniques can be pooled for several samples. In accordance with the requirements defined in the ISO standards, PCR in a 96-well plate can enable the simultaneous analysis of 23 samples for identification and eight samples for quantification. Preparation can take 30 minutes to one hour, while analysis lasts between 90 minutes and two hours. Several weeks can thus be required for the analysis of complex samples containing several species that may be genetically modified. For example, the SCL, whose mandate encompasses all plant species, is likely to detect up to 52 different GMOs in a sample containing soya, maize and rapeseed.

Limitations: The use of real-time PCR requires prior knowledge of the sequence or polymorphism to be detected. The validation of a real-time PCR method requires relatively pure positive material. Without knowledge of the sequences or material, it is not possible to develop or validate a method.

PCR quantification requires the use of positive controls, extracted from certified reference materials, if possible at several concentrations, in order to generate a series of standards and verify the accuracy of the measurements. Such controls are available for genetically modified plants that have been covered by a marketing authorisation application in the EU. It is more complicated to obtain reference materials for unauthorised plants.

One of the limitations of real-time PCR is the use of a specific method for each target. Indeed, the rise in the number of GMOs to be detected has led to an increase in the number of methods to be implemented. Given the throughput of the platforms used, this increasing number of methods is already problematic.

Development: According to Directive 2001/18/EC, applicants must submit an application for authorisation to use a genetically modified plant. This application must be accompanied by a detection, identification and quantification method that has been developed by the applicant and whose minimum performance criteria have been established by the EURL and the network of European NRLs [14]. The method must be submitted to the EURL, which validates it and makes it available to the NRLs. Therefore, the NRLs are not currently responsible for developing methods for GMOs covered by an authorisation application.

If the presence of an unknown GMO is suspected, a sequencing step is essential in order to develop a specific method. Sequencing methodologies have been published by laboratories working on this theme [15].

4 Current analytical capacities of the NRLs based on real-time PCR

The NRLs currently detect, identify and quantify transformation events using real-time PCR methods targeting known sequences with lengths of around 70-80 base pairs. These analytical procedures are the same as those to be used for the detection, identification and quantification of SDN-3 products with knowledge of the targets to be detected.

The platforms used by the NRLs also enable the specific detection, identification and quantification of DNA with a known polymorphism in a single nucleotide (single nucleotide polymorphism), corresponding to the cases of SDN-1 and SDN-2. Indeed, real-time PCR techniques can be used with hydrolysis probes specific to one nucleotide (e.g. TaqMan-MGB probes). This approach is suitable for the detection of a small number of SNPs and may be used to detect the adventitious presence of a few dozen SNP modifications with the same model as that currently used for monitoring the adventitious presence of GMOs, using the same real-time PCR equipment. If a large number (hundreds or thousands) of SNPs needed to be monitored, the use of real-time PCR equipment would no longer be competitive in relation to the technological alternatives described in Section 5.2, "SNP genotyping".

Real-time PCR can also be used to detect residual vector or effector sequences.

Performance: Real-time PCR enables the detection, identification and quantification of products obtained by SDN-1, SDN-2 or SDN-3 as well as vector or effector sequences. For SDN-3, and for vector and effector sequences, the expected level of performance is that currently obtained with the analysis of transformation events. The expected limit of detection of SNP assays for the analysis of SDN-1 or -2 products remains extremely low (around 20 copies of target DNA per reaction [16]). However, as the specificity of the reaction lies only in a base, the background noise obtained in the presence of non-mutant DNA will prevent mutant DNA under the 1% threshold from being detected in non-mutant DNA [17].

Accreditation: These methods are no different from those commonly used by the NRLs and can be accredited on the basis of existing quality systems.

Cost: The necessary equipment (real-time thermocycler) costs between €20,000 and €40,000. It is already in place in the laboratories. The number of targets that can be analysed with this type of technology is limited. However, it can be increased considerably through the use of automated preparation platforms (€60,000 to €75,000).

The unit cost of SDN-1, -2, -3, effector or vector detection by real-time PCR corresponds to the current cost of detecting a transformation event. If multiple SNPs are detected in the same sample, a reaction is necessary for each SNP tested.

Duration: These methods are no different from those commonly used by the NRLs and the analysis times are similar (except if there is a sharp increase in the number of SNPs to be tested).

Limitations: These techniques cannot be used without prior information about the sequence or polymorphism to be detected. Moreover, for modifications likely to occur naturally, the technology will not provide information regarding the conditions in which this polymorphism occurred or was obtained. The analysis of samples containing non-mutant versions mixed with the target raises the relative limit of detection for the target. Lastly, real-time PCR platforms require the preparation of a reaction mixture for each target and are not suitable when multiple targets are to be tested for each sample.

Research and development: Although performance data for real-time PCR tools targeting SNPs are available in the literature, the GMO NRLs do not have access to such data in their work environments. The implementation of specific validation procedures with suitable equipment would enable the laboratories to prepare for the possibility of such analyses.

The analysis of commonly used vector and effector sequences should enable the definition of one or more real-time PCR methods capable of ensuring that these genetic elements are not contained in the analysed product. Such a method will not be able to determine whether portions of these genetic elements have been introduced into the genome.

5 Medium-term possibilities for technical analyses

There are various technologies that are routinely used and may be implemented by the NRLs if necessary to carry out analyses:

5.1 Real-time PCR via a microfluidic platform

As mentioned above, the detection of SDN-1 and SDN-2 corresponds to the analysis of single mutations (SNPs). Real-time PCR is entirely suitable for the detection, identification and quantification of such genome modifications, provided that there are not too many targets. There are platforms, such as microfluidic chips, which are more suitable than conventional real-time PCR devices for the implementation of several hundred TaqMan-MGB genotyping tests. For example, the Fluidigm® microfluidic chips support up to 4608 simultaneous reactions for the genotyping of 192 SNPs in 24 samples.

Accreditation: The reaction is still a real-time PCR reaction and can therefore be accredited with no major changes to the quality system in place in the laboratories.

Cost: Microfluidic platforms are significant investments. For example, the Biomark Fluidigm® platform costs around €200,000. Each reaction occurs within a chip enabling the simultaneous analysis of 48 targets in 48 samples, 96 targets in 96 samples, or 192 targets in 24 samples. The cost of this chip ranges from €500 to €1000. As for the reagents, they are the same as those currently used; however, since the reactions take place in very small volumes, the reagent costs are much lower than the current costs.

Duration: The duration of the reaction is the same as with conventional real-time PCR. It takes much longer to prepare a chip than to prepare a conventional plate. However, a microfluidic plate enables the equivalent of 96 conventional plates to be processed in just one reaction.

Limitations: Aside from the opportunity to reduce the number of reactions and thus reduce analysis times, the limitations are the same as for conventional real-time PCR. Indeed, this technique requires prior knowledge of the sequence or polymorphism to be detected. During SNP detection, it does not provide any information regarding the conditions in which this polymorphism occurred or was obtained.

Research and development: Transferring a real-time PCR method from a conventional platform to a microfluidic platform is not expected to pose any technical problems. Such a platform is available in an ANSES laboratory and may be used for preliminary testing if necessary without any major investments.

5.2 SNP genotyping

Other technologies are already used for the genotyping of individuals via the analysis of numerous pre-identified SNPs. Genotyping technologies using fluorescent primers (KASP™), for example, enable several hundred SNPs to be analysed in the same sample. There are also Affymetrix® hybridisation chips enabling tens of thousands of SNPs to be genotyped.

Performance: These technologies are suitable for genotyping and enable a sample or individual to be identified based on a very large number of SNPs analysed simultaneously. However, the modified target must make up at least 50% of the sample, which means that these technologies are not suitable for detecting the adventitious presence of an SNP occurring at a low level. These

techniques may be used to determine the presence or absence of a target sequence (SDN-3, vectors, effectors). However, they do not enable a target to be quantified.

Accreditation: Genotyping with KASP™ fluorescent primers is not significantly different from real-time PCR. Undertaking such analyses under accreditation is feasible subject to adaptation of the quality system and validation of all of the amplifications required for SNP analysis. Genotyping via chips is used for medical diagnoses by laboratories complying with quality commitments. However, the quality management associated with this type of platform is very different from that currently in place in the laboratories.

Cost: KASP™ genotyping can be performed using equipment dedicated to real-time PCR. However, in order to make the most of the technology and increase the number of targets, it is necessary to invest in equipment achieving higher throughput such as an automated plate-preparation machine (€75,000), a hydrocycler (€50,000) and a fluorescence reader combined with dedicated genotyping software (€100,000). The consumables associated with KASP™ genotyping are PCR consumables. Furthermore, since it is not necessary to have a fluorescent probe produced for each target, costs are reduced. The genotyping of 380 SNPs in a sample costs around €50 (extraction, assay, normalisation, PCR and analysis).

To analyse the chips, it is necessary to invest in dedicated hybridisation and reading devices. The cost of these platforms varies significantly depending on the level of automation, the throughput and the related applications (GeneAtlas™: around €350,000; GeneChip®: around €1,800,000; and GeneTitan®: around €3,000,000). The cost associated with the use of chips is mainly related to their development: for each new SNP to be tested, a probe has to be custom-created and tested in order to undergo quality checks performed by the supplier. For information, excluding equipment and development costs, a genotyping programme for 70,000 SNPs in 200 samples using chips available in the supplier's catalogue costs €20,000 (reagents and analysis).

Duration: If all SNPs are tested simultaneously, an analysis with fluorescent primers can be undertaken by simple PCR. For an analysis using hybridisation chips, the entire protocol (nucleic-acid preparation, hybridisation and scanning) lasts around three days.

Limitations: Although they can be used for the identification of SDN-3 products and vector or effector sequences, these technologies are particularly recommended for the identification of SDN-1 and SDN-2 products. In all cases, the polymorphisms must have been identified beforehand and the supplied samples must be pure. With KASP™ genotyping, it is necessary to simultaneously analyse a minimum of 24 samples or controls to reliably predict the identity of the SNPs.

The use of chips requires expertise that is not currently available in the laboratories. Moreover, aside from the cost of development, the optimisation of chip-production costs requires batch manufacturing, which reduces the flexibility of the targets that can be analysed. To read the chips, it is also necessary to invest in dedicated hybridisation and reading devices.

Research and development: The shift to genotyping technologies is feasible but will require extensive methodological work before any routine implementation. A dedicated platform must first be created. Then, for each SNP, a PCR method or hybridisation probe must be developed and validated. Lastly, the analysis of results requires appropriate bio-computing and statistical tools. GEVES already uses genotyping by fluorescent primers as part of its missions for the characterisation and monitoring of plant varieties.

5.3 Sequencing of fragments of interest

From a nucleic-acid extract, it is possible to isolate a fragment of interest by amplification or capture and sequence it. In the event of genetic modifications, differences may be identified between the sequence of the fragment of interest in the modified plant and the expected sequence.

Depending on the number of sequences of interest, the necessary sequence length or the number of samples to be analysed, Sanger sequencing or high-throughput sequencing (next-generation sequencing, NGS) may be used. For each target sequence, Sanger sequencing will give a consensus sequence for each sample while NGS will produce numerous sequences for each sample. In addition, high-throughput sequencing enables numerous samples to be multiplexed (analysed simultaneously) by identifying each sample with a barcode (tagging).

Performance: Sequencing is a highly effective technique enabling the exact sequence of a fragment of interest to be determined. This technique is able to highlight single mutations (SNPs) (SDN-1 or -2) or more significant modifications for several bases (SDN-2 or -3). The generation of a chromatogram with Sanger sequencing and the monitoring of NGS data with FastQC software are useful for ensuring the quality of the data obtained. Sanger sequencing can only be used for identification whereas NGS can be used for detection. The performance level of this detection depends on the sequencing depth, which is related to the technology used, the number of multiplexed samples, and therefore the cost of sequencing. For information, analyses have been undertaken with a relative limit of detection of 1%.

The size of the fragments of interest to be sequenced also varies depending on the sequencer used and can range from a few dozen bases (Sanger, Illumina) to 100,000 bases (MinIon™).

Accreditation: The accreditation of these approaches is feasible provided that the quality of the sequencing steps is controlled (by a service provider or the laboratory). In light of the development and deployment of small MinIon™ laboratory sequencers, the laboratory may be able to carry out and control the whole analysis in the medium term.

Cost: For outsourced sequencing, the cost depends on the technology used. It is around €5 per targeted sequence and per sample for Sanger sequencing. For NGS, the high cost of a sequencing reaction is offset by the ability to multiplex a large number of samples and targets. This multiplexing lowers the cost to around €20 per target and per sample. To simultaneously detect polymorphisms in several genome regions, it is necessary to perform multiple rounds of sequencing or increase the use of multiplexing.

Sequencing by the laboratory will require the acquisition of a suitable sequencer (for example, Illumina MiniSeq™, which costs around €50,000). The cost per sample will also be higher due to the lower analysis throughput.

Duration: When sequencing is performed by private providers or on platforms, the minimum analysis time is around one week for Sanger sequencing. It can reach several months for certain platforms. These analysis times will be considerably reduced if the full analysis can be controlled within the laboratory.

Limitations: This approach requires prior knowledge of the region to be sequenced in order to be able to amplify or capture the fragment of interest using PCR. The amplification of each fragment is relatively easy to develop but in terms of handling, it requires a reaction for each fragment. Conversely, capture PCR is still technically uncertain (lack of specificity) but should, in the long run, enable a large number of sequences of interest to be selected simultaneously. Another limitation of this approach is the rate of errors introduced during the PCR step and during sequencing. However, techniques using barcodes (tagging) enable true mutations to be distinguished from those related to PCR and sequencing errors [18]. Depending on the type of technology used, the analysis of results requires appropriate bio-computing and statistical tools.

Research and development: The shift to technologies for the routine sequencing of fragments of interest will require some degree of methodological work, depending on the approach considered both for the selection of sequences of interest and for sequencing itself.

As mentioned earlier, the use of PCR does not particularly require any adjustments but will rapidly become limiting. The system for capturing sequences of interest would enable a large number of targets to be selected and then sequenced. However, the tests undertaken by ANSES lack

specificity and do not allow the extract to be sufficiently enriched with target DNA. ANSES would like to sustain research efforts focusing on this sample-preparation method.

Regarding the sequencing step, Sanger technology can only be used after a PCR reaction that has been developed and validated for each genome portion to be tested. In this case, the analysis of results requires relatively accessible bio-computing tools. The use of outsourced NGS requires expertise in terms of sample preparation and the bio-computing tools to be used are also more complex. Lastly, the management of the entire analysis requires the acquisition and control of a sequencing platform as well as proficiency in the appropriate bio-computing and statistical tools. ANSES currently has IonProton™ and MinIon™ NGS platforms as well as a dedicated bio-computing department. The ANSES GMO NRL has a MinIon™ platform, is experimenting with its use for the identification of transformation events, and would like to sustain research efforts focusing on this sample-analysis method.

5.4 DNA methylation analysis

DNA methylation involves the covalent addition of methyl groups to certain DNA nucleotides (primarily cytosines). Modification of DNA methylation is associated with gene regulation in eukaryotes [for review, see: 19].

The main technology used to study DNA methylation is bisulphite conversion combined with sequencing. During this treatment, unmethylated cytosines are converted into uracils. The converted DNA can then be amplified and sequenced. If there are cytosines in the obtained sequence, this means they contain methyl groups. Data are analysed to determine the methylation of a site as well as the methylation level of an analysed genome region.

Bisulphite treatment does not pose any major technical difficulties and can be incorporated into a traditional analysis scheme with amplification and/or sequencing.

Performance: Bisulphite conversion is a method of DNA preparation. The level of performance depends on the subsequent technology. For the detection of an epigenetic mutant identified at a set position, real-time PCR can be used to detect this mutant with a limit of detection of 0.01%. Conversely, unless a complex and expensive strategy is used, the sequencing of the converted DNA will not enable the detection of mixed individuals but simply the identification of separate individuals.

Accreditation: The accreditation of such a protocol seems feasible under the same conditions as for the subsequent technologies (PCR, sequencing, etc.).

Cost: The cost of DNA conversion is very low (around €2 per extract), while the cost of analysis depends on the technology applied to the converted DNA.

Duration: The duration of bisulphite treatment is not limiting in the activity framework of the NRLs.

Limitations: As indicated above, bisulphite treatment is merely a method for preparing DNA. Combined with real-time PCR, it easily enables the specific detection of a given sequence with a given methylation level. Combined with sequencing, it easily enables the methylation levels of targeted sequences to be compared. The limitations are the same as for conventional analyses: knowledge of the potentially modified site and its expected status is a prerequisite for any analysis. Without prior knowledge, only whole-genome sequencing can be implemented. The data analysis could reveal major anomalies in the methylation profile of a genome. A more in-depth analysis would require a reference epigenome that is not currently available.

Research and development: This approach is commonly used for research purposes. The development of partnerships between analytical and research laboratories should enable skills to be transferred. The acquisition of data regarding the methylation levels of genomes for species of interest is a prerequisite for any whole-genome analysis.

6 Long-term possibilities for technical analyses

6.1 Whole-genome sequencing and analysis of exogenous genetic elements

The insertion of a sequence whose nature and position have not been identified cannot be detected via the targeted approaches currently used. In order to identify this type of modification, it is necessary to analyse the full genetic information of the individual or test sample. New high-throughput sequencing technologies (NGS) enable the acquisition of a full sequence for an individual or a large portion of the genetic information contained in a sample. Using bio-computing tools, these sequences can then be compared with databases of available sequence data for various organisms.

Performance: By analysing a whole genome, it is possible to verify the presence of SDN-1, -2 or -3 modifications at pre-determined locations. However, this approach is excessive for this type of operation. This approach can also enable exogenous sequences to be identified without any prior information and is therefore applicable to SDN-3 products and the detection of vector and effector sequences. Its performance level directly depends on the sequencing depth and analytical procedures. In the context of identification, the full coverage of a large plant genome such as that of maize requires extensive and expensive sequencing. In the context of detection in a complex sample, the required sequencing depth is too large for it to be feasible to date.

The analysis of results also influences the performance level of this approach. The simplest and most effective strategy involves creating a database of the sequences potentially used for transformation and comparing each raw sequence (read) from the sequencing with those listed in this database. However, this approach only enables the identification of a sample containing one of the pre-identified sequences included in the database. Another approach would be to compare each read with all of the known data. However, this would require computing power that is not readily accessible.

Accreditation: As mentioned above, the accreditation of this approach requires quality control for the sequencing steps. In light of the development and deployment of small laboratory sequencers, the laboratory may be able to control the whole analysis in the medium term.

Cost: The systems likely to offer an appropriate sequencing depth for this type of approach (for example, Illumina NextSeq™ and HiSeq™) can be accessed from certain sequencing platforms. Many companies also offer sequencing services using this type of equipment. For information, the purchase cost of an Illumina NextSeq™ is around €250,000, while that of an Illumina HiSeq™ is around €650,000.

An Illumina HiSeq™ sequencing service providing sufficient coverage for a whole maize genome currently costs around €4000.

Duration: Depending on the system used and the desired sequencing depth, sequencing lasts between 10 hours and six days.

Limitations: Equipment availability, the cost of the analyses and expertise in the techniques are some of the barriers to implementation of this type of approach. Moreover, this strategy relies on the identification of a sequence from the genome of a different organism via the creation of an appropriate database. It does not enable modifications in the sequence to be detected without prior knowledge. Lastly, this approach requires technical expertise and bio-computing skills that the

laboratories do not necessarily have. With lower sequencing costs and the possible acquisition of sequencers by the laboratories, this type of approach should become feasible.

Research and development: As indicated above, the use of NGS requires expertise in terms of sample preparation, sequencing operations and data processing. These approaches are widely deployed in various fields of research and are starting to be used for analyses. However, their transfer requires that the laboratories acquire new skills.

The use of a specific database of the sequence data potentially used in the event of SDN-1, -2 or -3 modifications (vectors, effectors, promoters, genes of interest, terminators, etc.) requires prior compilation and validation work. Such a database then needs to be regularly updated with data from the scientific literature, patents, etc. Such a database is maintained by the EURL based on the sequence data submitted by applicants with authorisation applications. However, this database cannot be made available to the NRLs for data-ownership reasons.

6.2 Whole-genome sequencing and statistical approaches

On the basis of whole-genome sequencing results, it is possible to consider an approach that does not seek to compare sequences from a sample with sequences from the genomes of other organisms but rather seeks to identify anomalies in a sequence via statistical approaches. These anomalies can be attributed to the insertion of an exogenous sequence.

Performance: This would be the only approach enabling certain SDN-3 modifications to be identified without prior information and without the introduced sequence coming from a known organism. The performance level of this type of approach for the identification of SDN-3 modifications has not yet been assessed. However, with a such a tool being based on a statistical difference between the gene pool and the introduced sequence, it is understood that it would not be capable of detecting all SDN-3 modifications.

Accreditation: This is a prospective approach and is not intended for routine use in the short term. Moreover, given the type of result obtained (probability of having a positive sample), the accreditation of such an approach is not currently feasible.

Cost: The sequencing costs are those given in Section 6.1, "Whole-genome sequencing and analysis of exogenous genetic elements".

Duration: The sequencing duration is that given in Section 6.1, "Whole-genome sequencing and analysis of exogenous genetic elements".

Limitations: Such a tool should enable certain genome portions to be identified as likely to have been introduced. This result would not be evidence of the origin of this sequence or of the method used to introduce it into the sequenced organism.

Research and development: This approach is currently being developed at ANSES as part of a thesis project on bio-computing undertaken jointly by the GMO NRL and the bio-computing platform. A first tool will be made available at the end of 2019. Depending on the results obtained, this approach may be developed much more broadly.

6.3 Chromatin conformation

Histone proteins associated with DNA in chromatin can undergo various modifications when acetyl or methyl groups or small peptides are added to their N-terminal tails. All of these modifications impact the three-dimensional conformation of chromatin.

Specific antibodies against certain protein epitopes can be used to immuno-precipitate protein-bound DNA in the desired conformation. However, instead of directly studying histone modifications, some technologies such as ATAC-seq [20] and Hi-C [21] are based on sequencing

and propose to directly study the 3D structure of chromatin. The most accessible of these seems to be ATAC-seq based on the accessibility of certain chromatin sites with transposases, which will insert adapters into accessible sites and enable sequencing to be carried out from these regions of the genome.

Performance: This technology enables open chromatin regions to be mapped for an individual at a given time.

Accreditation: At this point in time, ATAC-seq and other similar technologies are fundamental research tools and are hardly compatible with the constraints associated with accreditation. The main obstacle will be the availability of known reference material with known and controlled conformation DNA.

Cost: The main cost of the technique involves the use of high-throughput sequencing platforms. Depending on the desired resolution, it is possible to multiplex samples and reduce the cost to around €100 per sample.

Duration: Large structures for epigenetic research are capable of processing up to 60 samples per day with ATAC-seq. However, the bio-computing resources required for data processing should not be neglected.

Limitations: Three-dimensional chromatin conformation analysis requires intact cells. It can thus only be used with fresh tissues and never with processed material. Moreover, the physiological status of the plant and its development stage heavily influence the three-dimensional structure of chromatin. Lastly, given the plasticity of the dominant non-coding regions of genomes, the structure changes significantly from one variety to the next. The use of this type of technology to undertake analyses would thus require cultivating the variety suspected of having been modified in parallel with this same variety known as not having been modified. The implementation of such analyses is hardly feasible for the time being.

Research and development: Approaches enabling open chromatin to be described have mainly been used for medical research thus far. The resources required for their implementation are not yet compatible with analyses for the detection or identification of SDN plants.

Date of validation of the report: 31 July 2018

7 References

1. *Avis sur les nouvelles techniques d'obtention de plantes (New Plant Breeding Techniques - NPBT)*. 2017, Conseil scientifique du Haut Comité des Biotechnologies.
2. Waddington, C.H., *The epigenotype.* Endeavour, 1942. **1**: p. 18-20.
3. Richards, E.J., *Inherited epigenetic variation—revisiting soft inheritance.* Nature Reviews Genetics, 2006. **7**(5): p. 395.
4. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nature Reviews Genetics, 2016. **17**(8): p. 487.
5. Shafiq, S. and A.R. Khan, *Plant epigenetics and crop improvement*, in *PlantOmics: the omics of plant science*. 2015, Springer. p. 157-179.
6. Brunaud, V., et al., *T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites.* EMBO reports, 2002. **3**(12): p. 1152-1157.
7. Kyndt, T., et al., *The genome of cultivated sweet potato contains Agrobacterium T-DNAs with expressed genes: An example of a naturally transgenic food crop.* Proceedings of the National Academy of Sciences, 2015. **112**(18): p. 5844-5849.
8. *NF EN ISO 24276 - Produits alimentaires - Méthodes d'analyse pour la détection des organismes génétiquement modifiés et des produits dérivés - Exigences générales et définitions*.
9. *NF EN ISO 21569 - Produits alimentaires - Méthodes d'analyse pour la détection des organismes génétiquement modifiés et des produits dérivés - Méthodes qualitatives basées sur l'utilisation des acides nucléiques*
10. *NF EN ISO 21570 - Produits alimentaires - Méthodes d'analyse pour la détection des organismes génétiquement modifiés et des produits dérivés - Méthodes quantitatives basées sur l'utilisation des acides nucléiques*
11. *NF EN ISO 21571 - Produits alimentaires - Méthodes d'analyse pour la détection des organismes génétiquement modifiés et des produits dérivés - Extraction des acides nucléiques*
12. *GA V03-042 - Produits alimentaires - Manuel d'application aux normes NF EN ISO 24276 , NF EN ISO 21569 , NF EN ISO 21570, NF EN ISO 21571 et leurs amendements 1 - Méthodes d'analyses pour la détection des organismes génétiquement modifiés et produits dérivés*
13. Trapmann, S., et al., *European technical guidance document for the flexible scope accreditation of laboratories quantifying GMOs.* Publications Office of the European Union, Luxembourg, 2013.
14. Mazzara, M., et al., *Definition of minimum performance requirements for analytical methods of GMO testing.* European Network of GMO Laboratories. JRC, European Commission, European Network of GMO Laboratories, Brussels, 2008.
15. Fraiture, M.-A., et al., *An innovative and integrated approach based on DNA walking to identify unauthorised GMOs.* Food Chemistry, 2014. **147**: p. 60-69.
16. Van Ert, M.N., et al., *Strain-specific single-nucleotide polymorphism assays for the Bacillus anthracis Ames strain.* Journal of clinical microbiology, 2007. **45**(1): p. 47-53.
17. Clément, J., et al., *Specific detection and quantification of virulent/avirulent P hytophthora infestans isolates using a real-time PCR assay that targets polymorphisms of the Avr3a gene.* Letters in applied microbiology, 2013. **56**(5): p. 322-332.
18. Ståhlberg, A., et al., *Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing.* Nucleic acids research, 2016. **44**(11): p. e105-e105.

19. Law, J.A. and S.E. Jacobsen, *Establishing, maintaining and modifying DNA methylation patterns in plants and animals.* Nature Reviews Genetics, 2010. **11**(3): p. 204.
20. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nature methods, 2013. **10**(12): p. 1213.
21. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* science, 2009. **326**(5950): p. 289-293.

ANNEXES

Annex 1: Request letter

2018 -SA- 0 0 9 1

COURRIER ARRIVE

1 0 AVR. 2018

DIRECTION GENERALE

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTERE DE L'AGRICULTURE ET DE L'ALIMENTATION

Direction générale de l'alimentation

Service des actions sanitaires en production primaire

Sous-direction de la qualité, de la santé et de la protection des végétaux

Bureau des semences et de la protection intégrée des cultures

251, rue de Vaugirard
75732 PARIS CEDEX 15

Dossier suivi par :

Tél. : 01 49
Fax : 01 49

Réf. :

Le Directeur général de l'alimentation
à
Monsieur le Directeur général de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail
14 rue Pierre et Marie Curie
94701 MAISONS-ALFORT CEDEX

Paris, le - 6 AVR. 2018

Objet: Demande d'appui scientifique et technique du Laboratoire National de Référence OGM de l'ANSES concernant les méthodes de détection des produits issus de NPBT.

Les nouvelles techniques de sélection (NBT pour New Breeding Techniques, ou NPBT pour New Plant Breeding Techniques) comprennent un ensemble de techniques récentes visant à modifier le génome de manière ciblée, à moduler l'expression des gènes ou à appliquer la transgénèse dans des situations particulières. Les techniques de modifications ciblées du génome (mutation ou insertion de gène en un site précis du génome) sont les plus innovantes et sont en évolution rapide.

Les NBT ne sont pas mentionnées dans les listes de techniques qui définissent le champ de la réglementation sur les OGM et dont la rédaction est antérieure à l'apparition de ces techniques. De ce fait, il n'est pas aujourd'hui juridiquement établi si l'utilisation de ces techniques doit ou non respecter le cadre réglementaire tel qu'il a été conçu pour les OGM : évaluation, autorisation, traçabilité, étiquetage et contrôle.

Le Conseil d'Etat a interrogé la Cour de Justice de l'Union européenne (CJUE) sur le statut des nouvelles techniques de mutagénèse dirigée vis-à-vis de la directive 2001/18/CE et sur la possibilité pour les Etats membres d'adopter des mesures nationales pour encadrer les techniques exclues du champ d'application de la directive 2001/18/CE. Les conclusions de la CJUE pourraient être rendues d'ici la fin du printemps 2018.

Le Haut conseil des biotechnologies (HCB) a publié le 2 novembre 2017 un avis sur les NPBT, en réponse à une saisine des ministères chargés de l'environnement et de l'agriculture. Les méthodes d'analyse des produits issus de NPBT font partie des questions étudiées. Ainsi, l'avis du Comité scientifique du HCB examine les possibilités de détection en fonction de la technique utilisée et des informations disponibles sur le produit.

Dans le prolongement de l'avis du HCB, sans préjuger des décisions futures concernant l'encadrement réglementaire des produits issus de NPBT, il serait utile de disposer d'un avis du Laboratoire national de référence (LNR) pour la détection des OGM de l'ANSES sur les techniques actuellement disponibles pour la réalisation d'analyses relatives aux produits issus de NPBT.

1

A partir des différentes situations examinées dans l'avis du Comité scientifique du HCB (NPBT utilisée, informations disponibles sur le produit), il est demandé au LNR d'identifier les techniques actuellement disponibles qui permettraient au LNR de procéder à la détection, l'identification et la quantification des produits issus de NPBT. Il devra être tenu compte des exigences en termes d'accréditation et de performance des méthodes d'analyse liées aux analyses officielles ainsi que des contraintes telles que le coût et la durée de l'analyse.

S'agissant des cas où la détection, l'identification et la quantification ne seraient pas possibles actuellement dans le cadre d'analyses officielles de routine, il est demandé au LNR d'indiquer les possibilités et les perspectives de développement de méthodes adaptées et le cas échéant les besoins de mise au point ou de recherche.

Les deux autres laboratoires désignés LNR pour la détection des OGM seront informés et, s'ils le souhaitent, associés à ce travail.

Je vous saurais gré de bien vouloir me transmettre votre analyse pour fin juillet 2018.

2

Annex 2: Summary table

	Modification type	Real-time PCR	Microfluidic real-time PCR	SNP genotyping	Sequencing of fragments of interest	Whole-genome sequencing	DNA methylation analysis	Chromatin conformation
Prior knowledge of modifications	SDN-1	D – I – Q	D – I – Q	I	I	I		
	SDN-2	D – I – Q	D – I – Q	I	I	I		
	SDN-3	D – I – Q	D – I – Q	I	I	I		
	Epigenetics						I	I
No prior knowledge of modifications	SDN-1							
	SDN-2							
	SDN-3					I		
	Epigenetics						I	I
	Vectors and effectors	D – I – Q	D – I – Q	I	I	I		

D: detection; I: identification; Q: quantification

Technically mature

Mature for other uses, uncertain performance level

Under development, uncertain performance level

Cannot be routinely implemented to date