Reconsideration of the molecular characterisation criteria for marketing authorisation of GM crops

COGEM Report CGM/140929-02

1. Introduction

Before a genetically modified (GM) crop can be approved for use in the European Union, it must first be assessed for any risks it may pose to human health and the environment. This environmental risk assessment must pay particular attention to the molecular characterisation of the GM crop and the inserted sequences. Information about the introduced DNA, the regulatory sequences used, and the bioinformatics analysis of the insertion and the adjacent regions provide indications of the characteristics that have or may have been altered.

The molecular characterisation is followed by a comparative analysis of the relevant characteristics of the GM plant, such as its composition and phenotypic and agronomic characteristics. The analysis involves, among other things, the probability of dispersal by pollen or seeds, the possibility of outcrossing of a GM crop with wild or other relatives, possible changes in persistence, invasiveness and establishment in the wild, as well as potential negative impacts should the inserted gene be disseminated in the environment and any effects on non-target organisms.

In 2008 COGEM drew up several criteria for the molecular characterisation of GM crops:[1]

1. It must be known which elements have been inserted into the plant genome and in how many copies. The function of these elements must be known.
2. The GM plant must be analysed for the presence of DNA from the transformation vector (the backbone DNA).
3. All insertions must be fully characterised by means of a sequence determination and this characterisation must extend into the flanking sequences of the genome.
4. The sequences spanning the insertion or insertions to the genomic DNA of the plant must be bioinformatically analysed. However, there may be situations in which there are justifiable reasons for omitting this analysis.
5. The analysis of the theoretical fusion ORFs must extend from stop codon to stop codon.

In the light of recent advances in scientific understanding, the experience that has since been gained with permit applications for the deliberate release into the environment of GM crops, and the discussions held within COGEM when assessing permit applications, it has been decided to expand on and elucidate the criteria from 2008 and publish this in an update to the topic report of 2008.

2. Molecular characterisation: the different elements

There are two main techniques for creating a GM crop: transformation with *Rhizobium radiobacter* (previously *Agrobacterium tumefaciens*) and particle bombardment. The latter procedure involves bombarding cells with tiny metal balls coated with the DNA to be inserted. Nowadays the most frequently used method is the transformation procedure using the bacterium *R. radiobacter*.

R. radiobacter is a Gram-negative bacterium found in soil. It is able to insert a DNA fragment (T-DNA or Transfer DNA) into the genome of a plant cell. The T-DNA is part of a plasmid and is flanked by a left border and a right border (LB and RB). During infection with *R. radiobacter*, the DNA fragment that lies between the LB and RB is transferred to the plant cell and incorporated into the plant genome. This property is why *R. radiobacter* is frequently used to transform plants through the insertion of the desired DNA fragment.[2]

2.1 The insertion

For the environmental risk assessment it is important to know which genes and sequences have been introduced into the plant and what their functions are. A description of the transformation vector, the insertion cassette, the functions of the various genes, the regulatory sequences of the vector used and the sequence order of the insertion cassette must all be available.

2.1.1 Characterisation of the insertions

Copy number

During the transformation several copies of the desired DNA fragment may be inserted into the plant genome. These copies may be inserted at the same insertion site (in tandem) or distributed throughout the genome. Sometimes a copy will not contain the entire insertion because a part is missing. COGEM emphasises that the DNA sequences of all the elements inserted into the plant genome must be known. The inserted fragments can be detected using Southern blot analysis in which the T-DNA fragment can be used as a probe to detect the fragments.

It is also possible to determine the number of copies and partial copies of the insertion by whole genome sequencing. The possibility of routine whole genome sequencing at low cost is already within reach for plants with a small genome, and genome sequencing is expected to become an alternative to Southern blot analysis within a few years. COGEM points out that permit applicants are not required to hand over the full genome sequence. They do have to provide a bioinformatics analysis containing information on how the analysis was carried out and on the number of copies and partial copies of the insertion that are present.

Characterisation of the insertions

COGEM is of the opinion that the insertions must be fully characterised by means of sequence determination. The full sequence of the inserted DNA must be determined extending into the flanking sequences of the plant genome. It must be shown that these flanking sequences belong to

the genome of the host. This can be determined by means of a bioinformatics analysis, by comparing the obtained sequences with known plant sequences, or by means of a PCR analysis using DNA from the non-modified parental line or from a non-modified line with a genetic make-up as close as possible to that of the GM plant.[*,3]

2.2 Analysis to detect fusion open reading frames

Genetic modification can lead to the creation of new open reading frames at the insertion site, which are called 'fusion ORFs'. A fusion ORF consists of a sequence from the parental organism and a sequence from the inserted fragment. The fusion ORF may be located in a different reading frame than the reading frame of the introduced transgene. In theory, therefore, no more than 12 fusion ORFs can be created.

It should be noted that the chance of a fusion ORF leading to the synthesis of a functional protein is very small indeed. First, the proper regulatory signals must be present for transcription, and in addition a start codon, the poly(A) tail and the 5'CAP structure must be present for translation .

Nevertheless, the synthesis of new (fusion) proteins cannot be ruled out. The literature contains descriptions of various molecular mechanisms that can lead to the production of new genes. In many cases new genes arise from existing genes, for example through mutations, exon shuffling, gene duplication and recombination. New protein coding genes can also arise *de novo* from non-coding DNA sequences.[4] A limited number of genes that were formed in this way have been found amongst others in *Drosophila*, plants and humans.[5,6] For example, in *Drosophila melanogaster* five 'new' genes (younger than 5 million years old) have been found that originated from non-coding sequences.[7] It should be pointed out that such new genes are usually relatively simple and often code for short, poorly structured proteins.[6]

The coding sequence for a fusion protein will be a combination of DNA sequences from the plant and the insert. By searching sequence databases it is possible to find out whether such sequences are homologous to proteins that have specific biological properties, such as toxins and allergens.

As it is difficult to link genes to specific ecological effects, there is little point in performing database searches with unknown sequences to identify changes in ecological characteristics. This may be different for gene sequences with clear adverse effects, such as toxins. Except for the assessment of food safety (which in almost all cases is assessed by other agencies, such as the European Food Safety Authority and the RIKILT Institute of Food Safety in the Netherlands), the presence of a potential toxin also has implications for the environmental risk assessment, for

[*]The EFSA GMO Panel has, to date, required the use of non-GM lines with comparable genetic background as comparators. In the case of vegetatively propagated crops, these are the isogenic lines. In the case of sexually propagated crops these are non-GM lines as close as possible genetically to the GM plant under assessment.…In the case of a GM plant, its isogenic line is the non-GM line from which the GM plant is derived.'

example because of possible impacts on non-target organisms. If a homology is found, this is an indication that a potential toxin may be synthesised, in which case further investigation is necessary.

COGEM considers that the chance of finding a significant homology with a toxin in a database will be very small indeed. Until now COGEM has not observed a single instance of such a homology. Even if a sequence does resemble that of a known toxin, it does not automatically mean that a protein synthesised from the sequence will also have toxic properties. And even if no similarities are found, the sequence in question may still be part of a toxic protein that has not yet been entered into the database.

Based on the above, COGEM considers that the chance of a fusion ORF being transcribed and leading to the production of a protein is very small, but cannot be ruled out. Nor is it impossible that the protein thus formed could have toxic properties. COGEM is therefore of the opinion that the border regions should be analysed for the presence of ORFs that are situated between stop codons and which code for peptides larger than eight amino acids, which is the minimum size for an allergen epitope.[8]

2.3 Analysis for the presence of vector (backbone) DNA

As described earlier, *R. radiobacter* is used to insert desired DNA sequences into the plant genome. These DNA sequences are cloned between the left and right border of the T-DNA. However, plasmid DNA (all sequences outside the borders) may sometimes inadvertently be inserted into the plant genome along with the transgene. This plasmid DNA is also called 'backbone DNA'. The backbone DNA may contain unwanted vector sequences, such as an antibiotic-resistance gene. Particle bombardment can also lead to the introduction of unintended vector sequences as a result of insufficient purification of the fragment to be inserted before the particles are coated.

To be able to assess the risks of a GM crop it is important to know whether any of the bacterial backbone DNA has been transferred to the GM crop, and if so which parts, so that the characteristics coded for on the inserted backbone DNA can be included in the risk assessment. COGEM therefore argues that the molecular characterisation of a GM crop must also show whether any backbone DNA is present in the GM crop. One method for doing this is Southern blot analysis. If any backbone DNA is present, the sequence of this DNA must be identified until it extends into the flanking plant DNA sequences in order to determine with certainty which sequences and genes have been inserted and an analysis of fusion ORFs can be made.

2.4 Rearrangements are a natural phenomenon in the plant genome

The process of introducing DNA into the plant genome facilitates genomic rearrangements. Particle bombardment in particular causes numerous breaks in the genomic DNA, which leads to rearrangements. However, rearrangements are also a natural phenomenon in the plant genome.

In recent years it has become increasingly clear that the genome is not static, but is continually subject to change, rearrangements, deletions and insertions.[9,10] Each meiotic and mitotic division causes rearrangements (translocations, inversions, duplications, single nucleotide polymorphisms (SNPs), etc.) in the genome.[11,12,13] These genomic rearrangements are a natural and frequently occurring phenomenon and a driving force behind evolution.

It is almost impossible to predict whether a rearrangement of the genomic DNA will lead to a change in the plant's characteristics. This means that a bioinformatic analysis of any rearrangements will have limited predictive value. Field trials will give more information about the biological characteristics of a GM crop. Because any additional rearrangements resulting from the introduction of DNA into the plant genome will be small compared with naturally occurring genomic changes, and given the limited predictive value of a bioinformatic analysis, COGEM considers that the risk assessment of GM crops does not have to include an analysis of genomic rearrangements.

However, COGEM makes an exception to this for cases in which the introduced gene is inserted into a known functional region of the plant genome. The predictive value of the consequences of the disruption of an endogenous gene is greater and should therefore be taken into account in the environmental risk assessment. In the next section we explain the reasoning behind this.

2.4.1 Analysis of the disruption of endogenous genes

The insertion of T-DNA via *R. radiobacter* transformation is a seemingly random process. The insert can be introduced at any place in the plant genome. In most of the permit applications for marketing authorisation reviewed by COGEM over the years the T-DNA was found to be present in non-coding regions of the genome. In just one case indications were found that the insert had been introduced into a functional region of the plant genome.[14] In this case, based on the molecular data provided, COGEM was not able to rule out the possibility that the T-DNA had influenced the expression of an endogenous gene. However, the phenotypic analysis showed that the GM plant was a phenotypic equivalent of the parent plant.

The disruption of an endogenous gene can have a desirable effect (targeted inactivation) or an unwanted effect (unintended inactivation). Insertion into a regulatory sequence (for example a promoter) can also lead to activation or inactivation of genes. Endogenous genes can have diverse functions and the consequences of inactivation resulting from a T-DNA insertion can therefore also differ. Inactivation can have major effects that will become apparent during the developmental process, such as the inactivation of genes that play a part in development, growth, harvestable products, etc. However, it is possible that a disruption will have smaller effects that will not be directly discernible during the development process. For instance, disruption of metabolic pathways by the inactivation of enzymes with a function in these pathways could influence crop characteristics that are directly connected to the presence or absence of specific combinations of metabolites, such as food quality, taste, nutritional value, toxicity and allergenicity. Some

metabolites work at often low concentrations as signalling molecules, conveying signals to other organisms, such as fungi, herbivores and pollinators.

COGEM points out that the functions of most plant genes are not yet known. For this reason it cannot be ruled out that the unintentional inactivation of an endogenous gene will lead to unexpected changes in the behaviour of the plant. In view of this, COGEM is of the opinion that the potential consequences of insertions in an endogenous gene or in a regulatory sequence (such as a promoter) should be considered during the environmental risk assessment and investigated during the phenotypic analysis.

In COGEM's experience, most permit applications contain information about potential disruptions to endogenous genes.

2.5 Insertions of chloroplast DNA occur frequently in nature

The transformation of plants, especially by means of particle bombardment, often leads to cointegration of DNA from the chloroplast. Integration of chloroplast DNA into the nuclear genome is a process that occurs naturally on a large scale. During the course of evolution this process has led to the creation of numerous plant genes.[15,16,17,18] Research has shown that the integration of chloroplast DNA into nuclear DNA occurs at a high rate (once in every 16,000 to 49,000 pollen grains or seedlings formed).[19,20]Alongside the constant integration of chloroplast DNA is a continual process of removal of these sequences, leading to a balance between the 'entrance' and 'exit' of chloroplast sequences. [21]

In the light of the above, COGEM concludes that insertion of chloroplast DNA into the nuclear DNA is a natural and frequently occurring phenomenon. Any theoretical risks associated with the cointegration of chloroplast DNA during genetic modification therefore do not exceed the threshold of 'naturally occurring risks'. COGEM is therefore of the opinion that a sequence analysis and a bioinformatic analysis of the chloroplast DNA integration site is not necessary.

2.6 Epigenetic signal sequences

The insertion of a gene into the plant genome can lead to epigenetic effects. These effects could be caused by the insertion of a transgene that disrupts the normal epigenetic regulation and thus influences the expression of an endogenous gene. In addition, endogenous epigenetic mechanisms can influence the intended expression of the transgene, for example by suppressing this expression.

To detect transgene-associated differences in epigenetic mechanisms it is necessary to know the normal epigenetic structure of the DNA under different conditions (the baseline) and to define the changes that should be considered deviant. However, at the moment it is not clear where the boundaries of the 'normal' variation in cells, tissues or organisms lie. It is known that environmental factors and growing conditions can cause epigenetic variation, but here too it is not

clear to what degree this happens and what the natural variation is. Setting a 'normal' baseline is therefore not yet possible.

COGEM notes that potential effects due to epigenetic changes will be identified in the studies carried out for the current GMO risk analysis, including greenhouse trials, field trials and compositional analyses. On this basis, COGEM considers that at this time epigenetic effects should not have to be considered in the molecular characterisation.

3. Conclusions

COGEM has reviewed the criteria for the molecular characterisation of GM crops it adopted in 2008. COGEM sees no scientific reasons to amend these criteria, but has added an additional criterion requiring an analysis to determine whether the insertion has occurred in an endogenous gene. If this is the case, particular attention should be paid to this in the subsequent environmental risk assessment, for example in field trials.

References

1. COGEM (2008). Heroverweging criteria voor de moleculaire karakterisering bij markttoelatingen van gg-Gewassen. COGEM advies CGM/081219-01
2. Gelvin SB (2003). *Agrobacterium*- mediated plant transformation: the biology behind the "Gene-Jockeying" tool. Microbiol. Mol. Biol. Rev. 67: 16-37
3. European Food Safety Authority (EFSA). Guidance on selection of comparators for the risk assessment of genetically modified plants and derived food and feed. EFSA Journal 2011;9(5): 2149
4. Long M *et al,* (2003). The origin of new genes: glimpses from the young and old. Nat Rev Genet. 4(11): 865-875.
5. Ding Y *et al,* (2012). Origins of new genes and evolution of their novel functions. Annu. Rev. Ecol. Syst 43:345-63
6. Neme R & Tautz D (2013). Phylogenetic patterns of emergence of new genes support a model of frequent *de novo* evolution. BMC Genomics. 21; 14:117
7. Levine MT *et al,* (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A. 27: 9935-9939.
8. Thomas K *et al,* (2005). *In silico* methods for evaluating human allergenicity to novel proteins: International bioinformatics workshop meeting report, 23-24 February 2005. Toxicological sciences 88: 307-310
9. Nowacki M *et al,* (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature 451: 153-159
10. Eckardt N (2006). Genomic hopscotch: gene transfer from plastid to nucleus. Plant Cell 18: 2865-2867
11. Cai X & Xu S (2007). Meiosis-Driven Genome Variation in Plants. Current Genomics, 8: 151-161
12. Hamant O *et al,* (2006). Genetics of meiotic prophase in plants. Annual Review of Plant

Biology 57: 267-302
13. Petes T (2001). Meiotic recombination hot spots and cold spots. Nature Reviews Genetics 2: 360-369
14. COGEM (2013). Import of genetically modified soybean DAS-44406-6 with three herbicide tolerance Traits. COGEM advice CGM/130627-01
15. Leister D (2005). Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends in genetics 21:655-663
16. Timmis J *et al,* (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews 5: 123-135
17. Shahmuradov I *et al,* (2003). Abundance of plastid DNA insertions in nuclear genomes of rice and Arabidopsis. Plant Molecular Biology 52:923-934
18. Stegemann S *et al,* (2003). High frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. U S A. 100: 8828-8833
19. Huang CY *et al,* (2004) Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. Proc. Natl. Acad. Sci. U S A. 101: 9710-9715
20. Sheppard AE *et al,* (2008). Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiology 148: 328-336
21. Matsuo MY *et al,* (2005). The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. Plant Cell 17:665-675