Annex 6: Nucleotide Sequence Information System

BEETLE

Biological and Ecological Evaluation
towards Long-Term Effects

Long-term effects of genetically modified (GM) crops on health, biodiversity and the environment: prioritisation of potential risks and delimitation of uncertainties

Reference:ENV.B.3/ETU/2007/0007

Bundesamt für Verbraucherschutz und Lebensmittelsicherheit

BLAU UMWELTSTUDIEN

Genius SCIENCE & COMMUNICATION

Contributors (in alphabetical order)

Detlef Bartsch
Hans-Jörg Buhk
Achim Gathmann
Petra Heinze
Christiane Koziolek
Anja Vaasen

Table of contents

1. Introduction

GMOs to be released into the environment or to be placed on the market have to undergo a rigorous risk assessment followed up by strict monitoring programmes according to Directive 2001/18/EEC and Regulation (EC) No. 1829/2003. Detailed information on the introduced genetic modification on the basis of sequence data is required for the risk assessment of GM plants and development of detection and identification methods. In the future, with increasing knowledge of gene regulation and the development of bioinformatic tools, *in silico* sequence analysis will contribute further to risk assessment of GM plants. There are currently two databases on the EU level:

- The **Commission GMO Register** as defined in Article 31(2) and further described in Annex IV of Directive 2001/18/EC is currently being developed by the JRC[1]. The register must record and make publicly available, when not confidential, the information on GMOs authorized in the EU, including information on genetic modifications. The types of information that must be recorded in the register are described at a high level of detail in Commission Decision 2004/204/EC of 23 February 2004 laying down detailed arrangements for the operation of the registers for recording information on genetic modifications in GMOs provided for in Directive 2001/18/EC of the European Parliament and of the Council.

- The **Central Core DNA Sequences Information System** (CCSIS)[2] is the molecular database where the GMO sequence data submitted to the Community Reference Laboratory for GM Food and Feed (CRL) by applicants is stored to run homology searches in order to assess the specificity of the proposed GMO detection method as required by the Commission Regulation (EC) No 641/2004. The sequences and their biological metadata have been manually encoded and annotated following international sequence standard formats according the rules of the International Nucleotide Sequence Database Collaboration's (INSDC DDBJ/EMBL/GenBank) Feature Table.

However, for the time being it remains unclear if CCSIS will be made accessible to the Competent Authorities of EU Member States. One of the difficulties will be to comply with requirements for the protection of confidential data. In order to support identification of GM plants which are not approved for placing on the market in the EU it would be important to gather

[1] see: http://biotech.jrc.it/home/ict/dir-2001-18.htm
[2] see: http://biotech.jrc.it/home/bioinformatics/ccsis.htm

the relevant information on GM plants approved in countries outside the EU in such a database. It remains unclear to what extent such information will be incorporated even though such information should be available to the Competent Authorities of the Member States for the purpose of enforcing the European legislation on GMOs. If the detailed information is confidential, it could be distributed via the established channels for the exchange of confidential information within the EU.

Whilst the accessibility of CCSIS is not yet clear, the BVL (German Competent Authority) is developing its so-called 'Molecular Register' (MR) for 2009/2010. It will be realised as a web-based application integrating regulatory, phenotypic and molecular information on GMOs, and allowing the use of established bioinformatics tools to support Competent Authorities in their responsibilities according to Directive 2001/18/EC and Regulation (EC) No. 1829/2003. As a second step it is considered to advance existing bioinformatic tools with the aim of analysing or predicting genetic interaction at the level of gene regulation including epigenetics.

For the purpose of supporting risk assessment and future identification of potential long-term effects of GM plants it does not matter which system will be used, e. g. CCSIS or systems established by Competent Authorities like the MR, but it is essential to have the relevant data available, e. g. nucleotide sequence data including those on GM plants not approved in the EU but outside.

The essential features based on the example of the MR (no detailed information available on CCSIS) are described as follows for a better understanding of the requirements of the aforementioned nucleotide sequence information system including a database (CCSIS and MR).

2. Web-based Application of the Molecular Register

As sequence data are often classified as confidential business information, the *in silico* analysis will have to be performed in a secured IT environment providing not only an up-to-date local mirror of the commonly used public sequence databases, but also a local installation of currently established bioinformatics tools, and, finally, a confidential sequence database including all available sequence information on notified GMOs world-wide.

Additionally, with respect to the growing number of GMOs world-wide, providing a user-friendly data retrieval system is essential to keep track of regulatory affairs aspects and the classification of GMOs and their detailed characterisation.

2.1 World-wide Regulatory Affairs Tracking Tool

Data retrieval for a variety of regulatory purposes (especially with respect to the European refinements concerning various directives, regulations and decisions) is to be provided with respect to

- GMO unique identifier (and event synonyms),
- country of approval,
- filed documents (application/petition/notification)
- applicant or holder of the authorisation,
- legal basis,
- legal status,
- legal decision,
- intended use, and
- approval expiration date.

2.2 GMO Characterisation Tool

A taxonomic characterisation, a general classification of the event and its traits, and a detailed molecular characterisation (i.e. overview of introduced genetic elements, constructs, integration sites and their sequences) combined with information on detection methods for GMOs is to be provided with user-friendly data retrieval.

2.3 Bioinformatics Tools for the Development of Screening Methods

The *in silico* development of screening methods needs essentially to concentrate on the identification of PCR primer sequences among all available sequence information on notified GMOs world-wide.

2.4 Bioinformatics Tools Supporting Risk Assessment

Nucleic acid sequences are an essential requirement for the analysis of potential interactions (DNA/DNA, DNA/protein, protein/protein) of genetic modifications and their probability of occurrence. For example, the analysis of potential nucleic acid interactions on the basis of homology may be necessary, as the combined presence of transgenes might influence

expression: gene silencing that involves transgene/transgene interactions might occur between homologous DNA sequences (e.g. regulatory elements).

In silico sequence analysis can support risk assessment by providing additional information on the basis of

- homology analysis,
- protein motifs and domain prediction (pattern recognition),
- homology modelling (protein structure prediction)
- allergenicity prediction
- characterisation of open reading frames (ORF) in the insert,
- identification of potential chimaeric ORF in the junction region, etc.

3. Summary

Nucleotide sequence information systems including a database with relevant information on GM plants have been developed or are under development in the EU and Member States. However they are not available to Competent Authorities yet. Nucleotide sequence information systems have the potential to support the risk assessment of GMOs by combining established bioinformatic tools and relevant nucleotide sequence data on GM plants. They have the potential to allow the identification of potential long-term effects at an early stage (e. g. effects related to gene regulation including epigenetics) if existing bioinformatic tools are extended accordingly.