COMMISSIE
COGEM
GENETISCHE
MODIFICATIE

State Secretary for
Infrastructure and the Environment
Mr J.J. Atsma
P.O. Box 30945
2500 GX The Hague
The Netherlands

**DATE** 09 July 2012
**REFERENCE** CGM/120709-04
**SUBJECT** Accompanying letter research report 'The potential of applying bioinformatics approaches in the risk assessment of genes that lack a function annotation'

Dear Mr Atsma,

The Ministerial Regulation on Genetically Modified Organisms lays down an assessment methodology for carrying out risk assessments of laboratory activities involving genetically modified organisms (GMOs). This methodology helps both the applicant and the licensing authority to determine the appropriate containment level for activities involving GMOs. Assessing the potential risks associated with these types of activities requires knowledge about the function of the nucleotide sequence used for the genetic modification. This function may be derived from the function performed by the sequence in the source organism.

New developments, such as synthetic biology, are increasingly leading to activities involving GMOs that contain nucleotide sequences for which it is not possible to refer back to a source organism from which the sequence has been obtained when compiling the risk assessment. This means the sequence to be assessed cannot be related to an empirically demonstrated trait or function.

Bioinformatics analysis methods may be of assistance in these types of risk assessments. This has been investigated in a research project commissioned by the GMO office (Bureau GGO) and the Netherlands Commission on Genetic Modification (COGEM). The results of this study are presented in the enclosed research report, 'The potential of applying bioinformatics approaches in the risk assessment of genes that lack a function annotation' (CGM 2012-03). The research was carried out by Dr C. Francke of the Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre.

*Approach and findings of the research project*

In his report Dr Francke states that one bioinformatics analysis method has been sufficiently developed to allow the possible function of an as yet uncharacterised nucleotide sequence to be

identified. This method is based on a comparison of the sequence under investigation with sequences that have already been analysed and characterised. The research report contains a systematic review of algorithms and databases that can be used to make these sequence comparisons. It also describes a number of examples that provide practical indications on how a bioinformatics analysis can be designed and evaluated.

The databases that can be consulted are available via internet and grouped according to specific characteristics. They contain data on sequences that have already been characterised and to which specific functions and characters have been attributed, the 'function annotations'. In his report Dr Francke stresses that these databases are not always reliable. Some are not kept up to date and contain old, uncorrected data. In addition, the function annotations are not always backed up by references to the scientific literature, making it impossible to ascertain how these annotations were determined and whether or not they have been confirmed by empirical investigation. Neither all the function annotations have been peer reviewed. He emphasises that the annotation of a function is only definitive once it has been empirically demonstrated.

Dr Francke also notes that the way in which the algorithms present the final outcomes of the sequence comparisons is often complex and difficult to interpret. In his report he recommends that for a quick interpretation of the final data, computer programs are needed that can pull together the relevant data in a comprehensive and transparent manner.

In summary, the report provides a critical analysis of the workable algorithms and databases currently available that can be used to determine the probable function of an as yet uncharacterised sequence. It also gives practical guidance and contains two charts that clearly set out the steps for conducting a bioinformatics analysis of an as yet uncharacterised sequence.

*COGEM's reflections on the report*

COGEM commends the clear and readily comprehensible structure of the research report and considers it to be a welcome additional reference work for licence applicants and risk assessors. The Commission endorses the recommendations on carrying out a bioinformatics analysis elaborated in the report. The Commission also endorses the recommendation on improving the visualisation of the data generated by the algorithms to assist data interpretation. It also points out that the bioinformatics analysis and interpretation of the results should be performed by people with the relevant specialist knowledge.

COGEM emphasises that reliable databases are needed for a good analysis and shares Dr Francke's concern about the dubious quality of many of the databases available on the internet. Licence applicants and risk assessors need to be fully aware of this.

Bioinformatics is a complex and rapidly changing international discipline. Many algorithms and databases are being developed through international cooperative arrangements based in America, Japan and, in the case of the European Molecular Biology Laboratory (EMBL), in Europe. COGEM is of the opinion that international coordination and alignment of criteria and standards for bioinformatics analysis would be a welcome benefit to applicants and risk assessors. To support marketing authorisation (release into the environment) a European guidance document on the conduct of bioinformatics analyses could be prepared, for example by the European Food

Safety Authority (EFSA). This document could also be used to establish quality standards for the algorithms and databases to be used. To this end it would be helpful if existing available databases were combined. The accompanying report can provide a useful input to the development of such a European initiative. A copy of this letter and the research report will also be sent to EFSA for their information.

Yours sincerely,

Professor Bastiaan C.J. Zoeteman
Chair of COGEM

c.c. Dr I. van der Leij
Drs. H.P. de Wijs