European Commission

Guideline for the submission of DNA sequences derived from genetically modified organisms and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No 1829/2003

European Union Reference Laboratory for Genetically Modified Food and Feed

(EURL GMFF)

2016
with explanations of April 2017

Joint Research Centre

Contact information
European Union Reference Laboratory for GM Food & Feed
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 201, 21027 Ispra (VA), Italy
E-mail: jrc-eurl-gmff@ec.europa.eu

JRC Science Hub
https://ec.europa.eu/jrc

How to cite: European Union Reference Laboratory for Genetically Modified Food and Feed; Guideline for the submission of DNA sequences and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No 1829/2003

Guideline for the submission of DNA sequences derived from genetically modified organisms and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No 1829/2003

February 2016

(with explanations of April 2017)

Introduction

The aim of this guideline is to define minimum requirements and recommendations for the generation and provision of sequences of the insert(s) and flanking regions included in any submission by applicants in the context of their obligations under Directive 2001/18/EC and Regulation (EC) No 1829/2003.

This version of the guideline replaces the previous one (see http://gmo-crl.jrc.ec.europa.eu/doc/Guidelines_Applic_sequence.pdf) and should be fully respected for all submissions of sequence data. It will be updated whenever needed, e.g. for incorporating technical progress of sequencing technologies.

A reporting form is attached to this guideline.

Guideline

1. Definitions

- Material: organ/tissue from which the DNA is extracted; e.g. leaves, roots, seeds.
- Sample: purified DNA solution extracted from a defined material and used for sequencing experiments.
- Final sequence: the sequence of the insert(s) and flanking regions obtained by assembling the results of the sequencing experiment(s).

2. Material and Sample preparation

- The sample has to be extracted from a well-defined material. The applicant shall specify the genetically modified organism (GMO) and the material (organ/tissue) from which the DNA was extracted (e.g. leaves, roots, seeds) and if this sample came from an individual (e.g. a single plant) or if it was pooled from a number of individuals, in which case the number of individuals shall be indicated.
- For stacked GM events, the individual or the pooled individuals from which the sample is derived, must all carry the complete stack of events.

2.1 Sample reporting requirements

A full report on the sample has to be provided. It has to include, at least, the description of the GM plant (single line or stack, name according to nomenclature, species, traits), the type of material used (e.g. leaves, roots, seeds, etc.), the number of individuals

used/pooled and their origin (e.g. from the elite GM plant), and the protocol for DNA extraction that was used for the sample preparation.

If several samples are sequenced, the information described above shall be provided for each sample. It must include a description of the overall strategy to obtain the DNA fragment(s) used for sequencing (e.g. sub-cloning, long-run PCR).

An amount of material or sample sufficient for at least three further sequencing exercises should be stored until authorisation of the respective GM food or feed and has to be made available to the EURL GMFF upon request.

3. Sequencing

The final sequence can be generated with a strategy and sequencing technology at the discretion of the applicant, provided that these are clearly described in the accompanying information and that the following requirements are respected:

- For each event the final sequence has to include all complete and partial insert(s) as well as their genomic flanking regions. Regarding the size of flanking regions, we recommend to follow the EFSA GMO Panel guidance document or the Implementing Regulation EC 503/2013 as appropriate[1].
- If the sequencing involves the amplification of the template with a Polymerase Chain Reaction (PCR), the final sequence shall be generated from at least two independent PCR products covering every position of the sequence.
- For Sanger-based sequencing, the sequence shall be produced by bi-directional sequencing (i.e. each base should be sequenced on the forward and reverse strand) and in at least 2 independent PCR amplicons. The reported coverage or number of times that a nucleotide is represented in a collection of random raw sequence should thus be at least 4.
- For Next-Generation Sequencing (NGS) of all complete and partial insert(s) and their flanking regions, the optimal sequencing depth (i.e. the number of times each base of the produced sequence is sequenced) will depend on the technology used and the obtained value shall be reported and justified. It should normally not be less than 40.

In case of stacks, all sequencing experiments must be carried out with the same sample (see above), even if the different events are sequenced separately.

3.1 Sequencing reporting requirements

A full report on the sequencing strategy and the details of the experiment has to be provided for each submitted final sequence. It has to include, at least, the description of the technology used (Sanger, NGS platform), the sequencing method (e.g. targeted sequencing, amplicon sequencing, primer walking sequencing, whole genome sequencing), and the details of the experimental design.

---

[1] - EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2015. Guidance for renewal applications of genetically modified food and feed authorised under Regulation (EC) No 1829/2003. EFSA Journal 2015;13(6):4129, 8 pp. doi:10.2903/j.efsa.2015.4129

- Commission Implementing Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006. OJ L 157, 8.6.2013, p. 1–48.

The raw data shall be provided for the experiments performed to derive the sequences of the inserts and their flanking regions as described below.

4. Bioinformatics analyses

The Bioinformatics analyses can be performed with software and tools at the discretion of the applicant, provided that the software and tools (name and version), options, and parameters used are clearly described in the submission. Results of the analysis, however, shall be reported as indicated in section 4.1 below.

For authorisation applications submitted after July 2016, the applicant has to compare the sequence of each GM event contained in the application with all sequences of the event in question that were submitted after July 2016 to the European Commission, EFSA or the EURL GMFF, and has to report all differences found.

4.1 Bioinformatics reporting requirements

The submission shall be in an electronic format and has to contain at least:

- A clear description of the bioinformatics tools and software used (name, version, etc. - see attached reporting form) to generate the final sequence(s).
- The final sequence(s) in the correct format as described in point 5 below.
- For Sanger experiments:
    - The details of the base-calling procedure, including, at least, the software tools (name and version), options and parameters used, and any uncertainty observed during the base-calling.
    - The individual sequences in ABI or FASTQ format.
    - The sequences aligned/mapped to and used to generate the final sequence(s), in CLUSTAL or FASTA format.
- For NGS experiments:
    - The details of the base-calling procedure, including, at least, the software tools (name and version), options and parameters used, and any uncertainty observed during the base-calling.
    - Raw NGS-reads in FASTQ format, already filtered, and eventually trimmed off the used adapters. The software and parameters used for the filtering and trimming should be also described.
    - The sequences aligned/mapped to and used to generate the final sequence(s), in Sequence Alignment/Map (SAM) format, Binary Alignment/Map (BAM) format, or CRAM format.
    - If possible, an additional ACE file.

For taxon-specific reference gene sequences taken from a public database and used for the detection method assay, the relevant references to the source database and record have to be given.

5. Final sequence format

The final sequence(s) has/have to be provided as electronic ASCII text files using either the EMBL/GenBank format, or, preferably, the NCBI's Sequin (ASN.1) format.

*Please note that PDF and image-files will **not** be accepted.*

The final sequence(s) shall be annotated according to the INSDC Feature Table Definition Document[2] with at least the following descriptors and features, including their location on the sequence:

- "DEFINITION" (Title describing the sequence record),
- "SOURCE" , "ORGANISM" (according to the NCBI Taxonomy database),
- "SIZE" (in base pairs),
- "MOLECULE TYPE" (DNA),
- "TOPOLOGY" (linear / circular),
- "REFERENCE" (References with Authors, Title, Journals, etc.),
- "Source" (regions / sources of GMO insert and host organism),
- "STS" (Sequence Tagged Site corresponding to the PCR amplicon of the detection method),
- "Primer bind" (with primer name and sequence for Forward, Reverse Primer and Probe),
- All genetic elements ("gene", "promoter", "terminator", etc.)
- All coding sequences ("CDS"), including their translation.

If the sequence of a taxon-specific reference gene is included in the submission, the full sequence of the taxon-specific target and its GenBank accession number shall also be submitted, to the extent possible in the same format.

6. Reporting form

Please find at the following web address the reporting form (Annex 01) to be filled in, signed and provided by the applicant together with the submitted sequence(s).

http://gmo-crl.jrc.ec.europa.eu/doc/Annex01sec.pdf

In order to assist in the submission of sequencing information and data in accordance with this Guideline and for enhancing the efficiency of the compliance checks, the applicant is requested to implement a harmonised structure of such information and data as described in Annex 02.

http://gmo-crl.jrc.ec.europa.eu/doc/Annex02sec.pdf

[2] http://www.insdc.org/files/feature_table.html