# The use of bioinformatic analysis in support of HGT from plants to microorganisms

Meeting with applicants

Parma, 26 November 2015

efsa
European Food Safety Authority

www.efsa.europa.eu

# WHY WE NEED TO CONSIDER HGT IN GM PLANT RA

- Directive 2001/18/EC

  - As general obligation for the Member States and European Commission (EC), the Directive 2001/18/EC requests the assessment of gene transfer. This should be conducted in line with the principles for the ERA laid down in Annex II

- Implementing Regulation 503/2013

  - The implementing regulation requests to assess the probability of horizontal gene transfer (HGT) from the product to human, animals and micro-organisms

# WHY WE NEED TO CONSIDER HGT IN GM PLANT RA

## HGT is one of the areas of concern for the ERA

- HGT from GM plants to microorganisms
  - HGT refers to the transfer of a DNA sequence from a GM plant to a microorganism and its stable integration into the recipient genome

DONOR (GM plant) → RECIPIENT (Microorganism)

# HOW THE POTENTIAL FOR HGT IS ASSESSED

EFSA Guidance 2010 (ERA) implements the 6 steps approach for ERA

Plant to micro-organisms gene transfer

**Step 1: Problem formulation**

- problem formulation should focus on (among other)
  - "*detailed molecular characterisation of the DNA sequences inserted in the plant*"
  - "*presence of inserted plant DNA sequences showing similarities with DNA sequences from relevant microbial recipients*"

# HOW THE POTENTIAL FOR HGT IS ASSESSED

How can the similarity between the GM event and the microbial genomes be evaluated?

**Narrative description**

- qualitative
- useful
- no longer in line with available tools and with the current knowledge
- not reproducible
- error prone

**Bioinformatic analysis**

- quali-quantitative
- accurate
- reproducible
- easily updatable
- clear thresholds for the sufficient similarity to support homologous recombination (HR) can be posed on the basis of experimental data

**Bioinformatic analysis** is the most efficient way to evaluate the extent of similarity between GM event sequences and microbial genomes

# AIM AND CONTENT OF THE NOTE TO THE GD

## Aim and content of the explanatory note

- to expand on the scientific rationale for using the bioinformatic analyses to support the assessment of the potential for HGT
- to provide more detailed recommendations on how to perform such analyses
  - description of the query sequences
  - algorithms and parameters
  - sequence databases
  - length and sequence identity

# CONTENT OF THE NOTE TO THE GUIDANCE

- the key role of HR in HGT from plant to microorganisms
- efficient homologous recombination depends on nucleotide sequence identity (% identity + length)
  - a lower HGT rate is expected in case of sequence divergences
- a lower HGT rate is expected in case of longer non-homologous insert.
  - a $10^{-1}$ reduction was observed in case of increase from 1 to 2kb of the non-homologous insert.
  - no recombination was detected in case of a 6kb non-homologous insert, flanked by 1kb homologous sequences

# CONTENT OF THE NOTE TO THE GUIDANCE

- Bioinformatic analyses are considered necessary to:
  - estimate the possibility of HGT facilitated by double HR events
  - perform a proper problem formulation, including hazard identification
  - identify any hazard that may be associated with the GM sequences by informing on similarity with microbial sequences encoding known functions

# CONTENT OF THE NOTE TO THE GUIDANCE

- Recommendations to perform the bioinformatic analyses
  - **query**: full insert (flanking regions not to be included)
    - 'filler DNA' can be interspersed in the vicinity of the full/main insert (e.g. in case of particle bombardment or PEG-mediated transformation). These stretches should be part of the analysis
  - **algorithm**: local alignment (e.g. BLAST or FASTA)
  - **parameters**: default settings should be used, except for the low complexity filter that should be 'off'. Any deviation from these recommendations should be indicated and justified
  - **sequence databases**:
    - bacteria and Archaea
    - sequence patents and sequence vectors
  - **threshold values for reporting:**
    - All sequences with at least **95% identity** over a **length of 200 bp**