Bioinformatic analyses: methodology for allergen similarity search

Zoltán Divéki, Ana Gomes
EFSA GMO Unit

European Food Safety Authority

EFSA info session on applications - GMO
Parma, Italy
28 October 2014

BIOINFORMATIC ANALYSES

- **Analysis of the genomic flanking regions** of the insert with the aim to identify interruption of known genes
- **Analysis of ORFs** present within the insert and spanning the junctions with the aim to assess similarity to known allergenic or toxic proteins. For similarity searches to allergenic proteins, use EFSA GD (2011) and the Scientific opinion on assessment of allergenicity (2010)
- **Analysis of newly expressed proteins** with the aim to assess similarity to known allergenic or toxic proteins. For similarity searches to allergenic proteins, use EFSA GD (2011) and the Scientific opinion on assessment of allergenicity (2010)

- **Potential for horizontal gene transfer** to identify regions with similarity to microbial sequences

SEARCH FOR ALLERGEN SIMILARITY - GUIDELINES

International guidelines for newly expressed proteins

- FAO/WHO expert consultation report (2001)
- Codex Guidelines (2003/2009) ⇒ refers back to the FAO/WHO report

Current EFSA guidelines for newly expressed proteins and ORFs

- Guidance for risk assessment of food and feed from genetically modified plants (2011) ⇒ refers back to:
  - **Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed (2010)**
  - **Codex (2009) guidance**

FAO/WHO EXPERT CONSULTATION REPORT (2001)

Algorithm for allergenicity assessment

Step 1: *obtain the amino acids sequences of all allergens in the protein databases. […] Let this be data set (1).*

Step 2: *prepare* ***a complete set of 80-amino acid*** *length sequences derived from the expressed protein […]. Let this be data set (2).*

Step 3: *[…] compare each of the sequences of the data set (2) with all sequences of data set (1), using the FASTA program on the web site for alignment with the default settings for gap penalty and width.*

Cross-reactivity between the expressed protein and a known allergen (as can be found in the protein databases) has to be considered when there is: ***more than 35 % identity in the amino acid sequence*** *of the expressed protein […], using a window of 80 amino acids and a suitable gap penalty […]*

EFSA SCIENTIFIC OPINION ON ALLERGENICITY (2010)

Allergenicity potential of newly expressed proteins and ORFs

- *For ORFs and/or newly expressed proteins ≥ 80 amino acids:*

The alignment-based criterion involving 35% sequence identity to a known allergen using a sliding window of 80 amino acids as proposed by FAO/WHO is still accepted as adequate for allergenicity/IgE-cross reactivity assessment (p. 14)

Recommendation: the alignment-based criterion involving 35 % sequence identity to a known allergen over a window of at least 80 amino acids is considered a minimal requirement for risk assessment, although the identity threshold is conservatively set (p. 15)

BIOINFORMATIC ANALYSES NOT IN LINE WITH GUIDANCE DOCUMENTS

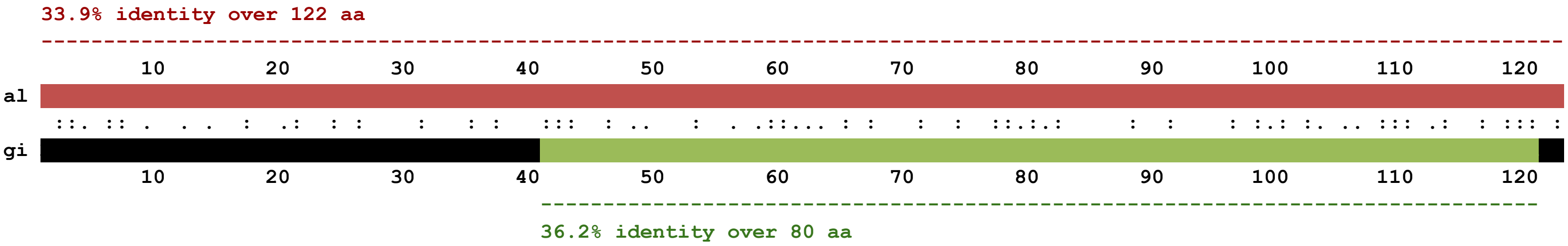
- A full-length FASTA search of concatenated ORFs is used instead of an 80-mer sliding window applied to individual ORFs
- An E-value cut-off is chosen without taking into account the size of the database used
- Following a search with 80-mer window, shorter hits are not recalculated to an 80-mer window

RECOMMENDATIONS

- An 80-mer sliding window should be used in all cases
- An E-value cut-off should be appropriate to the size of the allergen database (e.g. Allergen Online database default E-value is 100)
- Following a search with 80-mer window, shorter alignments should be re-calculated to an 80-mer window

EXAMPLE FOR DISCREPANCY

Alignment



Results

program	alignment type	query	database	positive hits	best identity [%]	appendix
fasta	local	full length	full length	0	33.9	a1
fasta	local	80aa window	full length	1	36.2	a2
fasta	local	full length	80aa window	1	36.2	a3
fasta	local	80aa window	80aa window	1	36.2	a4
allergen online	local	full length + 80aa window	full length	1	36.3	a5
needle	global	80aa window	80aa window	1	36.2	a6

EFSA GMO Unit

gmo@efsa.europa.eu

Thank you for your attention!