04 December 2018

Technical notes on the 'detectability problem' of GE

[redacted], [redacted], [redacted], [redacted], [redacted]

A recent ruling of the European Court of Justice (C528/16ECJ) has stated that organisms created by novel gene editing techniques (such CRISPR/ Cas9 and related methods, here referred to as gene editing, GE) are GMOs, and are thus covered by Directive 2001/18/EC. A full discussion of these methods and how they differ from 'conventional' genetic modification and conventional breeding and mutagenesis techniques is found in SAM (2018).

However, the Directive requires that organisms that have been produced by GE are detectable as such by testing laboratories. The EURL - GMFF has produced a draft Explanatory Report in this area; this draft is not for wider release, but within it, it discusses detectability issues and suggests potential pathways to detection of GE organisms.

In this technical report, we summarise why GE detection is not feasible in most scenarios, and why proposed pathways to GE detection will not resolve this detectability problem (again in most scenarios).

Detectability

As discussed elsewhere, the DNA sequence changes introduced through GE methods will be indistinguishable from DNA sequence changes produced by natural processes or conventional mutagenesis (SAM 2018). The exception may be where GE could be used to introduce more than two base pair changes in the DNA at any one location; these are less likely to be natural or mutagenic occurrences and can therefore be inferred to be gene edits with far greater confidence. In this technical document we will focus on GE methods that introduce single or double base pair changes or delete sections of DNA.

The current principle of GMO detection methods is to detect the transgenic or synthetically derived DNA that has been inserted by GM, but this approach cannot be used to detect GE.

Pathways to detection

A number of proposed pathways to GE detection have been suggested.

1) Where any GE organism has been through a documentation process (either an approval, or a scientific publication) which details the modified DNA sequence of the GE region and provides comprehensive details of the organism itself (e.g. the variety, in the case of cultivated plants), it would be possible to identify whether an unknown sample matched the varietal type and GE region of a known GE organism. This would provide strong evidence that the unknown sample was a GE organism. Additional evidence, such as the likelihood of the mutation occurring naturally (mutations in highly conserved gene regions occur very infrequently) and the route of import, could be used to complement the overall diagnosis. However, this is not direct evidence whether the sample was from a GE organism *per se*.

2) Where the suspect sample is from a cultivated plant (or animal), it has been suggested that the sample genome could be compared to a well-curated, comprehensive reference genome database of non-GE varieties of that plant (or animal), looking for DNA sequence differences

that are not found in any of the known varieties. The assumption would then be that these are GE modifications.

- This is a weak assumption because new mutations could plausibly have arisen (naturally or by mutagenesis) in any variety in each generation, or the putative GE mutation could be part of the natural variability within a variety, or natural variability from varieties not represented in the database;
- Where the DNA sequence difference matches to a known GE modification produced in that species, there is a certain likelihood that the DNA sequence difference is due to GE. Where it does not match to any known GE modification, there is no certainty at all that the DNA sequence is not due to natural variation or mutagenesis. Uncharacterised GE DNA sequence differences would not be detectable.

In addition to the weakness of the evidence it would produce, this proposal (database and genome comparisons) would be an unmatched scientific endeavour. As an illustration of this:

- A conservative estimate of the cost of sequencing and assembling a plant variety genome database would be £323 million[1] , excluding the substantial infrastructure costs required for a project of this scale. This database would have to be regularly updated to capture new variation.
- Assuming it is possible to create this database, the per sample cost would hypothetically be between £2 000 and 10 000 depending on the complexity of the sample genome, excluding the initial infrastructure costs to equip reference laboratories.
- No UK NRL currently has the infrastructure, technical capacity and computational resource to deliver this solution, and we would question whether any other EU NRL currently has this capability.
- For species such as bread wheat with large and highly repetitive genomes, obtaining a single near-complete genome is at the leading edge of what is currently possible (Zimin et al 2017). Sequencing 2478 wheat genomes in a small timeframe (5 years) would require facilities and capabilities the EU does not have, without halting or impacting existing sequencing projects.

  In summary, our assessment is that the proposed pathway would be almost unfathomably costly, may be technically unfeasible, and would provide relatively weak indirect evidence that GE had taken place.

[1] There are 14,442 varieties of Bread Wheat, Durham wheat, Maize, Soya bean, Barley (2 and 6 row), Swede rape, Turnip rape and Potatoes registered in the EU (EC plant variety database). From Wikipedia, there are 7500 varieties of apple and 10000 varieties of tomato. This does not account for any within variety variation. If we assume a genome cost of £10, 000 (sequencing and bioinformatic costs), that gives the figure of £323,420,000. These are speculative numbers: the number of varieties could be an underestimate, as for many species the variety number is for the EU alone and for a small list of species, but it could be an overestimate, as it may be possible to target some but not all varieties. The sequencing cost is higher than the oft quoted $1000 genome as: plants have very large genomes; sequence depth would have to be greater to reliably identify single nucleotide changes; complex plant genomes often require sequencing on multiple different sequencing platforms and the $1000 figure does not include staff costs for analysis. This number may be an underestimate.

Bibliography and references

Zimin, Aleksey V., Daniela Puiu, Richard Hall, Sarah Kingan, Bernardo J. Clavijo, and Steven L. Salzberg. "The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum." *Gigascience* (2017).

European Commission Plant variety database
http://ec.europa.eu/food/plant/plant_propagation_material/plant_variety_catalogues_databases/search/public/index.cfm?event=SearchForm&ctl_type=A

SAM 2018 Statement by the Group of Chief Scientific Advisors: A Scientific Perspective on the Regulatory Status of Products Derived from Gene Editing and the Implications for the GMO Directive.
https://ec.europa.eu/info/sites/info/files/2018_11_gcsa_statement_gene_editing_1.pdf