Quantification of GMO at a Level of 0.1 %
A Statistical Approach Using Frequency Distribution

Lars Gerdes, Ulrich Busch, and Sven Pecoraro
Bavarian Health and Food Safety Authority (LGL), Veterinärstr. 2, 85764 Oberschleißheim, Germany

Background and Motivation
- According to Regulation (EC) No 689/2008, trace amounts of non-approved genetically modified organisms (GMO) in feed are tolerated within the EU if certain prerequisites are met. Amongst others, an application for authorisation pending with the European Food Safety Authority (EFSA), a specific detection method validated by the EU Reference Laboratory for GM Food & Feed (EURL-GMFF) and the availability of certified reference material are required. Tolerable traces of such genetically modified material must also not exceed the so-called 'minimum required performance limit' (MRPL), which was defined to correspond to 0.1 % mass fraction. Currently, 13 GMO events fulfill these requirements (EU Register of authorised GMOs 2013).
- Because of this new regulation, trace amounts of not yet authorised GMO (and some GMOs whose approvals have expired) in feed have to be quantified following the qualitative detection of such in genetically modified material in extracted genomic DNA from those samples. As the results of quantitative analysis can imply severe legal and financial consequences for producers or distributors of feed, the quantification results need to be utterly reliable.

What actually is GMO percentage?
- A genetic modification is a sequence of foreign DNA integrated stably at a defined position in the host genome; such a unique insertion is known as event. Presence and absence of the integrated DNA can be interpreted as two forms (alleles) of this specific genetic locus (Fig. 1).
- In case of homozygosity all alleles for the considered locus are the same, in case of hemizygosity only half of the alleles are identical. Crossing of homozygous parent plants results in hemizygous hybrids (compare Fig. 1).
- GMO percentage can either be based on genome copies (cp/cp) or on mixed fractions (w/w). For homozygous plants (e.g. soy) there is no difference between cp/cp and w/w (Fig. 1 at top). A correction factor is needed for hemizygous plants (e.g. hybrid maize) because even 100 % GMO plants exhibit only 50 % of transgene alleles (Fig. 1 bottom).
- Plant seeds consist of several tissue types (simplified in Fig. 2). As a result of double fertilisation, the embryo is diploid (2n), the endosperm is triploid (3n) from maternal and paternal cells, and the integument is diploid (2n) exclusively from maternal cells. When analysing putative hybrid material, different GMO proportions in seeds and products thereof have to be considered, depending on either maternal or paternal origin of the transgene (Fig. 2). These circumstances can distort measurement results, especially because in processed food or feed the ancestry of the hybrid plants is usually unknown. In order to deal with this uncertainty, a medium-sized correction factor is assumed (e.g. for maize in Fig. 1: 50 % cp/cp = 100 % w/w [2]).
- For several independent transgenere (so-called stacked events) the genetic situation is way more complicated.

The developed novel statistical approach
- In the context of a scientific project sponsored by the Bavarian State Ministry of the Environment and Public Health (SMU), we developed a statistical approach in order to visualise the experimental measurement variability within one 96-well PCR plate as a graphical frequency distribution (Fig. 4).
- The impact on statistical parameters, for example the relative standard deviation within one laboratory (RSDR), can be investigated by increasing the number of replicates used for calculation (Fig. 5).

How to quantify GMO percentages
- GMO are quantified via real-time PCR. Therefore, both the transgene at the insertion site (allele g in Fig. 1) and the chosen reference gene for the given plant species are detected. The reference gene substitutes for both alleles (denominator g+w in Fig. 1) because the wild-type allele (w in Fig. 3) is not readily available for direct measurement.
- Copy numbers for transgene and reference gene in unknown samples are extrapolated from separate standard curves (Fig. 3). The ratio of copy numbers directly results in copy-based GMO percentage (cp/cp). For weight-based amounts (w/w), a correction factor has to be considered, where necessary.
- Software parameters (e.g. baseline correction, threshold settings) can have a notable influence on measured copy numbers and thus the final results. Selection and documentation of these parameters are therefore crucial to our approach.

References
3. WSSD. DNA-copy number on the x-axis of the G2M DNA-copy number in relation to hybrid maize seeds. DNA copy number calculated as linear function of hybrid gmo frequency, adjusted to real-time by single mouse click. The spreadsheet indicates classes and their respective values, all values, arithmetic mean and corresponding standard deviation (SD), and relative standard deviation (RSD) for all generated values. The green columns contain a certain percentage of all values (e.g. at least 95 %). green are open, and the class in the middle plus values deviation to max/min value. In case of a Gaussian normal distribution, the green data should allow a statistical prediction of the population (represented by the parameters in white).

Fig. 1 Zygosity and GMO percentage
Schematic illustration and calculation of GMO percentages (copy- and weight-based) resulting from mixing of either ground flours or isolated DNA derived from GMO plants. homozygous case (e.g. soy) at the top, hemizygous case (e.g. hybrid maize) below. The two possible alleles (GMO or wildtype) at the insertion site are depicted as red or green boxes, respectively.

Fig. 2 Constituents of the maize kernel and their origin
Schematic illustration of a maize kernel and its three tissue types. Ovule and placenta of the tissue types are listed in the table, together with the consequence of maternal or paternal GMO parentage (red). Only the case of a single transgene (considered two or more independent transgenere (so-called ’stacked events’) would complicate the situation further [3].

Fig. 3 Relative quantification of GMO
Standard curves for transgene (e.g. MON 87701) and reference gene (e.g. hmgA) are generated from dilution series of adequate plasmids and analysed in triplicates via real-time PCR with hypergenic probes (genomic (at top), normalised fluorescence vs. PCR cycles: top left: semi-logarithmics, top right: linear display). Once the curve crosses the green threshold, the corresponding cycle is noted as Ct. Arithmetic means of the Ct values are depicted vs. common logarithmic of copy numbers (bottom panel). Unknown numbers for transgene and reference gene can be extrapolated via linear regression.

Fig. 4 Course of action of the novel statistical approach
Schematic illustration of the sequence of procedures. transgene and reference gene are analysed via real-time PCR in the same 96-well plate but in separate wells. The light blue wells contain reactions for 32-32 replicates of the sample DNA to be measured. transgene and reference gene copy numbers are extrapolated from corresponding standard curves. For each number of possible combinations, the experiments are conducted by chance. The resulting 1000 GMO percentages are divided into frequency classes and depicted in a frequency diagram. For each number of replicates, a corresponding frequency distribution with its statistical parameters can be visualised by mouse click.

Fig. 5 Frequency distributions
Exemplary screenshots taken from the developed dynamic MS Excel spreadsheet. Issolated genomic DNA from certified reference material for maize event MON 810 (0.1 % w/w) was quantified with the novel statistical approach (Fig. 4). Both screenshots are based on the same experimental dataset from a single 96-well plate, on the left with 4 replicates, on the right with 12 replicates; additionally, different threshold and baseline correction parameters were chosen (refer to bottom in bottom right corner). via control elements (mostly semi-logarithmics vs. exponential, clone size, replicate number, data source, relative deviation) of the graphic display can be dynamically adjusted in real-time by simple mouse clicks. The spreadsheet indicates classes and their respective values, all values, arithmetic mean and corresponding standard deviation (SD), and relative standard deviation (RSD) for all generated values. The green columns contain a certain percentage of all values (e.g. at least 95 %). green are open, and the class in the middle plus values deviation to max/min value. In case of a Gaussian normal distribution, the green data should allow a statistical prediction of the population (represented by the parameters in white).