Quantitation using informative zeros (QUIZ): Project G03030

Final Report

CONTENTS

EXECUTIVE SUMMARY

Background

The Foods Standards Agency's objective to "protect public health risks which may arise in connection with the consumption of food, and otherwise to protect the interests of consumers in relation to food" requires the Agency to:

- Provide advice to the public and to Government on food safety, nutrition and diet;
- Protect consumers through effective enforcement and monitoring; and
- Support consumer choice through accurate and meaningful labelling.

This project seeks to apply a novel approach to GM quantification that is different from the standard method of real-time quantitative PCR (RT-PCR, also known as TaqMan™ PCR). The method, named QUIZ ("Quantitation Using Informative Zeros"), applies the standard PCR method to amplify target DNA sequences from very dilute DNA samples, and uses the presence or absence of these to determine their likely copy numbers using most probable number (MPN) statistics. Values, obtained independently for the GM sequence and a reference gene, allow an estimation to be made for the GM content. The results from such tests can be expressed as %GM per haploid genome (%GM DNA), as favoured by the European Commission (EC Recommendation 2004/787).

Current testing using RT-PCR for GM quantification requires certified reference materials (CRMs), a set of accurate calibrators from which results obtained for test samples can be interpolated to obtained GM estimations. The availability of CRMs is limited and QUIZ offers a feasible alternative to quantify the GM contents of any event in a sample as long as suitable markers are available.

This project seeks to show QUIZ as a feasible alternative to GM quantification by:

- Measuring GM contents from CRMs in two events (GTS 40-32-4, 'RoundUp Ready™' soya and MON 810 maize)
- Measuring the intra- and inter-laboratory variations by repeating the work in two separate laboratories using a common set of DNA samples
- Demonstrating the application of the technique by measuring the GM content of highly processed food.

In parallel to the laboratory work, computer simulations and modeling, allow us to quantify the bias and precision of QUIZ so that appropriate software can be developed that can process raw data to generate estimations using most probable number (MPN) statistics. The availability of this software allows experimenters, for example Public Analysts, to use QUIZ without the need to understand the statistical concepts behind the method and interpret results obtained. Furthermore, it negates the need to reinvent the wheel for those interested in applying the method.

Rationale and Objectives

The major advantage of QUIZ for GM quantification is that CRMs are not required. Another advantage is expected to be the ability to quantify from highly processed material such as food. Although GM crops have to undergo a rigorous safety assessment before they are allowed to be used for food and/or feed, adverse media reaction and ignorance have fuelled suspicion among some consumers. It is therefore important that people are allowed to choose whether they consume, or not, food containing GM. As such there are laws that require any food containing GM to be labeled. The threshold levels are 0.9% for an authorized GM to be labeled and zero tolerance for any unauthorized events. The enforcement of such laws requires the ability to detect and quantify GMs, not only in the raw unprocessed material but also in the final products.

DNA extracted from highly processed food products often contain high levels of contaminants that are co-extracted along with the DNA. These serve to inhibit the enzyme that performs the DNA replication, required for quantification. QUIZ can get round this problem because it requires extremely small amounts of DNA, often requiring many fold dilutions of the samples: the diluting of DNA also dilutes out the inhibitors allowing the enzyme to perform effectively.

Approach

Since the duration of the project was ten months, three major targets were set to:
(i) test QUIZ as a feasible alternative to RT-PCR for GMO quantification;
(ii) perform the groundwork in developing and testing the statistical methods so that it is possible to build from the work of this project; and
(iii) test that quantification can be performed on processed food.

The approaches were:
(i) to show that QUIZ can be universally applied to GM quantification, two GM events, found in different crop species, were selected for testing. CRMs were commercially available for both as well as sequence information for the design of primers to amplify event-specific markers.
(ii) a strategy for GM quantification was developed whereby, smaller numbers of reactions that are required to determine the dilutions for QUIZ are used to give estimates of GM content. Data, with confidence limits, may show that certain samples have GM contents that are likely to be above, or below, statutory thresholds without further testing. For those that fall into a 'grey' area, more accurate testing is performed.
Software has been developed that allows operators to input testing data and derive likely GM estimates, without the need to understand the underlying statistical theory.
(iii) following successful delivery of (i), it was also important to test QUIZ GM quantification in processed food. Here QUIZ was applied to the GM determination of bread and tacos, containing known amounts of the two GMs, in blind testing.

Outcome /Key Results Obtained

Demonstration of QUIZ was achieved through measurements of DNA samples extracted from CRMs for 2 GMs RoundUp Ready™ soya and MON810 maize, using event-specific markers and corresponding reference genes for soya and maize, respectively. The values obtained showed good correlation with the expected values but systematic bias was observed between the two assays for RRS and MON810. This bias was consistently greater in one laboratory than the other. The bias was attributed to different lengths of the markers used for quantification and the greater amplitude due to greater DNA degradation. Both problems can be resolved by better design of the primers used to amplify the markers.
We were able to test likely outcomes of QUIZ data by modeling and simulations. Software was developed which has been released, that uses MPN statistics to estimate the most likely GM content, whilst accounting for the natural bias of MPN when dealing with small numbers.

What it means and why it's important

The availability of QUIZ for GM quantification allows the quantification of GM events for which no reference samples exist, i.e. breaks the links between GM quantification and CRMs. The prerequisite to dilute DNA to the point of stochastic presence/absence makes QUIZ theoretically very sensitive, with the ability to quantify GM from small amounts of DNA, whilst the dilution process makes it more likely that the amplification of the target amplicons will work as inhibitors are diluted out. These factors make QUIZ highly suitable for the GM quantification of food.

Objectives and expected achievements

01	**Experimental strategy and statistical modelling:** the experimental procedure will be designed (e.g. number of replicates and samples) to ensure meaningful data that will allow appropriately accurate measurements for the purposes of evaluating and comparing the different GMO testing procedures between the partners. Statistical models will be tested to determine the optimum testing strategy using QUIZ. Computer modelling will be used to analyse testing strategies to optimise the procedure in terms of time and cost savings. For example, it is possible that results from 5 replicates at ten-fold dilutions to warrant no further testing if the results were, for example, 100 ± 50% GMO content, a figure way above the permitted threshold and with sufficient confidence to not warrant further testing.
02	**Uniform testing material:** – a set of experimental material (EM) will be produced and used, consisting of commercially available certified reference materials (CRMs) and internal reference materials (IRMs), constructed using admixtures of DNA from GM and non-GM plants. EM will be used by all partners to eliminate sampling errors caused by independent DNA extractions.
03	**Development of assays for the GMO-specific and reference amplicons:** Primers will be designed and tested for RR soya and MON810 (maize) for their suitability in QUIZ studies. Reference gene markers for soya and maize will be identified for use in QUIZ.
04	**Assess the robustness of single molecule amplification by PCR:** microtitre plates containing single protoplasts from a hemizygous GM maize plant (MON810) in each well will be produced. Methods will be assessed for their ability to amplify from this material. The efficiency of amplification from single molecules will be tested by amplification of the GM-specific amplicons.
05	**Data analyses:** evaluate and compare data from testing the EM by partners using QUIZ and RT-PCR. This will determine the accuracy of QUIZ compared to the industry standard, and the reproducibility of the method both within and between labs.
06	**Software for GMO estimation:** develop software that will estimate GMO-content and precision based on experimental data. Software will be developed that can calculate the % GMO of samples from the results of QUIZ analyses. Together with an estimate of error, the programme will allow informed decisions to be made regarding testing results from any sample.
07	**Testing of commercial products:** once the methodology is proven, the GMO content of purchased products containing soya and maize will be tested using QUIZ. A number of soya-containing products have been found to contain Roundup Ready™ soya (NIAB, unpublished). The testing of these samples will demonstrate the ability of QUIZ to quantify the GMO contents of foods.
08	**Project output:** Due the short duration of the project, a final report will be submitted to the Agency at the end of the project. Furthermore, a Standard Operating Protocol will be produced at the end of the work for the methods used in this study, written following the Agency's guidelines. Worthwhile findings will be submitted to peer-reviewed journals for publication.

Approaches and research plan

Task 1. Design experimental workplan using statistical modelling

1.1. The aims of the project require a rigorous data set from which QUIZ, as a method of GMO quantitation, can be fully analysed. The data must allow both intra- and inter-laboratory comparisons so that the reproducibility and portability are assessed. The accuracy of QUIZ will be measured by comparison with the industry standard method of GMO quantitation (RT-PCR) using both CRMs and IRMs. An experienced statistical geneticist will determine the numbers of samples and replicates required for each experiment. (NIAB 0.5, CSL 0.5 man months in Month 1)

1.2. Optimise testing strategy. Optimum testing strategy (or strategies) will be developed from computer modelling and simulations. These may reflect the expected numbers of DNA fragments that are extractable from different materials e.g. low numbers from very highly processed foods or oils. (NIAB 1 man month in Month 1)

Task 2. **Production of experimental material**

A set of experimental material (EM) will be generated by one lab and distributed to all partners so that the data generated can be used to analyse the performance of the method and not variation in the materials. DNA will be extracted from a set of CRM for RR soya and MON810. Furthermore microtitre plates containing a protoplast in each well (produced by flow cytometry of protoplast suspension) will be used for Task 3 to assess single molecule amplification for the GMO target gene in maize. (CSL 2 man month in Months 1-2)

Task 3. Development of assays for the GMO-specific and reference amplicons

Previously published data in the detection and quantification using real-time PCR of each event will be a starting point in this task. Primers for the GM events and reference genes will be tested to ensure that they are able to amplify from single molecules and that their copy number in the respective genome is determined and stable. In our previous study, we have found that the soya has two lectin genes which can be co-amplified using the same primers. (NIAB and CSL 0.5 man months each in month 1)

Task 4. Assessment of single molecule amplification (dependent on Tasks 2 and 3)

Microtitre plates containing single nucleus in each well will be used to assess the efficiency of PCR from single copies of template. The plant used will be hemizygous for MON810 event to ensure that a single copy of the transgene is present in each well. Amplification of the transgene will be tested to optimize detection conditions. Both partners will amplify from a 96-well plate to measure the efficiency of single molecule amplification. (NIAB and CSL 0.5 man months each in Month 3)

Task 5. **Validate methodology** (dependent on Tasks 2 & 3)

QUIZ will be validated in two events in two crop species (RoundUp Ready™ - soya and MON810 - maize) using PCR

5.1. QUIZ analyses of GM soya samples (NIAB and CSL 1 man month each in months 4-6)

5.1.1. QUIZ analyses of GM maize sample (NIAB and CSL 1 man month in months 4-6)

5.2. GMO content of ERM will be measured using Taqman™. (CSL 0.5 man months in months 3-6)

Task 6. **Software development** Programme for determining GMO content and precision will be developed that can use the testing results to calculate GMO content and its precision. The programme will be part of the overall platform of GMO testing by QUIZ, allowing consistency in the analyses of testing data. (NIAB 1 man month in Months 2-7)
Task 7. **Statistical analyses of data** (dependent on Tasks 5 & 6) Analyses of results from Tasks 4 and 5 to assess the reproducibility of QUIZ using the different platforms and between laboratories, and its accuracy when compared with the industry standard RT-PCR, with both CRMs and IRMs. (NIAB 1 man month in Months 5-8)
Task 8. **Testing of commercial products** (dependent on Tasks 5, 7 & 8) Soya- and maize-containing products will be purchased and DNA extracted from them by each partner. These DNA samples will be shared and the GMO content measured by RT-PCR and PCR-QUIZ. (NIAB 2 man months and CSL 2.5 man months each in Months 8-10)
Task 9. **Project output.** **9.1**. SOP: Protocols used will be described following the Agency's guidelines on 'writing SOPs'. **9.2**. Final report: submitted at the end of the project. (NIAB 1 man month and CSL 0.5 man months in Months 7-10)

Project milestones

	0 months	Start project
01/01	Month 1	Experimental Workplan
01/02	Month 1	Production of Experimental Material
01/03	Month 2	Assays for the GMO-specific and the reference amplicons for RR soya and MON810 maize.
01/04	Month 3	Assessment of single molecule amplification
01/05	Month 4	**Meeting with FSA**
01/06	Month 4	Models developed for GMO testing using QUIZ
01/07	Month 6	Quantification data for experimental material using QUIZ and RT-PCR
01/08	Month 8	GMO-content calculating software using QUIZ data
01/0	Month 8	**Meeting with FSA**
01/10	Month 8	Testing of commercial products
01/11	Month 10	Reports to the Agency: Final Report and Standard Operating Procedures

Project deliverables

1	Month 2	Experimental protocol decided
2	Month 2	Dissemination of experimental material to all partners
3	Month 5	Document detailing strategies on GMO testing
4	Month 7	Data analysing software: working programme for calculating GMO content
5	Month 8	Validation of QUIZ: report with analyses of data generated by all partners
6	Month 10	SOP and Final report.

Quantitation using informative zeros (QUIZ): Project G03030
Final Report

David Lee[1], Maurizio La Mura[1] Theo Allnutt[2] Andy Greenland[1], Ian Mackay[1]
[1]John Bingham Laboratory, National Institute of Agricultural Botany, Huntingdon Road, Cambridge CB3 0LE
[2] Central Science Laboratory, Sand Hutton, York, YO41 1LZ, York,

1. Introduction

G03030 is a follow-on project from G03025, in which we demonstrated the application of QUIZ to determine the GM content of admixtures of samples containing RoundUp Ready™ soya.
G03030 addresses the key issues surrounding the development of a new method, quantitation using informative zeros (QUIZ), for robust and sensitive quantification of relative amounts of GM to non-GM DNA in a food sample, as specified in research requirement G03R0011.
Firstly, it takes account of the likelihood (recently reinforced by EU Commission advice) that policing/enforcement of GM thresholds will probably be based on a quantitative DNA assay rather than a protein or metabolic ratio. The very nature of the method proposed, which centres around determining numbers of molecules of interest (GM-specific sequences and endogenous reference genes), means it is set up to explicitly deliver the percentage GM-DNA copy number in relation to control taxon-specific DNA copy numbers. Since it actively seeks to exploit limiting dilutions, it is highly amenable to detection at extremely low limits, down to single molecules (<< 0.1% GM DNA). This leads to the possibility of quantifying GM content in highly processed foods and oils, for which no test exists currently.
Although QUIZ has been shown to work in a proof-of-concept study (G03025 report; Lee *et al*. 2008), many features of the method require further evaluation before it can be adopted as a routine testing procedure. This project aims to evaluate QUIZ in two GM events (RoundUp Ready™ soya and MON810 maize) for its accuracy, repeatability and transferability between different laboratories using a PCR-based method of amplifying the target sequences.

2. Materials and Methods

2.1. DNA samples

The material used for this study comprised of the sets of CRMs for MON810 and RRS. Each CRM contains 6 samples with very accurately determined GM contents based on % w/w of GM in a near isogenic non-GM background. DNAs were extracted from each sample using commercial kits and details are provided in Appendices 3-5. DNAs are produced by one partner (CSL, listed in **Table 1**) and supplied to NIAB to eliminate DNA extraction variables from the comparisons.
Three processed food samples were also tested for their GM content, these were: two tacos samples containing defined MON810 and RRS contents and DNA extracted from bread, containing MON810 (**Table 2**).

The procedures for determining the DNA dilutions for QUIZ empirically, have been described in the report entitled: 'GMO Quantification Using Limiting Template Detection.doc' (Appendix 1). In brief, dilutions at which QUIZ is performed were determined empirically from unknown samples. Serial ten-fold dilutions of the DNA were made, from which multiple samples were taken. Results from this pre-QUIZ test may be sufficiently accurate to determine whether a sample has exceeded, or not, a given threshold. Samples, where the results are not statistically robust, can be tested with more replicates to obtain more accurate GM estimates.

2.2. Real-time PCR for GM quantification

Real-time PCR quantification of the RRS and MON810 were performed using the primers and probes listed in **Table 3**. Calibration reactions were performed in duplicates, whilst four replicates of the samples for each of four dilutions were performed (1:5, 1:10, 1:20 and 1:40). SOPs can be found in Appendices 4 and 5.

2.3. Measuring Efficiency of PCR and Real-time PCR Reactions

In order for QUIZ to give accurate GM quantifications, the efficiency of the GM and reference assays must be equal. Ideally this means that their probability of detecting a single target molecule, when one is present in the reaction, is 100%. In reality no reaction will be 100% efficient and it will vary slightly between assays due to DNA sequence melting temperatures, secondary structure, PCR competition/inhibition and target length. It is therefore important to measure the efficiency of GM and reference assays prior to using QUIZ. This does not have to be done every time it is used but should be done for new primers, targets and sample matrices. Some variation in efficiencies is, of course, expected and it will be one source of QUIZ stochastic error.

2.4. Single Nuclei Preparations

Flow cytometry was used to sort single nuclei, free from any other particulate contaminants, into 96-well PCR plates. Each well therefore contained a single GM (Mon810) target molecule and two reference molecules (*adh1*). This method avoids uncertainty in sampling molecules from a solution and provides a direct relative estimate of PCR efficiency.

Method

Maize nuclei from fresh F1 Mon810 leaf were isolated using a simple procedure:

1. Place 5 g leaf in a petri dish.
2. Add 20 ml ice-cold Galbraith's solution (45 mM $MgCl_2$; 20 mM MOPS; 30 mM sodium citrate; 0.1% Triton X-100; pH 7.00).
3. Slice leaf with fresh scalpel or razor blade into ~0.5 mm strips (90° to leaf veins).
4. Transfer leaf and Galbraith's into 50ml Falcon tube, rotate at ~12 rpm for 20 minutes at room temperature.
5. Decant liquid into new tube in ice bath. Keep all subsequent steps at 4°C. Filter through three layers of 20 μm nylon mesh.

Nuclei were stained with propidium iodide (1mg /ml) for 10 minutes then separated by a Beckman MoFlo flow cytometer and cell sorter into each well

of 96 well PCR plates. Nuclei sorting and counting onto glass slides was confirmed by fluorescence microscopy. Plates were centrifuged for 2 minutes at 5000 × g and stored at 4°C. PCR or real-time PCR reagents were added directly to plate wells and reactions carried out normally.

2.5. Primer design and method validation

The design of the primer sets for QUIZ was based on the primers used in the CSL Taqman™ assays with some modifications to alter the annealing parameters and amplicon sizes. For example, the primers for MON810 assays differ by the removal one 1 nucleotide at the 5' end from two of the primers. However, both laboratories had problems with the Soya primer set (QUIZ RRS primer set 1 – **Table 3**) where estimated copy number of template, based on DNA amounts and dilution factors, did not give expected results. This was compounded by the inability to dilute to zero amplifications in the Pre-QUIZ testing to determine the correct dilutions for the GM target (**Table 4**).

To resolve whether the problem was with the DNAs or the primer set, we took sample B from the project G03025 (0.45% admixture) and tested it with the new primers. At the same time we used the 0.5% CRM sample (no. 1456) and performed QUIZ using the old primers (QUIZ RRS primer set 2). Reciprocal testing of the primers and DNA samples (**Table 5**) indicated that the problem resided with the primer set rather than the DNA: sensible results were obtained only when we changed primers but not the DNA samples. Consequently we continued the QUIZ process with the original primer set used in the proof of concept study (G03025).

2.6. Statistical analyses of QUIZ

The accuracy of QUIZ has been investigated through modelling and simulations (Appendix 1). In summary, modelling and computer simulations demonstrated the performance of QUIZ and the accuracy of GM values obtained was quantified: the precision is determined by sample size and the positive/negative ratio of each marker. Furthermore, better estimates can be obtained from the pre-QUIZ testing if smaller dilutions are used rather than the ten-fold dilution steps currently used. The larger numbers of dilutions needed to attain QUIZ data, that is to get to the point of stochastic presence/absence of template, may be offset by the smaller number of repetitions required at each dilution (currently 10). Such an approach is consistent with that described for SIMQUANT (Berdal *et al*., 2008).

2.7. QUIZ estimating software

Software code written in the R software package is available for the analyses of QUIZ data to estimate GM content of samples using MPN statistics. Description of the software and how to load and use it are provided in Appendix 1. The software is easy to use and the numbers of simulations performed to get estimates can be controlled by the operator. Estimates can be obtained for either pre-QUIZ testing data (amalgamation of data from different dilutions) or for QUIZ where all testing for each target is performed at a single dilution.

Since the computed estimations are calculated by simulations, the estimates for pre-QUIZ data have been generated using 1000 re-iterations, as a compromise between time required to calculate the value and precision. It is therefore possible that slightly different values can be obtained by separate calculations even with the same data. QUIZ estimations were performed using 100000 simulations. The large numbers ensure very little variation between simulations.

3. Results and Discussion

3.1. Assessment of single molecule amplification

If the sorted nuclei droplet contained DNA from lysed nuclei, then the experimental approach would have been void. Control PCRs (96) were therefore performed on supernatant from the nuclei preparations. All of these controls were negative, demonstrating that only whole nuclei from the Moflo provided positive amplifications.

For each assay, a separate 96 well single nuclei plate was used. GM and reference assays for standard PCR, simplex and duplex Taqman PCRs were tested (**Table 6**). In Taqman reactions, the Mon810 assay was less efficient than the *adh1*1 assay. It should be noted that it is difficult to judge the degree of this difference because the reference and target reactions contain different numbers of target molecules: a single hemizygous Mon810 and two *adh1*1 targets. However, we would expect Mon810 to at least amplify 50% (approximately) of the reactions amplified with *adh1*1 but only 31% was observed (27 out of 87 positives). Given this difference in efficiencies we would expect QUIZ to underestimate Mon810 by 38% where Taqman was used qualitatively to detect amplification. When Taqman reactions were run together on a single plate (duplex), the ratio and the numbers amplified were very similar (24 out of 78 = 30.7%), indicating that at least for the duplex system used here, there was no competitive inhibition.

For standard PCR, the absolute efficiency was much lower than Taqman, *adh11* = 12.5%, Mon810 = 15.6% but the ratio between Mon810 to *adh1*1 was much higher (80%) where again approximately 50% was expected, we would therefore expect QUIZ using PCR for detection to overestimate Mon810 by 60%. The observed overestimation for PCRs performed at CSL (same conditions as single nuclei plates) for CRMs was +29.4%, which is less than expected. Note that at CSL products were detected on agarose gels and both *adh1*1 alleles were scored (see section 3.9 ‘Effect of Genetic Variation on QUIZ Estimation’).

The results of the single nuclei amplifications for each of the two amplicons do match with other observations. The efficiency of detection for the GM amplicon is less than that for the reference gene, even when we consider that the former is present in one copy per nucleus (hemizygous) and the latter in two copies. This can be attributable to amplification efficiencies of the two targets. That these differ between RT-PCR and standard PCR also suggest differences between the two amplification techniques. It should be noted that

amplification from single nuclei does not reproduce the exact conditions in QUIZ reactions. The DNA is not purified and amplification may be inhibited by bound proteins. However, we would expect such effects to be similar for both GM and reference assays and therefore the relative efficiency estimate should still be applicable to QUIZ, or at least indicate where there may be problems. Nuclei sorting using Moflo is not practical for routine QUIZ development and validation, the equipment being scarce and very expensive, but it is a very useful tool for initial investigation of the dynamics of the QUIZ assays.

3.2. Quantification using pre-QUIZ testing data

As mentioned above, the process of determining the dilutions for performing QUIZ can generate GM estimations. We have applied multiple sampling within each possible dilution to obtain a better guidance on the dilution factors to use and to see if estimations are sufficiently robust.

Pre-QUIZ data for RRS and MON810 are shown in **Tables 7 & 8**. Data for RRS demonstrate that for 0, 0.1 and 0.5% CRMs were all below the 0.9% labelling threshold (>95% confidence). Similarly, the estimate for the 5% CRM suggests that it is above the 0.9% threshold (>95% confidence). Results obtained for the 1 and 2% CRMs did not provide sufficient clarity and would require more accurate testing.

Pre-QUIZ data for MON810 was not as clear. The amplification frequency from the dilutions of the DNA samples did not fit the expected pattern and there were cases where the numbers would actually increase following dilution. **Table 8F** illustrates the effect of one potential false negative on the GM estimation. By changing one negative to positive at the highest dilution of the reference gene, the GM estimation for the 5% GM CRM changes from 87.2% (31.5-239.8 - 95% range) to 4.6% (1.6-13.9 – 95% range).

3.3. QUIZ GM quantification of CRMs

The dilution factors determined by pre-QUIZ testing and used for QUIZ testing are summarised in **Table 9**: the numbers obtained in the QUIZ tests are given in **Tables 10 & 11**. TaqMan™ results for the two sets of samples are given in **Table 12**.

The performance of QUIZ method has been measured by comparison of obtained results with the industry standard method of GM quantification real time-PCR (Taqman). Both partners (NIAB and CSL) used QUIZ to estimate the GM content in the samples while Taqman quantification was performed by CSL using their assays for the two events.

3.4. QUIZ Estimation of GM content of RoundUp Ready™ Soya CRMs

Analyses of the RRS results (**Figure 1**) suggest that there is an inherent bias between the experimental data and the expected values of the CRMs. The best-fitting straight lines for each experiment show good correlation of fit (R^2), very close to 1, which demonstrate the consistency of the method for GM estimation. However, despite the excellent correlation, there is a general trend to underestimate the GM content, and the underestimation is more pronounced with data generated at NIAB.

3.5. QUIZ Estimation of GM content of MON810 CRMs

A similar but contrasting story is observed for the MON810 data (**Figure 2**). Again there is a good correlation of fit with the best-fitting straight line from each set of data with the expected. Again there is a general consistent bias of the result from the expected but this time the trend is to overestimate the GM contents of the samples. Again NIAB's results show greater deviation from the expected.

3.6. Estimation of GM content of food

The principles and practices of QUIZ quantification were applied to food testing. There is little merit at this stage to test food with unknown GM content since we have no means of evaluating test data. Food samples were produced with known amounts of GM flour and DNA extracted at CSL. Although the GM content of the three samples was known, this information was not made available to NIAB so in effect NIAB performed the QUIZ estimation of all samples as blind tests.

As such, the results (**Table 13**) show that the QUIZ estimations for MON810 were generally very good but continued the trend of overestimation. Only in one experiment did we obtain a lower than expected value for MON810, and the overestimation appears more pronounced at the higher GM values.

For RRS the measured data appear to greatly underestimate the expected values, with only one reasonable estimate for the 1498 (taco) sample. However, even though the values deviated greatly from the expected, they were on the whole consistent within samples.

3.7. Sources of bias

There are a number of factors that we have identified and believe to be responsible for biasing the results:

- Relative amplicon lengths of the reference and GM targets. Shorter molecules are more abundant. The prediction is that there will be a bias in GM towards the shorter amplicon, i.e. if the GM fragment is shorter than the reference, then the assay will overestimate the GM content and this bias will be exaggerated by DNA degradation.
- The effect of genetic variation of reference amplicon on molecule count. Polymorphisms that eliminate the amplification of some reference molecules will result in overestimates of GM content using QUIZ. Likewise, the scoring of multigene families as reference targets will result in underestimation of the GM content.

3.8. Amplicon length of the respective reference and GM targets

In the two assays used, though the observation that when the reference gene is smaller than the GM target, then there is an under estimation of the GM content and vice versa, is circumstantial, there is a compelling reason why this may be so. Since PCR is about amplifying DNA targets, and that the process of DNA purification/extraction shears DNA, the target size must impact on the available amount of target molecules. For example, if the average DNA molecule is 1 kb in length, then there would more fragments of

200 bp than 1.8kb, even of the same gene sequence. Such imbalance in degradation extent will impact on any method that assays different lengths of targets for the GM and reference markers.
The later RRS assay reference and GM targets have lengths of 156/287 bp respectively giving a potentially large skew in degradation frequency. The bias is more notable in NIAB data and DNA degradation is possibly responsible (see later discussion). In fact it is the bias in the NIAB data that has forced us to look for an explanation for the general bias.
To test the effects of relative amplicon lengths we designed an assay for QUIZ testing of RRS CRMs. In this assay we replaced one of the GM primers so that the assay amplifies a smaller fragment of 92 bp (QUIZ RRS set 2b – **Table 3**). This had the effect of making the GM target shorter than the reference target. In the second assay we made the length of both reference and GM amplicons 78 bp (QUIZ RRS set 3 – **Table 3**). Previous testing of the CRMs meant there were not enough of the samples 0 and 0.1% GM for further testing: the low GM contents of these sample required more of the samples to get QUIZ estimates. Only the samples with the highest GM contents were tested with the new primers. The results are shown in **Figure 3**. The results show that the new assay reversed the underestimation, and the results now show a slight overestimation of the GM contents, with the exception of the 5% CRM sample. This sample has consistently been measured to be <5% using QUIZ and RT-PCR methods.

When used for food testing, results obtained using the new primers for GM soya were more accurate (**Table 9**). Taken together with the general reversal of the underestimation for the CRMs, these results suggest strongly that relative amplicon length was a significant factor in GM quantification accuracy.

The combined effects of relative amplicon lengths and DNA degradation could explain why NIAB consistently obtained exaggerated results in the QUIZ. The fact that NIAB consistently used lower dilutions of the templates than CSL for QUIZ would support differences in template concentration through the loss of DNA in transit. Such practice would certainly have a greater impact when testing food DNA, as observed.

As a final test, we designed a new assay where the lengths of the amplicons of the GM and reference gene targets were identical, 78 bp in length (QUIZ RRS primer set 3 – **Table 3**). This primer set gave the closest results to the expected values of the tested samples of the CRM (**Figure 3**): 2.07 for the 2%, 1.07 for the 1% and 0.63 for the 0.5%, but the estimate for the 5% sample, which have been consistently lower than expected, was 2.96%.

3.9. Effect of Genetic Variation on QUIZ Estimation

We have noted that there are two fragments produced in the amplification of the *adh1*1 reference marker. Besides the expected 133 nucleotide fragment, a second smaller amplicon of 128 nucleotides is often co-amplified. In the analyses of the data, NIAB uses a DNA sequencer to detect and identify amplicon fragments and, as such, will only count specific-sized fragments. CSL on the other hand used an agarose gel system for detection of DNA

fragments and will not have discriminated the different alleles and counted either or both.

It is important to determine whether the two fragments are allelic or members of a gene family, as observed for lectin (in project G03025), since they will have different impact on the GM quantification estimates. If they are allelic, NIAB's method will effectively have doubled the GM estimates since only half of the reference molecules have been counted. If they are members of a gene family, CSL results will be underestimated due to the excess of reference markers.

To discriminate between allelic variation and gene family, DNA was extracted from MON810 (hybrid) seeds. QUIZ analysis was performed on the sample. The results from 96 replicates gave GM values of 114 ± 34% suggesting that the two *adh1* gene fragments are allelic since the hybrid seed contains 58% GM DNA, and only by the omission of one of the reference gene allele would the GM content be doubled.

3.10. Re-evaluation of data

Reassessment of the previous data by halving the GM values obtained by NIAB gives a more accurate and truer picture of the QUIZ method for assessing MON810 GM content, certainly for the CRMs (Figure 4). The recalculated results for the food samples (**Table 9**) are less clear. Results for the taco samples are improved but those of the bread sample are not. Since the heterozygosity of the *adh1* gene is responsible for the overestimation of GM content, the genotype of the non-GM source of maize DNA is very important. For example, the assay for the *adh1* genes distinguishes between the two parental alleles. NIAB's initial interpretation of the CRM QUIZ data, where only the maternal allele was scored, gives an overestimation of the GM content. This finding highlights one potential problem with the reference genes for GM quantification: that heterogeneity between cultivars may affect the quantitative aspects of any assay.

One strange consequence of GM mixing flour with flour from from the parent with the non-scored allele (in this case paternal flour) to dilute the GM content is that QUIZ analysis using the NIAB analysis method will generate data of 100% GM, regardless of the amount of dilution used as paternal flour will not contribute *adh1* alleles.

3.11. Conclusions

The experiments performed in this project have demonstrated the application of QUIZ for GM quantification both to CRMs and to food products. However, there appears to be two key factors that affect the accuracy of GM quantification using QUIZ. These are:

- the relative lengths of the reference and GM amplicons used, and that this bias is greatly enhanced by DNA degradation; and
- the effect that polymorphisms within reference genes can have in GM estimations.

The effect of amplicon length on GM estimation has been reported (Moreano *et al*., 2005), and there is a general consensus that for the testing of food products, where the DNA quality is suspect, then short amplicons (<200 bp) are best. The resolution of inaccuracies and inconsistencies of QUIZ results between labs and samples has required a detailed analyses of the markers used. However a number of uncertainties remain about choice of markers and the length of the amplicons. Issues of relative amplicon lengths and choice of markers for QUIZ GM estimations do not only impact on the accuracy of QUIZ *per se* but have a generally wider effect in GM quantification. If relative amplicon lengths of the two markers can have an effect on molecule count in QUIZ, then this may also affect RT-PCR.

This project has fulfilled its primary aims of demonstrating accurate QUIZ estimations of the GM content for CRMs and food for two events. Software has been developed to allow greater access to the statistical evaluation of practical data, though there are certain aspects of the software that could be improved, such as the ability to highlight 'data deficiencies'. Even though there is a move to produce commercial reference materials that provide calibration using '%DNA per haploid genome units', QUIZ already does this so long as the markers used have been carefully assessed for their copy number and are of equivalent lengths to remove bias in the testing. With a small amount of refinement, QUIZ promises to be a robust and accurate method of GM quantification without requirement for certified reference material.

4. References

- Berdal KG, Bøydler C, Tengs T and Holst-Jensen A (2008). A statistical approach for evaluation of PCR results to improve the practical limit of quantification (LOQ) of GMO analyses (SIMQUANT). Eur Food Res Technol, doi: 10.1007/s00217-008-0830-1.
- Lee D, La Mura M, Greenland A and Mackay I (2008). Quantitation using informative zeros (QUIZ): application for GMO detection and quantification without recourse to certified reference material. Food Chem, doi: 10.1016/j.foodchem.2008.05.118.
- Moreano F, Busch U and Karl-Heinz Engel K-H (2005). Distortion of genetically modified organism quantification in processed foods: influence of particle size compositions and heat-induced DNA degradation. J Ag Food Chem **53**: 9971-9979.

Table 1. List of CRM samples, with relative DNA concentration and %GMO content, used to test QUIZ.

Event	Code	[DNA] (ng/µL)	Nominal %GMO
RRS	1453	50	5
	1454	34	2
	1455	57	1
	1456	129	0.5
	1457	135	0.1
	1458	91	<0.03
MON810	1459	54	2.9
	1460	32	1.16
	1461	29	0.58
	1462	30	0.29
	1463	29	0.058
	1464	54	<0.01

Table 2. List of processed samples used in this study. The %GMO (mass/mass) content and the proportion of that GMO in relation to the overall dry mass of the products are shown in A; and the constituents of processed samples in B.

A.

Processed Food Code	RRS		MON810	
	%GMO	GMO as % weight of product	%GMO	GMO as % weight of product
1494 (bread)	0	0	1.5	0.45
1497 (taco)	10	0.5	10	9
1498 (taco)	0.5	0.025	0.5	0.45

B.

Taco 1497				Taco 1498			
	Flour mass (g)				Flour mass (g)		
	Maize	Soya	OSR seed		Maize	Soya	OSR seed
non-GM	81	4.5	4.5	non-GM	89.55	5	5
GM	9	0.5	0.5	GM	0.45	0.025	0.025
% GM w/w	10	10	10	% GM w/w	0.5	0.5	0.5
% ingredient	90	5	5	% ingredient	90	5	5

Table 3. Primers used in the study.

Primer Set	Method	Code	Targeting Region		Sequence (5'-3')	Amplicon Length (bp)
RoundUp Ready™ Soya Quantification	Taqman™	RR1F RR1R RR1Probe	5' transgenic junction	CaMV P-35S Plant CaMV P-35S	CCTTTATCGCAATGATGGCATTTG CATACAGGTTAAAATAAACATAGGGA 9-AGGAGCCACCTTCCTTTTCCATTTGGGT0	82
	Taqman™	Le1r Le1f Le1Probe	Lectin gene	Plant Plant Plant	CAACGAAAACGAGTCTGGTGAT GGATTTGGTGGATCCCA 6-TCGCTGTTGAGTTTGACACTTTCCGGAA0	78
MON810 Quantification	Taqman™	Mon810f Mon810r MON810ZProbe	5' transgenic junction	CaMV P-35S Plant CaMV P-35S	ATTTGTAGGAGCCACCTTCCT AAGGACGAAGGACTCTAACGT 9-AGTGACAGATAGCTGGGCAATGGCA0	97
	Taqman™	ADHF3-4 ADHR3-4 Adh1-Probe	ADH gene	Plant Plant Plant	CCACTCCGAGACCCTC CGTCGTTTCCCATCTCTTCC 6-AATCAGGGCTCATTTTCTCGCTCCTCA0	134
QUIZ RRS set 1	QUIZ PCR	RR1aF RR1aR	5' transgenic junction	CaMV P-35S Plant	**6-**CGCAATGATGGCATTTG CATACAGGTTAAAATAAACAT	75
	QUIZ PCR	Lec1F Lec1R	Lectin gene	Plant Plant	**6-**CAACGAAAACGAGTCTGGTGAT TGTGTGGATTTGGTGGA	83
QUIZ RRS set 2	QUIZ PCR	RR2F RR2R	3' transgenic junction	T-Nos Plant	**6-**ACTGCTTCTCCCAGAATGATC TCGAGCTTCTTCACGAACTT	286
	QUIZ PCR	LecF LecR	Lectin gene	Plant Plant	**6-**ATGGGCTTGCCTTCTTTCT CCGATGTGTGGATTTGGTG	157
QUIZ RRS set 2b	QUIZ PCR	LecF LecR	Lectin gene	Plant Plant	**6-**ATGGGCTTGCCTTCTTTCT CCGATGTGTGGATTTGGTG	157
	QUIZ PCR	RR3F RR3R	3' transgenic junction	T-Nos Plant	**6-**TTTGGGATCGGAGAAGA AGGCAACAGCATGAAAAA	92
QUIZ RRS Set 3	QUIZ PCR	Lec3F Lec3R	Lectin gene	Plant Plant	**6-CAACGAAAACGAGTCTGGTG** **GGATTTGGTGGATCCCAAGA**	78
	QUIZ PCR	RR4F RR4R	3' transgenic junction	T-Nos Plant	**HEX-CGACCAGGCCATTC** **ACATATAGCTTCTCGTTGTTAG**	78
QUIZ MON810	QUIZ PCR	Mon810f Mon810r	5' transgenic junction	CaMV P-35S Plant	6-TTTGTAGGAGCCACCTTCCT AAGGACGAAGGACTCTAACGT	96
	QUIZ PCR	ADHF3-4 ADHR3-4	ADH gene	Plant Plant	6-CCACTCCGAGACCCTC GTCGTTTCCCATCTCTTCC	133

HEX and the numbers 6 and 9 denote HEX, FAM and TET fluorophores respectively, used for labeling the DNA primers/probes; 0 denotes TAMRA quencher moiety used to detect the degradation of probe in Taqman™ assays.

Table 4. Pre-QUIZ testing results obtained using primer set QUIZ RRS set 1 for: A) 5% CRM and B) 1% CRM. Unusual sequence of numbers obtained after 10-fold serial dilutions of template suggest that assay was not working.

A.

Dilution Factor	Copy Number Expected Lectins	Copy Number Expected GMOs	Numbers of Positive Lectins	Numbers of Positive GMOs
10^{-1}	5000	250	9	10
10^{-2}	500	25	9	5
10^{-3}	50	2.5	0	1
10^{-4}	5	0.25	0	0
10^{-5}	0.5	0.025	1	0
10^{-6}	0.05	0.0025	2	1

B.

Dilution Factor	Copy Number Expected Lectins	Copy Number Expected GMOs	Numbers of Positive Lectins	Numbers of Positive GMOs
10^{-1}	5700	57	3	3
10^{-2}	570	5.7	3	3
10^{-3}	57	0.57	10	2
10^{-4}	5.7	0.057	2	3
10^{-5}	0.57	0.0057	0	2
10^{-6}	0.057	0.00057	2	2

Table 5. Results from reciprocal testing of primers and DNA with DNA and primers that have been shown to work for QUIZ. Pre-QUIZ data obtained for A) DNA sample from proof of concept study (0.45% GMO admixture) using primers set QUIZ RRS set 1; and B) DNA sample 1456 (0.5% CRM) using primer set RRS set 2.

A)

Dilution Factor	Copy Number Expected Lectins	Copy Number Expected GMOs	Numbers of Positive Lectins	Numbers of Positive GMOs
10^{-1}	2000	9	5/5	5/5
10^{-2}	200	0.9	4/5	2/5
10^{-3}	20	0.09	5/5	3/5
10^{-4}	2	0.009	4/5	2/5
10^{-5}	0.2	0.0009	1/5	0/5
10^{-6}	0.02	0.00009	1/5	0/5

B)

Dilution Factor	Copy Number Expected Lectins	Copy Number Expected GMOs	Numbers of Positive Lectins	Numbers of Positive GMOs
10^{-1}	5700	57	5/5	3/5
10^{-2}	570	5.7	5/5	1/5
10^{-3}	57	0.57	5/5	0/5
10^{-4}	5.7	0.057	2/5	0/5
10^{-5}	0.57	0.0057	0/5	0/5
10^{-6}	0.057	0.00057	0/5	0/5

Data give an estimate of 0.2% RoundUp Ready™ soya (0.0-0.7% 95% range).

Table 6. PCR and Taqman single molecule amplification efficiency measured using single nuclei.

Assay	Positives	Number of reactions	%
Simplex Taqman			
*adh1*1	87	96	91
Mon810	27	96	28
Duplex Taqman			
*adh1*1	78	96	81
Mon810	24	96	25
PCR			
*adh1*1	12	96	13
MON810	15	96	16

Table 7. Pre-QUIZ data for RRS CRM samples ranging from 0-5% GM content. GM estimations are derived from the results obtained testing the dilution series using most probable number statistics.

A. Sample 1458 (0% GM content)

Dilution Factor	Copy Number Expected Lectins	Copy number Expected GMs	Numbers of Positive Lectins	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	9100	2.73	10	3	
10^{-2}	910	0.273	10	0	
10^{-3}	91	0.0273	10	0	0.1 (0-0.3)
10^{-4}	9.1	0.00273	4	0	
10^{-5}	0.91	0.000273	2	0	
10^{-6}	0.091	0.0000273	0	0	

B. Sample 1457 (0.1% GM content)

Dilution Factor	Copy Number Expected Lectins	Copy number Expected GMs	Numbers of Positive Lectins	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	13500	13.5	10	4	
10^{-2}	1350	1.35	10	0	
10^{-3}	135	0.135	10	0	0 (0-0.1)
10^{-4}	13.5	0.0135	6	0	
10^{-5}	1.35	0.00135	2	0	
10^{-6}	0.135	0.000125	0	0	

C. Sample 1456 (0.5% GM content)

Dilution Factor	Copy Number Expected Lectins	Copy number Expected GMs	Numbers of Positive Lectins	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	12900	64.5	10	4	
10^{-2}	1290	6.45	9/9	1	
10^{-3}	129	0.645	9	0	0.1 (0-0.5)
10^{-4}	12.9	0.0645	5	0	
10^{-5}	1.29	0.00645	1	0	
10^{-6}	0.129	0.000645	0	0	

D. Sample 1455 (1% GM content)

Dilution Factor	Copy Number Expected Lectins	Copy number Expected GMs	Numbers of Positive Lectins	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	5700	57	10	10	
10^{-2}	570	5.7	10	8	
10^{-3}	57	0.57	10	0	2.7 (0.9-7.9)
10^{-4}	5.7	0.057	3	0	
10^{-5}	0.57	0.0057	0	0	
10^{-6}	0.057	0.00057	0	0	

E. Sample 1454 (2% GM content)

Dilution Factor	Copy Number Expected Lectins	Copy number Expected GMs	Numbers of Positive Lectins	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	3400	68	10	9	
10^{-2}	340	6.8	10	4	
10^{-3}	34	0.68	7	0	2.1 (0.8-6.5)
10^{-4}	3.4	0.068	2	0	
10^{-5}	0.34	0.0068	0	0	
10^{-6}	0.034	0.00068	0	0	

F. Sample 1453 (5% GM content)

Dilution Factor	Copy Number Expected Lectins	Copy number Expected GMs	Numbers of Positive Lectins	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	5000	250	10	10	
10^{-2}	500	25	10	8	
10^{-3}	50	2.5	10	2	3.7 (1.3-10)
10^{-4}	5	0.25	3	0	
10^{-5}	0.5	0.025	0	0	
10^{-6}	0.05	0.0025	0	0	

Table 8. Pre-QUIZ data for MON810 CRM samples ranging from 0-5% GM content. GM estimations are derived from the results of the dilution series using most probable number statistics.

A. Sample 1464 (0% GM content)

Dilution Factor	Copy Number Expected *adh*	Copy number Expected GMs	Numbers of Positive *adh*	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	5400	1.08	10	7	1.3% (0.5-3.5)
10^{-2}	540	0.108	9	1	
10^{-3}	54	0.0108	9	0	
10^{-4}	5.4	0.00108	7	0	
10^{-5}	0.54	0.000108	0	0	
10^{-6}	0.054	0.0000108	0	0	

B. Sample 1463 (0.1% GM content)

Dilution Factor	Copy Number Expected *adh*	Copy number Expected GM0s	Numbers of Positive *adh*	Numbers of Positive GMOs	% GM0 Estimation (95% range)
10^{-1}	2900	2.9	10	8	1.8% (0.6-5.2)
10^{-2}	290	0.29	10	7	
10^{-3}	29	0.029	8	1	
10^{-4}	2.9	0.0029	3	0	
10^{-5}	0.29	0.00029	0	0	
10^{-6}	0.029	0.000029	0	0	

C. Sample 1462 (0.5% GM content)

Dilution Factor	Copy Number Expected *adh*	Copy number Expected GMs	Numbers of Positive *adh*	Numbers of Positive GMOs	% GM Estimation (95% range)
10^{-1}	3000	15	10	10	9.2 (3.2-23.9)
10^{-2}	300	1.5	9	5	
10^{-3}	30	0.15	9	1	
10^{-4}	3	0.015	7	0	
10^{-5}	0.3	0.0015	0	0	
10^{-6}	0.03	0.00015	0	0	

D. Sample 1461 (1% GM content)

Dilution Factor	Copy Number Expected *adh*	Copy number Expected GMs	Numbers of Positive *adh*	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	2900	29	10	9	2.6 (1.0-8.1)
10^{-2}	290	2.9	9	3	
10^{-3}	29	0.29	10	1	
10^{-4}	2.9	0.029	5	0	
10^{-5}	0.29	0.0029	0	0	
10^{-6}	0.029	0.00029	0	0	

E. Sample 1460 (2% GM content)

Dilution Factor	Copy Number Expected *adh*	Copy number Expected GMs	Numbers of Positive *adh*	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	3200	64	9	9	46.7 (16.5-135.2)
10^{-2}	320	6.4	6	6	
10^{-3}	32	0.64	7	2	
10^{-4}	3.2	0.064	5	0	
10^{-5}	0.32	0.0064	1	0	
10^{-6}	0.032	0.0027	0	0	

F. Sample 1459 (5% GM content)

Dilution Factor	Copy Number Expected *adh*	Copy number Expected GMs	Numbers of Positive *adh*	Numbers of Positive GMs	% GM Estimation (95% range)
10^{-1}	5400	270	9	10	87.2 (31.5-239.8)
10^{-2}	540	27	10	9	
10^{-3}	54	2.7	9	0	
10^{-4}	5.4	027	6	0	
10^{-5}	0.54	0.027	0	0	
10^{-6}	0.054	0.0027	0	0	

The value highlighted in red demonstrates the effect of a probable false negative on GMO estimations.
By 'correcting' the highlighted value to 10, pre-QUIZ estimate would give 4.6% (1.6-13.9)

Table 9. Summary of Pre-QUIZ testing data. The selection of the dilutions for QUIZ is determined empirically. The table shows the basis for the choice of the dilutions selected for full QUIZ based on the pre-QUIZ data for the two laboratories.

	Sample	Dilution Factor for reference target	No of positive (10 repetitions)	Dilution factor For GM target	Numbers of Positives for GM target (10 repetitions)
NIAB	1453	10^{-4}	3	10^{-2}	8
	1454	10^{-4}	2	10^{-2}	4
	1455	10^{-4}	3	10^{-2}	8
	1456	10^{-4}	5	10^{-1}	5
	1457	10^{-4}	6	10^{-1}	4
	1458	10^{-4}	4	10^{-1}	3
	1459	10^{-5}	2	10^{-3}	4
	1460	10^{-5}	1	10^{-3}	2
	1461	10^{-5}	1	10^{-2}	3
	1462	10^{-4}	0 (10^{-3}=8)	10^{-2}	2
	1463	10^{-4}	0 (10^{-3}=8)	10^{-1}	5
	1464	10^{-5}	2	10^{-1}	1
CSL	1453	10^{-4}	0 (10^{-3}=7)*	10^{-2}	5
	1454	10^{-4}	2*	10^{-2}	6
	1455	10^{-4}	6*	10^{-2}	3
	1456	10^{-4}	6*	10^{-2}	3
	1457	10^{-4}	4*	10^{-2}	2
	1458	10^{-4}	4*	10^{-1}	2
	1459	10^{-4}	6*	10^{-2}	9
	1460	10^{-4}	5*	10^{-2}	2
	1461	10^{-4}	5*	10^{-2}	3
	1462	10^{-4}	7*	10^{-2}	5
	1463	10^{-4}	3*	10^{-1}	7
	1464	10^{-4}	7*	10^{-1}	1

*Eight repetitions were performed. Sometimes dilutions that have given no positives may be chosen because the numbers in the previous dilution (shown) is deemed high.

Table 10. QUIZ GMO estimation data obtained for RR™ Soya CRMs:

A. Summary of NIAB data.GM estimations obtained using 96-well plate QUIZ format.

Primers	Samples	Dilution		Positives (96)		Expected	QUIZ values		
		Reference	GM	Reference	GM	%GM DNA	lower 2.5%	estimate	upper 2.5%
QUIZ RRS 1a (1st test)	1453	-4	-2	30/96	51/96	5	1.294	2.019	3.276
	1454	-4	-2	63/80*	56/80*	2	0.525	0.78	1.152
	1455	-4	-2	45/96	23/96	1	0.25	0.434	0.705
	1456	-4	-1	50/96	79/96	0.5	0.162	0.234	0.344
	1457	-4	0	54/96	69/96	0.1	0.0106	0.015	0.0353
	1458	-4	0	67/96	33/96	0 (<0.03)	0.00225	0.004	0.00533
QUIZ RRS 1a (2nd test)	1453	-4	-2	22/96	50/96	5	1.74	2.822	4.93
	1454	-4	-2	88/96	86/96	2	0.633	0.913	1.32
	1455	-4	-2	37/96	46/96	1	0.859	1.339	2.104
	1456	-4	-2	33/96	25/96	0.5	0.415	0.716	1.206
	1457	-4	0	66/96	76/96	0.1	0.00943	0.013	0.02971
	1458	-4	0	12/96	6/96	0 (<0.03)	0.00134	0.005	0.01264

* Some reactions lost to evaporation.

B. Summary of CSL data. GM estimations obtained using 96-well plate QUIZ format.

Primers	Samples	Dilution		Positives (96)		Expected	QUIZ values		
		Reference	GM	Reference	GM	%GM DNA	lower 2.5%	estimate	upper 2.5%
QUIZ RRS 1a (1st test)	1453	-4	-2	45/96	80/96	5	1.93	2.813	4.21
	1454	-4	-2	60/96	69/96	2	0.893	1.291	1.879
	1455	-4	-3	69/96	9/96	1	0.316	0.78	1.425
	1456	-4	-2	70/96	46/96	0.5	0.335	0.501	0.732
	1457	-4	-2	82/96	11/96	0.1	0.0289	0.064	0.1134
	1458	-4	-1	78/96	13/96	0 (<0.03)	0.00426	0.009	0.01499
QUIZ RRS 1a (2nd test)	1453	-4	-2	40/96	90/96	5	3.39	5.04	7.91
	1454	-4	-2	55/96	74/96	2	1.189	1.725	2.521
	1455	-4	-3	77/96	12/96	1	0.389	0.83	1.444
	1456	-4	-2	74/96	56/96	0.5	0.411	0.596	0.866
	1457	-4	-2	82/96	21/96	0.1	0.0737	0.129	0.2052
	1458	-4	-1	84/96	14/96	0 (<0.03)	0.004234	0.009	0.014993

Table 11. QUIZ GMO estimation data obtained for MON810 CRMs:

A. Summary of NIAB data. GM estimations obtained using 96-well plate QUIZ format.

Primers	Samples	Dilution		Positives (96)		Expected	QUIZ values		
		Reference	GM	Reference	GM	%GM DNA	lower 2.5%	estimate	upper 2.5%
QUIZ MON810 (1st test)	1359	Reference	-2	5/96	32/96	2.9	3.27	7.572	32.9
	1360	-4	-2	62/96	88/96	1.16	1.65	2.362	3.49
	1361	-4	-2	27/96	30/96	0.6	0.666	1.134	1.952
	1362	-4	-2	41/96	32/96	0.29	0.449	0.728	1.159
	1363	-4	-1	84/96	17/96	0.058	0.00507	0.009	0.01564
	1364	-4	0	20/96	57/96	0 (<0.01)	0.0236	0.038	0.0691
QUIZ MON810 (2nd test)	1359	-4	-2	33/96	84/96	2.9	3.23	4.884	7.74
	1360	-4	-2	53/96	79/96	1.16	1.492	2.143	3.132
	1361	-4	-2	43/96	44/96	0.6	0.671	1.032	1.592
	1362	-4	-1	22/96	49/96	0.29	0.1687	0.274	0.4755
	1363	-4	-2	25/96	13/96	0.058	0.2233	0.483	0.9355
	1364	-4	0	19/96	34/96	0 (<0.01)	0.01148	0.02	0.03682

B. Summary of CSL data. GM estimations obtained using 96-well plate QUIZ format.

	1364	-5	-1	45/96	16/96	0 (<0.01)	0.001498	0.003	0.004926
Primers	Samples	Dilution		Positives (96)		Expected	QUIZ values		
		Reference	GM	Reference	GM	%GM DNA	lower 2.5%	estimate	upper 2.5%
QUIZ MON810 (1st test)	1359	-5	-3	24/96	63/96	2.9	2.35	3.7	6.26
	1360	-5	-3	12/96	18/96	1.16	0.749	1.555	3.465
	1361	-5	-2	23/96	69/96	0.6	0.292	0.461	0.787
	1362	-4	-2	65/96	32/96	0.29	0.226	0.36	0.545
	1363	-4	-1	73/96	80/96	0.058	0.0879	0.125	0.1794
	1364	-5	-1	20/96	19/96	0 (<0.01)	0.00493	0.009	0.0181
QUIZ MON810 (2nd test)	1359	-5	-3	21/96	52/96	2.9	1.94	3.154	5.54
	1360	-5	-3	18/96	16/96	1.16	0.423	0.878	1.749
	1361	-5	-2	11/96	16/96	0.6	0.688	1.498	3.618
	1362	-4	-2	82/96	40/96	0.29	0.1828	0.282	0.4175
	1363	-4	-1	78/96	72/96	0.058	0.0582	0.083	0.1173

Table 12. Taqman™ estimates of the GM contents of CRMs. GM content estimations for A. RoundUp Ready soya and B. MON810 using TaqMan™.

A. RoundUp Ready soya.

RoundUp Ready™ soya (GTS40-3-2)	CRM (%GMO)				
	5	2	1	0.50	0.10
Expected %GM DNA	5	2	1	0.5	0.1
Mean %GM DNA	4.16	2.43	1.61	0.77	0.21
95% conf. Int	± 0.34	± 0.25	± 0.14	± 0.12	± 0.05
% RSDr	16.15	20.54	8.27	26.52	27.33
% Bias	-16.89	21.42	61.19	53.21	110.82

B. MON810 maize.

MON810	CRM (%GMO)				
	5	2	1	0.50	0.10
*Expected %GM DNA	2.9	1.16	0.58	0.29	0.058
Mean %GM DNA	2.39	1.12	0.64	0.3	0.07
95% conf. Int	± 0.12	± 0.12	± 0.1	± 0.05	± 0.01
% RSDr	10.01	20.32	12.34	18.14	21.24
% Bias	-17.75	-5.69	10	-2.22	7

*A conversion factor of 0.58 has been applied to CRM stated values to adjust for the hemizygous nature of the hybrid GMO and the triploid endosperm of maize seed.

Table 13. Summary of food testing and re-testing results. GM content estimations of processed samples using TaqMan™ and QUIZ, and the effects of using different primer combinations.

Food Sample	Primers	QUIZ % GMO DNA	QUIZ uncertainty	Taqman™ % GM DNA	Taqman™ uncertainty	Expected % GMO
1494 (bread)	QUIZ MON810	1st 0.5 2nd 0.46	± 0.08 ± 0.13	+ve 0.92	Na ± 0.12	0.87
1497 (taco)	QUIZ RRS Set 2a	1st 0.02 2nd 0.05	± 0.004 ± 0.007	3.09	± 1.04	10
	QUIZ RRS Set 2b	1st 2.21 2nd 8.56 (CSL)	± 0.68 ± 2.43			
	QUIZ MON810	1st 5.15 2nd 4.69	± 1.25 ± 0.84	5.6	± 0.67	5.8
1498 (taco)	QUIZ RRS Set 2a	1st 0.44 2nd 0.082	± 0.07 ± 0.01	0.27	Na	0.5
	QUIZ RRS Set 2b	3rd 0.68 4th 0.082 (CSL)	± 0.16 ± 0.01			
	QUIZ MON810	1st 0.27 2nd 0.13	± 0.06 ±0.03	0.76	± 0.20	0.29

Numbers highlighted in green denote restated MON810 data and blue boxes denote re-testing results. Unless stated otherwise, all QUIZ data were generated by NIAB and Taqman™ data by CSL. Na = not applicable.

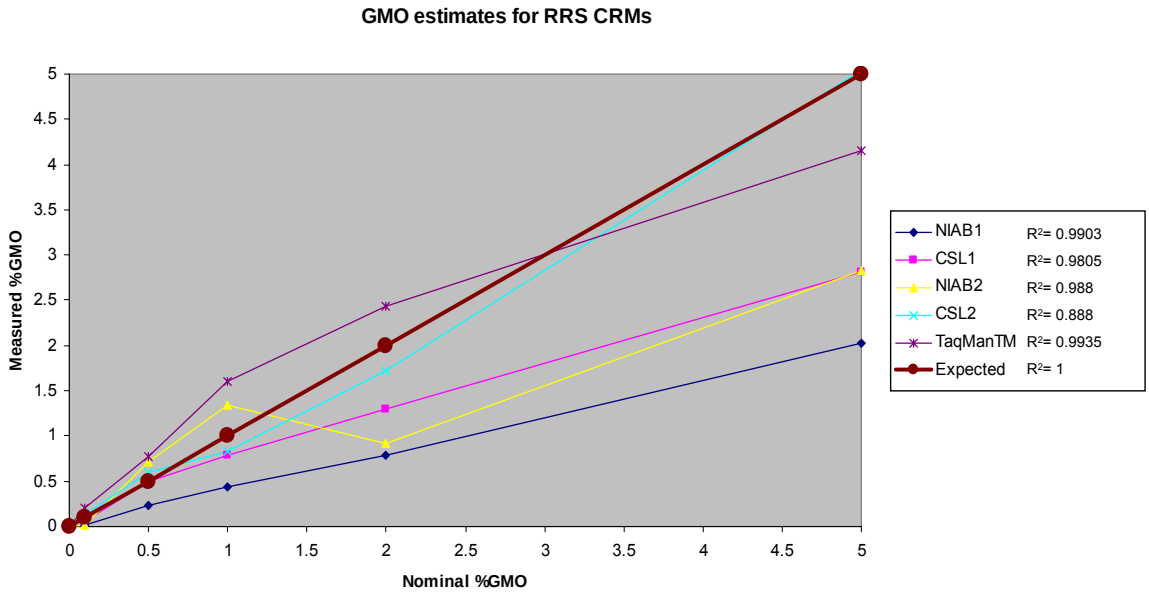


Figure 1. Summary of QUIZ and TaqMan™ results for RRS CRMs. The coefficient of fit (R^2) value of the best-fitting straight line through each set of results is shown.

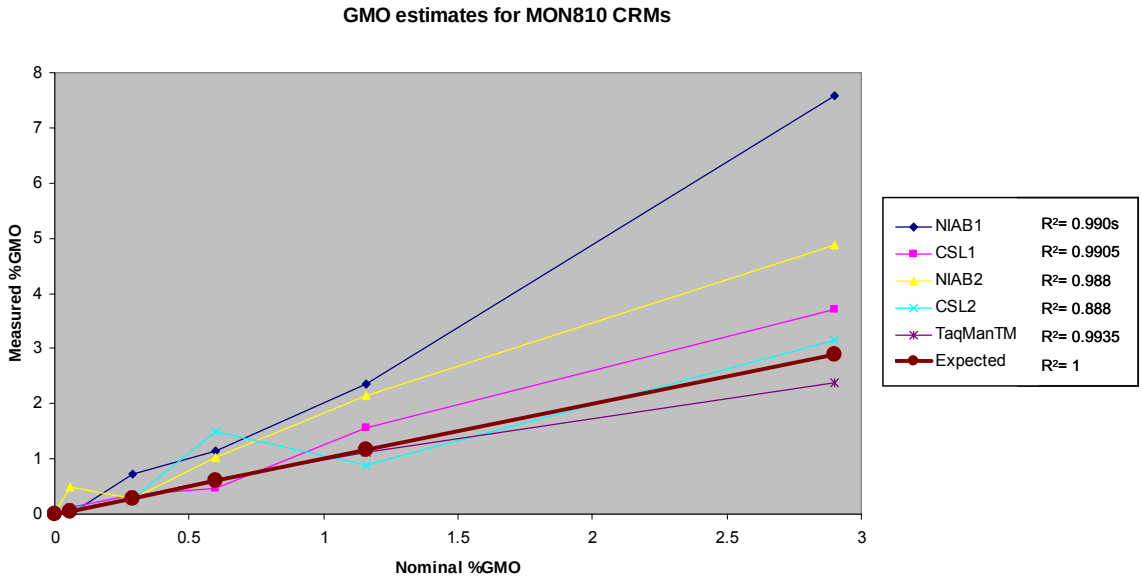


Figure 2. Summary of QUIZ and TaqMan™ results for MON810 CRMs. The coefficient of fit (R^2) value of the best-fitting straight line through each set of results is shown.

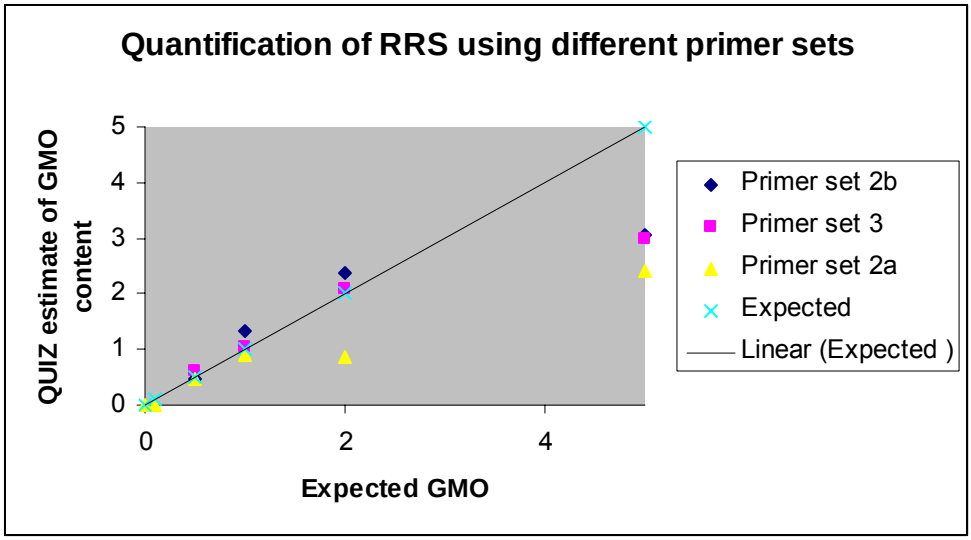
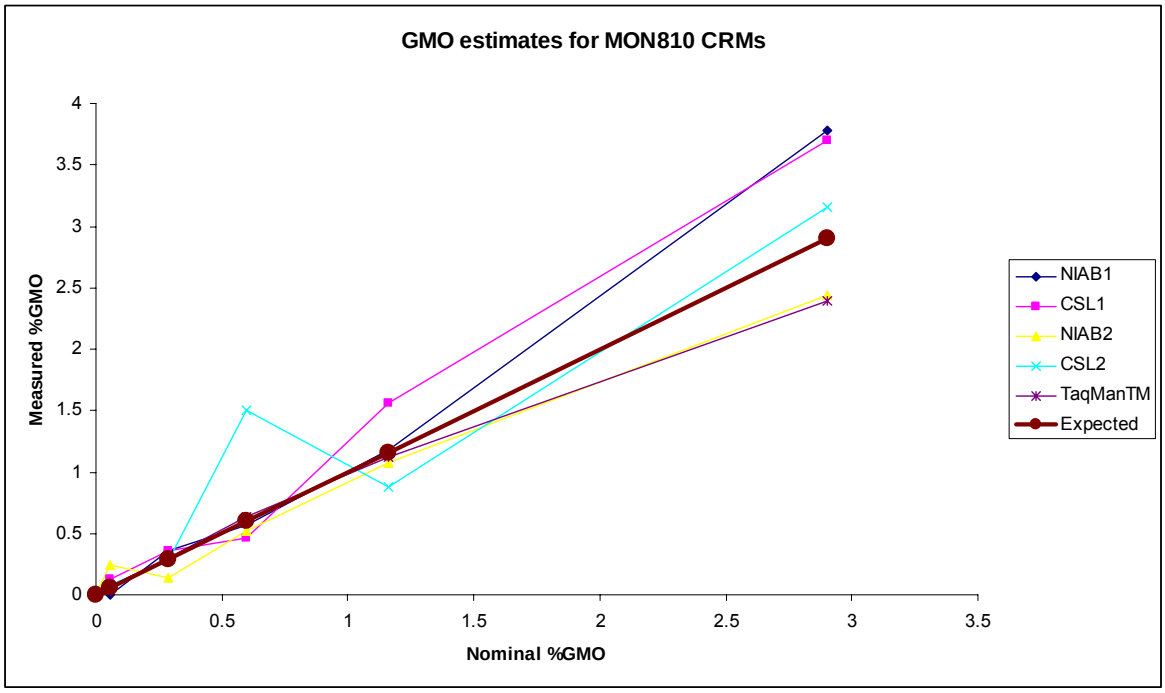


Figure 3. Effect of primer sets upon QUIZ quantification of RRS. GM estimations of RRS CRM samples using primer sets RRS 2a, RRS 2b and RRS 3. These have reference/GM indices of 0.55, 1.7 and 1, respectively or reference – GM length differences of -129, 65, and 0 bp, respectively.

Figure 4. Restated MON810 estimates. NIAB QUIZ estimates for MON810 CRM samples have been recalculated based upon one *adh* allele scored.



Appendix 1
Statistical Analysis of QUIZ data

Introduction

Our earlier reports have demonstrated that QUIZ can provide accurate estimates of GM% without the requirement for CRMs. However, there appeared to be a statistical overestimation of GM% in the demonstration cases we studied. Here we quantify and correct this bias while assessing the practicality of QUIZ as a method for routine assessment of GM contamination. In addition we present R scripts for the routine analyses of QUIZ data.

Data from QUIZ fall into two separate stages. First there is the estimation of most probable number (MPN) and of GM% from a serial dilution of GM and reference constructs. For practical reasons, we have used six, ten-fold dilutions experimentally with ten PCR amplifications (wells) at each dilution, though some alternative dilutions series have been examined here by simulation. The primary purpose of this first stage is to estimate suitable dilutions for the second stage in which more accurate estimation of GM% is carried out with larger numbers of wells at single dilutions of the GM and reference samples. The statistical analysis of this second stage is relatively straightforward, that for the first stage more complex, although GM% can still be estimated from these serial dilutions, often with sufficient accuracy to classify samples as lying within or outside statutory thresholds.

Bias in MPN estimates of microbial densities from serial dilutions is well known (Garthright and Blodgett 2003). However, its correction can be complicated. For example, Salama et al. (1978) presented a correction to the maximum likelihood estimator of density, u^* to produce an unbiased estimator u^b:

$$\mu^b = \mu^* - (1/2)\sum_{i=1}^{r}\left(\frac{\partial^2\mu^*}{\partial x_i^2}\right) n_i e^{-\mu^* V_i}(1 - e^{-\mu^* V_i})$$

with

$$\frac{\partial^2\mu^*}{\partial x_i^2} = \frac{V_i^2}{2(1 - e^{-\mu^* V_i})^2 D^3}\left\{\sum_{j=1}^{r}\frac{V_j^3 z_j \sinh(\mu^* V_j)}{[\cosh(\mu^* V_j) - 1]^2}\right\} - \frac{V_i^3}{(1 - e^{-\mu^* V_i})[\cosh(\mu^* V_i) - 1]D^2}$$

$$z_j = n_j(1 - e^{-\mu^* V_j})$$

$$D = \sum_{j=1}^{r}\frac{V_j^2 z_j}{2[\cosh(\mu^* V_j) - 1]} \qquad [1]$$

Haas (1989) discusses this further and Klee (1993) presents a computer program (“5000 lines of Fortran and 1000 of assembler code”) to provide

unbiased estimates and their standard errors. We have not pursued this method for three reasons. Firstly, it is clearly very complex. Secondly, although it may provide an unbiased estimate of MPN, there is no guarantee that it will also provide unbiased estimates of the ratio of two MPN. Finally, this work is targeted at the use of a single dilution series to estimate MPN, often with very small numbers of dilutions and wells. For QUIZ, we propose a two stage process in which the initial dilution series is used to calculate suitable dilutions for a larger assay at a single dilution. At this second stage the bias is expected to be lower (see below). We have used simulation experiments to assess the magnitudes of bias in this process and to justify a more simple correction. Some of the combinations of well numbers and dilutions that we have simulated are extreme. They have been included to test the method, not to advocate the use of QUIZ with these parameters. For example, some simulations are of only a single dilution with 10 wells. QUIZ would not be used with such a small number of wells: more wells at additional dilutions or additional wells at the same dilution are required.

The approach we have adopted is first to find suitable statistical procedures which work at the single dilution stage and then to test if these transfer acceptably to the serial dilution levels. For single dilutions, we have found that adding ½ to the observed number of zero wells virtually eliminates bias from the estimation of MPN even with very low numbers of wells. However, although biases are removed from the MPN estimates, the ratio of two MPNs, which provides the estimate of GM%, remains strongly biased upwards. For accurate determination of GM% using at least 96 wells, the bias is slight and can be ignored. For smaller numbers, and for the serial dilutions, most bias can be eliminated by working on log differences in MPN rather than on ratios.

Bias in MPN estimates at a single dilution

The estimated mean number of number of molecules per well at a single dilution is:

$$\text{No. mols} = -\ln(\text{proportion of zeros}) \qquad [2]$$

We propose, as a simple adjustment for bias, to inflate both the total number of zeros by ½ and the total number of wells tested:

$$\text{No. mols} = -\ln\left[(\text{no. of zeros} + ½)/ (\text{total} + ½) \right] \qquad [3]$$

A heuristic justification for this procedure is that the expected proportion of zeros should be a continuous variable in the range 0 – 1, yet the observed proportion goes up in steps of 1/(no of wells). The addition of ½ may correct for this in the same manner as the “Yates’ correction for continuity” in contingency chi-squared tests with low expected numbers (Yates 1934). At high numbers of wells, the correction will make little difference. Note, however, that the test will give an estimate of MPN even when no zeros are observed. For convenience, we refer to methods based on [3] as the QUIZ

method, those based on [2] as the standard method and those based on [1] as the bias corrected method.

To assess the QUIZ method, we simulated DNA concentrations in the range 0.005 to 4.0 molecules/ul for a total of ten 1ul PCR wells. Each concentration was simulated 10,000 times. Simulations were carried out in R.

Estimates from [2], from [3] and by averaging the count of molecules in each well (henceforth the exact method) are given in Table 1. The standard estimator is clearly biased but the QUIZ estimator removes virtually all bias up to a mean of 1 mol/ul and remains acceptable up to a mean of 2 mol/ul. For values from 0.5 and higher, a mean over all simulation is not possible for the standard estimator since for some simulations no zeros are observed. This is to be expected. With only 10 wells, with an expected mean of 1 molecule per well, we shall observe no zeros in 10 wells in 1% of experiments or simulations. With 2 molecules per well, we expect no zeros on 23% of occasions. In practice, we would never use such small numbers of wells at a single dilution for QUIZ. It is remarkable how well the QUIZ estimates perform, even in excess of 2 molecules per well, when the probability of observing no zeros is high. The exact method is working near perfectly, as is should: its role is to validate the simulation procedure.

Table 2 gives means across simulations, excluding those where no zeros occurs. This selection against instances where, by chance, the number of molecules per well is high, introduces a downwards bias in all estimation methods. This can be seen in the progressive downwards bias in the (no longer) exact method as the proportion of simulations which are discarded increases with the simulated number of molecules per well. For a true mean of up to 1.2, the QUIZ method has a lower overall bias than the standard method. At this value, roughly 3% of simulations give no zero counts. For values greater than 1.2, the standard method is better: the upward bias in the maximum likelihood estimator is compensated to some extent by the downward bias induced by rejecting the 100% positive simulations. At high mean values, with the majority of simulations giving 100% positive results, neither method is any good, and biases from the exact method become increasingly great too. Note that over the whole range, the QUIZ method with all simulations (Table 1) is more accurate than the standard method including only simulations with at least one zero (Table 2).

The low number of molecules at which either method is successful is due to the low number of PCR wells in these simulations. Such a low number of wells would not be used in a single stage estimation by QUIZ, but serve to exaggerate the bias in the estimation of MPN so we can better assess the validity of the method. As well number increases, the risk of observing 100% positive results is reduced and higher concentrations can be estimated. Table 3 provides results for 10,000 simulations at a series of concentrations with 96 wells, for all simulations and separately for simulations giving at least one negative result. As a result of increasing the number of wells, the probability of observing 100% positive wells is reduced. It is less than 0.01 at 3 molecules per well. QUIZ works well over the range 0 - 3 molecules per well, with

average estimates which are very close to the simulated values. At extremely low numbers of molecules per well, the exact method shows a large downwards percentage bias. This is spurious: a consequence of expressing bias as a percentage which grossly inflates its apparent effect at these extreme low values. Although the proportional bias is large, the absolute bias is negligible: the mean estimate agrees with the true value to the fourth decimal place. In line with the exact method, estimates from QUIZ and the standard method show the same apparent downwards bias although their mean estimates are near perfect. As before, the QUIZ method using all simulations is always less biased than the standard method excluding 100% positives. These results show that biases in estimation can still be substantial with higher number of wells but that the QUIZ method is effective in correcting for this.

In practice, for estimation of MPN alone, 0% or 100% positive results would generally require retesting of the sample, probably at a different dilution. However, to evaluate if GM% lies above or below some threshold, this is not always necessary. This is discussed in more detail later. If the true MPN is such that a 0% or a 100% positive result is a likely outcome, then retesting to generate a less extreme result will also introduce bias. If the retest is at a different dilution, this bias should be small, though we have not verified this.

In practice, the initial estimation of MPN in the first stage of QUIZ should ensure that 0% or 100% positive results in this second stage should be very rare occurrences.

Bias in MPN from multiple dilutions

The standard QUIZ protocol has been simulation in which five ten-fold serial dilutions give measurements at six concentrations. For each dilution, ten 1 ul wells are assayed, 3 ul are taken forward to be made up to 30 ul for the next dilution and 17 ul remain. The method of adjusting for bias has been modified slightly from that described above: here ½ is added to the number of observed zeros and to the total only for dilutions at which at least one zero is observed. To add ½ a zero to all dilutions showing no negative amplifications would greatly overcorrect the upwards bias in the MPN estimates. To balance the increased weight placed on the lower dilutions by the addition of ½, the contribution to the total likelihood from the higher dilutions is weighted by a factor (n+ ½) / n, where n is the number of wells. These adjustments are heuristic but work in practise (see below).

The simulations allocate molecules to PCR wells at different dilutions using the multinomial random number generator in R. This ensures that all molecules are accounted for. Estimation of MPN from these simulated data is by maximization of the likelihood of the observed number of positive and negative PCR reactions over all reactions simultaneously, assuming that the frequency of positive reactions follows an independent Poisson distribution at each dilution level. Before maximisation of the likelihood, ½ is added the observed numbers at some dilutions, as described in the previous paragraph.

10,000 simulations were carried out over a range of concentrations from 0.1 mols/ul to 1,000,000 mols/ul. Independent simulations were carried out for the two methods (standard and QUIZ). Results are in Table 4. At high concentrations, estimation fails if all wells are amplified at all dilutions. Results are therefore presented only where estimation was possible. Estimation was always possible at concentrations up to 10,000 mols/ul. This corresponds to a probability of no zeros of 0.99 in the second last dilution. A conservative heuristic for MPN estimation, therefore, is that at-least one zero must be observed in the second last dilution. Within this range, the average absolute bias the QUIZ method, is 2.8 % compared to 9.0 % from the original method. Standard deviations are also consistently lower over this range. Therefore, for estimating MPN, the simple adjustment of the QUIZ method works well.

Effect of dilution series on MPN precision and bias

Estimates of MPN are known to be "choppy" (Hass 1989) so it is probable that for any dilution series some MPN will be estimated with greater precision and less bias than others. The results in Table 4 show peaks in bias at 4, 40, 400 and 4,000 mols/ul. To study these effects, we have simulated alternative dilution series. These differ from those described earlier in assuming a volume at each dilution of 1000 ul rather than the 30 ul used earlier. This provides greater flexibility in combinations of well numbers and dilutions. For example, with only 30 ul, a dilution series with 30 wells of 1 ul each is impossible since some residual volume is required to create the next dilution in the sequence. The following combinations of wells/dilutions have been studied: 10/6, 6/10, 3/20, 20/3, 12/5, 5/12, 30/2, 2/30, 15/4 and 4/15. To facilitate comparisons among the series, dilutions were varied such that the final concentrations in all series were expected to be 10^{-5} that of the original. For the 10/6 series this is a simple 10 fold dilution at each stage. For 6/10 for example, it is a 3.594 fold dilution. (Since $1/3.594^9 = 10^{-5}$.) In practice, of course, dilutions would be kept to more simple, probably integer, numbers.

Results are in Table 5. Firstly note that the results for the 10/6 series are very similar to those in Table 4: altering the initial volume from 30 ul to 1000 ul has had no perceptible difference. Secondly, the peaks in bias seen in Table 4 at 4, 40 and 400 mols/ul disappear in the other dilution series, though other, often larger, biases emerge. High values of coefficient of variation (CV) and of bias have been highlighted. Lower numbers of dilutions (and therefore higher numbers of wells per dilution) are generally less accurate. The lowest average absolute bias (0.8) is with six wells and 10 dilutions. This also has the lowest maximum absolute bias. The lowest average standard deviation is with four wells and 15 dilutions. The pattern of results can loosely be related to the dilution factor. Figure 1 plots the maximum observed bias for any dilution series against the dilution factor on a $\log_{10}\log_{10}$ scale. This scale is required to see any detail over the very variable range of dilutions. 10 on the original scale gives a $\log_{10}\log_{10}$ value of zero. Values lower than 10 are negative after transformation. 100,000 on the original scale is 0.7 on the double log scale. It is clear that, for the combination of dilution series and MPN simulated here, provided the dilution factor is 10 or less, the maximum observed bias is <10% which is acceptable for the determination of suitable dilutions for the second stage of QUIZ. Additional simulations are needed covering a greater range of MPN and dilutions to confirm this rule. However, Tables 4 and 5 together demonstrate that current experimental practice, using 10 wells at each of six dilutions, is adequate for QUIZ.

Bias in ratios estimated at single dilutions.

Results for 100,000 simulations for 96 wells over a range DNA concentrations for both the GM and reference PCR products are given in Table 6. A 100 fold dilution of reference DNA relative to the GM was assumed, giving an expected GM% in the range 0.1% to 10%. Expected proportions of zero cells (e^{-mean}) were therefore in the range 0.05 to 0.90.

Results show little difference between the standard and QUIZ estimator over the range of simulations, measured by deviation of the mean from the true value, or by the root mean squared error (RMSE). Estimation by direct counting of molecules, were it possible, is more accurate, giving lower RMSE.

Although there is little difference between the standard and QUIZ estimator, presumably because the biases in MPN for GM and reference tend to cancel, there is a slight advantage in favour of QUIZ. This can be exaggerated by selection of high GM and low reference dilutions. However, taking a mean of three molecules per well for the GM and 0.1 molecules per well for the lectin gave a mean GM% of 34.01 for the QUIZ method with a RMSE of 17.01 compared to 35.1%, RMSE 18.08, for the standard. These results are for a common set of simulations (after removing all problems due to 100% and 0% positives). It seems therefore, that there is little bias in GM% estimation at this scale of testing and either method will give reasonably unbiased results.

0% or 100% positive results make the estimation of MPN problematic and generally require a repeat evaluation at different dilutions. However, if the prime concern is to establish that the GM% in a sample is greater or less than a statutory threshold, this may still be possible. In particular, if only the GM shows 0% or 100% positive results, then simulations of the estimated MPN for the reference but with the GM fixed to the threshold value will give the probability of observing 0% or 100% positive GM wells at the threshold value. These probabilities can then be used to accept or reject the sample. For example, simulating 1 mol/ul of reference (assumed to be the observed value) with 0.9% GM (the threshold) at the same dilution, shows that zero GM would be observed in 42% of occasions. This is too common an occurrence to confidently accept a sample with an observed value of zero GM as truly below the threshold. However, in 100,000 simulations of 1 mol/ul of reference and 0.9% GM at a 10 fold higher concentration we observe zero GM on only 0.011% of runs: we would accept the sample.

We emphasise however, that the initial estimation of MPN in the first stage of QUIZ should avoid these complications by fixing appropriate dilutions for this second stage.

Bias in ratios estimated from multiple dilutions

The standard QUIZ protocol described above (10 wells at each of six 10-fold dilutions) was repeated for a range of DNA concentrations. Pairs of simulations were then used to generate simulated sets of estimated GM%. As before, GM% was estimated by both the standard and QUIZ method. Results are in Table 7. It is clear that both methods give very biased estimates: >20% of the true value on average. Both methods give broadly similar standard deviations, though these are of no direct use in constructing confidence intervals – subtracting 1.96 times the estimated standard deviation always gives a lower bound less than zero.

These high biases seem counterintuitive at first, given the small residual bias present in the QUIZ method of estimating MPN. They result from a bias in the method of estimating GM% as a ratio of two MPNs, as can be simply demonstrated. Consider pairs of uniform random numbers in the range 0…1. Over a large series of simulations, the mean value of each member of the pair is 0.5. The ratio of the means is therefore 0.5 / 0.5 = 1. The mean value of the ratio is not 1.0 however. For example, simulating 1,000,000 pairs of random numbers in R gives:

mean A	0.5000
mean B	0.4998
mean(A)/mean(B)	1.0006
mean(A/B)	7.6386
exp(mean(log(A)-log(B)))	1.0009

The mean of the ratios is hugely different from the ratio of the means. The antilog of the mean difference in logs between the two pairs is, however, very close to 1.0. The bias also drops with sample size: if we simulate the ratio of means of 2, 5 and 10 uniformly distributed random numbers rather than of a single pair of numbers, mean (A/B) is 1.381, 1.085 and 1.036 respectively.

This simple example demonstrates that at least part of the bias in estimating GM% as a ratio of two MPN estimates is likely to be a result of the built-in bias from dealing in ratios. It also suggests that working with logs can remove this bias. Results for GM% from Table 7 are repeated in Table 8, with the addition of median estimates and of back-transformed estimates of the difference in logs:

$$e^{\frac{\sum(\ln(MPN_{gm})-\ln(MPN_{ref})}{number\,of\ simulations}}$$

This estimate is identical to taking the geometric mean of the individual ratios:

$$\sqrt[n]{\frac{MPNgm_1.MPNgm_2...MPNgm_n}{MPNref_1.MPNref_2...MPNref_n}}$$

Estimation from a difference in logs is only possible for MPN >0. For this reason, no results are given in Table 8 for simulations of the lower numbers of GM molecules.

It is clear from these results that the back-transformed log estimates and the median estimates show much reduced bias. In practice, a point estimate from a single experiment will always be taken as the ratio of the two MPNs. Table 8 demonstrates that it matters little which estimation method is used, though since the multiple dilution experiment is primarily intended to provide estimates of MPN for GM and reference DNA, we would advocate the routine use of the QUIZ method. However, to combine estimates from multiple experiments, these results show that it is best to take the mean of the difference in logs (probably weighted by 1/[empirical estimate of the variance of log MPN]). Unless very large amounts of data are to be combined,

estimation from medians will be too inaccurate. However, zero GM presents problems for the log based method. There are various simple alternatives which have not been studied here. If available the raw data can be pooled from multiple experiments and GM% across the pooled dataset (For counts, this is the ML estimator; Teunis et al. 1999). Equivalently, rather than taking the mean of the estimated GM%, one could take the mean of the estimated MPNs, then compute GM% as the ratio of these two means. (This latter estimate shows little bias – data not shown) .

Confidence intervals for estimates and rejection / acceptance thresholds. (1) MPN

The calculation of confidence intervals (c.i.) for estimates of MPN from dilution series are complicated. Methods are discussed in Garthright and Blodgett (2003). However, c.i. may also be obtained by simulation. Over a large number of simulations, the 95% c.i. is given by the 2.5% and 97.5 % quantiles. Table 9 compares results from some published tables of MPN and confidence intervals (Haas, 1989; Garthright & Blodgett, 2003; Klee, 1993) with simulated estimates using the new method described here.

The first four cases are from large scale dilution series, with 96 wells and between three and five ten-fold dilutions. Garthright and Blodgett (2003) report only the biased estimate of MPN together with c.i. obtained from an approximate method (Haldane, 1939). The QUIZ method gives slightly lower estimates of MPN, reflecting the slight bias in the standard method even with larger numbers of wells. Confidence intervals from the simulated method are comparable to Haldane's method, though the range is slightly reduced with the simulated method.

The second four cases are from an extreme dilution series – three 10-fold dilutions with five wells per dilution. Haas (1989) and Klee (1993) report estimates of MPN by two methods – the standard ML estimate and the bias free method given by formula [1]. The QUIZ method gives very similar estimates to the bias free method in three of the four cases, and lies between the ML and bias free estimates in the fourth. This validates the use of the QUIZ method. However, compared to the magnitude of the 95% confidence intervals, all biases are trivial: the cases considered here have too low replication to be considered seriously for QUIZ.

We have tabulated c.i.'s from Klee (1993) for two methods: a "standard" method and a method due to Loyer and Hamilton (1984). We have also simulate confidence intervals based on the QUIZ estimate. The Loyer and Hamilton method is based on the distribution of the observed pattern of results. Alternatives (eg Woodward 1957) use the distribution of the MPN itself. The Loyer and Hamilton method is stated to be more robust, give narrower confidence intervals, but to lead to "choppy" confidence bands – that is they tend to change in jumps as the pattern of zeros changes (Haas 1989). Garthright and Blodgett (2003) caution against its use for regulatory purposes. For the (4,2,0) and (5,4,1) cases, the simulated c.i.'s are very similar to the

Loyer and Hamilton values. For the (5,5,2) case, all three methods reported here give equally plausible but different results. The simulated c.i. has the smaller range. For the (5,5,3) case, the lower confidence limits from the Loyer and Hamilton and simulation method are similar, but the upper value from the simulations is very high. This is because >2.5% of the simulations have outcomes of (5,5,5) with an estimated MPN of 184.86. (NB the likelihood is very flat in this case and different estimates can be obtained with different settings of the maximisation procedure, nlme, within R). This could be regarded as a deficiency in the simulation method. However, we regard it as a clear indication that this small dilution series is inadequate for the task of estimating MPN at this concentration of DNA. A similar situation is possible with very low GM levels – the lower confidence limit could be zero since many simulated outcomes will detect no GM. This limit is misleading, since GM has been detected. In these circumstances, it is better to regard the lower limit as 0 + δx with x as some arbitrarily low value. The simulated c.i. are more honest. If desired, simulated results can easily be parsed to generate confidence intervals conditional on a least one positive or at least one negative result being detected, though in such cases, the estimates and confidence intervals will be biased.

More work is clearly possible in comparing existing methods of c.i. estimation with the simulations. We have not pursued this. Firstly, with more realistic numbers of wells our proposed methods performs well. Secondly, of much greater importance is the estimation of c.i. for GM%.

Confidence intervals of estimates and rejection/acceptance thresholds. (2) GM%

An evaluation of confidence intervals for estimates of % GM derived as scaled ratios between two MPNs does not exist. Lee et al. (2008), in an initial study of QUIZ, gave the following formula for the variance of %GM (not scaled for differences in dilution):

$$\text{Var(ratio)} \approx \left(\left(\frac{p_l q_l}{N_1} \right) \left(\frac{1}{p_l} \right)^2 \left(\frac{\mu_g}{\mu_l} \right)^2 + \left(\frac{p_g q_g}{N_2} \right) \left(\frac{1}{p_g} \right)^2 \right) \left(\frac{1}{\mu_l} \right)^2$$

However, there are some problems with the application of this formula. Firstly, the distribution of zeros is assumed to follow a binomial distribution (as for all other MPN estimation methods). However, more strictly it follows a hypergeometric distribution reflecting the sampling without replacement inherent in QUIZ, though the formula could be modified to account for this. Secondly, the distribution of GM% is non-normal and non-symmetric about its estimate, which renders the estimated variance of limited value in constructing a c.i. Finally, the variance does not take into account the bias in the estimate of GM%. For these reasons, we think the best approach is to simulate confidence intervals in the same manner described for MPN. Table 10

presents simulated confidence intervals for true GM% values of 0.5% and 0.9% at a range of concentrations and dilutions of reference molecule. 10,000 simulations were carried out. Results are presented only for the QUIZ method. The standard method gives similar results, but at high numbers of molecules, estimation can frequently fail.

Concentrations and dilutions were selected to illustrate the magnitude of the confidence intervals to be expected using QUIZ, rather than to search for trends. Nevertheless, some generalisations can be made. Firstly, the 95% confidence intervals are acceptably small for both of the simulated GM%, though they are narrower at 0.5% GM. If used to test for statistical significance of differences from thresholds, 90% rather than 95% c.i. might be more appropriate. In this case we may be interested only in testing whether the observed result is significantly higher than the threshold. Limits would therefore be even narrower. The c.i. also vary little from a 50-fold dilution of reference to a 200-fold dilution though 10-fold dilutions are less accurate.

The median of the simulated estimates of GM% can be used as a test of the robustness of the method. Generally, this is close to its expected value. However, at high concentrations of the GM molecule per well, the method breaks down. At the higher concentrations of three and four molecules per well of reference molecule, the method is reasonably robust even though, at these dilutions, with 96 wells, an increasing number of simulations will generate no negative PCRs. We would advise, therefore, that dilutions should aim for between 1 and 2 molecules per well for both GM and reference. A more exhaustive search might provide more accurate guidelines.

It is informative to study the power to reject samples as lying below or above the current statutory thresholds of 0.5% GM and 0.9% GM. For this, we have fixed the dilution of reference relative to GM at 100, and allowed GM% to vary from 0.2 to 1.55. We can then simulate and compute the proportion of cases in which the estimated GM% is greater than the threshold. We have done this for expected numbers of reference molecules per well of 0.5, 1, 2, and 3. Results for the QUIZ method are in Figure 2. Results from the standard method, not presented, are very similar. The power curves are very similar for all concentrations of the reference molecule, in line with the results for confidence intervals at varying dilutions. Power curves for 0.5% and 0.9% GM are well separated. Those for 0.5% GM rises more steeply, again in line with the small confidence intervals seen at lower percentage GM. Thresholds could be altered to reduce the risk of falsely rejecting or accepting samples to any specified value. Precision could be increased by increasing the sample size, from 96 to 384 for example, or by introducing a sequential testing strategy in which samples which cannot be confidently accepted or rejected on the existing data are retested. Such strategies are common in clinical trials and have also been advocated in germination testing (Whitehead 1981). They provide increased precision for a lower cost than simply increasing scale.

We can also assess the probability of accepting or rejecting samples at the initial, serial dilution, phase of QUIZ. Taking the same sets of simulations used to produce Table 7, we can count the number of simulations which

exceed the thresholds of 0.5% and 0.9% GM. Arbitrarily, we accept a sample if 99% of simulations have values which are less than or equal to the threshold and reject a sample if 99% of simulations have values which are greater than the threshold. Results are in Table 11. For the systematic set of concentrations used, roughly half the samples can be rejected or selected at this stage. Since one would hope that the majority of real samples to be tested will fall comfortably within the statutory limits, is seem likely that many samples will never require advancing to full testing in 96 or more wells.

Software

Two small programs have been developed for routine analysis: *QUIZest.R* and *dil_series.R*. Software has been written to work under R, which is free and available from http://cran.r-project.org/.

Running R.

The windows version of R should install with no problems. There are a number of excellent guides available to the package, for example Dalgaard (2002) and much information is available from the R CRAN web site. Here, we simply describe the bare minimum required to produce results from the supplied scripts. To run R under windows, click on the R icon, or select R from the Start/Programs menu.

Firstly, you will need to change to the directory in which the R scripts and data files are stored. Select the “File” pull down menu from the top left of screen and then “Change dir” then ”Browse” then select the required directory from the pop-up window in the usual manner. Then click OK.

To run one of the QUIZ scripts, select “File” then “Source R code”, then double click on the desired program.

QUIZest.R

This script estimates MPN for GM and reference, and hence GM%, for data from a single dilution. Data for analysis must be in the text file *QUIZest_data.txt*. R is case sensitive so the file names need to be an exact match. An example data file is given below

```
10      # observed positives in the GM plate
41      # observed postives in the reference plate
96      # total number of PCR wells, GM
96      # total number of PCR wells, reference
100     # dilution of reference relative to GM
100000  # number of simulations for c.i. ests. 10,000 is OK if your computer is slow.
```

In practice, the most simple procedure is to edit this file as required. The numbers in the first column are the required data. Anything following the # symbols is ignored by the program and provides a description of the data. Data must be in exactly this order. Do not put the number of simulations first, for example.

Note that it is the number of POSITIVE wells that the program expects.

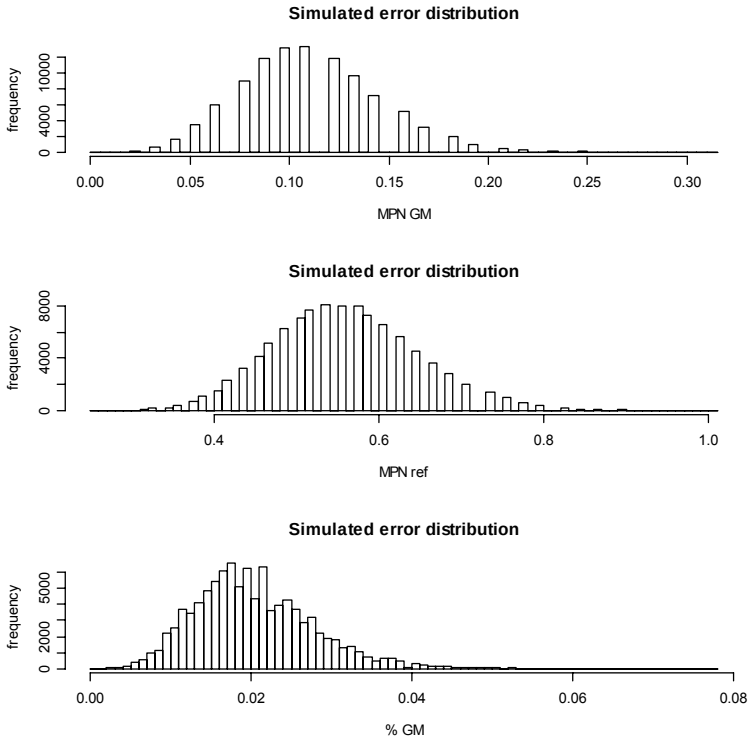
Selecting and running the example program with the example file above should give the following output in the console window within R.

```
[1]     Input data
[1] Observed positives, GM: 10
[1] Observed positives, ref: 41
[1]   Dilution of ref:GM: 1000
[1] Number of simulations : 100000
[1]
[1]
[1]  Estimated values
[1]         MPN GM: 0.109
[1]         MPN ref: 0.553
[1]     Estimated % GM: 0.02
[1]
[1]
[1]   Simulation results
[1]
[1]   Quantiles for GM 95% c.i.
   0%  2.5%   50%  97.5%  100%
0.00000 0.00822 0.01956 0.03715 0.07792
[1]
[1]   Proportion exceeding thresholds
[1]
[1] >0.5% 0
[1] >0.9% 0
```

Output first are the input data, then the estimates of MPN and GM%. Simulations are carried out with these estimates to produce empirical quantiles for GM%. The median value should be close to the estimated GM%. If not, there is a problem. The 2.5% and 97.5% quantiles provide the lower and upper 95% confidence limits respectively. Finally, the proportion of simulated values exceeding thresholds of 0.5% and 0.9% GM are given – these can be used to test if the estimated value is significantly greater or smaller than the threshold.

The script can easily be edited to print additional output if required – for example c.i. for the MPN estimates may be required.

In addition to the numerical output, graphs of the simulated distribution of the MPN and the GM% is produced. If R's graphical window is selected with the mouse, these graphs can be copied and pasted to other applications or can be saved by selecting "File" then "Save as". It is worth inspecting the graphs of MPN, since very high or very low numbers of observed zeros can give particularly "choppy" distributions which may indicate some caution is required in interpreting the estimate of GM% and its empirical confidence limits.



The current version of the script does not write numerical results to a file. This can be easily altered if required.

dil_series.R

This script estimates MPN for GM and reference, and hence GM%, for data from a series of dilutions using the QUIZ method. It can also be used to estimate from a single dilution, but *QUIZest.R* is better used for that. It is possible that the methods and algorithms for the two may diverge in the future. Data for analysis must be in the text file *dil_series.txt*. An example file is given below.

```
30              # pot_vol             Size of pot at each dilution.
1               # PCR_vol             Volume of alequot for PCR: volume in well.
10              # no_PCRs             No. of PCR tests at each dilution.
100             # lectin_ratio        Initial dilution (if any) of reference.
6               # no_dils             No. of dilutions, including the intial (min=1).
10              # dil_factor          Dilute ratio to create the series.
10 9 1 0 0 0    # GM                  GM data        NB - NO. OF POSITIVES READ IN.
10 10 10 2 0 0 # lectin               Ref. data      NB - NO. OF POSITIVES READ IN.
1000            # no_iterations       No. of sims for c.i. ests. etc. Min. 1,000
```

In practice, the most simple procedure is to edit the data in this file as required.

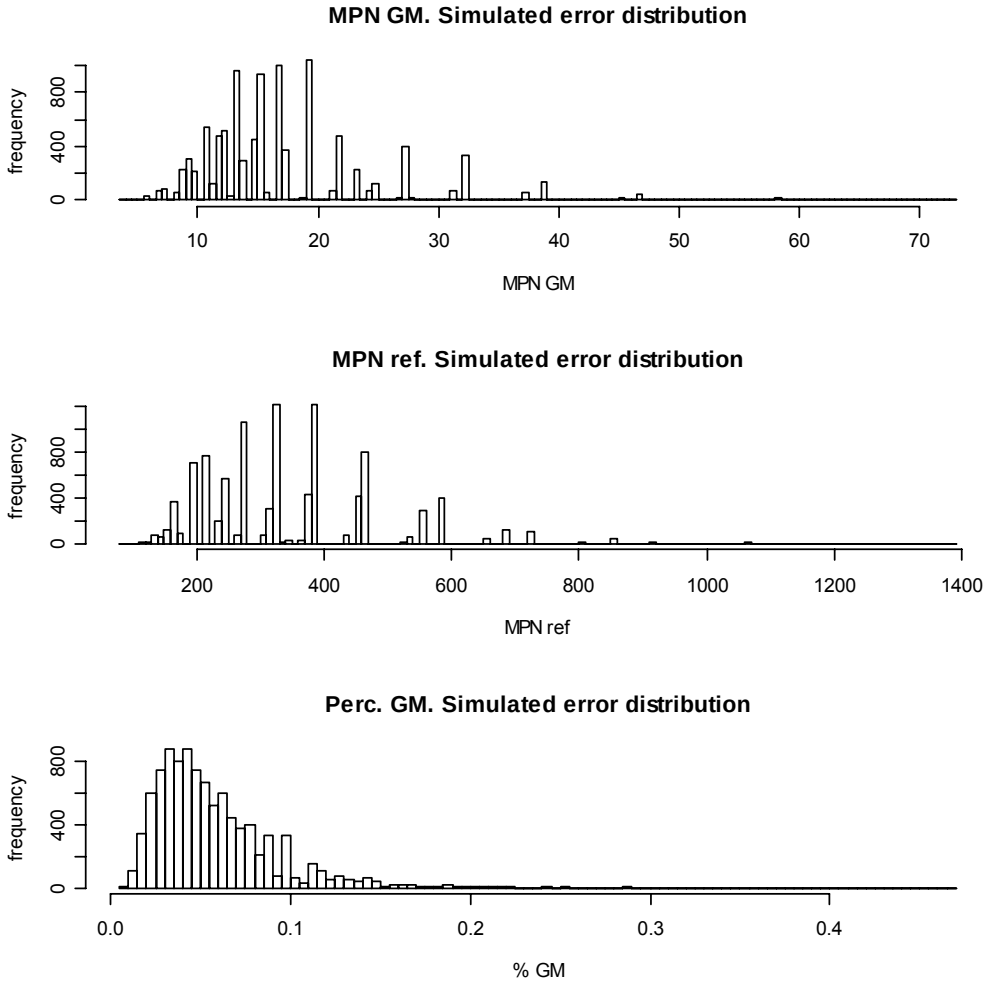
No. of simulations should be at least 1000. The program will then take about 10 seconds to run. Low numbers can be used initially to check the data file format. 10,000 is a reasonable number for reporting results.

Note that it is the number of POSITIVE wells that the program expects.

Selecting and running the example program with the example file above should give the following output in the console window within R.

```
[1]
[1]     Input data
[1] Observed positive GM :
[1] 10 9 1 0 0 0
[1] Observed positive ref :
[1] 10 10 10 2 0 0
[1]   Dilution of ref:GM: 10
[1]  Number of dilutions: 6
[1] Number of wells/diln : 10
[1]    Dilution factor: 10
[1]  Initial diln of ref: 100
[1]
[1] "Confirm that maximization OK"
[1] "Error codes should be 1:"
[1] GM : 1
[1] ref : 1
[1]
[1] "Estimates of MPN and GM%"
[1] GM molecules per well: 17
[1] ref molecules per well: 322
[1]     GM percentage: 0.1
[1]
[1] " Simulations results "
[1] Number of simulations : 10000
[1] "Maximum error code should be 1:"
[1] max error code : 1
[1]
[1]   Quantiles for MPN GM, MPN ref, %GM
[1]     0%  2.5%  50%  97.5%  100%
[1] GM  : 5.5  7.3  15.2  38.5  55.4
[1] ref : 91.8  151.8 321.7 729.7 1159.8
[1] % GM : 0   0   0   0.2  0.3
[1]
[1]
[1]   Proportion exceeding thresholds
[1]
[1] >0.5% 0
[1] >0.9% 0
[1]
```

In addition to the numerical output, graphs of the simulated distributions of MPN and GM% are produced.



The long tails of MPN and GM% distributions are typical and result from using series with small numbers of wells. The choppy distribution of MPN is also typical and has been commented on (Haas 1989). A consequence is that estimates of high GM% should preferably be confirmed in a larger scale experiment. Results from this early phase are, however, often suitable to pass samples.
The current version of the script does not write numerical results to a file. This can be easily altered if required.

Additional work

There are a number of areas where additional work is possible:

1. *Error checking*

Improbable data from dilution series can indicate errors. For example: the series:
9, 10, 10, 2, 0, 0 at 10-fold dilutions gives an estimate of 15 molecules / per well, but the pattern of results strongly suggests that the blank observed at the highest dilution is a false negative. If this were so, the estimate of MPN becomes 322. Equally, the pattern

10, 10, 10, 1, 0, 1 also seems unlikely, suggesting that the 1 in the last dilution is a false positive yet this gives a high estimate of MPN of 314. It is likely that false positives at low dilutions are less damaging to the analysis than false negatives at high concentrations.

These potential errors might be detected in various ways. Firstly for the estimated MPN, the distribution of possible dilution patterns can be studied. If the observed pattern is relatively rare, this is indicative of an error (reviewed in Garthright and Blodgett, 2003). Secondly, each observation could be dropped and the effect on both estimated MPN and the likelihood could be assessed. A big change in likelihood and/or estimated MPN could indicate an error. (But not necessarily at that data point.) Equivalently, changes in likelihood and/or MPN by flipping the results from single wells (0→ 1 and 1 → 0) may also identify errors.

Finally, it may be possible to include error rates in the estimation procedure. Such error rates could be supplied - for example an error rate of 0.001 per well could be assumed - but it may be possible to estimate the error rate by maximizing the likelihood for MPN and error rate simultaneously. Greater levels of sophistication could include differing positive and negative error rates.

One would expect, of course, that experimental errors are rare.

2. Experimental design

Only limited optimisation of experimental design has occurred so far. Most effort so far has gone into testing the robustness method, controlling for bias, and developing computer scripts. Practical expediency has restricted experimental methods to six ten-fold dilutions with ten tests at each dilution. Alternative dilution series using might prove more efficient and have not been tested exhaustively. At the second stage of testing, we have restricted simulation study to single 96 well plates. The sequential testing methods, discussed earlier, may be worth studying.

3. Computing

The R programs could be developed further:

1) Alter the programs to handle variable numbers of wells at each dilution: important for dealing with missing data.
2) Include data checking before running.
3) Simplified data input.
4) Batch runs for multiple data sets.
5) Testing for errors: see above.
6) Incorporate scripts into an R package for more general release
7) Develop a GUI version.
8) Offer variable dilutions at each stage of the dilution series.

Of these, (1) is the most important.

Conclusions & Summary

QUIZ offers an alternative to qPCR for the assessment of GM%, without the requirement for CRM. QUIZ is based on the ratio of two estimates of Most Probable Number (MPN), one for the GM target DNA and one for a crop specific reference DNA. QUIZ also provides confidence intervals and probabilities that sample GM% exceed statutory thresholds. In the application of MPN statistical methods to QUIZ, we have developed a very simple modification to the standard estimate of MPN which gives good correction for bias. The estimate of GM% as the ratio of two (unbiased) MPN is itself strongly positively biased. However, the log of GM% and the median of simulated datasets show little bias. This provides a simple means for pooling data from multiple QUIZ experiments. Simple R scripts have been developed for the estimation of MPN, GM%, confidence intervals, and the probability of exceeding thresholds. These could be developed further if required.

References

Dalgaard P (2002) Introductory Statistics with R. *Springer*, New York.

Garthright WE, Blodgett RJ (2003). FDA's preferred MPN methods for standard, large or unusual tests, with a spreadsheet *Food Microbiology* **20** 439-445.

Haas CN (1989). Estimation of microbial densities from dilution count experiments. *Appl Environ Microbiol*. 55 1934-42.

Haldane JBS (1939). Sampling errors in the determination of bacterial or virus density by the dilution method. *J Hygiene* **39** 289-293

Klee AJ (1993). A computer program for the determination of most probable number and its confidence limits. *Journal of Microbiological Methods* **18** 91-98

Lee D, La Mura M, Greenland A and Mackay I (2008). Quantitation using informative zeros (QUIZ): application for GMO detection and quantification without recourse to certified reference material. *Food Chemistry* (in press).

Loyer MW, Hamilton MA (1984). Interval estimation of the density of organisms using a serial-dilution experiment. *Biometrics* **40** 907-916

Salama IA, Koch GG, Tolley DH (1978). On the estimation of the most probable number in a serial dilution experiment. *Communications in Statistics - Theory and Methods* **7** 1267-1281.

Teunis PFM, Evers EG, Slob W (1999). Analysis of Variable Fractions Resulting from Microbial Counts. *Quantitative Microbiology* **1** 63-88

Whitehead J (1981). The Use of the Sequential Probability Ratio Test for Monitoring the Percentage Germination of Accessions in Seed Banks *Biometrics* **37** 129-136

Woodward RL (1957). How probable is the most probable number? *J. Am. Water Works Assoc.* **49** 1060-1068.

Yates F (1934). Contingency table involving small numbers and the χ^2 test. *Journal of the Royal Statistical Society (Supplement)* **1** 217-235.

Table 1 Estimation of MPN in 10 PCR wells by three methods. Means of 10,000 simulations.

true	exact	standard	QUIZ
0.005	0.005	0.005	0.005
0.01	0.010	0.010	0.010
0.1	0.100	0.106	0.100
0.2	0.202	0.214	0.202
0.3	0.299	0.320	0.300
0.4	0.401	0.431	0.403
0.5	0.499	Inf	0.498
0.6	0.596	Inf	0.597
0.7	0.702	Inf	0.703
0.8	0.799	Inf	0.801
0.9	0.899	Inf	0.899
1.0	1.000	Inf	1.001
1.1	1.104	Inf	1.100
1.2	1.197	Inf	1.198
1.3	1.295	Inf	1.297
1.4	1.403	Inf	1.402
1.5	1.501	Inf	1.497
1.6	1.603	Inf	1.590
1.7	1.702	Inf	1.674
1.8	1.796	Inf	1.772
1.9	1.903	Inf	1.869
2.0	1.996	Inf	1.935
2.5	2.496	Inf	2.287
3.0	2.992	Inf	2.550
3.5	3.503	Inf	2.740
4.0	4.002	Inf	2.850

Estimation methods:

True	Value input into simulations
Exact	Count of molecules per well
Standard	No. mols = -ln(proportion of zeros)
QUIZ	No. mols = -ln(no. of zeros + ½)/ (total + ½)
Inf	No zero wells observed, standard estimate is infinite.

Table 2. Estimation of MPN in 10 PCR wells by three methods, excluding simulations where no zeros were observed. Means of 10,000 simulations.

		Estimates			% bias		
true	Prop fail[†]	exact	standard	QUIZ	exact	standard	QUIZ
0.005	0	0.005	0.005	0.005	-2.0	3.1	-2.0
0.01	0	0.010	0.010	0.010	-4.5	0.7	-4.4
0.1	0	0.1	0.106	0.100	-0.2	5.6	0.0
0.2	0	0.2	0.214	0.202	1.2	7.1	1.1
0.3	0	0.3	0.320	0.300	-0.3	6.5	0.1
0.4	0	0.4	0.431	0.403	0.3	7.6	0.8
0.5	0.0001	0.5	0.534	0.498	-0.2	6.8	-0.4
0.6	0.0004	0.6	0.643	0.596	-0.7	7.2	-0.6
0.7	0.0011	0.7	0.760	0.701	0.2	8.6	0.1
0.8	0.0023	0.8	0.869	0.796	-0.3	8.6	-0.5
0.9	0.0052	0.9	0.975	0.888	-0.5	8.3	-1.3
1.0	0.0116	1.0	1.079	0.977	-0.7	7.9	-2.3
1.1	0.0162	1.1	1.187	1.068	-0.5	7.9	-2.9
1.2	0.0275	1.2	1.280	1.146	-1.4	6.6	-4.5
1.3	0.0410	1.3	1.373	1.223	-1.9	5.6	-6.0
1.4	0.0620	1.4	1.460	1.293	-2.0	4.3	-7.6
1.5	0.0795	1.5	1.547	1.363	-2.3	3.1	-9.1
1.6	0.1012	1.6	1.625	1.426	-2.6	1.6	-10.9
1.7	0.1321	1.6	1.675	1.466	-3.4	-1.5	-13.8
1.8	0.1656	1.7	1.743	1.519	-4.1	-3.2	-15.6
1.9	0.2031	1.8	1.807	1.570	-4.2	-4.9	-17.4
2.0	0.2305	1.9	1.850	1.603	-4.9	-7.5	-19.9
2.5	0.4195	2.3	2.026	1.739	-6.6	-19.0	-30.5
3.0	0.5927	2.8	2.148	1.831	-7.7	-28.4	-39.0
3.5	0.7394	3.2	2.210	1.877	-8.5	-36.9	-46.4
4.0	0.8297	3.6	2.244	1.903	-9.6	-43.9	-52.4

Estimation methods:

True	Value input into simulations
Exact	Count of molecules per well
Standard	No. mols = -ln(proportion of zeros)
QUIZ	No. mols = -ln(no. of zeros + ½)/ (total + ½)

[†] Prop fail is the proportion of simulations for which no estimate of MPN was possible by the standard method – because results were 100% positive.

Table 3. Estimation of MPN in 96 PCR wells by three methods, excluding simulations where no zeros were observed. 10,000 simulations.

	All simulations				Simulations with >0 zero counts only					
	estimates				estimates			bias (%)		
true	exact	standard	QUIZ	Prop. fail[†]	exact	standard	QUIZ	exact	standard	QUIZ
0.0001	0.0001	0.0001	0.0001	0	0.0001	0.0001	0.0001	-11.5	-11.0	-11.5
0.001	0.001	0.001	0.001	0	0.001	0.001	0.001	-2.4	-1.8	-2.3
0.005	0.005	0.005	0.005	0	0.005	0.005	0.005	0.3	0.9	0.3
0.01	0.010	0.010	0.010	0	0.010	0.010	0.010	1.5	2.0	1.5
0.1	0.100	0.101	0.100	0	0.100	0.101	0.100	0.3	1.0	0.4
1	1.000	1.010	1.001	0	1.000	1.010	1.001	0.0	1.0	0.1
1.5	1.501	1.520	1.501	0	1.501	1.520	1.501	0.1	1.3	0.1
2	1.999	2.036	2.001	0	1.999	2.036	2.001	0.0	1.8	0.0
2.5	2.500	Inf	2.503	0.0002	2.500	2.568	2.502	0.0	2.7	0.1
3	2.999	Inf	3.001	0.0076	2.998	3.102	2.984	-0.1	3.4	-0.5
3.5	3.500	Inf	3.492	0.0554	3.494	3.577	3.390	-0.2	2.2	-3.1
4	4.001	Inf	3.964	0.1689	3.986	3.952	3.695	-0.3	-1.2	-7.6
5	5.003	Inf	4.656	0.5211	4.968	4.340	3.996	-0.6	-13.2	-20.1
6	6.003	Inf	5.017	0.7883	5.941	4.481	4.103	-1.0	-25.3	-31.6

Estimation methods:

True	Value input into simulations
Exact	Count of molecules per well
Standard	No. mols = -ln(proportion of zeros)
QUIZ	No. mols = -ln(no. of zeros + ½)/ (total + ½)

[†] Prop fail is the proportion of simulations for which no estimate of MPN was possible by the standard method – because results were 100% positive.

Table 4. Estimation of MPN in 6 ten-fold dilutions of 10 PCR wells per dilution. Results of 10,000 simulations

		QUIZ estimate			Standard method			Percentage bias
mols / ul	No. successful estimates	mean	s.d.	No. successful estimates	mean	s.d.	QUIZ	standard
0.1	10000	0.10	0.077	10000	0.10	0.081	-0.4	3.5
0.4	10000	0.40	0.169	10000	0.42	0.187	-1.2	5.4
0.7	10000	0.69	0.250	10000	0.74	0.281	-1.7	5.7
1	10000	1.0	0.34	10000	1.1	0.39	-1.9	7.3
4	10000	4.3	1.69	10000	4.5	1.75	7.3	12.9
7	10000	7.1	2.72	10000	7.6	3.06	1.3	8.6
10	10000	10	4	10000	11	4	-0.8	7.9
40	10000	43	17	10000	45	18	7.5	12.9
70	10000	71	28	10000	76	31	1.6	8.2
100	10000	101	40	10000	108	44	0.5	7.7
400	10000	427	170	10000	451	182	6.6	12.7
700	10000	712	278	10000	756	307	1.7	8.0
1000	10000	990	394	10000	1085	444	-1.0	8.5
4000	10000	4316	1740	10000	4581	1896	7.9	14.5
7000	10000	7167	2819	10000	7659	3164	2.4	9.4
10000	10000	10086	4003	10000	10999	4548	0.9	10.0
40000	9999	46550	20389	10000	49408	21901	16.4	23.5
70000	9974	79421	32595	9978	86322	38822	13.5	23.3
100000	9817	108746	39595	9831	120044	48376	8.7	20.0
400000	1121	192146	11026	1116	226718	15266	-52.0	-43.3
700000	43	194591	0	40	230258	0	-72.2	-67.1
1000000	2	194591	0	2	230258	0	-80.5	-77.0

Estimation methods:

Standard	No. mols = -ln(proportion of zeros)
QUIZ	No. mols = -ln(no. of zeros + ½)/ (total + ½)

Table 5 Estimation of MPN in 96 PCR wells, QUIZ method, with differing dilutions series. 10,000 simulations.

no. wells	10	6	30	3	20	2	15	4	12	5
no. dilutions	6	10	2	20	3	30	4	15	5	12
dil factor	10.0	3.6	100000	1.8	316.2	1.5	46.4	2.3	17.8	2.8
ESTIMATE										
mols/ul										
1	1.0	1.0	1.0	0.9	1.0	0.9	1.0	0.9	1.0	0.9
4	4.3	4.0	9.2	3.9	6.7	4.0	5.0	3.9	4.5	4.0
7	7.1	6.9	19.6	7.0	10.5	7.1	7.9	6.8	7.4	6.9
10	9.9	10.0	21.1	9.9	13.1	10.1	10.5	9.8	10.0	9.9
40	42.8	40.2	57.7	39.4	41.0	40.5	39.9	39.4	42.7	39.7
70	71.2	69.8	72.7	69.5	70.6	71.2	70.7	69.0	78.5	69.0
100	99.3	99.2	103.5	98.7	100.3	100.9	108.7	98.1	107.8	99.8
400	428.9	402.2	416.9	396.0	404.2	402.1	430.5	393.9	400.6	395.0
700	708.6	698.0	692.7	692.8	653.9	708.1	701.4	688.4	744.3	690.5
S.D.										
mols/ul										
1	0.4	0.4	0.2	0.5	0.3	0.5	0.3	0.4	0.4	0.4
4	1.7	1.6	29.6	1.5	5.6	1.6	2.5	1.5	1.9	1.5
7	2.8	2.7	165.9	2.6	8.1	2.7	4.0	2.6	3.1	2.7
10	3.9	3.9	178.7	3.8	10.4	3.8	5.3	3.8	4.1	3.8
40	17.2	15.8	395.0	14.7	25.7	15.4	14.2	14.9	19.9	15.4
70	27.4	27.5	449.1	26.8	34.9	26.7	28.9	26.5	34.5	26.8
100	39.7	39.0	545.2	37.4	43.7	38.6	57.4	37.9	45.5	38.8
400	171.3	158.9	1173.6	151.0	147.9	149.5	218.8	150.2	153.1	153.2
700	272.5	270.9	1500.4	262.9	810.6	266.4	328.1	265.8	348.6	265.5
PERCENTAGE BIAS										
mols/ul										
1	-1.2	-3.6	0.1	-8.2	0.1	-6.4	0.1	-7.7	-0.5	-5.9
4	7.2	-0.5	131.0	-1.5	67.0	0.2	23.8	-2.3	12.2	-1.2
7	2.0	-1.0	179.8	-0.6	50.4	1.0	12.7	-2.4	5.1	-1.1
10	-0.5	0.1	111.3	-0.9	31.5	1.1	5.5	-1.6	0.3	-1.0
40	6.9	0.4	44.3	-1.5	2.4	1.2	-0.2	-1.5	6.6	-0.6
70	1.7	-0.3	3.8	-0.7	0.8	1.7	0.9	-1.4	12.2	-1.4
100	-0.7	-0.8	3.5	-1.3	0.3	0.9	8.7	-1.9	7.8	-0.2
400	7.2	0.6	4.2	-1.0	1.1	0.5	7.6	-1.5	0.1	-1.3
700	1.2	-0.3	-1.0	-1.0	-6.6	1.2	0.2	-1.7	6.3	-1.4
CV %										
1	41.1	45.2	23.8	50.5	30.7	52.7	34.8	48.1	38.4	46.2
4	39.9	39.0	320.7	39.0	84.1	39.4	49.6	38.8	43.0	38.9
7	38.9	39.3	847.2	37.8	77.4	38.4	50.2	38.4	42.1	38.9
10	39.5	39.0	845.3	38.8	79.2	37.7	50.0	38.6	41.3	38.2
40	40.1	39.3	684.4	37.2	62.8	38.0	35.6	37.9	46.6	38.7
70	38.5	39.5	618.2	38.6	49.5	37.5	40.9	38.5	43.9	38.8
100	40.0	39.3	526.6	37.9	43.6	38.2	52.8	38.7	42.2	38.9
400	39.9	39.5	281.5	38.1	36.6	37.2	50.8	38.1	38.2	38.8
700	38.5	38.8	216.6	38.0	124.0	37.6	46.8	38.6	46.8	38.5

Table 6. Estimation of GM% in 96 PCR wells by three different methods, excluding simulations for which % GM could not be estimated. Maximum 100,000 simulations.

dilution	100	100	100	100	100	100	100	100	100
Molecules GM	1	0.1	2	0.2	1	1	1	2	3
Molecules reference	0.1	1	0.2	2	1	2	3	1	1
E GM%	10	0.1	10	0.1	1	0.5	0.333	2	3
GM% QUIZ method	11.43	0.10	10.64	0.10	1.01	0.51	0.34	2.04	3.06
GM% standard method	11.43	0.10	10.76	0.10	1.02	0.50	0.33	2.06	3.14
GM% exact method	11.37	0.10	10.58	0.10	1.01	0.50	0.33	2.02	3.03
RMSE QUIZ method	11.84	0.037	3.288	0.028	0.197	0.096	0.069	0.387	0.644
RMSE standard method	11.88	0.037	3.372	0.028	0.198	0.096	0.069	0.399	0.683
RMSE by counting	11.67	0.034	2.912	0.024	0.148	0.063	0.040	0.256	0.365

Estimation methods:

Exact	Count of molecules per well
Standard	No. mols = -ln(proportion of zeros)
QUIZ	No. mols = -ln(no. of zeros + ½)/ (total + ½)

RMSE = $\sqrt{\sum (estimate - truevalue)^2 / n}$

Table 7. Estimation of GM% in 6 ten-fold dilutions of 10 PCR wells per dilution. Results of 10,000 simulations.

simulation values			QUIZ		standard		% Bias	
GM no. / ul	reference no. / ul	% GM	GM%	s.e.	GM%	s.e.	corrected	standard
0.1	10	1.000	1.147	1.043	1.111	1.030	14.7	11.1
0.1	40	0.250	0.269	0.246	0.263	0.238	7.7	5.2
0.1	70	0.143	0.161	0.146	0.157	0.143	12.5	9.9
0.4	10	4.000	4.545	2.712	4.539	2.844	13.6	13.5
0.4	40	1.000	1.067	0.642	1.072	0.643	6.7	7.2
0.4	70	0.571	0.638	0.385	0.641	0.386	11.6	12.1
0.7	10	7.000	7.987	4.420	7.978	4.579	14.1	14.0
0.7	40	1.750	1.869	1.020	1.883	1.033	6.8	7.6
0.7	70	1.000	1.114	0.599	1.127	0.629	11.4	12.7
1	100	1.000	1.134	0.607	1.150	0.646	13.4	15.0
1	400	0.250	0.265	0.147	0.272	0.146	6.2	8.6
1	700	0.143	0.159	0.083	0.164	0.090	11.1	14.8
4	100	4.000	4.942	2.783	4.875	2.824	23.5	21.9
4	400	1.000	1.161	0.677	1.151	0.650	16.1	15.1
4	700	0.571	0.692	0.384	0.697	0.408	21.1	21.9
7	100	7.000	8.234	4.609	8.214	4.920	17.6	17.3
7	400	1.750	1.922	1.101	1.943	1.122	9.9	11.0
7	700	1.000	1.153	0.643	1.174	0.695	15.3	17.4
0.1	100	0.100	0.115	0.105	0.112	0.104	15.0	11.6
0.1	400	0.025	0.027	0.024	0.026	0.024	6.9	5.5
0.1	700	0.014	0.016	0.015	0.016	0.014	12.2	10.8
0.4	100	0.400	0.455	0.274	0.453	0.282	13.8	13.4
0.4	400	0.100	0.106	0.065	0.107	0.064	6.2	7.1
0.4	700	0.057	0.063	0.037	0.065	0.040	11.1	13.2
0.7	100	0.700	0.796	0.435	0.798	0.455	13.7	14.0
0.7	400	0.175	0.187	0.104	0.188	0.102	6.6	7.4
0.7	700	0.100	0.111	0.060	0.113	0.062	11.5	13.2
1	1000	0.100	0.114	0.060	0.115	0.064	13.8	14.7
1	4000	0.025	0.027	0.015	0.027	0.014	6.2	8.5
1	7000	0.014	0.016	0.008	0.016	0.009	10.3	13.6
4	1000	0.400	0.497	0.284	0.488	0.286	24.3	21.9
4	4000	0.100	0.116	0.067	0.115	0.065	15.8	15.0
4	7000	0.057	0.069	0.040	0.069	0.040	20.9	20.6
7	1000	0.700	0.826	0.460	0.821	0.495	18.0	17.3
7	4000	0.175	0.192	0.109	0.194	0.112	9.7	11.0
7	7000	0.100	0.114	0.063	0.116	0.069	14.5	15.9

Estimation methods:

Standard	No. mols = -ln(proportion of zeros)
QUIZ	No. mols = -ln(no. of zeros + ½)/ (total + ½)

Table 8. Estimation of GM% in 6 ten-fold dilutions of 10 PCR wells per dilution by four methods. Results of 10,000 simulations.

GM no. / ul	reference no. / ul	true GM%	QUIZ GM%	standard GM%	QUIZ GM% (log)	standard GM%(log)	GM% median QUIZ	GM% median standard
0.7	10	7.000	7.987	7.978	6.945	6.879	6.994	6.950
0.7	40	1.750	1.869	1.883	1.626	1.638	1.652	1.654
0.7	70	1.000	1.114	1.127	0.973	0.974	0.991	0.992
1	100	1.000	1.134	1.150	0.998	1.001	0.999	0.999
1	400	0.250	0.265	0.272	0.232	0.238	0.231	0.239
1	700	0.143	0.159	0.164	0.140	0.143	0.144	0.146
4	100	4.000	4.942	4.875	4.308	4.226	4.356	4.198
4	400	1.000	1.161	1.151	1.002	1.004	0.999	0.999
4	700	0.571	0.692	0.697	0.604	0.604	0.607	0.604
7	100	7.000	8.234	8.214	7.183	7.067	7.100	6.914
7	400	1.750	1.922	1.943	1.671	1.679	1.674	1.704
7	700	1.000	1.153	1.174	1.008	1.010	0.999	0.999
0.7	100	0.700	0.796	0.798	0.694	0.688	0.699	0.695
0.7	400	0.175	0.187	0.188	0.161	0.163	0.165	0.165
0.7	700	0.100	0.111	0.113	0.097	0.098	0.099	0.099
1	1000	0.100	0.114	0.115	0.100	0.100	0.100	0.100
1	4000	0.025	0.027	0.027	0.023	0.024	0.023	0.024
1	7000	0.014	0.016	0.016	0.014	0.014	0.014	0.014
4	1000	0.400	0.497	0.488	0.433	0.422	0.435	0.420
4	4000	0.100	0.116	0.115	0.100	0.100	0.099	0.099
4	7000	0.057	0.069	0.069	0.060	0.060	0.060	0.061
7	1000	0.700	0.826	0.821	0.721	0.706	0.711	0.691
7	4000	0.175	0.192	0.194	0.167	0.168	0.170	0.168
7	7000	0.100	0.114	0.116	0.100	0.100	0.099	0.099

Estimation methods:

True Value input into simulations
Standard GM% MPNs estimated by ML. GM% is the scaled ratio of the two MPNs.
QUIZ GM% ½ added to the zero category prior to ML estimation of MPN. GM% is the scaled ratio of the two MPNs.
GM% logs average GM% calculated as $e^{\frac{\sum(\ln(MPN_{gm})-\ln(MPN_{ref}))}{number\ of\ simulations}}$

yellow highlights values with absolute values of bias >5% of the true GM value

Table 9 Comparison of published and simulated confidence intervals for MPN.

	observed +ve at given dilution [1]					MPN			95% c.i.		Loyer & Hamilton		simulated	
No. wells	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	standard [1]	no bias [1]	QUIZ	l ower[1]	upper[1]	lower [1].	upper [1]	lower	upper
96	92	26	3			3.17		3.10	2.51	4.00			2.46	3.88
96	96	92	26	3		3.17		3.09	2.51	4.00			2.46	3.90
96	96	92	26	2	1	3.16		3.08	2.5	3.98			2.45	3.83
96	92	26	12			3.6		3.51	2.86	4.54			2.77	4.49
5	4	2	0			0.22	0.17	0.17	0.09	0.56	0.08	0.51	0.06	0.46
5	5	4	1			1.72	1.42	1.41	0.7	4.5	0.43	4.27	0.40	4.56
5	5	5	2			5.42	3.69	4.96	2	20	1.17	16.61	2.17	12.99
5	5	5	3			9.18	7.49	7.93	3	29	1.82	25.14	1.77	184.86

[1] From Garthright & Blodgett (2003) and Haas (1989)

Estimates of MPN

standard	estimates taken from Garthright & Blodgett (2003) and Haas (1989)
no bias	estimates taken from Garthright & Blodgett (2003) and Haas (1989)
QUIZ	estimates following method described in the text.

Estimates of c.i.

standard	standard (biased) ML estimates
Loyer and Hamilton	Loyer & Hamilton (1984)
simulated	from 10,000 simulations of the QUIZ method.

Table 10. Simulated 95% confidence intervals for 0.9% and 0.5% GM estimated by QUIZ in 96 well plates

GM% 0.9

expected no. mols./ well GM	ref.	dilution of ref.	GM%	lower c.i	median	upper c.i	range
1.8	1	200	0.9	0.64	0.90	1.30	0.66
3.6	2	200	0.9	0.60	0.89	1.43	0.83
5.4	3	200	0.9	0.50	0.82	1.14	0.64
7.2	4	200	0.9	0.50	0.72	0.92	0.42
0.9	1	100	0.9	0.62	0.89	1.33	0.71
1.8	2	100	0.9	0.63	0.90	1.27	0.64
2.7	3	100	0.9	0.58	0.90	1.35	0.77
3.6	4	100	0.9	0.58	0.91	1.45	0.87
0.45	1	50	0.9	0.58	0.90	1.37	0.79
0.9	2	50	0.9	0.61	0.90	1.28	0.67
1.35	3	50	0.9	0.60	0.91	1.30	0.69
1.8	4	50	0.9	0.59	0.93	1.37	0.78
0.09	1	10	0.9	0.33	0.87	1.67	1.33
0.18	2	10	0.9	0.47	0.89	1.48	1.01
0.27	3	10	0.9	0.51	0.90	1.42	0.91
0.36	4	10	0.9	0.53	0.92	1.46	0.93

GM% 0.5

expected no. mols./ well GM	reference	dilution of ref	%GM	lower c.i	median	upper c.i	range
1	1	200	0.5	0.35	0.50	0.72	0.38
2	2	200	0.5	0.35	0.50	0.72	0.37
3	3	200	0.5	0.32	0.50	0.77	0.45
4	4	200	0.5	0.32	0.50	0.79	0.48
0.5	1	100	0.5	0.32	0.50	0.75	0.42
1	2	100	0.5	0.34	0.50	0.71	0.37
1.5	3	100	0.5	0.33	0.51	0.73	0.39
2	4	100	0.5	0.32	0.51	0.76	0.43
0.25	1	50	0.5	0.29	0.50	0.80	0.51
0.5	2	50	0.5	0.33	0.50	0.74	0.41
0.75	3	50	0.5	0.32	0.51	0.74	0.42
1	4	50	0.5	0.33	0.52	0.76	0.43
0.05	1	10	0.5	0.10	0.48	1.06	0.96
0.1	2	10	0.5	0.20	0.49	0.90	0.71
0.15	3	10	0.5	0.24	0.50	0.86	0.62
0.2	4	10	0.5	0.26	0.50	0.86	0.60

Table 11. Proportion of simulated values in 6 ten-fold dilutions of 10 PCR wells per dilution with estimated values >0.5% FM or < 0.9 GM%. Results of 10,000 simulation. Same combinations of GM and reference DNA as Table 6. All estimation by QUIZ methods.

GM no. / ul	reference no. / ul	GM%	median GM%	prop.>0.5%	prop.>0.9%
0.1	10	1.000	0.976	0.735	0.556
0.1	40	0.250	0.233	0.148	0.023
0.1	70	0.143	0.137	0.029	0.001
0.4	10	4.000	3.973	0.994	0.978
0.4	40	1.000	0.925	0.836	0.524
0.4	70	0.571	0.565	0.894	0.576
0.7	10	7.000	6.994	1.000	0.999
0.7	40	1.750	1.652	0.981	0.867
0.7	70	1.000	0.991	0.894	0.576
1	100	1.000	0.999	0.913	0.593
1	400	0.250	0.231	0.067	0.003
1	700	0.143	0.144	0.005	0.000
4	100	4.000	4.356	1.000	0.999
4	400	1.000	0.999	0.898	0.567
4	700	0.571	0.607	0.644	0.225
7	100	7.000	7.100	1.000	1.000
7	400	1.750	1.674	0.991	0.876
7	700	1.000	0.999	0.915	0.588
0.1	100	0.100	0.098	0.007	0.000
0.1	400	0.025	0.023	0.000	0.000
0.1	700	0.014	0.014	0.000	0.000
0.4	100	0.400	0.397	0.339	0.074
0.4	400	0.100	0.092	0.000	0.000
0.4	700	0.057	0.056	0.000	0.000
0.7	100	0.700	0.699	0.737	0.323
0.7	400	0.175	0.165	0.014	0.001
0.7	700	0.100	0.099	0.001	0.000
1	1000	0.100	0.100	0.001	0.000
1	4000	0.025	0.023	0.000	0.000
1	7000	0.014	0.014	0.000	0.000
4	1000	0.400	0.435	0.394	0.082
4	4000	0.100	0.099	0.001	0.000
4	7000	0.057	0.060	0.000	0.000
7	1000	0.700	0.711	0.763	0.338
7	4000	0.175	0.170	0.019	0.001
7	7000	0.100	0.099	0.001	0.000

green ≤ 0.01 (1%) of simulated vales > threshold
orange ≥ 0.99 (99%) of simulated vales > threshold

Figure 1. Relationship between dilution factor ($log_{10}log_{10}$ scale) and maximum observed bias for different dilution series. Same simulation set as reported in Table 5.

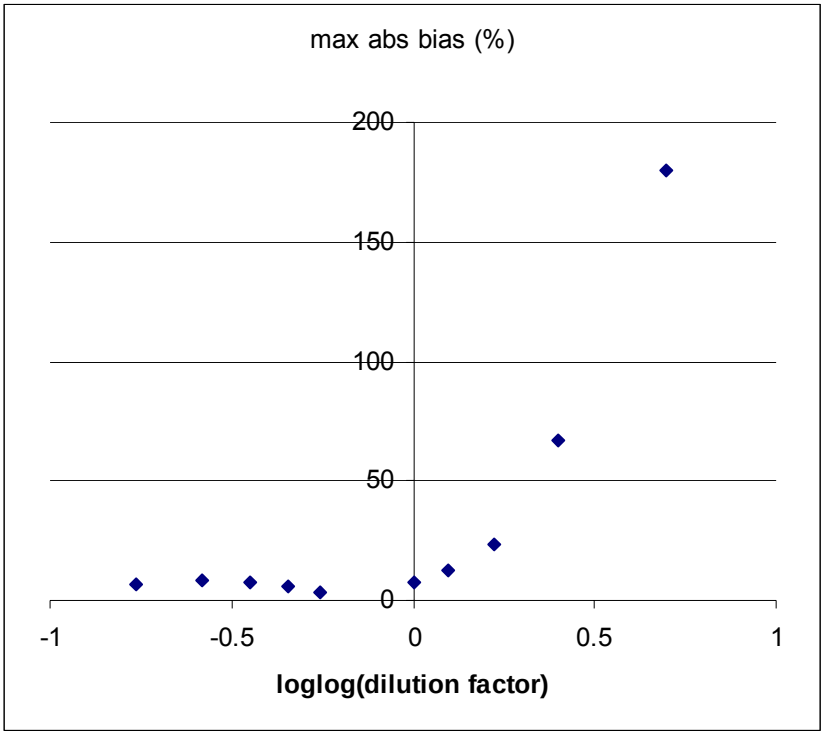
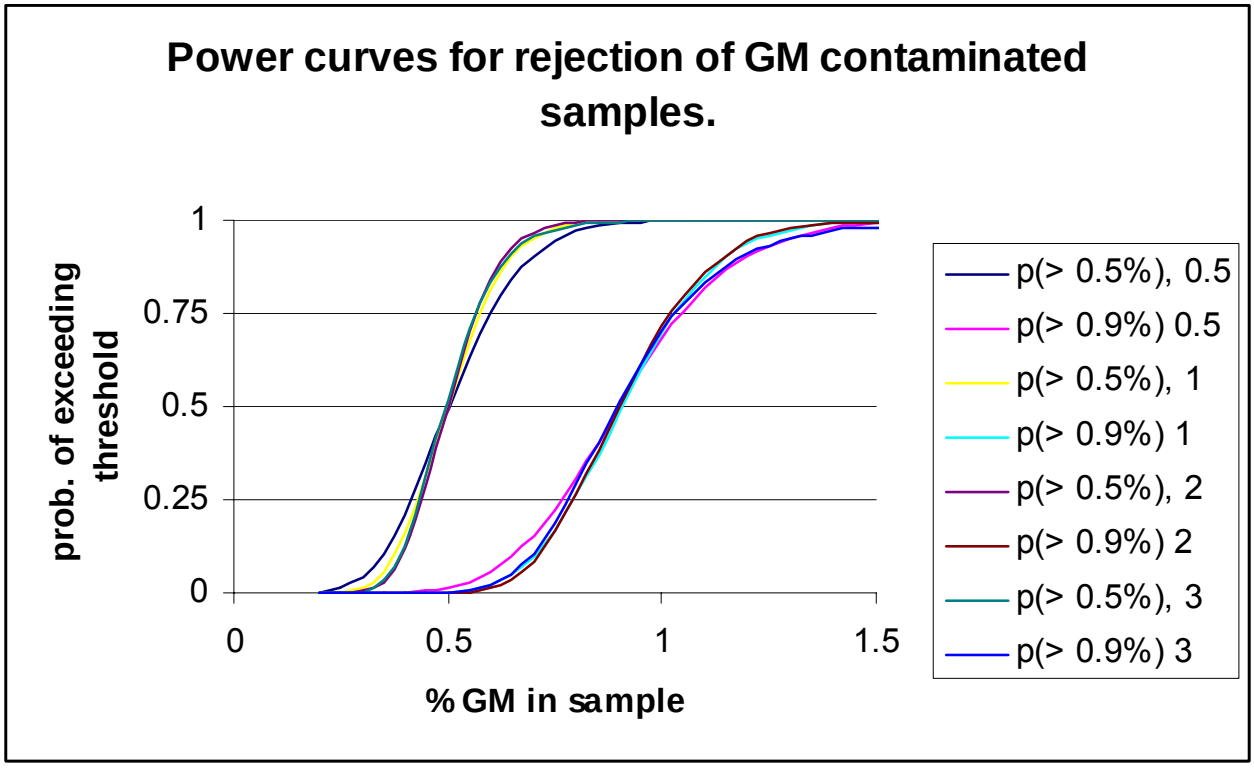


Figure 2. Probability of exceeding a threshold of 0.5% or 0.9% GM at different concentrations of reference (0.5, 1, 2, and 3 molecules per well) with GM content varied at each concentration between 0.2% and 1.55%. in steps of 0.05%. GM% estimated using the QUIZ method with 96 wells each for GM and reference MPN determination. Initial dilution of reference relative to GM of 100. Result of 100,000 simulations for each combination.



Appendix 2
GMO Quantification Using Limiting Template Detection

The application of QUIZ to GMO quantification relies on being able to determine the dilution range of the DNA template such that, when dispensed into a set of tubes, some reactions will yield negative results due to the absence of the template. This dilution for the reference gene target may be calculated if the DNA concentration of the sample and the genome size of the crop are known. However, in samples where the DNA cannot easily be measured e.g. in highly degraded DNA, and/or very small amounts of recovered template, the appropriate dilution for reference gene must be determined empirically. In the case of the GMO target, with the exception of 'threshold determination' (see below), the GMO target must always be determined by empirical means.

Empirical determination of the correct testing dilutions

The amplification of the respective targets through a serial dilution of template provides a quick and easy means of finding suitable dilutions for QUIZ. A single reaction for each serial dilution may not be sufficiently accurate to give the optimum results. For example, should we sample the dilution that gives the last positive or the first failure? A better assessment may be derived from multiple reactions at each dilution.

Since a dilution series must be performed, and that repetitions are required for determining the optimal dilutions, we explored the use of this process to derive useful data so that if a sample tested is at the extremes of GMO content, then confidence of the results obtained in this process may avoid the need for further testing. It is conceivable that for samples with less than 0.1% and greater than 5% GMO content could yield results that show that they were < or >0.9%, respectively with ≥95% confidence. Samples with lower confidences would require further, more accurate QUIZ testing.

Here we explore the use of 10 replications at each dilution both through simulated and practical data.

Threshold determination (TD)

The process of determining GMO content of a sample is more onerous than determining whether it exceeds a given threshold. To determine the GMO content it is necessary to measure and quantify the amount of each DNA molecule, reference and GMO targets, requiring empirical determination of both. TD requires determining the dilution required for the reference target. Then using an appropriate dilution of the template for a given GMO threshold. For example, using 111 times more DNA for a 0.9% threshold. If the sample exceeded 0.9% then the numbers of GMO target will be greater than the numbers of reference target; the reverse would be true for less than 0.9%. A simple count of the numbers of the two targets provides an answer as to whether the threshold has been exceeded. If the experiments yield both positives and negatives reactions, the absolute numbers of each may be used for quantification. The GMO contents where quantification is unlikely to be possible are the extremes where all GMO reactions are positive or negative within the dilutions used and therefore exceed greatly the threshold or are so low that they do not register.

Experimental approach

For this project we will only consider the quantification approach, where the testing dilutions of the GMO and reference targets are independently determined. Since it is unlikely that the DNA concentration extracted from testing material, other than CRMs, will exceed 20 ng/μl, serial dilutions over 5 orders should suffice.
GMO content determination is performed in two stages, denoted Part A and B below. We performed 10 replicates at each dilution, though for practical purposes, 8 replicates fit into a microtitre plate better.

PART A

1. Set up a reaction mixture for 150 reactions for each target without adding the primers:

H_2O	1020 μl
PCR buffer (10X)	150 μl
$MgCl_2$ (25 mM)	150 μl
dNTP's 20 mM	15 μl
Roche FastartTaq (5U/μl)	15μl

2. Split the reaction mixture into two tubes - 675 μl for testing reference genes and 675 μl for GM gene testing. Add 7.5 μl of appropriate primers pairs* (50 μM each) to each tube (one for the ref. gene and the other one for GM) and mix well.
* One primer for each marker must be labeled with a fluorophore if the amplicons are to be identified in an ABI genotyper.

3. Dispense 100 μl of reaction mix into six tubes for each marker**.** Use silanised tubes in all steps that deal with DNA to reduce DNA loss (steps 3 & 4).

4. Prepare dilutions of the template DNA
To prepare dilutions of the DNA, add 3 μl of the DNA's stock to 27 μl of water to get the first dilution (10^{-1}), repeat this process to get serial dilutions from 10^{-1}-10^{-6}.

5. Add 10 μl of each diluted DNA to a tube containing the reaction mix of step 3. Dispense 10 μl of each into 10 separate wells of microtitre plate.

6. PCR amplify using the following profile:

7 mins @ 92°C to denature template and activate enzyme
50 cycles of:
92°C for 30s
55°C for 30s
72°C for 30s
5 mins @72°C final extension
hold @15°C

7. After the PCR prepare the following mix for loading onto ABI (for 100 samples)

50 μl size standard (Rox or Liz 500)
750 μl HiDi formamide

8. Dispense 7.5 μl of the formamide mix into each well of a new 96-wells plate.

9. Mix each of the reference amplification with a GMO gene amplification together. It is simplest to mix the corresponding well from each plate.

10. Transfer 2 μl of PCR mix to the HiDi-containing plate. Load onto ABI genotyper and score data after run.

Empirical determination of QUIZ dilutions

Data from the dilutions series can be used to determine the optimum dilutions for QUIZ. Theory, verified by simulation, demonstrates that the probabilities of no DNA template within the range 0.1 to 0.6 give the lowest errors in estimates of DNA concentration. This represents approximately an average of 0.5 to 2 target molecules per reaction giving a fourfold range of DNA concentration where the error is minimal. Translated to numbers, this represents an expected 4 to 9 positives in a set of 10 reactions. Dilutions which give numbers in this range should be used for full plate QUIZ – Part B.

For values outside the range e.g. if the serial dilutions give 10, 2 and 0 positives, then it should be sufficient to double the DNA at the dilution that gives 2 positives.

PART B

Protocol for full plate QUIZ

1. Once the dilutions required have been determined for both the GMO and reference gene markers, make 2 sets of reaction mixture (100 reactions) for each test sample. To minimize DNA loss, use silanised tubes and the numbers of dilution steps should be minimised as follows:

	for dilution 10^{-2}	for dilution 10^{-1}	for dilution 10^{0}
	1 μl of stock	10 μl of stock	100 μl stock
H_2O	750 μl	740μl	650 μl
PCR buffer (10X)	110 μl	110 μl	110 μl
$MgCl_2$ (25 mM)	110 μl	110 μl	110 μl
dNTPs (20 mM)	11 μl	11 μl	11 μl
Primers pairs (50 μM)	11 μl	11 μl	11 μl
Roche FastartTaq (5 U/μl)	11 μl	11 μl	11 μl

If different dilutions are required e.g. 10^{-4}, dilute 1:100 and add 1 μl to the reaction mix.

2. Mix well and dispense 10 μl into each well of 96 well plate. Follow protocol as above for PART A (steps 6-10).

Establish numbers of positives/negatives for each class of marker. Input data into programme to get results.

Appendix 3
Measuring Efficiency of PCR and Real-time PCR Reactions

In order for QUIZ to give accurate GM quantifications, the efficiency of the GM and reference assays must be equal. Ideally this means that their probability of detecting a single target molecule, when one is present in the reaction, is 100%. In reality no reaction will be 100% efficient and the efficiency will vary slightly between assays due to DNA sequence melting temperatures, secondary structure, target competition / inhibition and target length. It is therefore important to measure the efficiency of GM and reference assays prior to using QUIZ. This does not have to be done every time it is used but should be done for new primers, targets and sample matrices. Some variation in efficiencies is, of course, expected and it will be one source of QUIZ stochastic error.

Single Nuclei

Flow cytometry was used sort single nuclei, free from any other particulate contaminants, into 96-well PCR plates. Each well therefore contained a single GM (Mon810) target molecule and two reference molecules (*adh*1). This method avoids uncertainty in sampling molecules from a solution and provides a direct relative estimate of PCR efficiency.

Method
Maize nuclei from fresh F1 Mon810 leaf were isolated using a simple procedure:

6. Place 5 g leaf in a petri dish.
7. Add 20 ml ice-cold Galbraith's solution (45 mM $MgCl_2$; 20 mM MOPS; 30 mM sodium citrate; 0.1% Triton X-100; pH 7.00).
8. Slice leaf with fresh scalpel or razor blade into ~0.5 mm strips (90° to leaf veins).
9. Transfer leaf and Galbraith's into 50ml Falcon tube, rotate at ~12 rpm for 20 minutes at room temperature.
10. Decant liquid into new tube in ice bath. Keep all subsequent steps at 4°C. Filter through three layers of 20 μm nylon mesh.

Nuclei were stained with propidium iodide (1mg /ml) for 10 minutes then separated by a Beckman MoFlo flow cytometer and cell sorter with single maize nuclei into each well of 96 well PCR plates. Nuclei sorting and counting onto glass slides was confirmed by fluorescence microscopy. Plates were centrifuged for 2 minutes at 5000 × g and stored at 4°C. PCR or real-time PCR reagents were added directly to plate wells and reactions carried out normally.

Results
If the sorted nuclei droplet contained DNA from lysed nuclei, then the experimental approach would be void. Control PCRs (96) were thererfore performed on supernatant from the nuclei preparation. All of these controls were negative, demonstrating that only whole nulcei from the Moflo provided positive amplifications.

For each assay, a separate 96 well single nuclei plate was used. GM and reference assays for standard PCR, simplex and duplex Taqman PCRs were tested (Table 1). In Taqman reactions, the Mon810 assay was less efficient than the *adh*1 assay. It should be noted that it is difficult to judge the degree of this difference because the reference and target reactions contain different numbers of target molecules: a single hemizygous Mon810 and two homozygous *adh*1 targets. However, we would expect Mon810 to at least amplify 50% (approximately) of the reactions amplified with *adh*1 but only 31% was observed (27 / 87 positives). Given this difference in efficiencies we would expect QUIZ to underestimate Mon810 by 38% it Taqman was used qualitatively to detect amplification. When Taqman reactions were run together on a single plate (duplex), the ratio and the numbers amplified were very similar (24 / 78 = 30.7%), indicating that at least for the duplex system used here, there was no competitive inhibition.

For standard PCR, the absolute efficiency was much lower than Taqman, *adh1* = 12.5%, Mon810 = 15.6% but the ratio between Mon810 to *adh*1 was much higher (80%) where again approximately 50% was expected, we would therefore expect QUIZ using PCR to overestimate Mon810 by 60%. The observed overestimation for PCRs performed at CSL (same conditions as single nuclei plates) for CRMs was +29.4%, less than expected. Note that at CSL products were detected on agarose gels and both *adh*1 alleles were scored (see QUIZ discussion section).

Single nuclei do not reproduce the exact conditions in QUIZ reactions. The DNA is not purified and amplification may be inhibited by bound proteins. However, we would expect such effects to be similar for both GM and reference assays and therefore the relative efficiency estimate should still be applicable to QUIZ, or at least indicate where there may be problems. Nuclei sorting using Moflo is not practical for routine QUIZ development and validation, the equipment being scarce and very expensive.

Table 1. PCR and Taqman single molecule amplification efficiency measured using single nuclei.

Assay	Positives	Number of reactions	%
Simplex Taqman			
*adh*1	87	96	91
Mon810	27	96	28
Duplex Taqman			
*adh*1	78	96	81
Mon810	24	96	25
PCR			
*adh*1	12	96	13
Mon810	15	96	16

Appendix 4
Quantification of Soybean GM Event, RR, by Real-time PCR

Scientific Basis

This protocol describes a soybean (*Glycine max* L. Merr.) DNA extraction and purification process and event-specific real-time quantitative TaqMan® PCR method for determination of the relative content of GM event RR (GTS40-3-2), from ground grain, flour, leaf and processed food.

The PCR assay has been validated for the specific detection of RR DNA with several real-time PCR machines: AB7900HT, AB7500, AB7700, and AB7500. Amplification using two specific primers is measured during each PCR cycle (real-time), by means of a target-specific oligonucleotide probe labeled with two fluorescent dyes: TET as a reporter dye at its 5′ end and TAMRA as a quencher dye at its 3′ end. The 5′-nuclease activity of the *Taq* DNA polymerase is exploited, which results in the specific cleavage of the probe, leading to increased fluorescence, which is monitored.

For relative quantification of RR DNA, a soybean-specific reference system amplifies *Lec*1, a single-copy soybean endogenous sequence, using a pair of *Lec*1 gene-specific primers and a specific probe labeled with FAM and TAMRA. The emission spectra peaks of the reporter dyes of the RR and *Lec*1 Taqman assay allow their simultaneous measurement in the same single reaction (a duplex Taqman assay), i.e. both primer and probe sets are used in the same reaction.

The method is calibrated with DNA solutions prepared in the laboratory and used as solutions. DNA for this purpose is extracted from GM and non-GM (independently tested with qualitative PCR) soybeans. The DNAs are accurately relatively quantified using Pico Green fluorescence (Molecular probes Inc.) and real-time PCR. They are then mixed in each other to give standards of 5, 2, 1, 0.5, 0.1 and 0 % GM DNA. RR is a homozygous GM event therefore %GM DNA units are equivalent to w/w in most cases, except for the very rare possibility of GM-nonGM cross having occured in cultivation.

Simultaneously to the quantification process, sample DNA quality and reliability of the quantification is assessed. This is achieved by using four sample DNA dilutions with four replicates of each in the quantification reactions.

Scope of the Method

This protocol has been tested using soybean flour, leaf, grain and cooked food products, giving a detection range from 5% to 0.05% RR DNA. The method has been validated for quantification down to 0.05% GM DNA, with the accuracy and precision better than 25% CV and 30% respectively.

Operational Characteristics of the Method

Equipment recommended for the operation of the method is AB7900HT, AB7500, AB7700, and AB7500 although it should operate acceptably with other real-time PCR machines given sufficient capacity and ability to use the two fluorescence channels specified. For DNA purification, the method has been optimised using the Qiagen (GMbH, D-40724 Hilden) DNeasy Plant Mini Kit (part 69104). All other equipment, centrifuges, water baths, blenders etc. have been interchanged between several models without affecting the DNA extraction and purification method.

Protocol

For genomic DNA standards, certified RR grain and non-GM grain are used. The DNA extraction method below is used, but the operator may wish to extract larger quantities and store the standards for future use (max. storage in 1 x TE buffer for 6 months). Larger quantities are obtained by combining several single extractions prior to quantification. The Pico Green genomic DNA quantification kit (Molecular Probes Inc.) and appropriate fluorometer are required in order to separately quantify the RR and non-GM DNA. Following quantification by pico-green, the RR and non-GM DNA solutions are further relatively quantified by real-time PCR, using the proportion of their mean Ct values to calculate a ratio between the DNA concentrations, this ratio is then used to adjust the relative quantification of the pico green method. We have found that quantification by pico-green is necessary in order to standardise the total DNA content of the solutions; and further quantification by real-time PCR enables the adjustment for any difference in degradation or other factors that can affect amplification in the solutions. %GM DNA standards: 5, 2, 1, 0.5, 0.1 and 0% are made as follows:

Volume of GM DNA, $G = (p_1 \times t \times v) / c_1$
Volume of non-GM DNA, $N = (p_2 \times t \times v) / c_2$
Volume of TE = v - G - N

where p_1 is the proportion of GM, e.g. in 0.1% GM DNA = 0.001; p_2 = proportion of non-GM DNA, e.g. in 0.1% GM DNA = 0.999; t = target concentration = 20 ng/µl; v = total volume, typically 1000 µl; c_1 = concentration of GM DNA; c_2 = concentration of non-GM DNA.

A. *DNA Extraction and Purification.*

This method is intended to be performed on homogenised 300 mg samples. For test samples, it must be previously determined how many 300 mg samples will be required to give accurate results for an entire analytical sample. Different sample types will require different forms of homogenisation or grinding. It is not the scope of this method to specify these here. Generally sample ground as finely as possible will give better results, provided the process does not create too much heat or take too long (<20°C, <2 minutes recommended) chilling of samples prior or during homogenisation may help.

Reagents:
Qiagen DNeasy Plant Mini Kit (part 69104)

100 × TE buffer: 1 M Tris base, 100 mM EDTA, pH 8

Prepare sufficient Qiagen DNeasy lysis buffer, AP1, for all samples beforehand, according to the kit instructions.

1. Add 800 μl buffer AP1 to 300 μl sample in 2 ml microcentrifuge tube. Mix well and incubate 65°C for 15 minutes.
2. Follow remaining Qiagen DNeasy instructions for purification, omitting the optional 'shredder column' steps. Elute DNA twice using 2 × 100 μl EB so that the total volume eluted is 200 μl. Add 2 μl 100 × TE to the eluate, mix and store at 4°C.

B. *Real-time PCR protocol*

Reagents:
ABsolute QPCR (+ROX) 2 × mix, part number: e.g. AB-1209 (ABgene, www.ABgene.com).
10 × dNTP mix solution = 2 mM each TTP, dATP, dCTP, dGTP in water.

Oligonucleotides:

***Lec*1 endogenous assay**	Sequence 5' - 3'
le1f	CAACGAAAACGAGTCTGGTGAT
le1r-3	GGATTTGGTGGATCCCA
le1p	FAM-TCGCTGTTGAGTTTGACACTTTCCGGAA-TAMRA

RR **assay**	Sequence 5' - 3'
RR1F	CCTTTATCGCAATGATGGCATTTG
RR1R	CATACATACAGGTTAAAATAAACATAGGGA
RR1P	TET-AGGAGCCACCTTCCTTTTCCATTTGGGT-TAMRA

Oligonucleotides obtained from Sigma Genosys (http://www.sigmaaldrich.com/Brands/Sigma_Genosys.html)
All solutions should be stored at 4 °C.

Reaction preparation

Three serial dilutions of the test sample DNA are measured in duplicate. Make the dilutions as follows:

Dilution	Test sample DNA (μl)	1 × TE (μl)
1:5	20 of neat	80
1:10	40 of 1:5	40
1:20	40 of 1:10	40
1:40	40 of 1:20	40

Mix each well before next dilution.

Each reaction has a 25 µl volume. For all reactions to be performed, prepare a master mix containing for each reaction:

12.5 µl ABsolute QPCR 2 × mix;
1 µl of 5 µM each primer stock (for all four primers = 4 µl)
0.5 µl of 5 µM each probe stock (for two probes = 1 µl)
2.5 µl 10 × dNTP mix stock (2 mM each TTP, dATP, dGTP, dCTP)
Mix well before use.

Therefore, mix for 28 reactions, one quantification run:

Component	Volume (µl)
AB-mix	375
dNTPs	75
*Lec*1-F-4	30
*Lec*1-R-4	30
*Lec*1-P	15
RR-A+4	30
RR-B+2	30
TQ-RR	15

Pipette 20 µl of the master mix into each well in the PCR plate, and add 5 µl sample or standard DNA. The layout of samples in the plate is not critical. Seal and then centrifuge the plate briefly.

Each real-time PCR run will consist of the following 28 reactions:

1. 5% GM DNA standard
2. 2% GM DNA standard
3. 1% GM DNA standard
4. 0.5% GM DNA standard
5. 0.1% GM DNA standard
6. 0% GM DNA standard
7. 5% GM DNA standard
8. 2% GM DNA standard
9. 1% GM DNA standard
10. 0.5% GM DNA standard
11. 0.1% GM DNA standard
12. 0% GM DNA standard
13. Sample 1:5
14. Sample 1:5
15. Sample 1:5
16. Sample 1:5
17. Sample 1:10
18. Sample 1:10
19. Sample 1:10
20. Sample 1:10
21. Sample 1:20
22. Sample 1:20

23. Sample 1:20
24. Sample 1:20
25. Sample 1:40
26. Sample 1:40
27. Sample 1:40
28. Sample 1:40

Program the real-time PCR machines for the following run parameters: 95°C, 15 minutes followed by 45 cycles of 95°C, 15 seconds; 60°C, 90 seconds. Default background setting are used on AB machines.

C. *Assay calibration and calculations*

The RR assay is calibrated using DNA extracted from independently tested GM and non-GM seedlings. RR and *Lec*1 assays are performed on the same standard series in a duplex reaction. Where a single standard ∆Ct value from a duplicated pair is judged to be anomalous, i.e. significantly deviating from the linear relationship to Log_{10} %GM concentration of the other standards, it should be excluded from the analysis. If more than one standard value significantly deviates, then the entire run should be repeated. Sample replicates should be excluded if they significantly deviate from the linear relationship of a plot of Ct vs. log of dilution factor. This enables detection of inhibition in neat samples and loss of linearity (out of range) in 1:40 samples. If more than three sample data points are excluded then the analysis should be repeated using new DNA extractions (allowing either a higher concentration and/or higher purity DNA to be obtained). The *Lec*1 assay provides a relative estimate of the total sample DNA concentration. For accurate quantifications, *Lec*1 Ct values should be below 28. If they are higher then DNA extraction should be repeated and / or diluted less. Similarly, if the higher (1:40) dilution RR Ct values are above 38 then DNA extraction should be repeated and / or diluted less.

A standard curve is constructed using wells 1- 12, of ∆Ct versus Log_{10} % GM DNA, where ∆Ct is the RR assay Ct value (TET channel) minus the *Lec*1 Ct value (FAM channel). A linear regression of the data points provides the basis for calculation of % GM DNA from sample ∆Ct values. The R squared value for the regression must be greater than 0.95, or the run should be repeated. The RSD_r is given by s.d./mean $\times$ 100, where s.d. is the standard deviation among the %GM DNA estimates (n = 12 if no data excluded) from the Ct value alone. RSD_r should be less than or equal to 25% after exclusion of outliers.

Calculations for ∆Ct *method*

1. Draw a graph of Ct Vs. Log_{10} dilution factor for samples' FAM channel (*Lec*1).
2. Draw a graph of Ct Vs. Log_{10} dilution factor for samples' TET channel (RR).
3. Perform a linear regression for both graphs. Examine data points and exclude any that are obvious outliers from all further analysis. The regression for both curves must have parallel gradients for the ∆Ct method to be valid. Both gradients should lie within the range -3.1 to -3.5 and the difference between gradients should not be more than 0.2.

4. Draw a calibration curve of ΔCt Vs. Log_{10} %GM DNA for standards. Perform a linear regression and remove any obvious outliers. The gradient of this curve should be in the range -3.1 to -3.8 and the correlation coefficient, R^2 should be $\geq$ 0.95. Record the gradient, m, and intersect, c.
5. For all sample replicates not excluded in step 3, calculate: %GM DNA = 10^((ΔCt - c)/m)).
6. Calculate the mean %GM DNA and %CV.

Appendix 5
Quantification of Maize GM Event, MON810, by Real-time PCR

Scientific Basis

This protocol describes a maize (*Zea mays* L.) DNA extraction and purification process and event-specific real-time quantitative TaqMan® PCR method for determination of the relative content of GM event MON810, from ground grain, flour, leaf and processed food.

The PCR assay has been validated for the specific detection of MON810 DNA with several real-time PCR machines: AB7900HT, AB7500, AB7700, and AB7500. Amplification using two specific primers is measured during each PCR cycle (real-time), by means of a target-specific oligonucleotide probe labeled with two fluorescent dyes: TET as a reporter dye at its 5′ end and TAMRA as a quencher dye at its 3′ end. The 5′-nuclease activity of the *Taq* DNA polymerase is exploited, which results in the specific cleavage of the probe, leading to increased fluorescence, which is monitored.

For relative quantification of MON810 DNA, a maize-specific reference system amplifies *adh*1, a single-copy maize endogenous sequence, using a pair of *adh*1 gene-specific primers and a specific probe labeled with FAM and TAMRA. The sequence for *adh*1 primers and probes in this method was provided by Y. Berteau (Hernandez M. *et al.,*. 2004. Development and Comparison of Four Real-Time Polymerase Chain Reaction Systems for Specific Detection and Quantification of *Zea mays* L. *J. Agric. Food Chem.*, *52,* 4632-4637).

The emission spectra peaks of the reporter dyes of the MON810 and *adh*1 Taqman assay allow their simultaneous measurement in the same single reaction (a duplex Taqman assay), i.e. both primer and probe sets are used in the same reaction.

The method is calibrated with DNA solutions prepared in the laboratory and used as solutions. DNA for this purpose is extracted from GM and non-GM (independently tested with qualitative PCR) maize seedlings. The seedlings' DNA are accurately relatively quantified using Pico Green fluorescence (Molecular probes Inc.) and real-time PCR. They are then mixed in each other to give standards of 5, 2, 1, 0.5, 0.1 and 0 % GM DNA. MON810 is a hemizygous GM event with complex zygosity in seeds. The use of seedling tissue only for standards means that GM DNA is precisely 50% and is used as such in standards' dilutions. The standards used are therefore %GM DNA, not w/w, and give a true % GM DNA quantification without need for any further conversion of units. The full extraction method for the genomic DNA standards is given below. A standard curve is constructed of ΔCt (MON810 assay Ct value minus *adh*1 endogenous DNA assay Ct value) versus Log_{10} of MON810 concentration (as % GM DNA) and used to calculate the %GM DNA of unknown samples.

Simultaneously to the quantification process, sample DNA quality and reliability of the quantification is assessed. This is achieved by using four sample DNA dilutions with four replicates of each in the quantification reactions.

Scope of the Method

This protocol has been tested using maize flour, leaf, grain and cooked food products, giving a detection range from 5% to 0.05% MON810 DNA. The method has been validated for quantification down to 0.05% GM DNA, with the accuracy and precision better than 25% CV and 30% respectively.

Operational Characteristics of the Method

Equipment recommended for the operation of the method is AB7900HT, AB7500, AB7700, and AB7500 although it should operate acceptably with other real-time PCR machines given sufficient capacity and ability to use the two fluorescence channels specified. For DNA purification, the method has been optimised using the Qiagen (GMbH, D-40724 Hilden) DNeasy Plant Mini Kit (part 69104). All other equipment, centrifuges, water baths, blenders etc. have been interchanged between several models without affecting the DNA extraction and purification method.

Protocol

For genomic DNA standards, seedlings of certified MON810 seed and non-GM seed are germinated in the laboratory (trays of tissue soaked in water is sufficient for approx. 3 days). The extraction method below is used, but the operator may wish to extract larger quantities and store the standards for future use (max. storage in 1 x TE buffer for 6 months). Larger quantities are obtained by combining several single extractions prior to quantification. The Pico Green genomic DNA quantification kit (Molecular Probes Inc.) and appropriate fluorometer are required in order to separately quantify the MON810 and non-GM DNA. Following quantification by pico-green, the MON810 and non-GM DNA solutions are further relatively quantified by real-time PCR, using the proportion of their mean Ct values to calculate a ratio between the DNA concentrations, this ratio is then used to adjust the relative quantification of the pico green method. We have found that quantification by pico-green is necessary in order to standardise the total DNA content of the solutions; and further quantification by real-time PCR enables the adjustment for any difference in degradation or other factors that can affect amplification in the solutions. %GM DNA standards: 5, 2, 1, 0.5, 0.1 and 0% are made as follows:

Volume of GM DNA, $G = (p_1 \times t \times v) / c_1$
Volume of non-GM DNA, $N = (p_2 \times t \times v) / c_2$
Volume of TE = v - G - N

where p_1 is the proportion of GM, e.g. in 0.1% GM DNA = 0.001; p_2 = proportion of non-GM DNA, e.g. in 0.1% GM DNA = 0.999; t = target concentration = 20 ng/µl; v = total volume, typically 1000 µl; c_1 = concentration of GM DNA; c_2 = concentration of non-GM DNA.

A. *DNA Extraction and Purification.*

This method is intended to be performed on homogenised 300 mg samples. For test samples, it must be previously determined how many 300 mg samples will be required to give accurate results for an entire analytical sample. Different sample types will require different forms of homogenisation or grinding. It is not the scope of this method to specify these here. Generally sample ground as finely as possible will give better results, provided the process does not create too much heat or take too long (<20°C, <2 minutes recommended) chilling of samples prior or during homogenisation may help.

Reagents:
Qiagen DNeasy Plant Mini Kit (part 69104)
100 × TE buffer: 1 M Tris base, 100 mM EDTA, pH 8

Prepare sufficient Qiagen DNeasy lysis buffer, AP1, for all samples beforehand, according to the kit instructions.

3. Add 800 µl buffer AP1 to 300 µl sample in 2 ml microcentrifuge tube. Mix well and incubate 65°C for 15 minutes.
4. Follow remaining Qiagen DNeasy instructions for purification, omitting the optional 'shredder column' steps. Elute DNA twice using 2 × 100 µl EB so that the total volume eluted is 200 µl. Add 2 µl 100 × TE to the eluate, mix and store at 4°C.

B. *Real-time PCR protocol*

Reagents:
ABsolute QPCR (+ROX) 2 × mix, part number: e.g. AB-1209 (ABgene, www.ABgene.com).
10 × dNTP mix solution = 2 mM each TTP, dATP, dCTP, dGTP in water.

Oligonucleotides:

***adh*1 endogenous assay**	Sequence 5' - 3'
forward primer: *adh*1-F-4	GGTCTCTCAGCAAGTGGG
reverse primer: *adh*1-R-4	TCGTCCCGAACTTCATCT
probe: *adh*1-P	TET-ATGAACCAAGACACAAGGCGGCTTCA-TAMRA

MON810 assay	Sequence 5' - 3'
forward primer: MON810A+4	TCAGCAAGATTCTCTGTCAACAATTGA
reverse primer: MON810B+2	TAGATTTCCCGGACATGAAGATCA
probe: TQ-MON810	FAM-TCCTTTTCTTGCCTTCGTATAAGCTTGTG-TAMRA

Oligonucleotides obtained from Sigma Genosys (http://www.sigmaaldrich.com/Brands/Sigma_Genosys.html)
All solutions should be stored at 4 °C.

Reaction preparation

Three serial dilutions of the test sample DNA are measured in duplicate. Make the dilutions as follows:

Dilution	Test sample DNA (µl)	1 × TE (µl)
1:5	20 of neat	80
1:10	40 of 1:5	40
1:20	40 of 1:10	40
1:40	40 of 1:20	40

Mix each well before next dilution.

Each reaction has a 25 µl volume. For all reactions to be performed, prepare a master mix containing for each reaction:

12.5 µl ABsolute QPCR 2 × mix;
1 µl of 5 µM each primer stock (for all four primers = 4 µl)
0.5 µl of 5 µM each probe stock (for two probes = 1 µl)
2.5 µl 10 × dNTP mix stock (2 mM each TTP, dATP, dGTP, dCTP)
Mix well before use.

Therefore, mix for 28 reactions, one quantification run:

Component	Volume (µl)
AB-mix	375
dNTPs	75
*adh*1-F-4	30
*adh*1-R-4	30
*adh*1-P	15
MON810-A+4	30
MON810-B+2	30
TQ-MON810	15

Pipette 20 µl of the master mix into each well in the PCR plate, and add 5 µl sample or standard DNA. The layout of samples in the plate is not critical. Seal and centrifuge the plate briefly.

Each real-time PCR run will consist of the following 28 reactions:

29. 5% GM DNA standard
30. 2% GM DNA standard
31. 1% GM DNA standard
32. 0.5% GM DNA standard
33. 0.1% GM DNA standard
34. 0% GM DNA standard
35. 5% GM DNA standard
36. 2% GM DNA standard

37. 1% GM DNA standard
38. 0.5% GM DNA standard
39. 0.1% GM DNA standard
40. 0% GM DNA standard
41. Sample 1:5
42. Sample 1:5
43. Sample 1:5
44. Sample 1:5
45. Sample 1:10
46. Sample 1:10
47. Sample 1:10
48. Sample 1:10
49. Sample 1:20
50. Sample 1:20
51. Sample 1:20
52. Sample 1:20
53. Sample 1:40
54. Sample 1:40
55. Sample 1:40
56. Sample 1:40

Program the real-time PCR machines for the following run parameters: 95°C, 15 minutes followed by 45 cycles of 95°C, 15 seconds; 60°C, 90 seconds. Default background settings are used on AB machines.

C. *Assay calibration and calculations*

The MON810 assay is calibrated using DNA extracted from independently tested GM and non-GM seedlings. MON810 and *adh*1 assays are performed on the same standard series in a duplex reaction. Where a single standard ΔCt value from a duplicated pair is judged to be anomalous, i.e. significantly deviating from the linear relationship to Log_{10} %GM concentration of the other standards, it should be excluded from the analysis. If more than one standard value significantly deviates, then the entire run should be repeated. Sample replicates should be excluded if they significantly deviate from the linear relationship of a plot of Ct vs. log of dilution factor. This enables detection of inhibition in neat samples and loss of linearity (out of range) in 1:40 samples. If more than three sample data points are excluded then the analysis should be repeated using new DNA extractions (allowing either a higher concentration and/or higher purity DNA to be obtained). The *adh*1 assay provides a relative estimate of the total sample DNA concentration. For accurate quantifications, *adh*1 Ct values should be below 28. If they are higher then DNA extraction should be repeated and / or diluted less. Similarly, if the higher (1:40) dilution MON810 Ct values are above 38 then DNA extraction should be repeated and / or diluted less.

A standard curve is constructed using wells 1- 12, of ΔCt versus Log_{10} % GM DNA, where ΔCt is the MON810 assay Ct value (TET channel) minus the *adh*1 Ct value (FAM channel). A linear regression of the data points provides the basis for calculation of % GM DNA from sample ΔCt values. The R squared value for the regression must be greater than 0.95, or the run should be repeated. The RSD_r is given by s.d./mean $\times$ 100, where s.d. is the standard deviation among the %GM

DNA estimates (n = 12 if no data excluded) from the Ct value alone. RSD_r should be less than or equal to 25% after exclusion of outliers.

Calculations for ∆Ct *method*

7. Draw a graph of Ct Vs. Log_{10} dilution factor for samples' FAM channel (*adh*1).
8. Draw a graph of Ct Vs. Log_{10} dilution factor for samples' TET channel (MON810).
9. Perform a linear regression for both graphs. Examine data points and exclude any that are obvious outliers from all further analysis. The regression for both curves must have parallel gradients for the ∆Ct method to be valid. Both gradients should lie within the range -3.1 to -3.5 and the difference between gradients should not be more than 0.2.
10. Draw a calibration curve of ∆Ct Vs. Log_{10} %GM DNA for standards. Perform a linear regression and remove any obvious outliers. The gradient of this curve should be in the range -3.1 to -3.8 and the correlation coefficient, R^2 should be ≥ 0.95. Record the gradient, m, and intersect, c.
11. For all sample replicates not excluded in step 3, calculate: %GM DNA = 10^((∆Ct - c)/m)).
12. Calculate the mean %GM DNA and %CV.