



# nature biotechnology

Letters may be edited for space and clarity. They should be addressed to:  
Correspondence  
Nature Biotechnology  
345 Park Avenue South  
New York, NY 10010-1707, USA  
or sent by e-mail to [biotech@nature.com](mailto:biotech@nature.com)  
Please include your telephone and fax numbers.

© 2001 Nature Publishing Group <http://biotech.nature.com>

## The limits of GMO detection

To the editor:

The European Union recently introduced legislation<sup>1</sup> stipulating the mandatory labeling of food products with a GMO content greater than 1%. (The regulation(s) don't specify whether this is weight per weight or any other unit.) Thus far, most discussions concerning the methods used for sampling have focused on sampling requirements outside of the laboratory; for example, how to procure GMO seeds from a grain shipment? Scant attention has been paid to sampling problems that lie further down the analytical chain—that is, variability in the proportions of GMO to non-GMO DNA in replicate “homogenized” laboratory samples. We believe this has serious implications for the practicability of GMO detection in foods.

The first problem in any DNA sampling protocol is defining the limits of detection. The amount of unreplicated haploid genome (i.e., the 1C value) present in a sam-

ple is useful for relating genome copy number to the amount of sample taken. For example, up to 36,697 copies of the haploid *Zea mays* genome (which we will use here for all examples below) are present in a typical 100 ng DNA analytical sample, given the 1C value of 2.725 picograms<sup>2</sup>. It follows that a single copy of the haploid *Z. mays* genome in a 100 ng DNA sample is present at a level of 0.0027% (wt/wt). Levels of DNA below this threshold simply cannot be detected reliably in samples of this size.

A second problem is sampling error. This occurs in a perfectly homogeneous preparation, even if a large amount (say, 50 µg) of DNA is extracted from a laboratory sample and simple random sampling procedures<sup>4</sup> are adopted. As the amount of DNA extracted from the sample becomes lower, sampling error becomes (proportionally) larger. Thus, replicate 100 ng DNA samples containing GMO material at a level of 0.1% (wt/wt) would produce GMO DNA estimates no better than ~30% of the mean value, 95% of the time—a poor level of accuracy, even if we ignore other types of error inherent in a real analytical system.

To illustrate this, we use the cumulative distribution function for the binomial distribution<sup>5</sup> to calculate the probable range of GMO genome copies that would be “sampled” in a single-step procedure—that is, from a (large) laboratory sample of “known” low content (0.1% GMO) into a series of 100 ng analytical samples.

Although on average, the analytical samples should contain 36.7 GMO genome copies, in fact the number of GMO copies ranges from 25 to 48, with a 94.3% probability. Thus, the actual DNA content that would be observed in a single sample, with an ~95% probability, would range from 0.068% to 0.131%; the probability of sampling exactly 36 GMO copies (i.e., 0.1% content) in a single analytical sample is only 0.066.

With lower levels of DNA, the problem is even more critical. For a laboratory sample containing DNA at a level of 0.01%, the 100 ng analytical sam-

ple would vary between 0.0027% and 0.0191% nearly 95% of the time. These calculations obviously refer to a “best possible” result, as they assume a single sampling step and a perfect analytical system.

When undertaking a dilution series, the assumption of simple random sampling may no longer be valid, as the number of copies available becomes strictly finite. Indeed, the number of copies used to prepare subsequent dilutions heavily influences the sampling error associated with the series. Consequently, the preparation of any dilution series must be undertaken in such a way as to minimize this bias; ideally, dilutions should be made from the primary laboratory sample. Unfortunately, we note that some equipment manuals actually encourage the construction of the series without recognizing this problem.

The classical solution to the issue of sampling error is to undertake repetitions and/or use appropriately sized (i.e., larger) analytical samples. We recommend that in the construction of a dilution series—for example, for determination of “limit of detection” of a method, or for the generation of standard curves—the nominal number of GMO copies in the weakest dilution of analytical sample should be set to ~20, thus providing good statistical probability that all repetitions contain relevant DNA (Table 1).

However, we are aware of important studies that seem to draw conclusions without such safeguards, despite explicitly working with copy numbers. Several international standards for PCR analysis of GMO in foodstuffs, currently under development, draw attention to sample sizes in the procurement of material for the laboratory sample, but in general do not address the issues of sampling associated with the analytical sample. We believe there is insufficient acknowledgment that repeated analytical samples drawn from a “homogenized” laboratory sample would *not* have identical proportions of GMO/non-GMO copies.

Simon Kay and Guy Van den Eede  
The Joint Research Centre of the European  
Commission  
Institute for Health and Consumer Protection  
Food Products Unit  
I 21020 Ispra, Italy  
([simon.kay@jrc.it](mailto:simon.kay@jrc.it)).

**Table 1. Expected probability of GMO copies in a 1000 ng DNA sample of *Zea mays*<sup>a</sup>**

Number of transgene copies	Cumulative binomial probability	Frequency	Actual content
8	0.57%	0.3%	0.002%
9	1.27%	0.7%	0.002%
10	2.55%	1.3%	0.003%
11	4.69%	2.1%	0.003%
12	7.96%	3.3%	0.003%
13	12.57%	4.6%	0.004%
14	18.61%	6.0%	0.004%
15	26.01%	7.4%	0.004%
16	34.49%	8.5%	0.004%
17	43.64%	9.2%	0.005%
<b>18</b>	<b>52.97%</b>	<b>9.3%</b>	<b>0.005%</b>
19	61.98%	9.0%	0.005%
20	70.24%	8.3%	0.005%
21	77.46%	7.2%	0.006%
22	83.48%	6.0%	0.006%
23	88.29%	4.8%	0.006%
24	91.96%	3.7%	0.007%
25	94.66%	2.7%	0.007%
26	96.56%	1.9%	0.007%

<sup>a</sup>Nominal content of 0.005%. Whereas the most probable number of copies would be 18 (bold), there is a 0.57% chance that 8 or fewer copies will be present.

1. European Commission. *Official Journal* L 006, 13–14 (2000).
2. Hübner, P., Studer, E. & Lüthy, J. *Nat. Biotechnol.* 17, 1137–1138 (1999).
3. <http://www.rbgkew.org.uk/cval/database1.html>.
4. Cochran, W.G. *Sampling techniques*. (Wiley, New York, NY; 1977).
5. Forthofer, R.N. & Lee E.S. *Introduction to biostatistics*. (Academic Press, San Diego, CA; 1995).