

SHORT COMMUNICATION

Use of hypergeometric distribution for estimating adventitious presence of GM traits in small seed lots may be misleading

Rod A. Herman* and Kelly R. Robbins

Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

(Received 15 February 2013; accepted after revision 24 April 2013; first published online 31 May 2013)

Abstract

Testing for the unintended or adventitious presence (AP) of genetically modified (GM) events in seed lots is a common practice to comply with regulatory requirements and good stewardship practices. A subsample of a seed lot is typically tested for AP levels, and then statistical methods are used to estimate the upper level of AP in the remainder of the lot with a given level of confidence. For large seed lots, a binomial distribution is typically assumed, but for seed lots where the tested sample is a substantial proportion of the overall seed lot, a hypergeometric distribution is typically assumed. Due to the destructive nature of AP seed testing, we suggest that this latter method may overestimate confidence of low AP in the remaining seed.

Keywords: adventitious presence, genetically modified, hypergeometric, transgenic seed

Introduction

Compliance with regulatory requirements on the unintended or adventitious presence (AP) of genetically modified (GM) traits in seed is a major objective when importing or exporting seed, or when conducting regulatory studies to support the safety assessment of GM crops (Lipp *et al.*, 2005). Most commonly, a subsample of a large seed lot is tested for AP, and if found free of AP, allows high confidence that AP in the large population is below a predetermined threshold. For example, finding that 2995 randomly sampled seeds contain no AP establishes, with 95% confidence,

that AP is below 0.1% in the larger lot from which the sample was drawn (Lipp *et al.*, 2005). This confidence level is calculated using probability theory assuming a binomial distribution.

Hypergeometric distribution

When the proportion of seed tested is substantial compared with the overall lot size (typically >10%) then use of a hypergeometric distribution is recommended if the seed is not replaced during the AP testing process (i.e. no repeat sampling) (Remund *et al.*, 2001). Because testing seed for AP is typically destructive, the assumption of not replacing seed during the testing process is usually met. However, estimation of AP using the hypergeometric distribution applies to the composite of the seed tested and the residual seed lot that is destined for inclusion in other studies. In the case of testing for defects in manufacturing, this type of estimation is justified if the samples tested for defects are replaced into the overall product lot after testing, but in the case of seed tested for AP, the destructive nature of testing makes this atypical.

The hypergeometric approach often requires testing fewer seeds to achieve the same confidence in a low AP level compared with a binomial approach, because we know with relative certainty the AP level in a substantial quantity of the seed lot being estimated. However, the destruction of the tested seed means that the tested seed will not be replaced into the seed lot for which we wish to estimate AP. For this reason, the use of the hypergeometric distribution overestimates our confidence in low AP in the residual seed lot. As a dramatic example, if one tests 990 seeds from a 1000-seed lot and finds 901 seeds free of AP and 89 seeds with AP, then one can be certain that <10% of the 1000-seed lot contains AP (901/1000 = 90.1% AP-free), but we are much less certain that the

*Correspondence
Email: raherman@dow.com

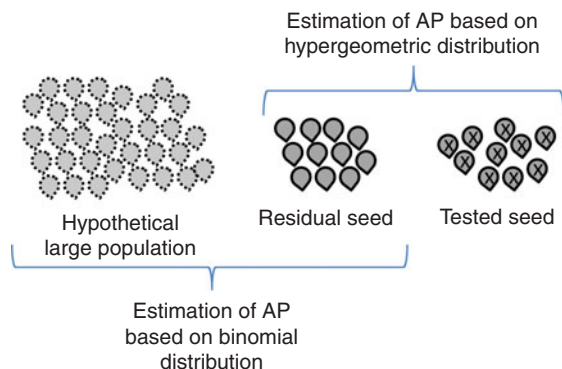


Figure 1. (colour online) Depiction of composite population of seeds for which AP is estimated based on the hypergeometric and binomial distributions (inference space).

remaining 10 seeds have $<10\%$ AP (i.e. no AP seeds), especially since the tested seed sample contained approximately 9% AP. Yet it is the AP level in these remaining 10 seeds that is relevant since this is the seed destined for use in a regulatory study.

A more germane example would be testing 3000 seeds from a 4000 overall seed lot and finding no AP in the tested seed. From this result, one can calculate the probability that the 4000 seed lot has 0.1% AP (4 seeds) as 0.004 (0.4%) using a hypergeometric distribution. This indicates a 99.6% probability that AP is below 0.1% in the 4000 overall seed lot, but does not address the probability that AP $<0.1\%$ in the remaining 1000 seeds. In fact, if there were one AP seed in the remaining 1000 seed lot, the threshold of $<0.1\%$ would be exceeded. In this example, if one AP seed was in the 4000 seed overall lot, then it would be found in the 1000-seed residual lot 25% of the time (i.e. 75% probability that AP in the remaining 1000 seeds is $<0.1\%$). For this reason, the former calculation of probability using the hypergeometric distribution (99.6%) has the potential to be misleading.

Discussion and conclusions

Therefore, even when residual seed lots are relatively small compared with the seed numbers tested for AP, it would appear more suitable and practical to assume a binomial distribution when estimating probabilities relative to AP testing of seed (considering the AP-tested seed and the residual seed lot as random samples from a larger potential population) (Fig. 1). Using the binomial approach, the number of seeds tested for AP is considered insignificant relative to the size of the overall seed lot, so the specific AP results found for the tested lot is also considered an insignificant contributor to AP in the larger seed lot (which is appropriate since these seeds are discarded). We therefore suggest extending the commonly recommended estimation of AP assuming a binomial distribution to situations where the size of a seed lot is small relative to the tested sample.

Acknowledgements

We thank Larry Freese of USDA GIPSA and Scott Ray, Pablo Valverde, Siva Kumpatla and John Cuffe of Dow AgroSciences for reviewing a draft of this paper.

References

- Lipp, M., Shillito, R., Giroux, R., Spiegelhalter, F., Charlton, S., Pinero, D. and Song, P. (2005) Polymerase chain reaction technology as analytical tool in agricultural biotechnology. *Journal of AOAC International* **88**, 136–155.
- Remund, K.M., Dixon, D.A., Wright, D.L. and Holden, L.R. (2001) Statistical considerations in seed purity testing for transgenic traits. *Seed Science Research* **11**, 101–120.