E

BMT/8/17
ORIGINAL: English
DATE: August 14, 2003

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA-PROFILING IN PARTICULAR

Eighth Session
Tsukuba, Japan, September 3 to 5, 2003

STATISTICAL ASPECTS OF ESSENTIAL DERIVATION

Document prepared by experts from the Netherlands and the United Kingdom

STATISTICAL ASPECTS OF ESSENTIAL DERIVATION

John Law[1] & Fred van Eeuwijk[2]

[1] NIAB, Huntingdon Road, Cambridge CB3 0LE, United Kingdom

[2] Lab of Plant Breeding, Dept of Plant Sciences, Wageningen University
P.O. Box 386, 6700 AJ Wageningen, The Netherlands

Introduction

1. Traditionally, plant breeding companies obtain proprietary rights on new varieties of crops when these varieties comply with the requirements of Distinctness, Uniformity and Stability (DUS). These requirements are crop specific and are described in special procedural documents under the responsibility of UPOV. A characteristic of the UPOV system for obtaining breeder's rights is that the criteria used to examine varieties are almost exclusively phenotypic.

2. In recent years, increasing concerns have been raised about the operation of the part of the UPOV system that is meant to stimulate the continuous development of new varieties, the so-called "breeder's exemption". According to the breeder's exemption; breeders are allowed to use any existing protected variety in the development of a new variety.

3. In the last decade, the process of breeding has been drastically changed by the incorporation of genetic techniques that produce new varieties from existing ones by the introduction of costly genetic constructs (i.e. genetic modification). These genetic constructs can for instance confer resistance and tolerance against diseases and stress conditions. Companies whose breeding strategies are mainly based on the introduction of genetic constructs have sought a plant variety protection system that is closer to a patent system. In this context, the concept of essential derivation (ED) has been proposed by the ISF (The International Seed Federation) and incorporated into the UPOV Convention. Essential derivation should cover those cases where an existing variety was copied genetically, but where sufficient phenotypic differences were present to pass the classical phenotypic test for D, U & S, thus assuring breeder's rights for the new (but essentially derived) variety.

4. An important point in the demonstration of ED is establishment of the genetic conformity between the initial, already protected, variety, and the potentially essentially derived variety (EDV). Molecular markers have an important role to play in this. The idea is that when a representative sample of DNA, in the form of molecular markers, is taken from an initial variety and an essentially derived variety, a similarity coefficient expressing the conformity between the two varieties will on average exceed a critical threshold. The challenge for the ISF lies in the formulation of useful thresholds for this reversal of the burden of proof. Beyond the thresholds, the breeder of the potential EDV will be asked to prove that the new variety was obtained without an intention of merely copying existing germplasm. The thresholds should be such that genuine ED is detected with sufficient power and precision, without flagging too many varieties that were not actually essentially derived.

5. Research has been undertaken in various crop species to investigate the feasibility of using of thresholds, based on genetic similarity, for use as a reversal of proof criterion. Statistical issues that are raised include the definition of the population of varieties to which an initial and derived variety belong; the distribution of the similarities in this population; the

coverage of the genome by the sample of molecular markers that is used for the estimation of the genetic similarity between initial and derived variety; the weighting regime for individual markers in the calculation of the similarity and the standard error of the similarity. In the current presentation, these statistical issues will be illustrated and discussed with data from part of an international project funded by the European Union (Molecular Markers for Essential Derivation, MMEDV).

An Example of an Approach to ED Determination in Barley

6. (a) Background and datasets. For a crop like barley, where varieties mostly consist of inbred lines which in principle should be largely homogeneous, an attractive molecular marker would be an easy to use, dominant marker system with good genome coverage and sufficient degree of polymorphism, e.g. AFLP.

7. The MMEDV project investigated a protocol for establishing possible ED in barley, using AFLPs as well as other markers. Forty-six barley varieties were fingerprinted with 10 AFLP primer combinations. The varieties were chosen as to be representative of the varieties on the UK Recommended List and the Finnish National List. This exercise produced 759 markers, of which 600 were polymorphic and 247 were mapped. Jaccard and simple matching similarities were calculated on the whole set of markers. The basis of the study was to investigate whether the relationships as generated by the similarity estimates agreed with breeders' knowledge of the relationships between the varieties. If the answer to this were positive, then it could be concluded that AFLPs are providing a more or less representative sample of the genome of the analysed varieties.

8. A simple concept that is useful for defining ED is the so-called tail principle. In this, it is assumed that the most extreme similarities, i.e. those in the tail of the similarity distribution, are the most likely candidates for ED. The issue then becomes where to place the threshold beyond which the reversal of proof should occur? This is not a question that the MMEDV project addressed *per se*. Rather, we were concerned to provide data that would help in choosing between various alternatives. One possibility would be to just use a particular percentile point, e.g., all variety pairs having similarities larger than the 95 percentile may be labelled as suspect EDVs. An alternative calculates the similarities for known relationships and places the threshold where the similarity for a particular type of relationship is surpassed. MMEDV studied both of these approaches to defining ED thresholds. In addition to the analysis of the 46 varieties above, we had a backcross family (Kustaa x Wanubet) from which it was possible to calculate similarities between parents and specific backcross generation offspring. This backcross programme was analysed in some detail with the explicit aim of providing information on similarity estimates for specific relationships.

9. In addition to estimating the similarities themselves, we were also interested in margins of error for the similarity estimates, as these errors will affect decisions on thresholds. This standard errors from the standard deviations for individual similarity estimates across primer combinations were calculated.

10. (b) Summary of Results

(i) Similarity Estimates: Genetic distances between 46 barley varieties based on AFLP were generally below 95 % similarity, but there were some pairs of varieties closer than this. Approximately 5 % of the set of n(n-1)/2 possible pair-wise variety similarities [where n =

number of varieties] were over 95 % similar. In other words, taking the ranked set of pair-wise similarities, the specific similarity corresponding to the 5th percentile of most similar pairs corresponds to a genetic similarity coefficient of 0.95. It has been suggested that similarities between 0.95 and 1.0 represents a 'zone' where potential ED cases could arise. Study of this set of pairs of varieties with similarity coefficients greater than 0.95 showed that they mostly comprised closely related 'sister lines', and material from the same breeder, varieties known to be morphologically alike.

(ii) Analysis of backcross families: It was necessary to calibrate the 0.95 similarity coefficient against systems where the relationship between the material is known. Examination of backcross lines showed that the 95 % similarity threshold was crossed at the BC1-BC2 boundary. This means that at BC2, the distribution of genetic similarities, generated from a number of individuals, all achieved a similarity greater than 0.95. Corresponding results at BC1 showed that, while a small proportion of the set of pair-wise similarities had a similarity exceeding 0.95, the majority had weaker similarity coefficients. Study of more back-cross families may allow additional precision to be extracted, but based on the present studies backcross family, a proposed EVD threshold of 0.95 similarity (based on grouped material, see (iv) below) translates to the BC1/BC2 boundary.

Data from the other marker systems used (S-SAP, IRAP/REMAP), while differing slightly in the detail, give comparable results.

(iii) Representation of known genetic relationships: A multi-dimensional scaling (MDS) plot showed that the barley varieties fell into four groups - Nordic varieties, UK varieties, a mixture group of Nordic and UK varieties and an outlier, which was the only winter type barley in the set of 46. These groups reasonably reflect breeders' information. Therefore, the AFLP-based genetic similarities seem to represent the DNA of the selected barley varieties to a sufficient degree.

(iv) Distribution of similarities: The multi-modal nature of the observed distribution of pair-wise distance/similarities was indicative of a structural or systematic feature of the data that required further investigation.

11. The quartile parameters (Table 1) for varieties classified as deriving from "Finland" or "UK" showed that the distributional properties of the two classes were different. Accordingly, material was partitioned according to country of origin and the resulting national distances or similarities calculated.

Table 1. Quartile pair-wise distances barley varieties partitioned into Finnish and UK material

Quartiles	Finland and UK	Finland	UK
Q0	**0.2279**	0.2279	0.2427
Q1	**0.4542**	0.4677	0.3971
Q2	**0.5273**	0.5692	0.4390
Q3	**0.6620**	0.6520	0.4626
Q4	**0.7524**	0.7617	0.5344
Q3-Q1	**0.2078**	0.1843	0.0655

12. A number of independent tests each showed a largely non-overlapping distribution of UK and Finnish AFLP (and S-SAP, IRAP/REMAP) pair-wise distances. Also of note was that, even within the Finnish material, the distribution of pair-wise distances was multi-modal, illustrative of a further clustering within this variety set. Work has been successfully undertaken to further cluster the Finnish material by row type, end-use and breeder.

13. The foregoing clearly shows that for meaningful operation of a similarity threshold system, it is essential to operate within an appropriate 'grouping' – such as end-use types, seasonal types and country of origin.

(v) Precision: Standard errors for Jaccard and simple matching (SM) coefficients were calculated and compared. As expected, the standard errors decreased with increasing similarity and those for Jaccard were larger than those for simple matching. Using 10 AFLP primer combinations, standard errors for higher similarities were of the order of 0.02 for Jaccard and 0.01 for simple matching. The 95% confidence intervals would thus be approximately 0.08 and 0.04, and least significant differences 0.056 and 0.028, respectively. This seems acceptable from most practical points of view. The beneficial effect of smaller errors for similarities based on the simple matching approach, as compared to Jaccard, are cancelled out by the smaller range in which simple matching coefficients appear.

General Conclusions

14. The marker systems chosen for use in barley (IRAP, REMAP, S-SAP and AFLP) all gave meaningful and useful levels of polymorphism when used to examine a representative set of varieties from Finland and the UK. The results from the different approaches taken to analyse barley varieties of various kinds confirm that molecular markers are suitable tools with which to indicate potential infringements of Plant Breeders' Rights. Comparison with morphological markers (used to examine the backcross families) has shown that molecular data are less subjective as well as being easier to produce and analyse in the context of ED and genetic conformity.

15. Grouping by type is required if effective ED triggers are to be operated. In the specific case of barley; material from Finland was separated from that of UK origin, as principal co-ordinate analyses and other statistical tests gave strong evidence of clustering. Potentially further sub-grouping by type (2/6 row) and season of growth and/or use may be required for practical and robust EDV assessment.

16. Other possible forms of EDVs in barley, e.g. involving somaclonal variants and other lines derived directly from a variety, may be less of a challenge to detect and molecular markers can provide an unambiguous source of data for this purpose.

17. Although not discussed in detail in this paper, essentially the same approach to similarity estimation using various markers also gave a workable EDV framework in roses and maize. For example, in rose, a total of 135 polymorphic SSAP markers were scored from a collection of varieties (83 varieties including 13 mutant groups). Maximum Jaccard distance for non-mutant pairs was 0.7705, and minimum Jaccard distance for mutant pairs was 0.9459. Thus, a very clear separation occurred between the mutant and non-mutant pairs. Results for S-SAP and AFLP in rose were comparable.

18. Thus in general, the MMEDV project unequivocally established that genetic conformity as a trigger for investigation of ED can be measured using molecular markers. Moreover, methods for its assessment in the three crops under examination were developed, along with a number of statistical tools for use in the measurement of ED and example sets of background data for comparison. In a more generic sense the project established a framework for the assessment of ED in any crop, giving clear principles of approach to the use of various analytical techniques. For more information, see http://www.niab.com/bbp/mmedv/.

Acknowledgements – the European Commission is thanked for funding the MMEDV project and we are pleased to acknowledge the hard work of all of the partners.

[End of document]