E

BMT/11/11
ORIGINAL: English
DATE: August 1, 2008

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

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

Eleventh Session
Madrid, September 16 to 18, 2008

A RESEARCH PROJECT CO-FINANCED BY THE
COMMUNITY PLANT VARIETY OFFICE OF THE EUROPEAN COMMUNITY (CPVO):
"MANAGEMENT OF WINTER OILSEED RAPE REFERENCE COLLECTIONS"

Document prepared by experts form the United Kingdom

A RESEARCH PROJECT CO-FINANCED BY CPVO:
"MANAGEMENT OF WINTER OILSEED RAPE REFERENCE COLLECTIONS"

Dr Carol Norris

Coordinator: NIAB, United Kingdom
Partners: Bundessortenamt (BSA), Germany
Danish Institute of Agricultural Sciences (DIAS - now transferred to The Plant Directorate and to The Faculty of Agricultural Sciences, University of Aarhus), Denmark
Groupe d'étude et de Contrôle des Variétés et des Semences (GEVES), France

INTRODUCTION

1. Oilseed rape (*Brassica napus* L.) (OSR) is an important oil and fodder crop, grown in many parts of Europe and world-wide. Variety registration and protection of OSR are carried out in several European Union (EU) Member States, requiring distinctness, uniformity and stability (DUS) testing of new varieties. A major problem for all countries carrying out DUS tests is the requirement to compare new varieties with an increasing number of varieties of common knowledge. In principle, every year newly listed varieties from all EU Member States should be added to each country's reference collection, and grown in the trials for DUS. In practice, to include this many varieties in a replicated field trial is logistically and financially prohibitive for any EU Member State. However, it is desirable that in order to maintain the strength of protection offered by plant breeders' rights (PBR), the principle of comparing new varieties with those of common knowledge should be upheld, and that variety reference collections should be as comprehensive as possible. Clearly, some means of "managing" reference collections is thus highly desirable. Attention has focused on pre-selection methods, comparing candidate varieties with the reference collection prior to sowing the field trial, in order to: on the one hand reduce the number of varieties that need to be grown; whilst on the other, maintain the quality of PBR.

Objectives addressed

2. One means of such management or pre-selection would be to use molecular markers (DNA-profiling) to compare new varieties with those already tested, eliminating those which are sufficiently distant and do not need to be compared in a field trial (according to pre-defined criteria) and then only growing the most similar varieties for detailed morphological DUS assessment. This was the approach examined in this project, which explored potential ways of improving the cost-effectiveness of WOSR DUS testing across the EU (and ultimately more widely), addressing the genotype x environment issue, and enabling increasing work-loads to be achieved within existing resources. The overall objective of this project was thus to examine the potential uses of DNA molecular markers (specifically simple sequence repeat microsatellites, SSRs) as a tool for the management of variety reference collections in winter OSR DUS testing, in the context of a UPOV Option 2 approach, i.e. *"Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics"* (see documents TC/38/14-CAJ/45/5 and TC/38/14 Add. – CAJ/45/5 Add.).

Project Outline

3. The experimental approaches used were to: (i) standardize conditions for the use of an agreed set of SSRs; (ii) analyze a large variety collection from different EU MS with these SSRs; (iii) analyze the data produced, including estimates of genetic and phenotypic distances, compare the distances in different ways; and (iv) validate these approaches in a field assessment.

SUMMARY OF RESULTS

4. In total, 410 varieties were analyzed using a set of 23 SSR markers, and morphological data for these varieties from four countries collated. After inspection of the data and taking into account missing data points, 335 varieties were analyzed with 18 SSRs and with sufficiently complete morphological data were used in the final consolidated dataset. The difficulties inherent in the DNA-profiling of a heterogeneous species such as OSR in different laboratories using different equipment were overcome by the development of a thresholding approach. This enabled good quality molecular data to be compiled.

5. The thresholding produced three datasets according to the concordance found between laboratories. The genetic distances using the three thresholded datasets – T1 (>90% concordance, 18 markers), T2 (>95%, 11 markers), and T3 (>90% + internal controls, 14 markers) were calculated using the software package DarWin (CIRAD).

Statistical Analysis of Molecular and Phenotypic Data

6. The basic objective of the statistical analysis was to calculate various estimates of distance (both genetic distance (GD), from the molecular data, and phenotypic distance, (PD)) and compare these estimates, to evaluate the fundamental UPOV Option 2 approach. A number of different analyses were used, in brief, for the GD estimates, NIAB converted the finalized and validated SSR data from band present/absence binary data into genotype-pattern profiles, and then computed GD with City Block, using the GenStat Software. It was thought that this would cope with the expected distribution and quantity of missing data. GEVES used the presence/absence binary data to compute a number of distances. The Nei & Li (or Dice) and Jaccard distances were calculated using LCDMV software, and Simple Matching, Ochiai and Sokal and Sneath distances with DarWin software. Once all GD matrices had been computed, the data were exchanged and their robustness validated using Mantel statistics.

7. For the PD estimates, again a range of approaches were undertaken, both for the data in Notes form, and for the measured values. In all cases a number of possible approaches were utilized. The one finally used for the Notes data was:

> Establishing where possible the MODAL note (with a maximum of 3 sites/centers each with a maximum of 3 years worth of data – 9 possible values). In cases where no unique mode exists (either due to too few data values, tied modal values or no defined mode) the median was used. See an example set below:

	Country1			Country2			Country3			Mode	Value Used
	Year1	Year2	Year3	Year1	Year2	Year3	Year1	Year2	Year3		
Example 1	3	4	3		3	3	4	3	3	3	Mode=3
Example 2	4	4	3	4	4	3	4	3	3	4	Mode=4
Example 3	1				2		3			N/A	Missing Value
Example 4	1	3	3	1	2	2	1	2	3	N/A	Median=2
Example 5	1	2	3	4	5	6	7	8	9	N/A	median=5 but missing value used

8. There were also issues with the morphological data that had to be overcome, mostly due to the adoption of different recording regimes in the countries involved. Nevertheless, a thorough statistical examination of the data showed that they were robust, with no evidence of any bias or clustering as a result of the country of analysis or other factors.

9. The data sets declared as definitive were those where the quantity and distribution of missing data were minimized, so as to retain the principal objective of a sufficiently large number of varieties to enable valid distance estimates to be calculated and for the operation of GAIA to be assessed effectively. This objective was achieved, and the final agreed sets of morphological data were fully fit for purpose.

10. A lot of effort in this project was put into the selection of SSR markers that could be analyzed successfully in different laboratories, and the validation of the resultant data sets, as well as into ensuring the selection of a sufficient number of appropriate varieties. The molecular analysis is particularly challenging in an out-crossing crop such as oilseed rape and where bulked samples of seedlings are being used to generate variety profiles in laboratories in different countries, utilizing different analytical equipment. However, in spite of these difficulties, the marker selection and validation methods developed within the project, coupled with the application of thresholding, were successful in producing a set of molecular data that were clearly fit for purpose, with “missing” data at a level of 1-2%.

11. An extensive statistical analysis of the data was conducted, which involved the computation of a wide range of distance (similarity) estimates applied to both the molecular and morphological data sets, and comparison of the resulting distances. For Option 2 to be applicable in its most straightforward form, there would need to be a relationship between the two methods of distance assessment, such that a threshold for Distinctness using molecular markers could be extrapolated from thresholds applied to traditional characteristics in such a way that the same decisions would be made, regardless of which method of assessing variety differences was used.

12. No evidence of any statistical correlation between molecular distances and morphological distances was found. However, other approaches to combining morphological assessments and molecular marker distances were investigated and found to produce promising results.

13. Such an approach has applications beyond the management of DUS reference collections, and could be used in any situation where molecular profiling data from different sources are being provided to populate a centrally held database of profiles. The production of the molecular dataset for this project can be seen as a practical example of the application of many of the principles enshrined in the draft UPOV document “Guidelines for DNA-Profiling: Molecular Marker Selection and Database Construction”

(the BMT Guidelines). The difficulties encountered in such an exercise should not be minimized, but as the project has shown, they can be successfully overcome.

14. Statistical analyses applied to the T1, T2 and T3 sets of data (the results of thresholding) did not indicate any significant influence of the number of markers on the reliability of the distance estimates. Using a larger number of markers, e.g. covering each arm of each chromosome, may be desirable, but this would require further study. It should be emphasized that for other applications of molecular markers (e.g. for studies of variety relatedness, essential derivation, genetic diversity, etc.) there is a good case for utilizing more, dispersed markers. In the present instance, the detailed statistical analyses performed on the molecular data sets clearly demonstrated that the data could be used with confidence for subsequent analyses. It was also shown that in the context of the project, there was no advantage in using a particular distance index.

Molecular markers in combination with GAIA

15. GEVES undertook a detailed analysis of the potential use of molecular markers in combination with the software programme GAIA. A summary of the main points is given below.

16. The overall purpose of this work was to compare different methods for selecting the pairs of varieties that should be compared in the field and to evaluate how molecular marker information could be combined with morphological data to reduce the number of these pairs.

17. To examine this, different thresholds for morphological and molecular distances were chosen, and the number of pairs of varieties to be tested in the field estimated, on the assumed use of (i) only morphological characteristics, (ii) morphological and electrophoresis characteristics, or (iii) morphological and molecular characteristics (Dice distances, calculated excluding the monomorphic markers). Phenotypic distances based on morphological and/or electrophoretic data were calculated by using the GAIA software. The GAIA threshold used to declare the varieties super-distinct (see below) was 6.

18. The general proposal for the combination of morphological and molecular data is illustrated in Figure 1. The first step is a selection on morphological characteristics, which leads to the following:

- if the GAIA distance is higher than 6 with morphological data, the varieties are considered super-distinct and do not need to be put in the field;
- if the GAIA distance is smaller than 2, the varieties are put in the field;
- if the GAIA distance is between 2 and 6, then the molecular distance between the varieties is used :
    - if the molecular distance is higher than a defined threshold (for example 0.2 in Figure 1), the varieties are considered distinct and do not need to be put in the field;
    - if the molecular distance is below the defined threshold, then the varieties have to be studied in the field.

19. Dice distance thresholds of 0.35; 0.3; 0.25; 0.2 and 0.15 were tested, in combination with minimal GAIA weights (distances) of 3, 4 and 5.

20. The common database contains 335 varieties, generating in theory and without selection 55,945 pairs of varieties to be compared in the field. Figure 2 presents the GAIA weight versus the Dice distance for the pairs with GAIA<6. Based on the varieties of this database and on the molecular markers used, no correlation can be observed between Dice distances and GAIA weights, which confirms the previous lack of relationship between molecular and morphological distances.

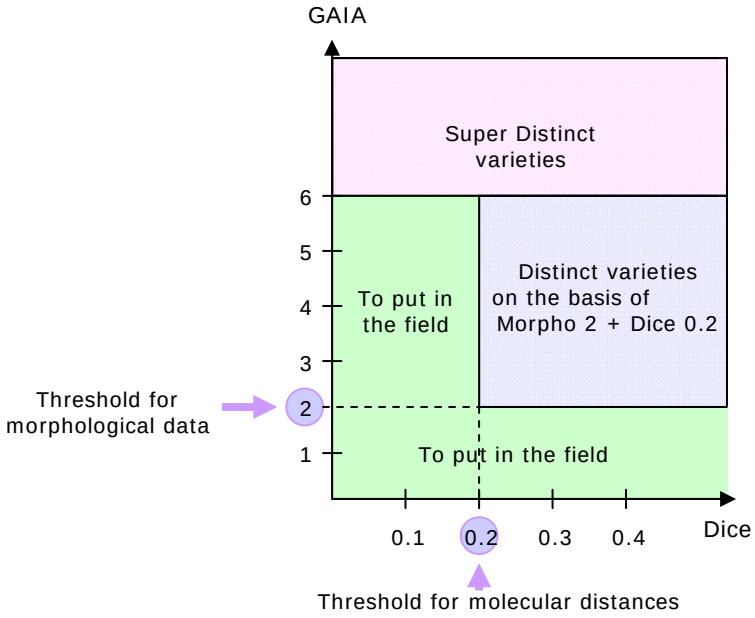

Figure 1: A summary of the GEVES proposal for the selection of the variety pairs to be compared in the field by using molecular data combined with morphological characteristics.

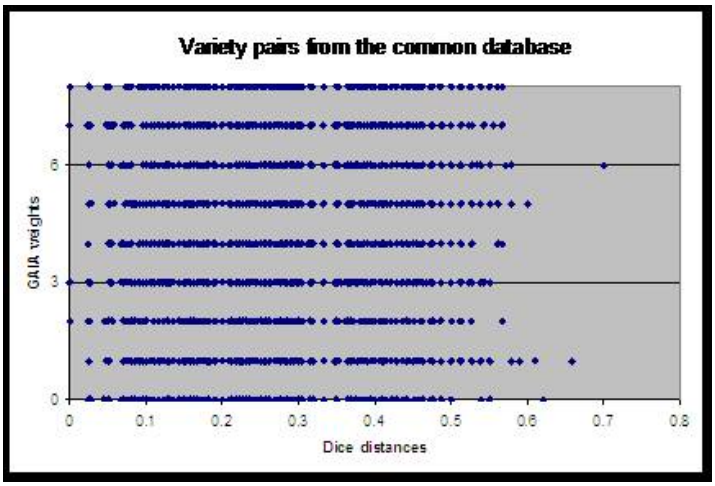

Figure 2: Dice distribution for the variety pairs from the consolidated database with GAIA weight <=8.

21. Figure 3 shows the number of variety pairs to compare in the field, selected using the three proposed methods for the different thresholds chosen. With this data set, the numbers of variety pairs to be put in the field on the basis of morphological data and on the basis of morphological and electrophoresis data are not substantially different (65% vs. 66%), and distinctness is essentially based on qualitative weights. Similar results were found when other data sets were used (see Annex 2 for details).

22. The mean of the GAIA qualitative morphological weights is c. 9.2, with a standard deviation of c. 7.5.

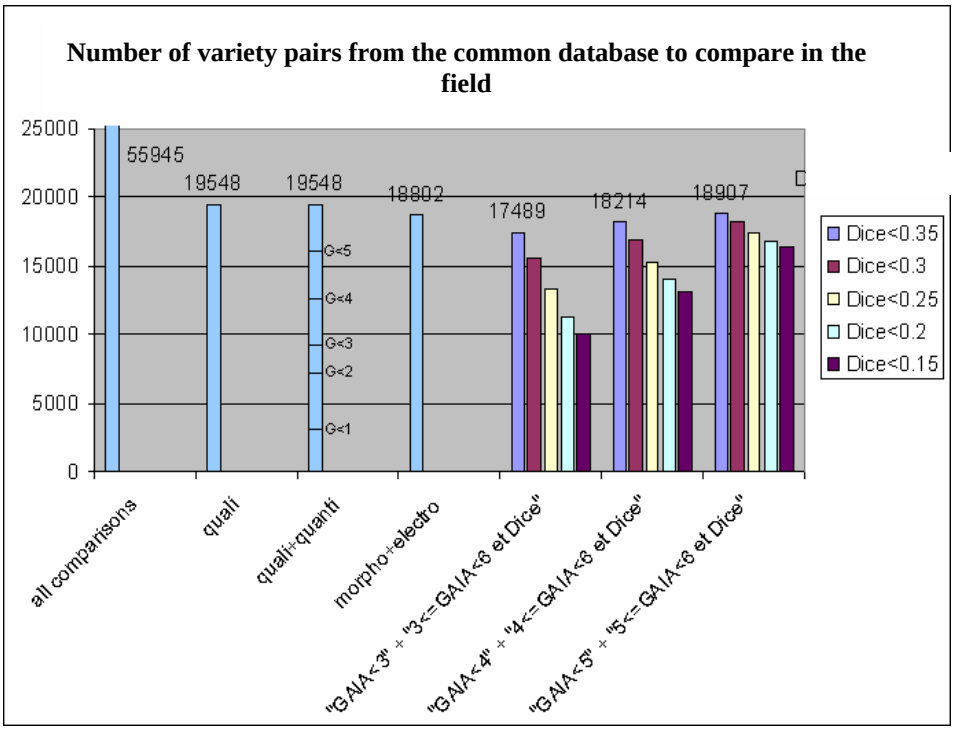

Figure 3: Number of variety pairs from the common database to compare in the field, selected according to the following criteria:

- quali : a GAIA weight<6 based only on qualitative data : qualitative morphological characteristics and quantitative morphological characteristics transformed into qualitative notes;
- quali+quanti: a GAIA weight<6 based on qualitative and quantitative data : qualitative morphological characteristics and quantitative morphological characteristics.
- morpho +electro: a GAIA weight<6 based on qualitative morphological characteristics and isoenzyme data (if qualitative weight>3) and quantitative morphological characteristics (if qualitative+electro weight <6)
- 'GAIA<3' + '3<=GAIA<6 et Dice': EITHER [a GAIA weight <3 based on qualitative and quantitative morphological characteristics] OR [a 3≤GAIA weight<6 based on qualitative and quantitative morphological characteristics AND a Dice distance < to the threshold given by the colors on the side of the graph];
- 'GAIA<4' + '4<=GAIA<6 et Dice': EITHER [a GAIA weight <4 based on qualitative and quantitative morphological characteristics] OR [a 4≤GAIA weight<6 based on qualitative and quantitative morphological characteristics AND a Dice distance < to the threshold given by the colors on the side of the graph];
- 'GAIA<5' + '5<=GAIA<6 et Dice': EITHER [a GAIA weight <5 based on qualitative and quantitative morphological characteristics] OR [a 5≤GAIA weight<6 based on qualitative and quantitative morphological characteristics AND a Dice distance < to the threshold given by the colors on the side of the graph].

23. From this and similar work on other datasets it can be seen that a combination of morphological distances (calculated as GAIA distances in this instance) and molecular distances (calculated as Dice distances in this example) could provide a framework for reducing the number of variety pairs that need to be grown in the field, i.e. managing the reference collection.

24. There is a pressing need to address the question of the management of the reference collection in WOSR DUS testing, and this project has demonstrated quite clearly the difficulties associated with this. Molecular markers still offer perhaps the best opportunities, but their application is by no means straightforward. In order to succeed in combining morphology and molecular distances effectively, it is necessary to define the threshold distances – both morphological and molecular – which produce satisfactory results, with an attendant level of risk which is acceptable to all stakeholders.

25. In order to achieve this, it is suggested that future work in this area should include: (i) the use of more and better quality (preferably single locus) SSRs; (ii) investigations of other types of markers, e.g. functional SSR markers, and/or SNPs.; (iii) continued investigation of distance measures and how best to score molecular profiles; (iv) analysis of the morphological characteristics used in WOSR DUS testing, to produce an agreed set that are robust, to enable data from different years to be combined with confidence.

ACKNOWLEDGEMENTS

CPVO for co-funding
BSA
DIAS
GEVES
NIAB

[End of document]