FINAL REPORT:

MANAGEMENT OF WINTER OILSEED RAPE REFERENCE COLLECTIONS

Research programme **CPV5766,** supported by the Community Plant Variety Office Research and Development Section, 1 January 2005 to 31 December 2007.

NIAB UK:

Address:	Huntingdon Road Cambridge CB3 0LE
Telephone:	+441223342200
Fax:	+441223342303
Email:	info@NIAB.com
Internet:	www.NIAB.com

04 February 2008

National Institute of Agricultural Botany

Bundessortenamt

G E V E S

Institution responsible for the project: NIAB
Huntingdon Road
Cambridge CB3 0LE
UK

Other institutes involved: GEVES
La Miniere
78285 Guyancourt Cedex
France

DIAS (Now transferred to The Plant Directorate and to The Faculty of Agricultural Sciences, University of Aarhus)
Ministry of Food Agriculture and Fisheries
Postbox 7
Teglvaerksvej 10
Tystofte
4320 Skaelskoer
Denmark

BSA
Osterfelddamm 80
30627 Hannover
Germany.

Co-ordinator & Responsible Scientist: Dr Carol Norris, NIAB

Other major contributors: Louise Bach Jensen, Anne Bernole, Francoise Blouet , Cecile Collonier, Robert Cooke, Laetitia Denecheau, Gerhard Deneken, Jöel Guiard, Sylvain Gregoire, Richard Horsnell, Huw Jones, Joelle Lallemand, John Law, Erik Lawaetz, Uwe Meyer, Sylvie Pinloche, Beate Rücker, Jürgen Ritz, Swenja Tams, Mustapha El Yakhlifi

Funding for this project was provided by the Community Plant Variety Office, Defra (UK), the Bundessortenamt (Germany), GEVES (France) and DIAS (Denmark).

TABLE OF CONTENTS:

EXECUTIVE SUMMARY.

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 EU MS, 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. Whilst it is axiomatic that the quality of the rights awarded depends on the ability to compare new varieties with as wide a collection of existing varieties as possible, strict adherence to the concept of common knowledge is impossible. The overall objective of this project was thus to examine the potential uses of DNA molecular markers (specifically 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"*.

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

In total, 410 varieties were analysed 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 analysed 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. 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.

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 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. 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.

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.

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.

~~~~~~~~
~~~~~~~~

1. INTRODUCTION

Oilseed rape (*Brassica napus* L.) (OSR) is an important oil and fodder crop, grown in many parts of Europe and world-wide. Both winter- and spring-sown types are common, and many hundreds of varieties of each seasonal type exist. Variety registration and protection of OSR are carried out in several EU MS currently, requiring DUS testing. Although the format of the DUS testing varies between MS (see below), in all cases replicated plots are grown in field trials and a range of phenotypic characteristics observed and/or measured. A major problem for all countries carrying out DUS tests is the requirement to compare new varieties with an increasing number of existing varieties, whether protected or not. Article 7 of the 1991 UPOV Convention says that a variety shall be considered Distinct "...if it is clearly distinguishable from any other variety whose existence is a matter of common knowledge at the time of the filing of the application". Testing on behalf of the CPVO requires adherence to these same principles, and CPVO protocols incorporate the same concept. Common knowledge is broadly defined elsewhere (e.g. the revised General Introduction to the Test Guidelines, TG/1/3, UPOV 2002) to include all known varieties, i.e. any variety entered into or subject to an application for PBR, varieties grown commercially, varieties held in publicly accessible reference collections, or of which there is a published description. It is axiomatic that the robustness of the rights awarded depends on the ability to compare new varieties of a given species with as wide a collection of existing varieties as possible. However, strict adherence to the concept of common knowledge is clearly logistically and financially impossible, especially in a crop such as OSR which is cultivated widely around the world. Thus DUS testing stations tend to take a somewhat pragmatic view of common knowledge, based on e.g. climatic factors and availability of material. Nevertheless, many hundreds of varieties should still be taken into account for OSR testing. This includes all those with European rights and/or listed on the Common Catalogue (well over 550 in total currently) and other varieties of common knowledge which are relevant in European climatic conditions and for which seed is available for testing stations. Again, the availability of the largest possible reference collection of known varieties is essential to ensure the efficacy of the system for granting PBR, and there is a danger that the quality and scope of protection offered by PBR schemes will be eroded if testing against varieties of common knowledge is not carried out adequately. This inevitably adds to the scale and costs of testing, and in order to reduce these costs, a way of managing the large number of reference varieties and of selecting those varieties most similar to candidates for inclusion in the field trials is needed. If at the same time the number of varieties taken into account could be increased, this would improve the quality of protection offered to breeders by PBR schemes.

Whilst in theory, the full reference collection to be used for comparison purposes for any candidate variety is the known world-wide collection of varieties of the species, in practice, the number of varieties to be included in a growing test can be reduced. UPOV TG/1/3 (2002) allows that "... a systematic individual comparison may not be required with all varieties of common knowledge. For example, where a candidate variety is sufficiently different, in the expression of its characteristics, to ensure that it is distinct from a particular group (or groups) of varieties of common knowledge, it would not be necessary for a systematic individual comparison with the varieties in that group (or those groups)." UPOV TG/1/3 (2002) continues by indicating that the

selection can usually be further narrowed down by using documented variety descriptions and the information on the most similar varieties supplied by the breeder in the Technical Questionnaire which accompanies the application for testing. Thus a testing authority can use a range of sources of information to limit the number of varieties from the reference collection which must be used in the field growing test (Barendrecht 1999).

Clearly then, there is much interest in approaches that could reduce the workload and costs of testing, by eliminating unnecessary comparisons between existing and candidate varieties prior to more formal testing. One possible way in which this might be approached is to use DNA profiling of varieties as a management tool. By comparing the profiles of candidate varieties with those of existing varieties maintained in a central database, it might be possible both to eliminate from further testing those varieties which do not require comparison in a field trial (according to an agreed set of criteria) and to select the varieties most similar to the candidate for close comparison in field tests (Jones *et al.*, 2003, Tommasini *et al.*, 2003). In order for such a scheme to work, it is necessary to have an agreed set of molecular markers to generate the DNA profiles, and an agreed means of utilising the profiling data.

The creation of DNA profile databases populated with data from different laboratories is not a trivial task, but recent research funded by the EU and others has identified the parameters that need to be considered and demonstrated that such an undertaking is feasible (Bredemeijer *et al.*, 2002, Röder *et al.*, 2002). Furthermore, UPOV has clarified the current options for the use of molecular markers in DUS testing, via discussions within the BMT and elsewhere. One of these options ("Option 2: Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics") has been supported by the BMT and subsequently by a special "BMT Review Group" as an approach which would be broadly in accordance with the UPOV Convention, would not erode the value of protection and which should be developed for use in the management of reference collections.

Thus it is now an appropriate time to investigate systematically the use of an "Option 2" approach in an important crop such as OSR, where the size of the reference collection is an issue of concern to MS. This requires (i) the generation of a sufficient quantity of DNA profiling data of good quality, and (ii) the subsequent analysis of these data to evaluate the usefulness for the management of reference collections. By creating a database of variety DNA profiles, which could ultimately be available to other MS, improvements in DUS testing across the EU can be achieved, whilst maintaining costs at no more than current levels. The overall rationale of the proposed work is thus to investigate approaches to OSR DUS testing in which the number of comparisons of candidate varieties with those of common knowledge is maximised whilst the subsequent number of comparisons in field tests is minimised, by using molecular markers. In this way, the strength and scope of the protection offered by PBR systems could be maintained and even enhanced, in a cost-effective and technically robust manner.

2. OBJECTIVES ADDRESSED

The overall objective of this project is to examine the potential uses of molecular markers (specifically microsatellites) for the management of variety reference collections in oilseed rape DUS testing. This will be done by (i) standardising conditions for the use of an agreed set of SSRs, (ii) analysing c. 410 OSR varieties from different EU MS with these SSRs, (iii) analysing the data produced, including estimates of genetic and phenotypic distances, and comparison of such distances in different ways, (iv) validation of these approaches in a field trial. If successful, the project would provide potential ways of improving the cost-effectiveness of OSR DUS testing across the EU, addressing the genotype x environment issue, and enabling increased work-loads to be achieved within existing resources.

3. CURRENT SITUATION

3.1. OSR DUS Testing in the EU

DUS testing of OSR is carried out in several EU MS (at least 8 currently). There is a UPOV Guideline for OSR (TG/36/6) and a CPVO Protocol (CPVO-TP/36/1). Although the characteristics to be recorded in OSR DUS are thus harmonised, there are varying approaches to the testing adopted in different MS, and various sets of "national" characteristics used. In the UK for example, the DUS field trial is grown at a single location (with a reserve site) and testing normally takes two years. Each variety (candidate or reference) is grown in three replicates, two of which are randomised and one of which is arranged so that close comparisons, reference varieties and example varieties are included where appropriate. Hybrid and "line" varieties are treated as separate types, but grown in the same trial. The parent lines of hybrids are also grown, although they are not routinely fully assessed for D,U and S purposes (but may need to be assessed if the parental formula is used for D purposes). A total of 29 characters are routinely observed/measured. A further 20 characters, many of them "combined" characters assessed using image analysis of cotyledons, can be recorded and used if necessary. A separate trial is grown to assess alternativity in winter OSR. For the measured characters, COYD at 2% is generally used for distinctness purposes and UNIF at 1% for assessing uniformity. It is also possible to enter candidates for a third year of testing if D is not established. In the third year, plots of the candidate and non-D variety (or varieties) are grown for side-by-side observations. Other countries e.g. Germany, Denmark, operate systems which are broadly similar to this, but vary in detail, for example the parent lines including maintainer lines are fully assessed for U & S routinely in the same trial.

A different approach is taken in France. DUS trials are grown at two locations, for two years. There are two types of trial. In the first type, grown in each year of testing, plots of candidates and reference varieties are grown for description purposes, with two replicates of each. A total of 16 characters are assessed, two of which are measured, along with a set of six isozymes (the genetic basis and chromosomal locations of which are well established). In these trials, all parent lines of hybrids, plus maintainer lines and restorers are grown, along with the conventional "line" varieties. Hybrids *per se* are also grown in this trial and described for U and S. D is determined by the parental formula, and close hybrids are directly compared when necessary. The data from this set of trials are analysed using a software program

known as "GAÏA". The details of GAÏA are outside the scope of this proposal, but the software has been made available to other UPOV member states. In essence, GAÏA estimates the degree of distance between varieties, based on weightings assigned by the crop expert to the characters measured. Once an established threshold distance has been exceeded, then a variety can be said to be D. The weightings consider the reliability of the character, and the difference required to provide evidence of distinctness. In this way, the D decision is constructed from the sum of varying degrees of difference. The GAÏA results from year 1 are used either to declare varieties D after one year of testing (unlikely in OSR, but happens in a small number of instances), or to plan the field trial for year 2. In a second trial, candidates are grown in replicated plots alongside the most similar variety or varieties, and recorded until a clear difference is found, which can be confirmed at both locations, at which point the variety is said to be D and recording is stopped.

In spite of these differences of approach, and the influence of the environment on the expression of morphological characteristics, which affects the comparability of variety descriptions, when it is possible to make comparisons then the same results are obtained. For example in 2003, the same candidate variety entered in the UK and in France was found to be non-D from the same existing variety (and went into a third year of tests in both countries). Again, the same problems exist whichever system is used – a large number of candidate varieties of different types, the need to include parent lines of hybrids in the trials and the existence of a large reference collection, not all of which can be accommodated within the trial (for both logistical and financial reasons). As an example, the winter OSR trial in the UK this year (2007/08) contains over 2000 plots, about 40% of which are parent lines. France has a large reference collection, with over 300 varieties/lines currently being grown, which represents about 3000 plots at each location.

Thus the outcomes of this project, if successful, will contribute not only to the management of this situation, allowing more reference varieties to be included in comparisons whilst reducing the number of field plots, but will also help to harmonise further DUS testing in EU MS, with consequent benefits to the CPVO.

3.2. Molecular Markers in OSR

In contrast to biochemical or morphological markers, molecular markers are numerous, polymorphic and unaffected by the environment or growth stage. Therefore, they offer several potential advantages for plant variety characterisation (Donini *et al.*, 2000). Various kinds of molecular markers have been used within *Brassica* species, including RAPD (randomly amplified polymorphic DNA) and RFLP (restriction fragment length polymorphism) (e.g. Lee *et al.*, 1996a, *b*). More recently however, DNA microsatellites (simple sequence repeats, SSRs), consisting of short tandem base repeats (2-8 bp units), have gained increasing importance in plant variety testing generally (Cooke 1999; Donini *et al.*, 2000) and are the marker of choice within the UPOV BMT group. SSRs have been studied in *Brassicas* (e.g. Kresovich *et al.*, 1995; SzewcMcFadden *et al.*, 1996; Plieske and Strauss, 2001; Tommasini *et al.*, 2002) and have been shown to be numerous, highly informative, technically simple, robust and suitable for automated allele detection and sizing using high throughput detection methods. Furthermore, due to the economic importance of cultivated *Brassica* species, large investments have been made in the

development of *Brassica* SSRs, many of which are available to the scientific community. For example, in the UK, many SSRs have been made available publicly via the BBSRC UK Cropnet Initiative (primer sequences can be found at http://ukcrop.net/perl/ace/search/BrassicaDB).

Both NIAB and GEVES have screened SSRs from this BBSRC set and from other sources and between them have selected about 30 SSRs that are of good quality. At NIAB, a set of 16 SSRs have been used to analyse levels of heterogeneity within 10 registered OSR varieties (by analysis of 48 individuals from each) and also to analyse discrimination between c. 160 OSR varieties from the UK, D and DK. GEVES have so far analysed a set of 15 varieties with 17 SSRs. Preliminary experiments have indicated that it should be possible to select a common set of SSRs that can be analysed successfully in both laboratories.

3.3. Molecular Markers and DUS Testing

The use of molecular markers for DUS testing has been discussed by UPOV and other interested parties for several years now. Whilst it is acknowledged that such markers have many potential advantages, there are also important issues that need to be addressed, including:

- the number of markers that should be used;
- whether or not the distribution of the markers within the genome is important;
- whether or not it is important that the markers are mapped;
- whether or not it is preferable to use markers that relate to expressed regions of the genome;
- are standardised methods of marker analysis available?
- are the suggested markers publicly available?
- whilst it is relatively well documented that markers can discriminate between varieties (and thus might be able to demonstrate D), what about the U and S aspects?
- would the use of markers inevitably reduce the "minimum distance" between distinct varieties?
- the necessity to develop a database of appropriate structure, to contain not only molecular but also other relevant data.

Following a meeting in 2002, a view has emerged within UPOV that in order to ensure that the value of protection is maintained in the event that molecular markers are used for DUS testing, a series of options for their use should be followed. In summary these are:

Option 1: Molecular characteristics as a predictor of traditional characteristics:

(a) Use of molecular characteristics which are directly linked to traditional characteristics (gene specific markers)

(b): Use of a set of molecular characteristics which can be used reliably to estimate traditional characteristics; e.g. quantitative trait loci

Option 2: Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics

Option 3: Development of a new system

Currently, with regard to the management of reference collections, most interest is being paid to Option 2 approaches. The aim is broadly to ensure that there would be no significant shift in the typical minimum distances as measured currently by "traditional" characteristics, if molecular markers are used. However, the problem is that in previous work, there is a lack of a clear relationship between molecular marker distances and differences in traditional characteristics, which would lead to the need to consider how to handle potentially different decisions on distinctness. The key is whether variety pairs, which are not distinct using traditional characteristics, would be judged as distinct using molecular characteristics (or *vice versa*) and the impact of such decisions on the value of PBR protection.

A major difficulty in pursuing such an option is the lack of sufficiently comprehensive datasets to be able to undertake the necessary statistical analyses in important crops. Thus the first stage of this project is to generate such a dataset for OSR using varieties from four EU MS. The data will then be used to examine the relationship between genetic and phenotypic distances in a number of ways.

4. EXPERIMENTAL APPROACHES

The work plan for the project was to:

(i) standardise conditions for the use of an agreed set of SSRs,
(ii) analyse c. 410 OSR varieties from different EU MS with these SSRs,
(iii) analyse the data produced, including estimates of genetic and phenotypic distances, and compare the distances in different ways,
(iv) validate these approaches in a field trial.

4.1 Selection and Standardisation of Markers.

At the start of the project, NIAB circulated DNA samples from 10 OSR varieties to GEVES and DIAS, along with the sequences of 29 SSRs and a draft analytical protocol. Both the SSRs and the protocol were derived from previous work undertaken by NIAB and GEVES. The microsatellite markers used were all obtained from publicly available sources (Kresovich. *et al*, 1995; SzewcMcFadden *et al*, 1996; Plieske *et al*, 2001; Tommasini *et al*, 2003, ukcrop.net/perl/ace/search/BrassicaDB). Participants from the three countries carried out analysis of these 10 samples and discussed their results. It was agreed that there were 14 markers that could be analysed and scored reliably. In addition, there were 11 that needed reviewing, and four markers were rejected.

Subsequently, NIAB circulated coded seed samples of 40 OSR varieties to the other laboratories, in order to test the usefulness of the markers, and the analytical and scoring methods of the laboratories. The samples were analysed using the markers as before, with an agreed protocol, and the results of these analyses were subsequently discussed. As a consequence, it was agreed that it would be desirable to use as standardised an approach to the genotyping as possible, e.g. to ensure that DNA of comparable quality is used in all laboratories, DNA extraction kits would be used. The PCR protocol e.g. in terms of the primer labelling strategy used and the source

of Taq polymerase would be agreed. Again, although the markers have been previously analysed using different platforms, it would improve comparability of data if all laboratories used the same equipment (this was achieved for the next phase of the project).

As a result of these preliminary analyses, the 23 SSRs detailed in Table 1 were selected for the genotyping elements of the project:

TABLE 1 – SSR markers chosen for analysis of WOSR varieties.

Original No.	Marker	5' primer sequence	3' primer sequence	Chromosome	No. of Alleles*
1	Ra2-E03	AGGTAGGCCCATCTCTCTCC	CCAAAACTTGCTCAAAACCC	10	3
2	BN12A	GCCGTTCTAGGGTTTGTGGGA	GAGGAAGTGAGAGCGGGAAATCA	13	2
3	BN26A	TAAACTTGTCAGACGCCGTTATC	CCCGTAAATCAAGCAAATGG	unknown	1
4	CLONE33	GTTTGTGTTGCAATTATTCCCA	CCTGCATTGCGAAAATATAATC	Unknown	3
5	LS107	GTTAAGTGTGGCGTTAGAGG	CCTTGGTACATGCCACTGAA	Unknown	3
6	MB5	AACATCTTTTTGCGTGATAT	AATAGCATTGAAGCCTTAC	Unknown	2
8	Na10-H03	GAGCTGGCTCATTCAACTCC	CACAATTTCTCAGACAAAACGG	Unknown	2
9	Na10-E02	TCGCGCATGTAATCAAAATC	TGTGACGCATCCGATCATAC	5	3
10	Na12-D04	ACGGAGTGATGATGGGTCTC	CCTCAATGAAACTGAAATATGTGTG	6	1
11	Na12-A02	AGCCTTGTTGCTTTTCAACG	AGTGAATCGATGATCTCGCC	16	5
12	Na12-E02	TTGAAGTAGTTGGAGTAATTGGAGG	CAGCAGCCACAACCTTACG	Unknown	4
14	Na14-H11	GGATGTTTTCACAGACCCTG	CTTTGCAGGTATGAACACGC	Unknown	4
15	Ol09-A06	TGTGTGAAAGCTTGAAACAG	TAGGATTTTTTTGTTCACCG	12	3
16	Ol10-B01	CCTCTTCAGTCGAGGTCTGG	AATTTGGAAACAGAGTCGCC	17	4
17	Ol10-BF11	TTTGGAACGTCCGTAGAAGG	CAGCTGACTTCGAAAGGTCC	11	2
19	Ol11-B05	TCGCGACGTTGTTTTGTTC	ACCATCTTCCTCGACCCTG	3	3
20	Ol11-G11	GTTGCGGCGAAACAGAGAAG	GAGTAGGCGATCAAACCGAG	3/13	3
21	Ol12-F02	GGCCCATTGATATGGAGATG	CATTTCTCAATGATGAATAGT	9	4
22	Ol13-C12	AGAGGCCAACAAAGAACACC	GAAGCAGCACCAGTGACAAG	13	3
23	Ra1-F06	ACCAAAATGTGTGAAGCCAC	CTTGTGGCCAGATTCATCAC	6	6
24	Ra2-A05	GCTAGTTACGGGCGG	AAACGACATCGGCAAGAAG	7	2
25	Ra2-A11	GACCTATTTTAATATGCTGTTTTACG	ACCTCACCGGAGAGAAATCC	9	4
27	Ra2-E11	GGAGCCAGGAGAGAAGAAGG	CCCAAAACTTCCAAGAAAAGC	3	6

* number of alleles found from analysis of whole set of 410 varieties, see below

The analytical protocol adopted can be summarised as follows:

DNA Preparation: 40 to 50 seeds of each variety were germinated on moist filter paper in the dark and harvested once the cotyledons had emerged from the testa and the seedlings were large enough to handle. The seedlings were cut from the roots, and 30 seedlings collected in a bulk to represent each variety were freeze dried. The dried seedlings were extracted using Qiagen DNeasy 96 Plant extraction kits in accordance with the manufacturer's instructions.

DNA Amplification: PCR reactions were prepared with 1 µl DNA template (nominally 10 ng), 1 µl x10 PCR buffer, 1 µl 25 mM MgCl2, MgCl2, 1 µl 5 mM primer pairs, 0.1 µl 20 mM dNTP, 0.1 µl 5U/µl TAq polymerase and water to 10 µl.

Use of Markers: The fluorescently labelled primers, suitable for the laboratory's instrument system, were synthesised for each laboratory. All fragments were amplified using the following PCR cycling conditions: 92ºC for 120 seconds, followed by 35 cycles of 92ºC for 30 seconds, then 55ºC for 30 seconds, then 72ºC

for 60 seconds followed by 72ºC for 600 seconds. Fragments were visualised using a MegaBace instrument (DIAS), Licor, and subsequently ABI 3130XL Genetic Analyser (GEVES) and an ABI 3100 Genetic Analyser (NIAB).

4.2 Selection and Molecular Analysis of Varieties.

It was agreed that the project should analyse only those varieties from the participating countries which were lines (i.e. no hybrids) and fertile (no male-sterile lines). Using these criteria, each of the partners produced a list of the relevant varieties from their country. The lists were collated at NIAB, and a set of 410 varieties compiled. Note that the names of the varieties are not supplied in this Report, but are available from NIAB if required. The molecular analyses were carried out on seed samples obtained from the reference collection at NIAB. If no seed of a variety was available at NIAB, seed samples were supplied by the partners from their collections. The samples were coded (to remove variety names) and the appropriate ones re-distributed to the laboratories undertaking the genotyping work. The appropriate permissions from the plant breeders to utilise some of the varieties in this final list for the experimental purposes within the project were obtained by CPVO where necessary.

The total list of 410 varieties to be analysed was divided between the three laboratories - 190 varieties to NIAB, 190 to GEVES and 70 to DIAS. In addition to these, 5 coded samples from the original set of 40 varieties were included, and of the 190 sent to NIAB and GEVES, 40 of these varieties were common to both, for quality control purposes. The varieties were analysed by each laboratory using the agreed set of SSRs and the protocol outlined above. It should be noted that by this stage, all laboratories were using a capillary-based platform for the molecular analyses. The raw data were compiled and sent to NIAB for collation and inspection. The data were collated into an Excel spreadsheet, containing the band molecular weights ("bins") of detected bands along with the associated peak heights, for each marker used (see Table 2, for example of the data format).

TABLE 2- Example of the collated molecular data.

		DK data		F data		UK data	
		Nominal Peak Size					
Sample	Marker	278	287	278	288	281	290
WOSR001	M2 Bn12A	12455	8366	787	659	7596	6348
WOSR002	M2 Bn12A	0	14386		1272	829	6772
WOSR003	M2 Bn12A	0	15254		1140		7174
WOSR004	M2 Bn12A	15878	11492	419	346	7362	6341
WOSR005	M2 Bn12A	49608	37177	682	550	7329	6268
WOSR006	M2 Bn12A	16199	11636	784	621	7751	7033
WOSR007	M2 Bn12A	16457	11516	1080	884	7853	7122
WOSR008	M2 Bn12A	44263	28614	1465	1120	7756	7177
WOSR009	M2 Bn12A	16053	11998	767	595	7580	6493
WOSR010	M2 Bn12A	12953	10251	839	629		5931

4.3 Phenotypic Analysis of Varieties.

It was agreed that each of the partners would provide the phenotypic data available from their national records for any of the 410 varieties from the selected variety set. Since it was thought that using data scored only as UPOV Notes would result in rather "clumped" data and consequently rather crude phenotypic distance estimates, it was further agreed that phenotypic data would be provided (i) in the form of UPOV Notes for each characteristic listed in the CPVO Technical Protocol, and (ii) as collated variety means for each appropriate continuously assessed characteristic from each of the years 2003, 2004 and 2005 for which data were available.

All of the data were supplied to NIAB for collation – Table 3 gives an example of the raw data format.

TABLE 3- Example of format of phenotypic data.

		Plant: height (at full flowering)	Plant: total length including side branches	Siliqua: length (between peduncle and beak)	Siliqua: width	Siliqua: length of beak	Siliqua: length of peduncle	Tendency to form inflorescences in year of sowing for spring sown trials	Tendency to form inflorescences in year of sowing for autumn sown trials	Cotyledon: length	Cotyledon: width	Leaf: number of lobes (fully developed leaf)	Time of flowering	Flower: length of petals
		Characteristic No. - UPOV Notes								Characteristic No. - Continuous Data				
		14	15	16	17	18	19	20	21	2	3	7	9	11
ID	NAME													
	CAPRICORN		5	6	6	6	7		2	13.6	25.1	5.03	200	18.3
347	Sparta													
348	Hobson													
1299	SILEX		6	5	5	5	5		1	*	*	5.47	198	17.9
536	ASKARI		7	4	5	5	5		1	13.8	24.5	4.73	201	17.3
460	Idol													
488	Samourai													
780	PRESTOL		4	4	5	4	4		3	10	18.3	4.57	190	15.5
541	VIVOL		6	4	5	6	4		3	12.2	19.6	5.93	200	17.6

Table 4 shows the amount of variety data supplied by each country, and the number of individual data points.

TABLE 4- Phenotypic Data Supplied by the Partners.

Partner	Number of varieties with phenotypic data supplied	Number of data points
BSA	290	24,333
DIAS	156	7,664
GEVES	255	10,044
NIAB	143	11,271

In all, data from four countries were supplied for 40 varieties, whilst 84 varieties had data from three countries, 141 varieties had data from two countries and the remainder had data from only one country. One variety had no data from any country. Some of the variety phenotypic data sets were not complete, with some characteristic data missing, or some years data missing, or both.

4.4 Statistical Analysis of Molecular and Phenotypic Data.

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, which are detailed in the Results section below and described in full in Annex 1.

In brief, for the GD estimates, NIAB converted the finalised and validated SSR data (see Results below) 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 exchanges and their robustness validated using Mantel statistics.

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 utilised. The one finally used for the Notes data was:

> Establishing where possible the MODAL note (with a maximum of 3 sites/centres 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

The approach finally used for considering measured data for a defined morphological trait converted to Notes form was:

> Undertaking a REML (restricted maximum likelihood) analysis of the note data and rounding (to integer) to give whole numbers for PD computation. These are referred to as NxM (Notes from Measured) and while included in specific analyses in the consolidated statistical report should not be considered as a main-stream feature of the statistical work.

In addition, methods for combining the data types were investigated. There are a number of characteristics for which data in both note and measured form was supplied. In these cases it is not sensible to combine the data, as essentially these derive from the same source, albeit summarised in a different way. There was not a 1:1 concordance between the data types from all centres and across all years, so it was necessary to consider the data which is available only as one specific type, and how to convert one type to another. Methods for computing PDs from mixed data types, e.g. Gower's method (GenStat procedure) were evaluated.

4.5 Variety Pairs Evaluated by Field Trials.

It was decided that a field trial would be sown which evaluated various variety pairs/groups. Because of time constraints, the choice of these was based on an initial analysis of distances (not from the full data matrix). The varieties selected were ones which appeared to be similar in phenotypic distance but were easily separable (i.e. dis-similar) in terms of genetic distance, and vice-versa.

On this basis, crop experts from each partner identified and selected appropriate varieties (Table 5). Seed of each of these variety pairs was exchanged, so that the pairs were replicated in all countries. The pairs were then sown in side-by-side plots in field trials in UK, F, DE and DK in autumn 2006, for visual examination and recording. All of the characteristics included in the CPVO Protocol were recorded in the following growing season. The data were analysed using the normal DUS testing procedures in each country.

TABLE 5- The variety pairs selected for field trials

Varieties Selected By			
DK	**UK**	**F**	**DE**
Caracas/Castille	Smart/Eclipse	Bellini/Caraco	Hektor/Zenith/Libea
Californium/Sansibar	Calvacade/Action	Bellini/Cannelle	Hektor/Casanova
WRG257/KW1097	Action/Caiman	Lewis/PR54W04	Casanova/Siska
SW9991097/WRG257	Action/Fortis	Capvert/Mohican	Apex/NKVictory
	Fortis/Limpet	Boston/Remy	
	Fortis/Licontent	Remy/Solomon	
	Licontent/Calvacade	Cannelle/Remy	
	Smart/Limpet	Remy/Splendor	
		Caraco/Mohican	
		Licorne/Mohican	

5. RESULTS

5.1 Validation and Thresholding of Molecular Data.

Before being able to compute any genetic distances, the raw molecular data from the three laboratories had to be in a condition which allowed the generation of an agreed matrix of band absence/presence.

The initial results showed microsatellite profiling of OSR to be a robust and rugged tool. However, the multi-allelic nature of the profiles was problematic when determining the agreed profile of a variety. The same allele could be identified with a different size (molecular weight), depending on the size standards and detection method used. Moreover, the relative response for each allele within a profile will depend on the proportion of individuals within the bulked sample possessing that allele. The relative response may also be affected by the efficiency of PCR for the fragments being amplified. The size of the fragment being amplified and presence of competing fragments have an effect on PCR efficiency, and these effects may not be consistent between laboratories. Whilst the three laboratories generated broadly similar profiles, the relative response for each allele within the profile could vary between laboratories, leading to minor peaks being called as alleles at one laboratory but not in another. This suggested the need for a set of objective "rules" for allele calling that would allow the differing profiles at each laboratory to be described in the same way. These allele calling rules - termed thresholding – were validated by analysis of data for a small number of varieties analysed at more than one laboratory and then applied to data for a much larger variety set, where varieties may have been assayed at one laboratory only. This process should allow microsatellite data from different laboratories to be unified in a database.

5.1.1 The Thresholding Process.

The options for thresholding include (i) absolute thresholding, and (ii) relative thresholding, either using a global threshold value or applying independent threshold values for data from each laboratory (see Figure 1).

Absolute thresholding entails rejecting all allele peaks below a certain threshold value (Figure 1A). All data generated in capillary electrophoresis genetic analysis systems will have been subject to absolute thresholding to a degree through pre-set threshold values in the data collection software and through inspection by the system operator. Both of these absolute thresholds are used to ensure that 'noise' in the detection system is not reported as data. Establishing rules based solely on absolute thresholding is complicated by a number of factors, including within and between batch variation in PCR efficiency, between batch variation in electrophoresis and the use of different measuring systems by instrument manufacturers. Hence it was considered that the use of absolute thresholding was not appropriate.

Relative thresholding requires that the allele with the largest response, for example peak height, within a variety profile is identified. All other peaks in the profile will then be scored as alleles if their response exceeds a predetermined percentage of this largest peak. Relative thresholding may be applied in two ways; the same predetermined global threshold is applied at all laboratories for all markers (Figure

1B), or empirically determined laboratory and/or marker specific thresholds are used (Figure 1C).

There are advantages and disadvantages to all of these approaches. When relative thresholding is applied using a global threshold, differences between laboratories in PCR efficiency for different sized fragments could result in different allele scores being recorded. On the other hand, if global thresholding is applied using a high threshold value (for instance 75%), the resultant allele calling produces a conservative, cautious set of allele data which does not exploit the full potential of these markers. Variation would also be introduced where the maximum peak observed in a particular analysis differs between laboratories. Global thresholding applied by including peak heights above a low threshold (for instance 15 % and above that of the maximum peak height, i.e. trim off the worst) results in a discriminating set of allele data, but increases the risk of potential variation between laboratories. When relative thresholding is applied using empirically determined laboratory and/or marker specific thresholds, considerable effort is required to determine the values that will be used.

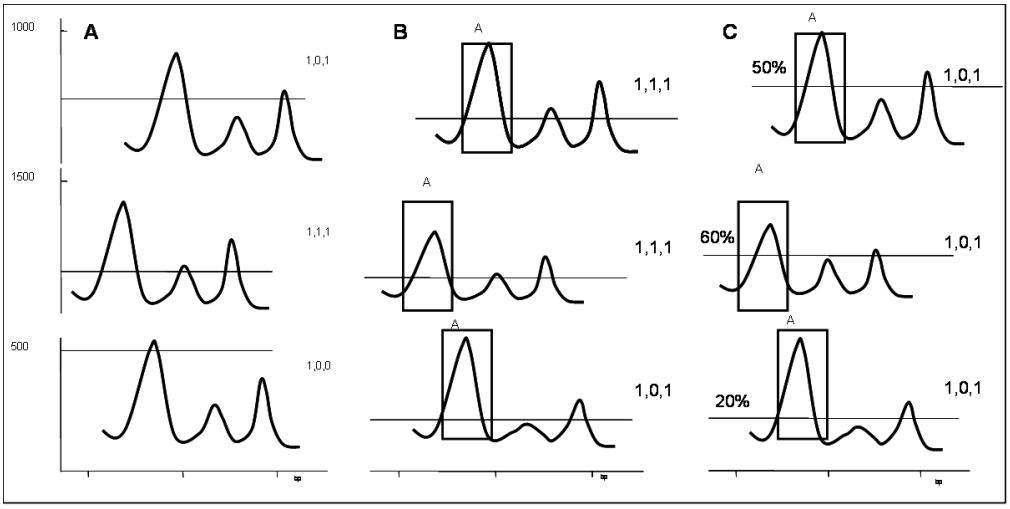
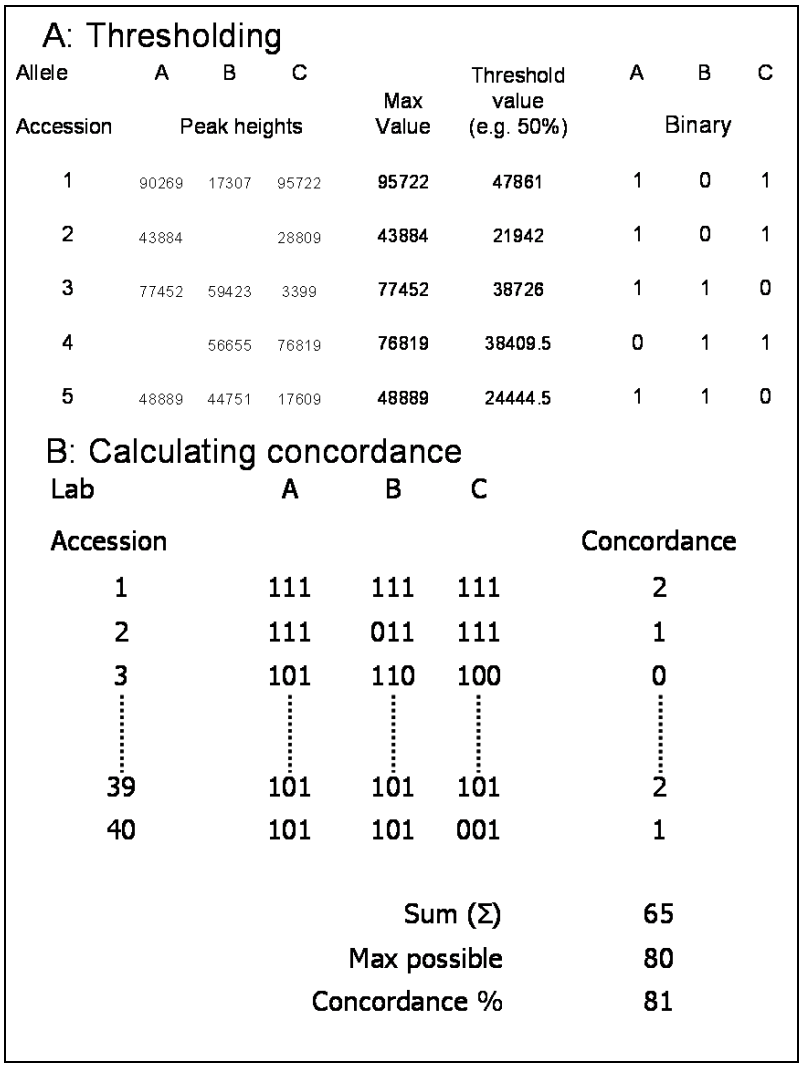


Figure 1: Strategies for thresholding. A) absolute thresholding where peaks are scored as present if they exceed a predetermined level of instrument response e.g. greater than 500 units in this example. B) Relative thresholding with a common threshold at all labs. The largest peak is identified for each sample (box). Alleles are scored in all labs if their peak height is greater than 25% (in this example) of the peak height seen in the largest peak in the sample trace C) Relative thresholding with an empirically derived threshold at each lab. The largest peak is identified for each sample (box). Thresholds are calculated for each marker at each laboratory. Alleles are scored if their peak height exceeds the threshold percentage of the peak height seen in the largest peak in the sample trace.

5.1.2 Production of Concordant Data Sets.

These different relative thresholding approaches were evaluated by their application to the raw data sets generated by the three partner molecular laboratories for the initial 40 varieties. The methods and results are described in detail in Jones *et al.*, 2008 (in press). Briefly, the data were subjected to the varying approaches using different levels of thresholding, and the degree of concordance (see Figure 2) achieved between laboratories for each of them calculated and compared.

Using this approach, it was demonstrated that it was possible to produce levels of concordance between the laboratories of any desired percentage, although at higher levels, the information content of the markers, and the number of markers that produced usable information, declined.



A: Thresholding

Allele	A	B	C		Threshold	A	B	C
Accession	Peak heights			Max Value	value (e.g. 50%)	Binary		
1	90269	17307	95722	95722	47861	1	0	1
2	43884		28809	43884	21942	1	0	1
3	77452	59423	3399	77452	38726	1	1	0
4		56655	76819	76819	38409.5	0	1	1
5	48889	44751	17609	48889	24444.5	1	1	0

B: Calculating concordance

Lab	A	B	C	
Accession				Concordance
1	111	111	111	2
2	111	011	111	1
3	101	110	100	0
⋮	⋮	⋮	⋮	⋮
39	101	101	101	2
40	101	101	001	1
			Sum (Σ)	65
			Max possible	80
			Concordance %	81

Figure 2: Thresholding and Concordance. A): Application of Thresholding. Set out the peak height data for each marker in an array, with variety data in a row. Identify the largest peak for each variety and tabulate the maximum peak heights in a second array. Calculate a threshold value as a percentage of the maximal peak height and tabulate this in a third array. Convert the peak height data for each allele into a binary form, scoring 1 if it exceeds the threshold value and scoring 0 if it does not. B): Calculating concordance. The data for each variety were compared across the three laboratories and scored according to the degree of agreement. Where all three laboratories agreed a variety profile, the result was scored as "2", where two laboratories agreed the result was scored as "1" and where there was no agreement a score of "0" was given. The total score for a combination of thresholds at the three laboratories was calculated and then expressed as a percentage of the maximum score (i.e. where all laboratories agree all variety profiles completely). This percentage was termed the concordance score.

It was agreed by the partners that levels of 90 and 95% concordance would be appropriate to accept as a working basis for the genetic distance estimates. This

meant that two datasets would be produced from the 410 variety set, utilising either 18 (at 90%) or 11 (at 95%) SSR markers. (In practice, a third set of data, with 90% concordance and comparable performance for the internal control varieties was also used – this utilised 14 markers). Although this procedure reduced the number of markers used for GD estimates, it was considered that the robustness and reliability of these estimates would be much improved, since the less useful (i.e. less informative) and/or more difficult markers will have been excluded, and problems arising from the use of bulked samples should be reduced.

On the basis of this approach, a final molecular data set for the 410 varieties (subsequently reduced to 335 varieties following scrutiny of the morphological data) was produced, consisting of the agreed allele scores for the appropriate SSR markers, in a binary format. These data were subsequently used by the statisticians for the estimations of genetic distances.

5.2 Statistical Analyses of Data.

Following scrutiny of the datasets by the statisticians within the project, a final set of agreed molecular and morphological data was produced, which took into account the availability of appropriate amounts of each type of data for a variety, the number and pattern of missing data values, etc. This final set comprised information from 335 varieties. A full report of all of the statistical analyses that were undertaken with the molecular and morphological datasets is included in Annexes 1 and 2. The following sections highlight the main approaches used and summarises the principal outcomes.

5.2.1 Genetic Distances - Introduction

There are a large number of methods that have been used to estimate genetic distances, some of the more common of which are outlined below. They generally differ in how the distance measure is computed, which can be illustrated as follows. Assume that we have two units (or varieties in the present context), called i and j, and that the distance between them is given by d_{ij} . Further, assume that in the allele descriptions, the absence of a band = 0 and the presence of a band = 1. The first step is to produce a matrix, thus:

		j	
		1	**0**
i	**1**	**a**	**b**
	0	**c**	**d**

The various distance indices then vary in how the different states (e.g. presence of a band in both i and j, presence in one, absence in another – the a, b, c, d notations above) are considered. Some examples are:

a. Dice (Nei & Li)

$$d_{ij} = \frac{b+c}{2a+(b+c)}$$

The Dice dissimilarity index corresponds to the percentage of different bands between the units i and j, without considering double absences (0/0), and gives more weight to bands present in both units.

b. Jaccard

$$d_{ij} = \frac{b+c}{a+(b+c)}$$

This dissimilarity index corresponds to the percentage of different bands between the units i and j, without considering double absences.

c. Ochiai

$$d_{ij} = 1 - \frac{a}{\sqrt{(a+b)(a+c)}}$$

d. Sokal & Sneath

$$d_{ij} = \frac{2(b+c)}{a+2(b+c)}$$

e. Simple Matching (Sokal & Michener)

$$d_{ij} = \frac{a+d}{a+b+c+d}$$

This index takes no account of any difference between presence and absence (i.e. it only takes into account the fact that a band is present in both units, or absent in both).

5.2.2 Genetic Distances – Results.

Using the software package DarWin (CIRAD), these distance estimates were calculated using the three thresholded datasets noted above – T1 (>90% concordance, 18 markers), T2 (>95%, 11 markers), and T3 (>90% + internal controls, 14 markers). For the initial analyses, an individual "band-by-band" approach was used, i.e. the absence/presence of each band recorded at all of the SSR loci was taken for computation. The detailed data are available in Annex 1, and only illustrative examples are shown below

For the T1 dataset, the distribution of the distances in the Dice analysis is given in Figure 3. The mean distance was 0.254.

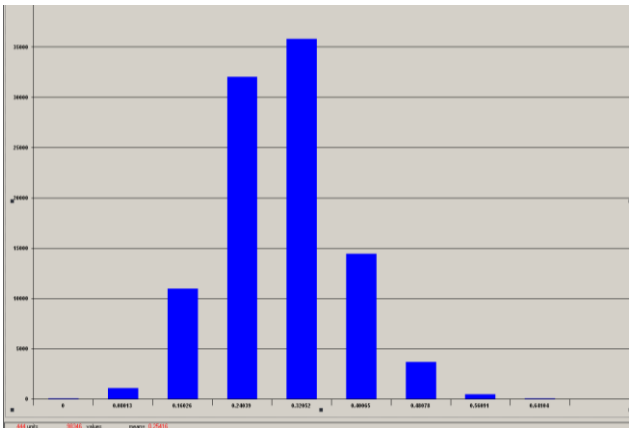


Figure 3: The distribution of the Dice genetic distances from the T1 dataset.

Although the exact shape of the distribution curves and the mean values for the other analyses differed slightly, all showed essentially the same features. The correlations between the various indices are shown in Table 6.

TABLE 6. Correlations between various genetic distance indices from the T1 dataset

	Dice	Jaccard	Ochiai	Sokal & Sneath	Simple Matching
Dice	1				
Jaccard	0.996	1			
Ochiai	0.999	0.996	1		
Sokal & Sneath	0.990	0.987	0.990	1	
Simple Matching	0.980	0.994	0.980	0.973	1

The high values of these correlations indicate that all of the approaches are broadly comparable – there does not seem to be any particular advantage in this instance to using one or other of the estimates.

These data were analysed in various other ways, including principal co-ordinates analysis (to check for any clustering of the data) and the production of "trees" of various kinds. Using the Dice distance set, the PCO analysis was as in Figure 4

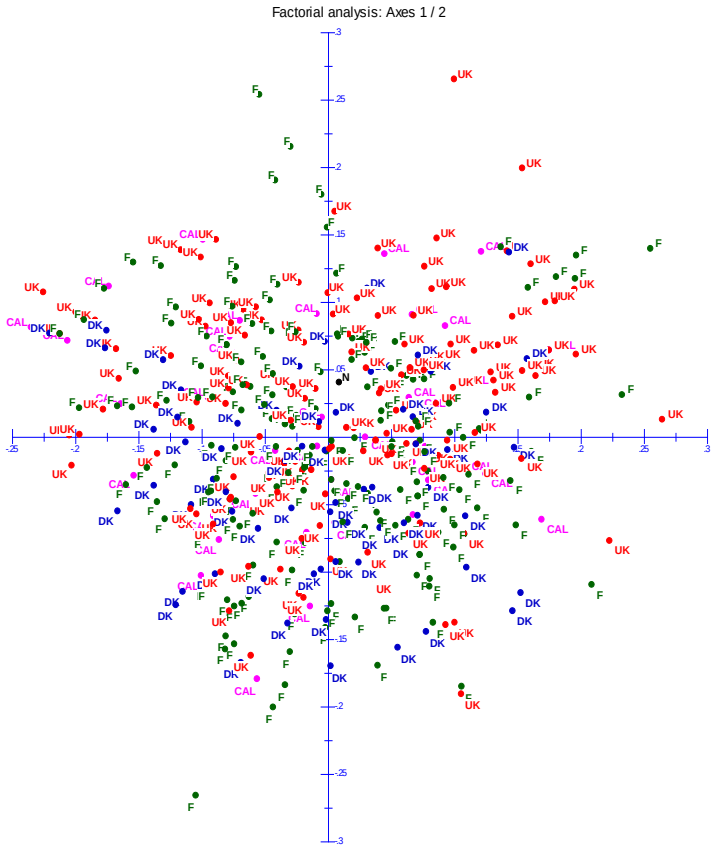


Figure 4: Principal co-ordinates analysis of the Dice distance set from the T1 data. The coloured spots represent data points from different sources:- pink –

calibration set, blue – results from DK, green – results from F, red – results from UK. Axes 1 and 2 represent 17 % of total variation.

There is no obvious clustering evident in this analysis, indicating that there is no apparent bias in the data arising from the seed source (country of origin of samples) or from the laboratory. A similar conclusion could be drawn from the various cluster analyses that were carried out – Figure 5 for instance shows a Ward hierarchical cluster tree of the Dice distance data from T1 dataset.

Similar results (not shown) were obtained for the analysis of the other datasets, although the absolute values of the mean distances varied (the mean for the Dice analysis of T2, for example, was 0.218).

From these initial analyses, it can be concluded that:

- the verified molecular data generated within the project were demonstrably "fit for purpose" and hence could be used for any subsequent analyses;
- the number of markers used affects the absolute values of the distances, but not their correlations;
- the choice of distance measure does not affect the utility of the data, and the choice of method in this instance is not critical.

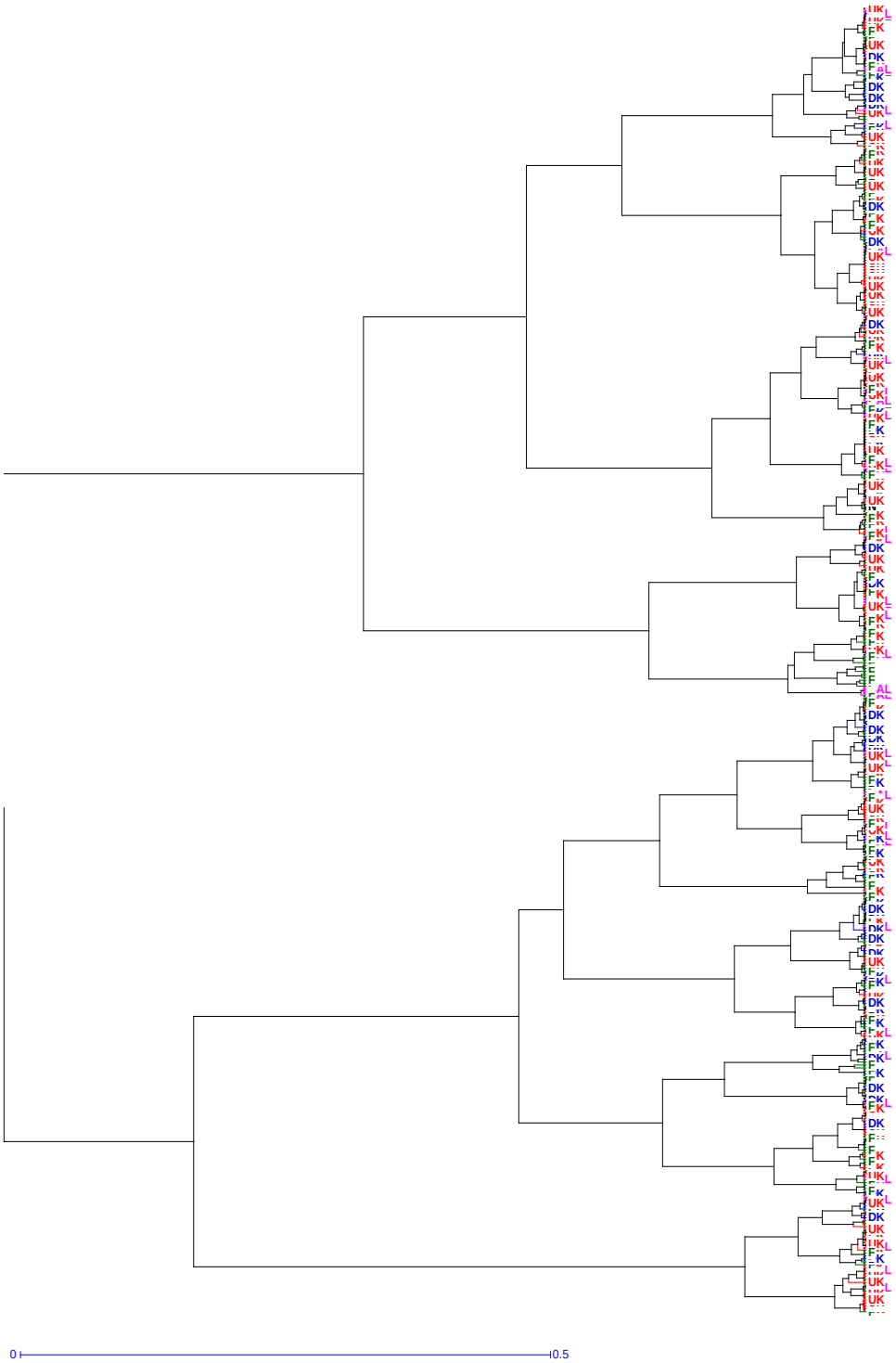


Figure 5: Ward hierarchical clustering. The Dice distance data from T1 were used. Colours as in Figure 4.

5.2.3 Genetic and Phenotypic Distances.

The thresholded molecular data (T1, 2 and 3 as above) and the finalised collated morphological data, provided by the four partners as Notes (N) or as Measured values (M) and treated as detailed in Annex 1, were analysed in a further series of computations. Data from a total of 335 varieties was used. Work was also carried out

on data in the form of Notes derived from measured values (NxM). This phase of statistical analyses concentrated on the assessment of the robustness of the similarity (distance) estimates. In essence, if any results/conclusions differed markedly depending on the technical detail of the statistical method used to compute the similarities or distances, then there is only low level of robustness. In an ideal case with high robustness, results and hence decisions are exactly the same, irrespective of the methods utilised to arrive at these decisions.

With regard to the overall project objective (to examine the UPOV Option 2 approach), the major comparisons with respect to robustness are to compare the similarities or distances derived from morphology with those from the molecular data, and to assess the robustness in terms of the method of computation of the similarities. A wide and diverse range of techniques was used to examine this. The salient points of the analyses are summarised below and full details can be found in Annex 1.

a. In all cases, similarities (1-distances) were used for all computations.
b. These analyses concentrated on the assessment of the robustness of the distance estimates and the correlations between the genetic and morphological distances.
c. A "marker-by-marker pattern" approach was used for the SSR data, and similarities based on both City Block (CB) and Euclidean (E) distance used for all data types. In addition, analyses based on Jaccard (J) distance were undertaken, to form a link with the analyses previously presented (above).
d. The similarities in all cases were high, with no values below 0.5 .
e. Attention was focussed on variety pairs in the tails of the similarity distributions, i.e. those which are either very similar or very dis-similar.
f. The results from the three sets of molecular marker analyses (T1, 2 and 3) correlated highly in all cases (>0.79), thus confirming the previous results with a different analytical approach.
g. Although there is no *a priori* reason to suggest that they should, the morphological notes and measured (N and M) data were also correlated (c. 0.67).
h. There was effectively no correlation (<0.1) between any of the morphological and molecular estimates of similarity.
i. An assessment of the situations where all morphology and molecular methods agreed in assessing variety pair similarity showed that for the top 1000 similarities, there were only 3 cases of agreement out of a possible 55000 pairs.
j. Shifts in the ranking order of similarity were also assessed, which again indicated that there were instances of large differences in morphological distance not being reflected in molecular distance (and *vice versa*).
k. The distributions of the morphological and molecular distances estimated in the various ways were plotted against each other; in all cases there were at best only very weak correlations observed.
l. All of these results were independent of the method of analysis, again in agreement with the previous conclusions.

Figure 6 below illustrates the general relationship found between the "phenotypic" similarity derived from the morphological data against the "genetic" similarity from the molecular data. Whilst the data were refined later in the project in consensus form, this Figure highlights the three main regions of interest: (i) potential for agreement between "morphology" and "molecular" (blue ellipse ~ top right-hand corner); (ii) cases where there is a very high level of observed similarity in terms of "Morphology" but weak when assessed by "Molecular" (red ellipse ~ top left-hand corner); (iii) cases where there is a very high level of observed similarity in terms of "Molecular" but weak when assessed by "Morphology" (green ellipse ~ lower right-hand corner).

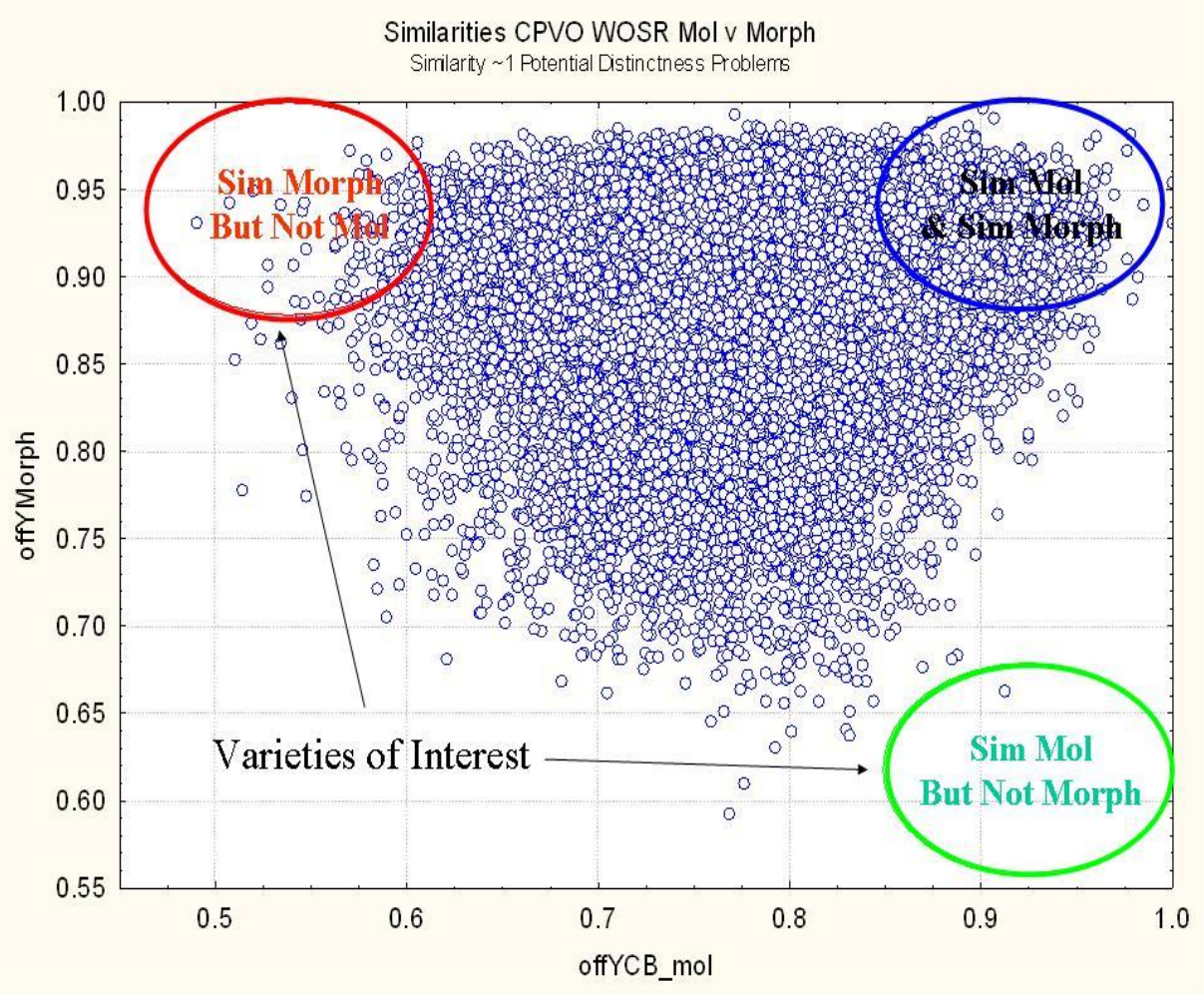


Figure 6: Comparison of Distance Estimates. The Relationship between "Phenotypic Distance" as Calculated from Morphological Data and "Genetic Distance" Calculated from SSR Data

The relationship observed follows the general triangular shape that has been reported previously (e.g. Dillmann and Guerin, 1998). It is evident that this similarity relationship is weak or virtually non-existent, and hence morphological similarity cannot, for the majority of variety pairs, be adequately predicted from the genetic similarity obtained from the molecular data.

Correlations between the various distance measures were also made and are summarised in Table 7. The results from the three sets of molecular marker analyses – see the *green block* - correlated highly in all cases (>0.79), thus confirming the results above with a different analytical approach. Of special note is the high correlation (0.95) when comparing the CB vs. E results. Thus, for the three sets of molecular data (T1, T2 and T3), the computational method applied (CB or E) gives highly correlated relationships. This indicates that the molecular pair-wise similarities are sufficiently robust and not unduly influenced by choice of method when establishing the similarity coefficients.

The relationships between morphological Note data by two computational methods (NCB and NE) and morphological measured data (MCB and ME) - the *yellow block*- were 0.95 and 0.93 respectively, showing no strong method-specific requirement and good robustness. The morphological notes and measured (N and M) data were also correlated (correlation coefficient c.0.67). Notes from Measured by City Block (NxMCB) similarities were only weakly related with NCB (correlation c.0.62), as were NxME v NE (correlations coefficients also of 0.62). NxMCB v NXME had a slightly weaker correlation of 0.59.

In the *orange* coloured block of the Table it can be seen that there was effectively no correlation (<0.1) between any of the morphological and molecular estimates of similarity. This result was consistent across data types and computational methods.

TABLE 7. Correlations between various measures of phenotypic and molecular similarity.

Correlations	MCB	ME	NCB	NE	NXMCB	NXME	T1CB	T1E	T2CB	T2E	T3CB	T3E
MCB	1	0.93	0.65	0.62	0.96	0.93	0.07	0.04	0.07	0.06	0.08	0.05
ME	0.93	1	0.66	0.67	0.89	0.95	0.08	0.06	0.08	0.07	0.08	0.07
NCB	0.65	0.66	1	0.95	0.62	0.63	0.06	0.03	0.08	0.06	0.08	0.05
NE	0.62	0.67	0.95	1	0.59	0.62	0.07	0.03	0.08	0.06	0.08	0.06
NXMCB	0.96	0.89	0.62	0.59	1	0.94	0.06	0.04	0.06	0.05	0.07	0.05
NXME	0.93	0.95	0.63	0.62	0.94	1	0.06	0.04	0.06	0.05	0.07	0.05
T1CB	0.07	0.08	0.06	0.07	0.06	0.06	1	0.95	0.83	0.79	0.93	0.87
T1E	0.04	0.06	0.03	0.03	0.04	0.04	0.95	1	0.82	0.86	0.89	0.93
T2CB	0.07	0.08	0.08	0.08	0.06	0.06	0.83	0.82	1	0.95	0.85	0.84
T2E	0.06	0.07	0.06	0.06	0.05	0.05	0.79	0.86	0.95	1	0.83	0.89
T3CB	0.08	0.08	0.08	0.08	0.07	0.07	0.93	0.89	0.85	0.83	1	0.95
T3E	0.05	0.07	0.05	0.06	0.05	0.05	0.87	0.93	0.84	0.89	0.95	1

Pair-wise Correlations 55000 points

MCB	Morphology Measured City Block
ME	Morphology Measured Euclidean
NCB	Morphology Notes City Block
NE	Morphology Notes Euclidean
NXMCB	Morphology Notes derived from REML Means of Measured Characteristics City Block
NXME	Morphology Notes derived from REML Means of Measured Characteristics Euclidean
T1CB	Molecular Marker Set1 City Block
T1E	Molecular Marker Set1 Euclidean
T2CB	Molecular Marker Set2 City Block
T2E	Molecular Marker Set2 Euclidean
T3CB	Molecular Marker Set3 City Block
T3E	Molecular Marker Set3 Euclidean

Very similar conclusions could be drawn from formal analysis of matrices of pair-wise similarities by Mantels tests and when using Spearman's Rank Correlation. Also, the use of the Jaccard coefficient (treating the molecular data as binary as opposed to a pattern) was shown to have no significant impact on the resulting comparisons of similarity – see Annex 1 for details.

From all of these analyses it can be concluded that – in agreement with the previous results- the method of assessment of distance is not important in this instance, although depending on data type some distance estimates are more appropriate than others**.** Furthermore, there is no evidence of any correlation between the genetic distances (from SSR data) and phenotypic distances (from morphological data) for the variety collection considered. From this**,** it can be concluded that it is not possible to apply a straightforward UPOV "Option 2" approach as originally conceived for the use of molecular markers to manage the reference collection in OSR. This leads to the conclusion that it is now clearly necessary to pursue other approaches for the application of markers in conjunction with morphological characteristics, e.g. within a GAIA-type or some other form of analysis.

5.2.4 Analysis of Morphological Characteristics.

This part of the statistical analysis was carried out by BSA and considered the evaluation of the morphological data. The aim was to identify appropriate statistical procedures for the analysis of morphological data both in the structure provided by the examination offices and the consolidated data developed by UK. Full details are given in Annex 1.

The consolidated data set consists of notes and measurements consolidated for countries and years. The characteristics were those in the UPOV Guideline, consisting of notes and measurements per country and year. An optimized dataset was prepared by dropping out those characteristics which are part of notes and measurements – the measurements were retained, but the corresponding notes not.

Since variables with large variances tend to have more effect on distance or similarity measures than those with small variances, it is recommended to standardise the variables (characteristics). In the SAS software package the DISTANCE procedure provides a convenient way to standardise each variable with it own method before measures are computed. Standardisation is not required if there is only one level of measurement, otherwise it is mandatory. Standardisation depends on the type of characteristic and scale level (nominal, ordinal, interval, ratio, see below).

The types of characteristics and their scale levels are well defined in the literature (UPOV2007: TGP/8/1 chapter 4). In summary:

Nominal scale: Nominal scaled qualitative data are qualitative data without any logical order of the discrete categories. Characteristics with only two categories (dichotomous characteristic) are a special form of nominal scales. The nominal scale is the lowest classification of the scales.

Ordinal scale: Ordinally scaled data are qualitative data of which discrete categories can be arranged in an ascending or descending order. They result from visually assessed quantitative characteristics. The distances between the discrete categories of an ordinal scale are not exactly equal. Therefore, an ordinal scale does not fulfil the condition to calculate arithmetic mean values, which is the equality of intervals throughout the scale. The Ordinal scale is higher classified than the nominal scale.

Interval scale: An interval scale is a quantitative scale without a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Interval scaled data may be distributed continuously or discretely. The interval scale is higher classified than the ordinal scale.

Ratio scale: A ratio scale is a quantitative scale with a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Ratio scaled data may be continuous or discrete. The ratio scale is the highest classification of the scales.

A wide range of distance and similarity measures were used to analyse the morphological data set (see Annex 1 for full details and definitions). A basic requirement for the application of a specific method is the type of data to be analyzed. It is important if the data set contains only one type of data or a combination of different types of data. In particular it is important if the dataset contains nominal scaled data. This is summarised in Table 8.

TABLE 8. Different types of characteristics and the appropriate methods for their analysis.

	Nominal		**Ordinal**	**Interval**	**Ratio**	**Combination nominal/ ordinal/ interval/ratio**
	two categories	>two categories				
Cityblock			X	X	X	
Euclidian			X	X	X	
Chebychev			X	X	X	
Cosinus	X					
Dice	X					
Jaccard	X					
M coefficient	X					
RR coefficient	X					
Kulczinski coefficient	X					
Gower's index	X	X	X	X	X	X

The essential objective of this work was to identify the most appropriate distance/similarity measure for variety comparisons. Before analysis, it was necessary to modify the data set somewhat, as follows:

The characteristics b1 (Seed: erucic acid; 1=absent, 9=present), b6 (Leaf: lobes; 1=absent, 9=present) and b13 (Production of pollen; 1=absent, 9=present)

were defined as ordinal characteristics instead of nominal. This is possible because of the absence of more than two categories. The characteristic b10 (Flower: Colour of petals; 1=white, 2=cream, 3=yellow, 4=orange-yellow) was dropped from the dataset. It is impossible (and indeed from the theoretical point of view forbidden) to handle this nominal scaled characteristic with four categories (colours) as ordinal, interval or ratio scaled characteristic.

The data were then analysed in a range of ways and the correlations between the results computed. The results are summarized in Table 9:

TABLE 9. Correlation s between various measures applied to the morphological data set.

Sample	Measure 1	Measure 2	Correlation Coefficient
Consolidated dataset	City block	Euclid	0.95687 (P<0.001)
		Chebychev	0.87801 (P<0.001)
		Gower	-0.92994 (P<0.001)
	Euclid	Chebychev	0.97336 (P<0.001)
		Gower	-0.81894 (P<0.001)
	Chebychev	Gower	-0.70844 (P<0.001)

The correlation coefficients varied from 0.71 to 0.97, indicating that the influence of the distance or similarity measure is appreciable. The best correlation was between "Euclidian" and "Chebychev" (0.97336), or between "City block" and "Euclid" (0.95687). From a theoretical point of view, "Gower's index" is the perhaps the best to use, allowing for the structure of the dataset. The best correlated measure to "Gower's index" was the "City block distance" (0.92994).

In conclusion:

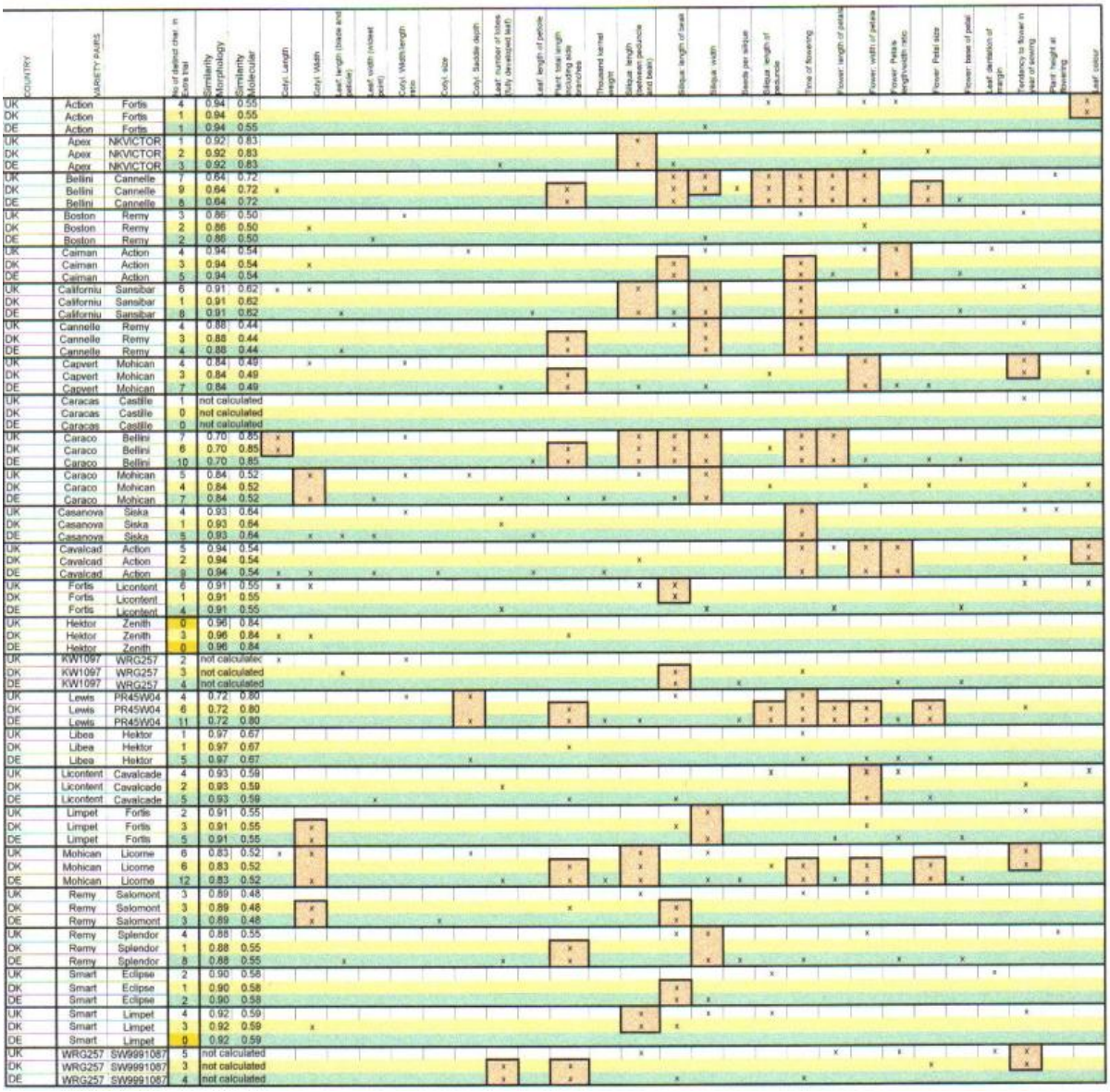
a) The different types of characteristics have to be taken into account, as there are nominal, ordinal, interval and ratio scaled characteristics
b) The "Gower's index" is the most appropriate procedure for the structure of the consolidated morphological dataset, because it is the only one which allows a combination of the present data types.
c) It is not allowed to use nominal scaled characteristics such as characteristic b10 (Flower: colour of petals; 1=white, 2=cream, 3=yellow, 4=orange-yellow) with more than two categories for evaluation of "City block distance"
d) For comparison of different distance measurements, dichotomous characteristics (e.g. b1, b6, b13) can be handled as ordinal characteristics. Nominal characteristics with more than two categories (b10) should be dropped.
e) The best correlated measure to "Gower's index" appears to be the "City block distance"

5.2.5 Analysis of Field Trial Data

5.2.5.1 Recording of Characteristics Individually.

Analysis of the side-side variety pairs in each of the four countries produced the results summarised in Tables 10A and B. The data from France is presented separately in Table 10B, as the use of GAIA does not require the same programme of recording as in the other countries. The varieties had been selected on the basis that they appeared to be similar in phenotypic distance, but were easily separable in terms of genetic distance, and *vice-versa*. Note that the UK data are based on differences at 2%, whereas the DK and DE data are on 1%. All are based on only one year of side by side data and not using COYD.

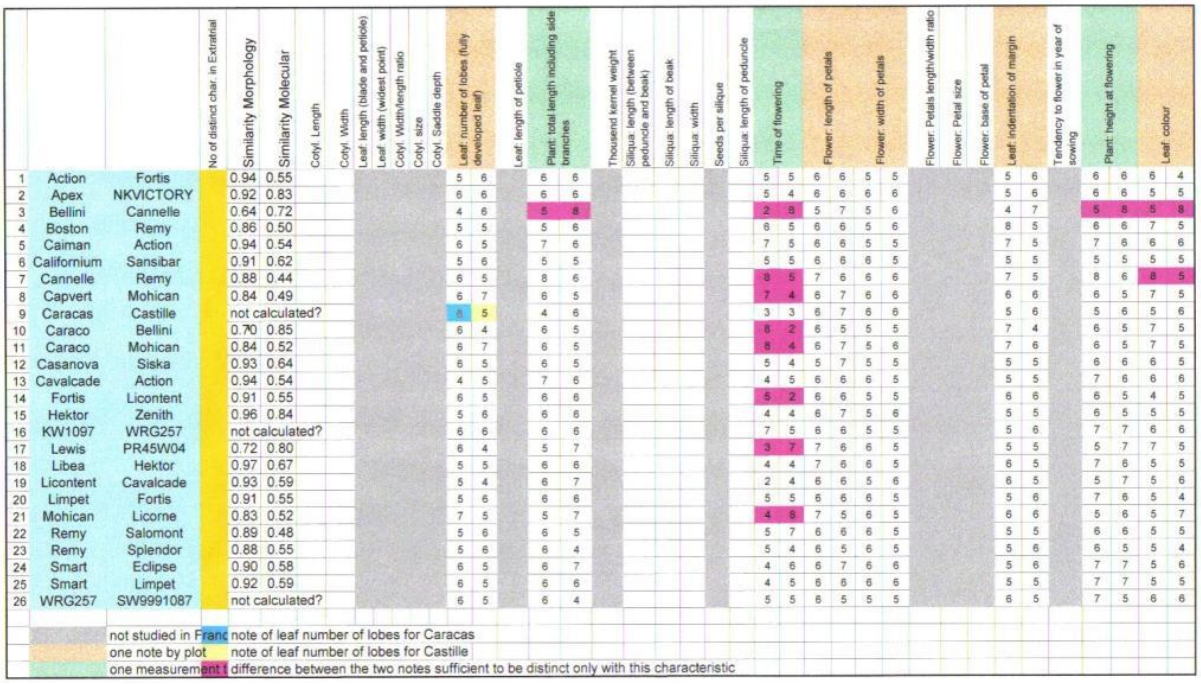
TABLE 10A. Summary of the field trial data from UK, DK and DE.



COUNTRY	VARIETY PAIRS		No. of distinct char. in side trial	Similarity Morphology	Similarity Molecular
UK	Action	Fortis	4	0.94	0.55
DK	Action	Fortis	1	0.94	0.55
DE	Action	Fortis	1	0.94	0.55
UK	Apex	NKVICTOR	1	0.92	0.83
DK	Apex	NKVICTOR	2	0.92	0.83
DE	Apex	NKVICTOR	3	0.92	0.83
UK	Bellini	Cannelle	7	0.64	0.72
DK	Bellini	Cannelle	9	0.64	0.72
DE	Bellini	Cannelle	8	0.64	0.72
UK	Boston	Remy	3	0.86	0.50
DK	Boston	Remy	2	0.86	0.50
DE	Boston	Remy	2	0.86	0.50
UK	Caiman	Action	4	0.94	0.54
DK	Caiman	Action	3	0.94	0.54
DE	Caiman	Action	5	0.94	0.54
UK	Californiu	Sansibar	6	0.91	0.62
DK	Californiu	Sansibar	1	0.91	0.62
DE	Californiu	Sansibar	8	0.91	0.62
UK	Cannelle	Remy	4	0.88	0.44
DK	Cannelle	Remy	3	0.88	0.44
DE	Cannelle	Remy	4	0.88	0.44
UK	Capvert	Mohican	4	0.84	0.49
DK	Capvert	Mohican	3	0.84	0.49
DE	Capvert	Mohican	7	0.84	0.49
UK	Caracas	Castille	1	not calculated	
DK	Caracas	Castille	0	not calculated	
DE	Caracas	Castille	0	not calculated	
UK	Caraco	Bellini	7	0.70	0.85
DK	Caraco	Bellini	6	0.70	0.85
DE	Caraco	Bellini	10	0.70	0.85
UK	Caraco	Mohican	5	0.84	0.52
DK	Caraco	Mohican	4	0.84	0.52
DE	Caraco	Mohican	7	0.84	0.52
UK	Casanova	Siska	4	0.93	0.64
DK	Casanova	Siska	1	0.93	0.64
DE	Casanova	Siska	5	0.93	0.64
UK	Cavalcad	Action	5	0.94	0.54
DK	Cavalcad	Action	2	0.94	0.54
DE	Cavalcad	Action	9	0.94	0.54
UK	Fortis	Licontent	6	0.91	0.55
DK	Fortis	Licontent	1	0.91	0.55
DE	Fortis	Licontent	4	0.91	0.55
UK	Hektor	Zenith	0	0.96	0.84
DK	Hektor	Zenith	3	0.96	0.84
DE	Hektor	Zenith	0	0.96	0.84
UK	KW1097	WRG257	2	not calculated	
DK	KW1097	WRG257	3	not calculated	
DE	KW1097	WRG257	4	not calculated	
UK	Lewis	PR45W04	4	0.72	0.80
DK	Lewis	PR45W04	6	0.72	0.80
DE	Lewis	PR45W04	11	0.72	0.80
UK	Libea	Hektor	1	0.97	0.67
DK	Libea	Hektor	1	0.97	0.67
DE	Libea	Hektor	5	0.97	0.67
UK	Licontent	Cavalcade	4	0.93	0.59
DK	Licontent	Cavalcade	2	0.93	0.59
DE	Licontent	Cavalcade	5	0.93	0.59
UK	Limpet	Fortis	2	0.91	0.55
DK	Limpet	Fortis	3	0.91	0.55
DE	Limpet	Fortis	5	0.91	0.55
UK	Mohican	Licorne	6	0.83	0.52
DK	Mohican	Licorne	6	0.83	0.52
DE	Mohican	Licorne	12	0.83	0.52
UK	Remy	Salomont	3	0.89	0.48
DK	Remy	Salomont	3	0.89	0.48
DE	Remy	Salomont	3	0.89	0.48
UK	Remy	Splendor	4	0.88	0.55
DK	Remy	Splendor	1	0.88	0.55
DE	Remy	Splendor	8	0.88	0.55
UK	Smart	Eclipse	2	0.90	0.58
DK	Smart	Eclipse	1	0.90	0.58
DE	Smart	Eclipse	2	0.90	0.58
UK	Smart	Limpet	4	0.92	0.59
DK	Smart	Limpet	3	0.92	0.59
DE	Smart	Limpet	0	0.92	0.59
UK	WRG257	SW9991087	5	not calculated	
DK	WRG257	SW9991087	3	not calculated	
DE	WRG257	SW9991087	4	not calculated	

Notes: Characteristics which are "Distinct" for a variety pair are marker x; shaded cells indicate where the data from two or more countries agree.

TABLE 10B. Summary of the field trial data from France.



			No of distinct char. in Extrafrial	Similarity Morphology	Similarity Molecular	Cotyl: Length	Cotyl: Width	Leaf: length (blade and petiole)	Leaf: width (widest point)	Cotyl: Width/length ratio	Cotyl: size	Cotyl: Saddle depth	Leaf: number of lobes (fully developed leaf)		Leaf: length of petiole	Plant: total length including side branches		Thousend kernel weight	Silique: length (between peduncle and beak)	Silique: length of beak	Silique: width	Seeds per silique	Silique: length of peduncle	Time of flowering		Flower: length of petals		Flower: width of petals		Flower: Petals length/width ratio	Flower: Petal size	Flower: base of petal	Leaf: indentation of margin		Tendency to flower in year of sowing	Plant: height at flowering		Leaf: colour	
1	Action	Fortis		0.94	0.55								5	6		6	6							5	5	6	6	5	5				5	6		6	6	6	4
2	Apex	NKVICTORY		0.92	0.83								6	6		6	6							5	4	6	6	6	6				5	6		6	6	5	5
3	Bellini	Cannelle		0.64	0.72								4	6		5	8							2	8	5	7	5	6				4	7		5	8	5	8
4	Boston	Remy		0.86	0.50								5	5		5	6							6	5	6	6	5	6				8	5		6	6	7	5
5	Caiman	Action		0.94	0.54								6	5		7	6							7	5	6	6	5	5				7	5		7	6	6	6
6	Californium	Sansibar		0.91	0.62								5	6		5	5							5	5	6	6	6	5				5	5		5	5	5	5
7	Cannelle	Remy		0.88	0.44								6	5		8	6							8	5	7	6	6	6				7	5		8	6	8	5
8	Capvert	Mohican		0.84	0.49								6	7		6	5							7	4	6	7	6	6				6	6		6	5	7	5
9	Caracas	Castille		not calculated?									8	5		4	6							3	3	6	7	6	6				5	6		5	6	5	6
10	Caraco	Bellini		0.70	0.85								6	4		6	5							8	2	6	5	5	5				7	4		6	5	7	5
11	Caraco	Mohican		0.84	0.52								6	7		6	5							8	4	6	7	5	6				7	6		6	5	7	5
12	Casanova	Siska		0.93	0.64								6	5		6	5							5	4	5	7	5	5				5	5		6	6	6	5
13	Cavalcade	Action		0.94	0.54								4	5		7	6							4	5	6	6	6	5				5	5		7	6	6	6
14	Fortis	Licontent		0.91	0.55								6	5		6	6							5	2	6	6	5	5				6	6		6	5	4	5
15	Hektor	Zenith		0.96	0.84								5	6		6	6							4	4	6	7	5	6				5	5		6	5	5	5
16	KW1097	WRG257		not calculated?									6	6		6	6							7	5	6	6	5	5				5	6		7	7	6	6
17	Lewis	PR45W04		0.72	0.80								6	4		5	7							3	7	7	6	6	5				5	5		5	7	7	5
18	Libea	Hektor		0.97	0.67								5	5		6	6							4	4	7	6	6	5				6	5		7	6	5	5
19	Licontent	Cavalcade		0.93	0.59								5	4		6	7							2	4	6	6	5	6				6	5		5	7	5	6
20	Limpet	Fortis		0.91	0.55								5	6		6	6							5	5	6	6	6	5				5	6		7	6	5	4
21	Mohican	Licorne		0.83	0.52								7	5		5	7							4	8	7	5	6	5				6	6		5	6	5	7
22	Remy	Salomont		0.89	0.48								5	6		6	5							5	7	6	6	6	5				5	5		6	6	5	5
23	Remy	Splendor		0.88	0.55								5	6		6	4							5	4	6	5	6	5				5	6		6	5	5	4
24	Smart	Eclipse		0.90	0.58								6	5		6	7							4	6	6	7	6	6				5	6		7	7	5	6
25	Smart	Limpet		0.92	0.59								6	5		6	6							4	5	6	6	6	6				5	5		7	7	5	5
26	WRG257	SW9991087		not calculated?									6	5		6	4							5	5	6	5	5	5				6	5		7	5	6	6

not studied in Franc note of leaf number of lobes for Caracas
one note by plot note of leaf number of lobes for Castille
one measurement t difference between the two notes sufficient to be distinct only with this characteristic

Although it is not possible to draw too many conclusions from this exercise, given that only one year of data is available, it is clear that, where direct comparisons are possible, there is general agreement between the countries in:

- variety pairs that are difficult or impossible to distinguish, e.g. Hektor/Zenith; Action/Fortis
- variety pairs that are readily distinguished, e.g. Mohican/Licorne, Caraco/Bellini.

5.2.5.2 Observation of Variety Pairs Globally.

In addition to the foregoing analyses, GEVES undertook a study of the side-by-side variety pairs evaluated according to a "global appreciation" of the varieties as opposed to a recording of the individual characteristics (see Annex 2 for more detailed information). In brief, the crop experts were asked to observe the pairs of varieties at two stages (leaves stage and flowering stage) and to give a note using the following "scale of similarity":

1 the two varieties are similar or very close
3 the two varieties are distinct but close
5 the comparison was useful, but the varieties are clearly distinct
7 the comparison should have been avoided because the varieties are very different
9 the comparison should have been avoided because the varieties are totally different

The set of 26 variety pairs planted in each country were visually evaluated for their degree of morphological similarity/difference by two to four crop experts, depending on the country. The experts' notes were then compared to the corresponding Dice distances.

The results from France are shown in Table 11 – the results from the other countries were broadly similar and are given in detail in Annex 2.

Table 11: **Correlation between French experts' notations and Dice distances.**

	Expert1	Expert2	Expert3	Expert4	Dice_dist
Expert1	1.00	0.88	0.73	0.46	0.35
Expert2	0.88	1.00	0.83	0.41	0.34
Expert3	0.73	0.83	1.00	0.58	0.54
Expert4	0.46	0.41	0.58	1.00	0.23
Dice_dist	0.35	0.34	0.54	0.23	1.00

In general, as might be expected, for all countries, although there was a degree of good agreement between experts, there was no correlation between experts' notations and Dice distances. Use of the mean (or modal, or median) value for the note from all countries did not improve the situation markedly. It should be noted that the number of variety pairs studied and the number of experts were rather limited. Moreover, very few selected variety pairs were found very similar (note = 1), which would be necessary to define properly the molecular threshold above which morphological distances were high enough to declare the varieties distinct. One point of interest however, was that for the few very similar variety pairs that were included, it was noted that the Dice distances did not exceed c. 0.3 (see Figure 7). More work would be required to investigate this further.

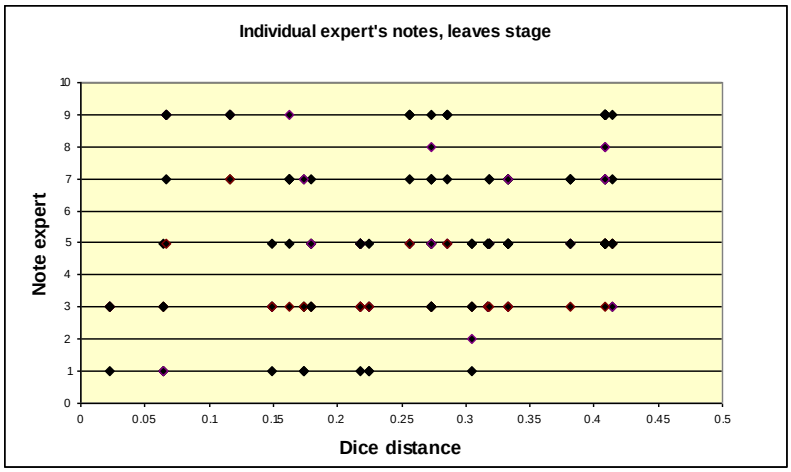


Figure7. An example of the distribution of variety pair-wise Dice distances for the various global notes.

5.2.6 Use of Molecular Markers in Combination with GAIA

GEVES undertook a detailed analysis of the potential use of molecular markers in combination with the software programme GAIA. The whole set of methods used and results are detailed in Annex 2, and a summary of the main points is given below.

The overall purpose of this work was to compare different methods for selecting the pairs of varieties that have to 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.

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.

The general proposal for the combination of morphological and molecular data is illustrated in Figure 8. The first step is a selection on morphological characteristics, which leads to the following:

- if the GAIA distance is higher than 6, 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 8), 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.

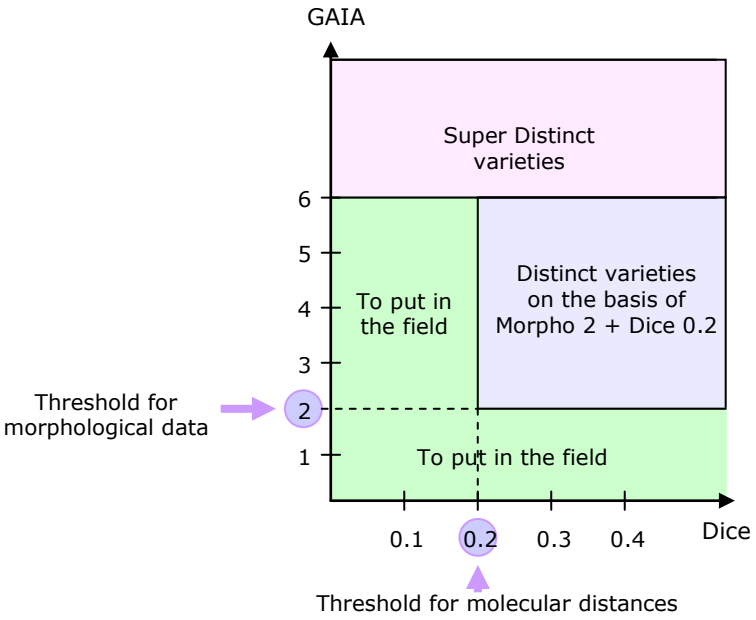


FIGURE 8: **A summary of the GEVES proposal for the selection of the variety pairs to be compared in the field by using molecular data combined with morphological characteristics.**

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.

The common database contains 335 varieties, generating in theory and without selection 55,945 pairs of varieties to be compared in the field. Figure 9 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.

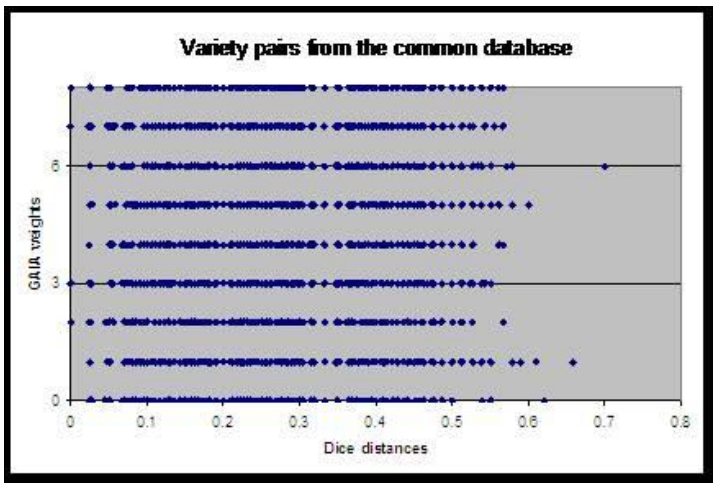


Figure 9: Dice distribution for the variety pairs from the consolidated database with GAIA weight<6.

Figure 10 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).

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

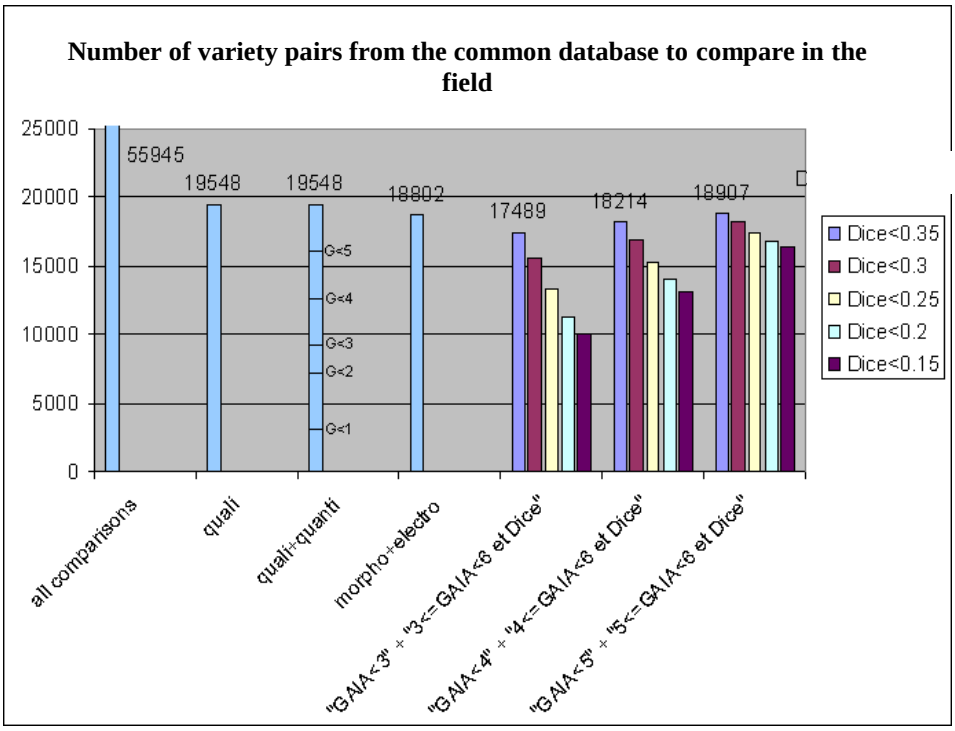


Figure 10: Number of variety pairs from the common database to compare in the field, selected according to the following criteria:

- **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 transformed into qualitative notes;
- **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 colours 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 colours 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 colours on the side of the graph].

From this and similar work on other datasets (see Annex 2) 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. This is considered further below (see section 6.4).

6. DISCUSSION AND CONCLUSIONS.

The overall objective of this project was to examine the potential uses of DNA microsatellites (SSRs) for the management of variety reference collections in oilseed rape DUS testing (i) standardising conditions for the use of an agreed set of SSRs, (ii) analysing c. 410 OSR varieties from different countries with these SSRs, (iii) analysing the data produced, including estimates of genetic and phenotypic distances, and comparison of such distances in different ways, and (iv) validation of these approaches in a field trial.

6.1 UPOV Option 2 – Background

The essential premise of the project was to examine the use of a UPOV "Option 2" ("Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics") approach for the use of molecular markers. In the document which describes and considers the UPOV Options (TC/38/14- CAJ/45/5, 2002), it is stated that .." *The [BMT] Crop Subgroups developed this option with the aim to ensure that there would be no significant shift in the typical minimum distances as measured by traditional characteristics. However, they noted that the lack of a clear relationship between molecular marker distances and differences in traditional characteristics would lead to the need to consider how to handle potentially different decisions on distinctness. The framework of an impact analysis was developed: the comparison of decisions by traditional characteristics with those by molecular characteristics and the analysis of different decisions using molecular characteristics on the value of protection. The key is whether variety pairs, which are not distinct using traditional characteristics, would be judged as distinct using molecular characteristics and whether such decisions would be acceptable for maintaining the value of protection…[These proposals] would be on the basis of a genetic distance assessment, rather than a characteristic by characteristic approach... and would be presented for use in the management of reference collections."*

The calibration of threshold levels for differences in molecular characteristics against differences in traditional characteristics would be more or less straightforward if there were a strong correlation between these two ways of measuring the differences (distances) between varieties. In such a situation, a plot of the variety distances assessed by the different methods would produce a result as in figure 11.

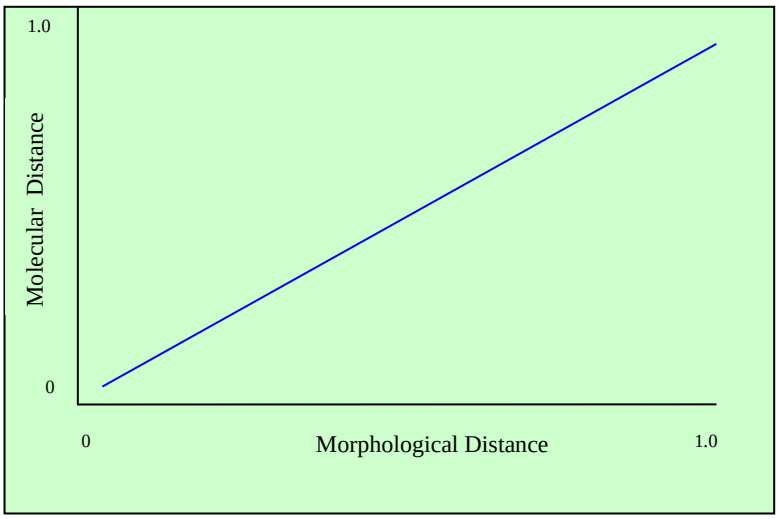


Figure 11: An idealised plot of distance/similarity estimates. A close correlation between molecular and morphological distances would facilitate the application of an Option 2-type approach.

In such a situation, 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.

6.2 Quality of Data

Clearly in order to be able to test such a model system, it is necessary to have available good quality data and to have analysed a sufficiently large number of varieties. Hence a lot of effort in the current project was put into the selection of SSR markers that could be analysed successfully in different laboratories, and the validation of the resultant data sets, as well as into ensuring a 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, utilising 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% (see Annex 1).

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 so-called BMT Guidelines. The difficulties encountered in such an exercise should not be minimised, but as the project has shown, they can be successfully overcome.

It could be argued that ideally a higher number of SSR markers should be used to produce genetic distance estimates. The direct examination of this point was not an objective of the project, but the existence of the T1, T2 and T3 datasets does allow

some analysis. These three sets of data contained information from varying numbers of markers, although it must be acknowledged that they are all sub-sets of the overall data. Even so, statistical analyses applied to the T1, T2 and T3 sets of data 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 emphasised 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 utilising 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.

With regard to the phenotypic data, it had not been anticipated that such a high level of resource input would be required in collating and harmonising morphological data from the four partners. In principle, there was an agreed set of characteristics (in "Note form" and as "Measured" values), and the partners had been requested to supply annual data for specified years. In practice, not all characteristics were recorded in each country in each year. And some were found to recorded but on differing scales and with a different (unspecified) baseline. Also, data were not always available annually, and consolidated national data was not utilized. Unlike the molecular data, which consisted of a common calibration set and then largely non-overlapping data from varieties of interest provided between the partners, the phenotypic data from each partner should ideally have covered the full variety set. In the end, it was necessary to consolidate the data, as described previously, to produce an agreed final data set. The Notes data had missing values randomly distributed throughout the set, at a maximum level of c. 3%. As the consolidated Measured data were derived for the appropriate REML analyses, the definitive working data matrices had REML estimates and were complete - based on annual/country data sets, there was a maximum of 5% missing values, which was agreed to be within acceptable limits.

The data sets declared as definitive were those where the quantity and distribution of missing data were minimised, 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.

6.3 Assessment of Option 2

As indicated above (Figure 11), in order for Option 2 to be applicable in its most straightforward form, there would ideally need to be a strong correlation between the distance estimates produced using molecular markers and those produced using traditional DUS characteristics. Unfortunately, it was clear from the analyses carried out (see Section 5 above, and Figure 6, Table 7, for example), that in practice the correlation was very weak (<0.1), regardless of the method of analysis or the number of markers included. This means that it is impossible to extrapolate any threshold for phenotypic distinctness directly from a measurement of molecular distance. From this, it can be concluded that the application of Option 2 in this rather simplistic form, at least in the case of winter oilseed rape varieties, is not possible.

Whilst this is clearly disappointing, the current project also points to the fact that research into other Option 2-type approaches may be profitable. According to the relevant UPOV documents, Option 2 requires "Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics". Results from the project indicate that it may be possible to define such thresholds, given further study.

The use of molecular markers in combination with phenotypic characteristics has many attractions as a means of reducing the number of varieties grown in field trials. One such approach investigated in the project is to use markers in combination with GAIA, and this is considered below. In addition, there are other avenues for exploiting the potential of molecular markers in DUS testing and related applications.

6.4 Molecular markers in combination with GAIA

In essence, the use of GAIA estimates the degree of distance between varieties, based on weightings assigned by the crop expert to the characters measured. Once an established threshold distance has been exceeded, then a variety can be said to be D. The weightings consider the reliability of the character, and the difference required to provide evidence of distinctness. In this way, the D decision is constructed from the sum of varying degrees of difference.

One objective of this study was to demonstrate that WOSR distinctness could be based on morphological and molecular data, without weakening the protection of PBR, and that this approach could allow a better management of the reference collection to save time, money and resources. GAIA allows such a combination of distances, and the results obtained were very preliminary, but promising. They showed the difficulties inherent in the application of this kind of approach to a crop like WOSR, which is very sensitive to the environment.

As largely expected, no direct correlation between GAIA weights (largely morphology) and Dice distances (molecular) was observed (Figure 9, see also Annex 2). The work thus focused on determining the optimum thresholds for GAIA weights and Dice distances in order to combine them and identify the variety pairs that should be compared into the field and the variety pairs that could be excluded from the field test. Several thresholds were tested for GAIA weights (3, 4 and 5) in combination with different thresholds for the Dice distances (0.35; 0.3; 0.25; 0.2 and 0.15). As the GAIA weights 3, 4 and 5 + Dice gave very similar results (see Figure 10), a GAIA threshold at 4 could be suggested as a first possibility. With respect to the Dice threshold distance, this would have to be defined according to the level of risk that is deemed acceptable, and the requirements of the resources management.

To continue this work, the appropriate thresholds would need to be defined with more precision by testing varieties side by side in the field. The analysis based on the crop expert's notations (see section 5.2.5) should be repeated, using a larger set of varieties and including pairs that would equate to note = 1. Also, more experts and more replications would improve the resultant data, along with a more harmonized procedure. Particular attention could perhaps be paid to variety pairs with distances close to the thresholds, to check the D decisions and the level of acceptable risk. In

addition, the computation of the molecular distances could perhaps be improved by increasing and improving the set of molecular markers used. New public and private markers have become available since the start of the project and would be worth exploring.

6.5 Potential Applications of Molecular Markers in WOSR DUS Testing

In the future, the potential applications of molecular markers in WORS DUS testing could be:

1) to combine the morphological distance with the molecular distance to reduce the number of comparisons required during **the second year** of DUS test. During the first year of testing, all the reference collection would still have to be put in the field to describe the material and the applications in the same place and the same year.

2) more interesting would be to use the combination of the morphological and molecular distances during **the first year** of DUS testing, to identify which varieties of the reference collection need to be compared to the candidate varieties in the field. This method would be a very useful tool for the management of the WOSR reference collection and would allow a significant reduction in the area required for DUS tests in the field and in workload.

The second approach would however require some important additional work to be carried out, including:

-several robust morphological characteristics would have to be defined. Robust in this context would mean "stable" in different places and in different years, and could be defined at different levels:

- For several countries;
- For one country, for several places or years;
- For one country, in one place and during one year, for several repeats.

-the thresholds for the minimum morphological distances and the minimum molecular distance would have to be consolidated by testing in the field.

- the possibility of using this approach in routine would have to be studied in terms of timing and cost, since the delay between the filing of the applications, the constitution of the reference collection and sowing of the field trials is very short in WOSR (less than one month) and the costs of the molecular analyses would have to be taken into consideration.

6.6 Other Potential Applications of Molecular Markers in Variety and Seed Testing

In addition to the application of molecular markers in DUS testing as considered in this project, there are several other ways in which markers could be used in variety and seed testing. These have not been studied to any extent in the current work, but should be borne in mind. Examples include, but are not limited to:

b) variety identification
c) confirmation/verification of varietal identity
d) hybrid purity analysis
e) checking hybrid formulae
f) assessing genetic diversity
g) assessing potential EDV situations
h) marker-assisted breeding.

7. FINAL REMARKS AND FUTURE POSSIBILITIES.

Although there is a pressing need to address the question of the management of the reference collection in WOSR DUS testing, this project has demonstrated quite clearly the difficulties associated with this. The use of molecular markers perhaps still offers the best opportunities, but their application is by no means straightforward. The perhaps rather simplistic interpretation of the UPOV Option 2-type approach, involving a direct relationship between morphological and molecular distances and allowing one to be inferred from the other, is clearly not achievable in practice. However, the results of the project have demonstrated that there is potential in combining morphological characteristics and molecular distances in other ways. In order to succeed in this, it is necessary to define carefully the threshold distances – both morphological and molecular – which produce satisfactory results, with an attendant level of risk which is acceptable to all stakeholders.

Hence it is suggested that future work in this area should concentrate on:

(i) The use of more and better quality (preferably single locus) SSRs, which are dispersed throughout the genome and can be reliably scored in more than laboratory;
(ii) Investigations of other types of markers. This might include functional SSR markers, and/or SNPs. The development and availability of the latter in large numbers would have the potential to transform the molecular aspects of this work;
(iii) Investigation of distance measures other than Dice;
(iv) Investigation of the issue of scoring SSR results as binary band absence/presence (0/1) *vs*. whole pattern analysis and the effect of this on distances;
(v) The use of functional markers in other approaches, e.g. within Option 1 – a marker for disease resistance for instance used to assess the resistance status of varieties;
(vi) 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.

This project has demonstrated the value of the harmonisation of approaches to both molecular analysis and morphological recording and indeed further efforts in both of these areas would facilitate progress. A greater degree of harmonisation –and the use of a robust character set - would simplify the creation of more centralised databases

of variety descriptions and the exchange of descriptions, and would perhaps enable the use of other sources of data, e.g. from breeders, to be investigated.

In conclusion, this project has clearly demonstrated the problems inherent in developing tools for the management of reference collections, but has also highlighted the areas where progress is possible in the future.

8. Literature Cited

1. Barendrecht, C. J., 1999. The concept of grouping in UPOV Guidelines focused on ornamentals. Plant Varieties & Seeds 12: 143-148.
2. Tommasini L, Batley J, Arnold GM, Cooke RJ, Donini P, Lee D, Law JR, Lowe C, Moule C, Trick M, Edwards KJ (2003). The development of multiplex simple sequence repeats (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus L.*) varieties.*)* Theoretical and Applied Genetics 106, 1091-1101.
3. Jones, H, Jarman, R J, Austin, L, White, J and Cooke, R J (2003). The management of variety reference collections in distinctness, uniformity and stability testing of wheat. Euphytica 175-184.
4. Bredemeijer, G.M.M., R.J. Cooke, M.W. Ganal, R. Peeters, P. Isaac, Y. Noordijk, S. Rendell, J. Jackson, M.S. Röder, K. Wendehake, M. Dijcks, M. Amelaine, V. Wickaert, L. Bertrand and B. Vosman. (2002). Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theoretical and Applied Genetics 105, 1019-1026.
5. Röder, M.S., K. Wendehake, V. Korzun, G. Bredemeijer, D. Laborie, L. Bertrand, P. Isaac, S. Rendell, J. Jackson, R.J. Cooke, B. Vosman and M.W. Ganal. (2002). Construction and Analysis of a Microsatellite-based Database of European Wheat Varieties. Theoretical and Applied Genetics 106, 67-73.
6. Donini, P., Cooke, R.J. and Reeves, J.C. (2000). Molecular Markers in Variety and Seed Testing. *Proceedings of the International Symposium on Plant Genetic Engineering*, Havana, Cuba 1999, pp.27-34.
7. Jones, H., Bernole, A., Bach-Jensen, L., Horsnell, R.A., Law, J.R., Cooke, R.J. & Norris, C.E. (2008). Minimising inter-laboratory variation when constructing a unified molecular databases of plant varieties in an allogamous crop. Euphytica (in press).
8. Lee, D, Reeves, J C and Cooke, R J (1996). DNA profiling and plant variety registration: 1. The use of random amplified polymorphisms to discriminate between varieties of oilseed rape. Electrophoresis 17, 261-265.
9. Lee, D, Reeves, J C and Cooke, R J (1996). DNA profiling and plant variety registration: 2. Restriction fragment length polymorphisms in varieties of oilseed rape. Plant Varieties & Seeds 9, 181-190.
10. Cooke, R. J. (1999). Modern methods for cultivar verification and the transgenic plant challenge. Seed Science & Technol. 27, 669-680.
11. Kresovich S, SzewcMcFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* (L.) (Rapeseed). Theor Appl Genet 91:206-211.
12. SzewcMcFadden AK, Kresovich S, Bliek SM, Mitchell SE, McFerson JR (1996) Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. Theor Appl Genet 93:534-538.
13. Plieske J, Strauss D (2001) Microsatellite markers for genome analysis in *Brassica*. I. Development in *Brassica napus* and abundance in *Brassicaceae* species. Theor Appl Genet 102:689-694.
14. Rogers, J.S. (1972) measures of genetic similarity and genetic distance. In: Wheeler, M.R. (ed.) Studies in genetics VII, pp. 145-173. University of Texas Publ. 7213.
15. Dillman C., Charcoset A., Goffinet B., Smith J.S.C. and Y. Dattée (1997) Best linear unbiased estimator of the molecular genetic distance between inbred lines. Advances in Biometrical Genetics. Proceedings of the 10th Meeting of the EUCARPIA Section Biometrics in Plant Breeding. Poznan, 14-16 May 1997, P. Krajewsky and Z. Kaczmarek (eds.), pp105-110.
16. Burstin, J. & Charcosset, A. (1997) Relationship between phenotypic and marker distances: theoretical and experimental investigations, Heredity, 79, 477-483.

17. Nuel, G., Robin, S., & Baril, C.P. (2001) Predicting distances using a linear model: the case of varietal distinctness, Journal of Applied Statistics 28, 607-621.
18. Dillmann C., & Guerin D. (1998) Comparison between maize inbred lines : genetic distances in the expert's eye. Agronomie 18, 659-667.