ORIGINAL PAPER

Minimising inter-laboratory variation when constructing a unified molecular database of plant varieties in an allogamous crop

Huw Jones · Anne Bernole · Louise Bach Jensen · Richard A. Horsnell · John R. Law · Robert J. Cooke · Carol E. Norris

Abstract The construction of large-scale databases of molecular profiles of plant varieties for variety identification and diversity analyses is of considerable interest. When varieties of an allogamous species such as oilseed rape are analysed and described using molecular markers such as microsatellites, care is needed to represent the variety in a meaningful yet useful way. It is possible to characterise such heterogeneous genotypes by analysing bulked samples comprising more than one individual seed or plant, but this approach may result in complex microsatellite profiles. Intuitively it would be reasonable to represent a variety by the common 'major alleles' in a profile, but how to define these 'major alleles' remains problematic. This paper describes methods of analysing DNA microsatellite data that will allow independent and objective data production at a number of laboratories. Methods for establishing allele scoring rules (thresholding) are described and the effect of these rules on the utility of the data is discussed.

H. Jones · R. A. Horsnell · J. R. Law · R. J. Cooke · C. E. Norris
NIAB, Huntingdon Road, Cambridge CB3 0LE, UK

A. Bernole
BioGEVES, Le Magneraud, Saint Pierre d'Amilly,
BP 52, 17700 Surgères, France

L. B. Jensen (✉)
Department of Genetics and Biotechnology,
University of Aarhus, 4200 Slagelse, Denmark
e-mail: LouiseB.Jensen@agrsci.dk

Introduction

There is considerable and increasing interest in the development and construction of large-scale databases of the DNA profiles of agriculturally and horticulturally important plant species. The effort and resources required to populate these databases frequently involve collaboration between laboratories. Such databases have a number of important applications not only for diversity analysis in breeding programs (Ford et al. 2002) but also in the establishment and management of core collections (Hao et al. 2006), establishment of population structure and evaluation of linkage disequilibrium (Maccaferri et al. 2005) and not least in the future registration and distinctness, uniformity and stability (DUS) testing of new varieties (Bredemeijer et al. 2002). The potential benefits of using molecular markers for DUS testing include freedom from environmental interactions, meaning that testing could be carried out more objectively and more rapidly, resulting in cost reductions (Donini et al. 2000). Also, more centralised testing, within Europe for instance, would become feasible.

Oilseed rape (OSR) (*Brassica napus* L.) is an important oil and fodder crop in Europe and indeed world-wide. Crops such as OSR however present considerable difficulties in the context of DUS testing. The number of candidate varieties entered for DUS testing is large and increasing annually, and the existence of different types of varieties (lines, synthetics, hybrids of various kinds, GMOs) complicates the DUS test field trial design and increases its complexity. Another set of problems arises from the increasing size of variety reference collections and the requirement (UPOV Convention 1991) to compare new varieties with those whose existence is a matter of 'common knowledge' at the time of application. This means that in principle, each DUS test annually should make comparisons between new

varieties and several hundreds if not thousands of existing varieties. It is desirable that in order to maintain the strength of protection offered by Plant Breeders' Rights schemes, the principle of comparing new varieties with those of common knowledge should be upheld, and variety reference collections should be as comprehensive as possible. Clearly then, some means of 'managing' the size of the reference collections is highly desirable (Barendrecht 1999).

One option would be to compare newly submitted candidate varieties with the reference collection prior to sowing the field trial, in order to reduce the number of varieties that need to be grown. An attractive means of such management would be to use molecular markers such as microsatellites (DNA profiling) to compare new varieties with the profiles of those in a database, eliminate those which do not need to be compared in a field trial (according to pre-defined criteria) and then only grow the most similar varieties for detailed DUS testing (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 using the profiling data.

The construction of molecular databases of crop varieties using DNA microsatellite markers have been reported previously, e.g. in wheat (Röder et al. 2002) and tomato (Bredemeijer et al. 2002). These studies established several important principles for the construction of such databases, and the approaches contained in them have been largely confirmed by the International Union for the Protection of New Varieties of Plants (UPOV) (9th Session of the Working Group on Biochemical and Molecular Techniques, Washington, DC, United States of America, 21–23 June 2005). In short, for construction of molecular databases the agreed variety collection should be analysed using a marker system with high repeatability, high information content, in different laboratories using different detection systems and preferably including reference samples in all analyses. If disagreements occur, these should be solved by exchanging samples.

Compared to many other important crops, OSR presents a further level of difficulty, due to the tetraploidy of the species and to its reproductive system, both of which can result in heterogeneity at many microsatellite loci. This heterogeneity can be minimised at the analytical level by using a bulked sample (e.g. of 30 plants, seeds or seedlings) to generate the DNA profile.

The present study was carried out as part of a wider project to examine the possibilities of using microsatellite profiles as a way of managing the size of the reference collections in OSR DUS testing. Although the international body responsible for DUS testing, UPOV has not yet recommended the use of molecular markers, it has recognised three possible options for their future application: Option 1: Molecular characteristics as predictors of traditional characteristics, Option 2: Calibration of molecular characteristics against traditional characteristics, Option 3: Development of a new system followed by impact assessment. The 'Option 2' type approach, which was taken in this study, required the profiling of a large number of varieties, using an agreed set of microsatellite markers, in accordance with a harmonised protocol, in three different laboratories. In an allotetraploid crop such as OSR, several primer pairs probably amplified products from more than one locus. Despite this, the terms marker and allele will still be used for each primer pair and amplification product, respectively. Initial results showed that microsatellite profiling of OSR is a robust and rugged tool. However, potential sources of error in microsatellite profiles are widely recognised (Heckenberger et al. 2002; Pompanon et al. 2005) and the multi-allelic nature of the profiles from pooled samples was problematic when determining the agreed profile of a variety. The relative response for each allele within the 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 amplified fragment and the presence of competing fragments each have an effect on PCR efficiency, and the magnitude of these effects may not be consistent between laboratories. While different laboratories generated broadly similar profiles, the relative response for each allele within the profile varied 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 'rules' for allele calling that would allow the differing profiles at each laboratory to be described in the same way. Previously, the result of analysis of a test set of accessions has been used to 'calibrate' a system to determine allele frequencies in pooled samples of maize (Dubreuil et al. 2006; LeDuc et al. 1995). Here, a test set of individuals and pooled samples derived from the same individuals were used to quantify variation due to differential amplification and for overlap with stutter bands in complex profiles. Rules were devised to modify the peak heights obtained for pooled samples so that the peak heights accurately reflected the allele frequencies quantified by genotyping the individual samples. The peak heights for individual alleles in test samples were modified using these rules to compensate for variation. This system is intended to minimise within laboratory errors when generating allele frequency profiles from microsatellite assays in pooled DNA representing samples of heterogeneous germplasm. Here, we devised allele calling rules—termed thresholding—that 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 into a centrally maintained database.

Materials and methods

Experimental design

The study was carried out in two phases. In summary, the first phase was used to select and validate microsatellite markers for use in the second phase. The second phase was used to assess between-laboratory variations, where each laboratory examined an independent sub-sample of seed taken from a bulk. In the first phase ten OSR varieties were selected, germinated and DNA extracted at the co-ordinating laboratory. Aliquots from these extracts were distributed to the participating laboratories. PCR amplifications were carried out at each laboratory for all markers and the results compared. Markers were validated where all laboratories were able to 'call' the same alleles and the results were clear and robust. In the first phase 40 OSR varieties (termed the thresholding set of varieties) were selected and their seeds were sub-sampled at the co-ordinating laboratory. The seed sub-samples were distributed to the participating laboratories where they were extracted, amplified by PCR for the selected markers and the results collated. The data generated showed the variability introduced when laboratories examine independent seed sub-samples from a heterogeneous cultivar. Alternative thresholding strategies were applied to the data from this phase and the results were compared.

Plant material

Seed of the ten OSR varieties used in the initial validation phase (Apex, Artus, Askari, Bienvenue, Bonar, Express, Falcon, Orlando, Samurai and Toucan) came from authenticated stocks held at NIAB. A list of the 40 varieties subsequently used in the thresholding assessments is available from the authors.

DNA preparation

A total of 40–50 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.

For the initial validation of the markers, ten solutions of extracted DNA were distributed to the three participating laboratories (referred to as laboratory X, Y and Z).

DNA amplification

PCR reactions were prepared with 1 μl DNA template (nominally 10 ng), 1 μl 10 × PCR buffer, 1 μl 25 mM $MgCl_2$, 1 μl 5 mM primer pairs, 0.1 μl 20 mM dNTP, 0.1 μl 5U/μl *Taq* polymerase and water to 10 μl.

Markers used in this study

The microsatellite markers used were all obtained from publicly available sources (Kresovich et al. 1995; Szewc-McFadden et al. 1996; Plieske and Struss 2001; Tommasini et al. 2003, see http://ukcrop.net/perl/ace/search/BrassicaDB). Full details of the markers are given in Table 1. The fluorescently marked primers, suitable for the laboratory's instrument system, were synthesised for each laboratory. All fragments were amplified using the following PCR cycling conditions: 92°C for 120 s, followed by 35 cycles of 92°C for 30 s, then 55°C for 30 s, then 72°C for 60 s followed by 72°C for 600 s.

Fragments were visualised using a MegaBace instrument at Laboratory X, Licor and ABI 3130XL Genetic Analyser instruments at Laboratory Y and an ABI 3100 Genetic Analyser instrument at Laboratory Z.

Marker validation

Initial marker validation was carried out by inspection of electropherograms and gel images. Comparisons between data from the three laboratories allowed polymorphisms to be identified. The polymorphisms for each marker were assigned an allele identity and tabulated with the fragment size (bp) at each laboratory, to allow simple cross-referencing between laboratories. The differences in fragment sizes between laboratories were always small and showed systematic variation; the differences are thought to be due to different 'size standards' and 'sizing algorithms' used by the different instruments and to the use of 'tailed primers' at some laboratories. Additional alleles, compared to the alleles agreed upon when looking only at the ten varieties in the first phase, were included from the thresholding data set containing 40 varieties only when they were seen at all laboratories and their fragment size made identification unequivocal. Where comparison between laboratory data did not allow for clear and robust identification of polymorphisms, markers were not considered for further use in the thresholding process (Table 1).

Thresholding

The options for the application of a thresholding approach include absolute thresholding and relative thresholding,

Table 1 Microsatellite markers used in the project were selected from publicly available sources. Chromosome locations, primer sequences and number of alleles are shown for each marker

ID	Marker	Chromosome	Primers (F, R)	Alleles
1	Ra2-E03	10	AGGTAGGCCCATCTCTCTCC, CCAAAACTTGCTCAAAACCC	3
2	BN12A	13	GCCGTTCTAGGGTTTGTGGGA, GAGGAAGTGAGAGCGGGAAATCA	2
3	BN26A	No info	TAAACTTGTCAGACGCCGTTATC, CCCGTAAATCAAGCAAATGG	1
4	CLONE33	No info	GTTTGTGTTGCAATTATTCCCA, CCTGCATTGCGAAAATATAATC	3
5	LS107	No info	GTTAAGTGTGGCGTTAGAGG, CCTTGGTACATGCCACTGAA	3
6	MB5	No info	AACATCTTTTTGCGTGATAT, AATAGCATTGAAGCCTTAC	2
7	Na10-F06[a]	3/9/14	CTCTTCGGTTCGATCCTCG, TTTTTAACAGGAACGGTGGC	–
8	Na10-H03	No info	GAGCTGGCTCATTCAACTCC, CACAATTTCTCAGACAAAACGG	2
9	Na10-E02	5	TCGCGCATGTAATCAAAATC, TGTGACGCATCCGATCATAC	3
10	Na12-D04	6	ACGGAGTGATGATGGGTCTC, CCTCAATGAAACTGAAATATGTGTG	1
11	Na12-A02	16	AGCCTTGTTGCTTTTCAACG, AGTGAATCGATGATCTCGCC	5
12	Na12-E02	No info	TTGAAGTAGTTGGAGTAATTGGAGG, CAGCAGCCACAACCTTACG	4
13	Na14-E08[a]	14	TTACTATCCCCTCTCCGCAC, GCGGATTATGATGACGCAG	–
14	Na14-H11	No info	GGATGTTTTCACAGACCCTG, CTTTGCAGGTATGAACACGC	4
15	Ol09-A06	12	TGTGTGAAAGCTTGAAACAG, TAGGATTTTTTTGTTCACCG	3
16	Ol10-B01	17	CCTCTTCAGTCGAGGTCTGG, AATTTGGAAACAGAGTCGCC	4
17	Ol10-F11	11	TTTGGAACGTCCGTAGAAGG, CAGCTGACTTCGAAAGGTCC	2
18	Ol10-H02[a]	12/13	AACAGGAAGAAACGACGAGG, AGAGAGCCATGAGAAGCACC	–
19	Ol11-B05	3	TCGCGACGTTGTTTTGTTC, ACCATCTTCCTCGACCCTG	3
20	Ol11-G11	3/13	GTTGCGGCGAAACAGAGAAG, GAGTAGGCGATCAAACCGAG	3
21	Ol12-F02	9	GGCCCATTGATATGGAGATG, CATTTCTCAATGATGAATAGT	4
22	Ol13-C12	13	AGAGGCCAACAAAGAACACC, GAAGCAGCACCAGTGACAAG	3
23	Ra1-F06	6	ACCAAAATGTGTGAAGCCAC, CTTGTGGCCAGATTCATCAC	6
24	Ra2-A05	7	GCTAGTTTACGCGGCGG, AAACGACATCGGCAAAGAAG	2
25	Ra2-A11	9	GACCTATTTTAATATGCTGTTTTACG, ACCTCACCGGAGAGAAATCC	4
26	Ra2-D04[a]	6	TGGATTCTCTTTACACACGCC, CAAACCAAAATGTGTGAAGCC	–
27	Ra2-E11	3	GGAGCCAGGAGAGAAGAAGG, CCCAAAACTTCCAAGAAAAGC	6
28	Ra2-F11[a]	13/14/15	TGAAACTAGGGTTTCCAGCC, CTTCACCATGGTTTTGTCCC	–
29	sORA26[a]	No info	TGTTTACCTGTTGGAGAT, AACCCTAAGCATCTGCGA	–

[a] Markers were rejected following the initial marker evaluation

either using a global threshold value or applying independent threshold values for each laboratory's data (see Fig. 1).

Absolute thresholding would entail rejecting all allele peaks below a certain threshold value (Fig. 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.

Relative thresholding requires that the allele with the largest response, say peak height, within a variety profile is identified. All other peaks in the profile will be scored as alleles if their response exceeds a pre-determined percentage of the maximal peak. Relative thresholding may be applied in two ways; either the same pre-determined global threshold would be applied at all laboratories for all markers (Fig. 1b) or empirically determined laboratory/marker specific thresholds are used (Fig. 1c).

When relative thresholding is applied using a global threshold, differences between laboratories in PCR efficiency for different sized fragments may result in different allele scores at each laboratory. Where global thresholding is applied using a high threshold (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 may also be introduced where the maximal peak differs between laboratories, and this was seen to be the case for 20 of the markers considered in this
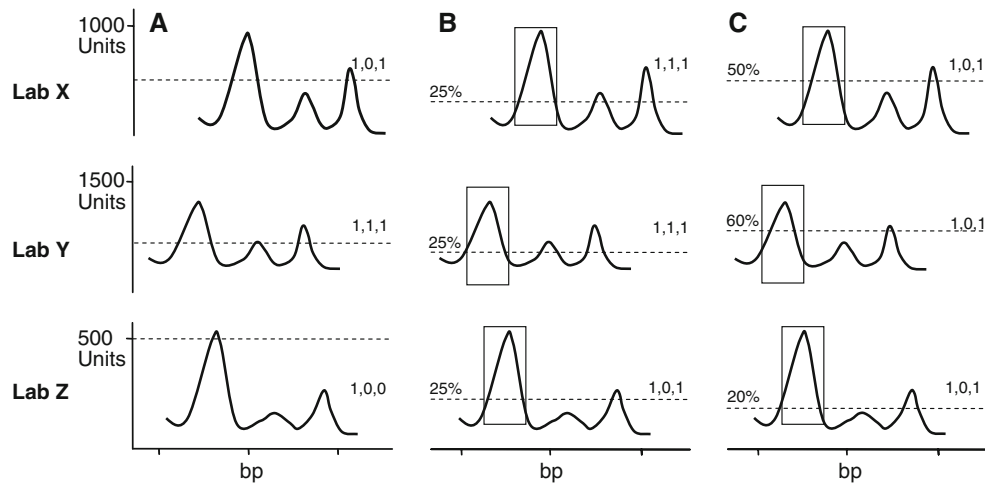
Fig. 1 Strategies for thresholding results from three different laboratories (Laboratory X, Y and Z). **a** Absolute thresholding where peaks are scored as present if they exceed a pre-determined level of instrument response e.g. >500 units in this example. **b** Relative thresholding with a common threshold at all labs. The largest peak is identified for each sample (*box*). Alleles are scored in all labs if their peak height is >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

study. Where global thresholding is applied by including peak heights above a low threshold (for instance 15% and above that of the maximal peak height, i.e. trim off the worst) the result is a discriminating set of allele data, but at the risk of maximising potential variation between laboratories.

When relative thresholding is applied using empirically determined laboratory/marker specific thresholds (see the following paragraph for details) considerable effort is required to determine the values that will be used.

Empirical determination of relative threshold values

In order to examine the effects of the different systems of thresholding, the microsatellite profiles for the thresholding set of varieties were tabulated in Microsoft Excel. The data for each marker was set out in an array where variety data was kept in a row, with each laboratory's data appearing in turn. The same alleles were tabulated for each laboratory with the alleles appearing in the same order. The alleles described during marker validation were always included and additional alleles were only included where all laboratories included them in their data and their common identity was unequivocal. Where laboratories declared partial data with some alleles present for a variety the missing data points were recorded as 0. Where no data appeared for any alleles in a variety the missing data were marked in the array with the string 'NA'. A simple programme was written in the Excel macro language Visual Basic for Applications. The macros identified the maximal peak for each variety at each laboratory and tabulated their peak heights in a second array. Relative thresholding was then applied to

the peak height data in the first array by reference to the maximal peak height data in the second array to generate a third array of binary data (Table 2). The programme was written to apply an arbitrary set of nine thresholding values (15, 25, 35, 45, 55, 65, 75, 85 and 95%) and to iterate through these values at each of the three laboratories such that all 729 combinations (9^3, nine thresholds and three laboratories) were used to generate binary data in turn.

Concordance

A score was calculated showing the extent of agreement between the participating laboratories for each thresholding treatment. The binary data for each variety were compared across the three laboratories and scored according to the degree of agreement (Table 3). The comparison was made

Table 2 Thresholding: the peak height data for each marker was set out in an array where variety data was kept in a row

Accession	Allele Peak height			Max value	Threshold value (e.g. 50%)	Allele Binary		
	A	B	C			A	B	C
1	90,269	17,307	95,722	95,722	47,861	1	0	1
2	43,884		28,809	43,884	21,942	1	0	1
3	77,452	59,423	3,399	77,452	38,726	1	1	0
4		56,655	76,819	76,819	38,410	0	1	1
5	48,889	44,751	17,609	48,889	24,445	1	1	0

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 scoring 1 if it exceeds the threshold value and scoring 0 if it does not

Table 3 Calculating concordance: the data for each variety were compared across the three laboratories and scored according to the degree of agreement

Accession	Laboratory			Concordance
	X	Y	Z	
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

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 labs agree all variety profiles perfectly). This percentage was termed the concordance score

for the variety profile rather than assessing agreement on a per allele basis. Where all three laboratories agreed a variety profile the result was scored as 2, whilst 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 labs agree all variety profiles perfectly). This percentage was termed the concordance score (Table 3). The thresholds at the three laboratories and their resulting concordance score were tabulated in a fourth array.

Using the data from this fourth array the thresholding combinations giving the best concordance could be identified and their data subject to further analysis. The number of combinations giving the best concordance was counted for each marker. The extreme ranges of threshold combinations (highest and lowest combined thresholds) giving the best concordance were identified. The between laboratory concordance for un-thresholded data was also calculated.

Additional statistical analysis

In addition to comparing the concordance between laboratories at each level of thresholding, frequency based genetic distances (Rogers 1972) were calculated and the correlation between the distance matrix for each laboratory was measured. Inspection of the distance matrices showed that one variety with a high incidence of missing data was having a deleterious effect on the correlations. This variety was subsequently excluded from the data for all laboratories.

Polymorphism information content (PIC) values (Botstein et al. 1980) were also calculated for all microsatellites markers.

Calculating distance matrices, tables of PIC values and average number of alleles for all possible combinations of thresholding at the three laboratories would be onerous. Tables were calculated for five thresholding treatments: (1) no thresholding applied, (2) uniform threshold of 15% applied at all laboratories, (3) uniform threshold of 95% applied at all laboratories, (4) differing thresholds applied at each laboratory giving optimum concordance and the lowest combined thresholds and (5) differing thresholds applied at each laboratory giving optimum concordance and the highest combined thresholds. Two further sets of tables were calculated from the above 'Treatment 5', where only markers giving an optimum concordance >90% (18 markers) or >95% (11 markers) were included.

PowerMarker 3.25 (Liu and Muse 2005) was used to calculate summary data for each allele table, to compute allele frequencies and to compute frequency based distances (Rogers 1972) between the 40 validation varieties. Between-laboratory correlations among genetic distance tables were assessed using Mantel's test (Manly 1991) using PopTools (Hood 2005).

Results

Concordance data are shown in Table 4 for different methods of thresholding. The data show the effects of no thresholding, and global thresholding at two different levels (15 and 95%) and also show the range of concordance levels generated in variable thresholding. The number of combinations of thresholds yielding optimum concordance is also given. A wide range of concordance was seen in the un-thresholded data. Global thresholding improved concordance at both 15 and 95%. Variable thresholding produced the widest range of results; the minimum concordance set showed the lowest average concordance with a minimum of 2% for one marker, demonstrating that variable thresholding must be applied with care. The optimum concordance set showed the highest average concordance and the narrowest range of results. The number of combinations of thresholds producing the optimum concordance had a mean of 75 out of a possible 729 combinations (all possible combinations for nine thresholding values applied independently in three laboratories), varying from one marker where all combinations gave optimum concordance to three markers where only one combination gave optimum concordance. The concordances generated by a range of global thresholds are shown in Table 5. The optimum concordance was seen in 14 out of the 22 markers, but the data showed that no single global threshold could be applied to all markers.

Table 4 Concordance values for combinations of thresholds

Marker ID	Thresholding treatment					Combinations
	None	15%	95%	Min	Max	
1	100	100	100	100	100	729
2	99	97	97	55	97	181
3	100	100	100	100	100	*
4	18	94	53	50	93	9
5	80	96	80	80	96	8
6	67	93	89	89	94	162
8	60	91	79	78	91	12
9	49	92	92	90	93	45
10	100	100	100	100	100	*
11	62	58	58	28	62	3
12	22	53	35	22	65	1
14	90	44	90	23	92	20
15	65	69	64	40	85	3
16	92	97	95	95	97	108
17	89	98	96	96	98	36
19	69	44	68	34	68	18
20	85	71	88	53	90	1
21	90	97	60	41	97	21
21	70	97	48	47	97	21
22	89	93	94	79	100	120
23	96	92	98	2	98	1
25	27	94	65	64	96	72
27	72	75	74	40	94	6
Mean	74	85	79	61	91	
Max	100	100	100	100	100	
Min	18	44	35	2	62	

None: laboratory data with no thresholding, 15 and 95% show data where uniform thresholds have been applied at all laboratories, Min and Max show the highest and lowest concordance achievable by variable thresholding. The number of thresholding combinations giving optimum concordance is shown. Markers with an asterisk (*) were monomorphic in this variety set

The effect of thresholding on the utility of the data is demonstrated in Table 6. The tabulated data show PIC values and the average number of alleles, which give an indication of the discrimination power made available by the markers under each of the thresholding treatments: (1) no thresholding applied, (2) uniform threshold of 15% applied at all laboratories, (3) uniform threshold of 95% applied at all laboratories, (4) differing thresholds applied at each laboratory giving optimum concordance and the lowest combined thresholds and (5) differing thresholds applied at each laboratory giving optimum concordance and the highest combined thresholds. As expected, datasets produced with no thresholding or a low uniform threshold (Treatments 1 and 2) produce the least conservative (average allele number = 3.62 and 3.57 respectively) and most informative (PIC = 0.39 and

0.40) datasets. The dataset produced with a high uniform threshold is, as expected, the most conservative and least informative (average allele number = 2.70, PIC = 0.32). The 'optimum concordance' datasets (Treatments 4 and 5) fall in between these extremes of performance (average allele number = 3.09 and 3.05, PIC = 0.36 and 0.35). Treatment 4 might be expected to produce a less conservative, more informative dataset than Treatment 5 but the average allele numbers and PIC values show little evidence to support this. This pattern is mirrored in the data for average Rogers' genetic distance, where Treatments 1 and 2 > Treatments 4 and 5 > Treatment 3. The correlations between the three laboratories' genetic distance matrices for each thresholding treatment showed a different pattern and, as might be expected, followed the same trend as that for concordance. The different treatments will produce different relationships between varieties and this is shown by the differing average genetic distances calculated for each treatment. Systematic changes in relatedness were tested by calculating Spearman's rank correlation between the distance matrices for each treatment. In all cases strong positive correlations were obtained, but in no case was the correlation perfect.

Improved correlations between distance matrices might be expected from two laboratories using instruments supplied by the same manufacturer. However, this effect is not seen in the inter-laboratory correlations in Table 6.

Table 5 shows that the optimum concordance achievable for some markers is relatively low. Markers where the optimum concordance was <90 and <95% were removed from thresholding Treatment 5 and the effects examined. In both cases the inter-laboratory correlations between distance matrices improved, but the PIC and average genetic distance decreased.

Discussion

The data presented demonstrate that, in heterogeneous material such as OSR varieties, microsatellite profiles for small bulks of individuals produced in independent laboratories may not always concur. The reasons for differences within the data may be an effect of sampling, differential efficiency in PCR reactions, use of different measurement systems, the settings used in peak detection software at the start of allele calling and subjective decisions made by operators while allele calling. For the thresholding set of varieties discussed in this paper it must be assumed that laboratories submitted a good set of data, yet when the data were compared with no thresholding the concordance was as low as 15% for one marker with an average of 74% for the set of 22 markers used. The problem of unifying data from more than one laboratory has been acknowledged by other workers (Vosman et al. 2006).

Table 5 Comparison of concordances generated by a range of global thresholds (identical thresholds applied at all laboratories)

Threshold	Marker ID																					
	1	2	3	4	5	6	8	9	10	11	12	14	15	16	17	19	20	21	22	23	25	27
15	100	97	100	53	80	89	79	92	100	58	35	90	64	95	96	68	88	60	94	98	65	74
25	100	97	100	53	83	91	79	92	100	46	47	90	71	95	96	67	78	56	97	93	81	70
35	100	97	100	52	83	92	80	92	100	48	50	61	72	95	96	53	81	63	100	24	86	74
45	100	97	100	50	81	93	82	92	100	50	50	44	75	95	96	55	82	81	100	24	94	87
55	100	97	100	50	81	93	80	92	100	50	51	40	75	95	96	39	81	88	100	44	93	84
65	100	95	100	50	87	94	80	92	100	47	50	38	77	95	98	40	85	93	100	53	93	80
75	100	68	100	50	88	94	82	92	100	46	50	35	72	97	98	41	82	96	95	76	93	76
85	100	70	100	56	94	94	83	92	100	47	50	34	69	97	98	43	78	96	89	88	94	76
95	100	97	100	93	96	93	91	92	100	58	53	44	69	97	98	44	71	97	93	92	94	75
Optimum	100	97	100	93	96	94	91	93	100	62	65	92	85	97	98	68	90	97	100	98	96	94

The optimum concordances achieved using empirical thresholds at each laboratory are shown at the bottom for comparison (optimum). Optimum concordance using empirical thresholds are matched by concordances from global thresholding in 14 out of 22 markers (dark grey). For the eight remaining markers the maximum concordance achieved using different global thresholds are shown in light grey

Table 6 Correlations between genetic distance matrices for each laboratory and marker performance indicators (average of Rogers' genetic distance, PIC and average allele number) for each thresholding treatment

Thresholding treatment	Interlab distance correlations				Genetic distance (mean)	PIC	Average allele number
	X vs. Y	X vs. Z	Y vs. Z	Mean			
1	0.78	0.68	0.67	0.71	0.44	0.39	3.62
2	0.78	0.75	0.74	0.76	0.44	0.40	3.57
3	0.80	0.63	0.65	0.69	0.37	0.32	2.70
4	0.82	0.78	0.78	0.79	0.41	0.36	3.09
5	0.80	0.77	0.78	0.79	0.40	0.35	3.05
5 (>90%)	0.85	0.81	0.83	0.83	0.36	0.31	
5 (>95%)	0.93	0.84	0.84	0.87	0.30	0.25	

All correlations are significant ($P < 0.05$). Thresholding treatments: (1) no thresholding applied, (2) uniform threshold of 15% applied at all laboratories, (3) uniform threshold of 95% applied at all laboratories, (4) differing thresholds applied at each laboratory giving optimum concordance and the lowest combined thresholds and (5) differing thresholds applied at each laboratory giving optimum concordance and the highest combined thresholds

Considerable improvement in concordance is possible by applying global thresholds but the analysis shows that no one level of thresholding produces optimum concordance for all markers. The effort needed to assess a set of markers to produce variable thresholds for each marker in each laboratory and to produce a set of global thresholds for each marker is similar so there appears to be no benefit to adopting a global thresholding approach.

The benefit of variable thresholding is that it offers the potential to generate datasets that give the least disagreement between laboratories. The method uses a thresholding set of varieties, assayed at all laboratories, which can be used to develop a set of thresholding rules to be applied to each marker at each laboratory when larger sets of varieties are analysed. The effects of thresholding on marker quality can be assessed, and an informed judgement made on whether to include all markers in the final database. Only

one polymorphic marker (Ra2-E03) showed a concordance value of 100 despite the thresholding treatment. For this project only a limited number of markers were used, testing a larger number of the publicly available markers would increase the possibility of identifying markers that give a high level of concordance regardless of the thresholding conditions. The effects of thresholding are clearly demonstrated in our data. The improvements in concordance are obtained at the expense of reduced discrimination between varieties. The issue that must then be addressed is whether the reduced discrimination renders the data set unfit for its intended purpose. This will vary from case to case.

Our study attempted to group varieties using molecular markers in order to manage DUS reference collections. The thresholding method was applied to a data set for 450 OSR varieties and the discrimination power provided by the markers individually, as measured by the number of

distinct pairs generated in comparison, ranged from 3 to 82% (Research programme CPV5766 2007: Management of winter oilseed rape reference collections). This compared with discrimination power ranging between 7 and 80% when the UPOV grouping characters (characteristic 1: seed: erucic acid; characteristic 5: leaf: lobes and characteristic 11: Time of flowering) are used (UPOV 1996). When used in combination the morphological grouping characters achieved 83% discrimination, a total readily achieved by combinations of three or four markers.

Thresholding may not be appropriate where the dataset will be used to estimate heterogeneity within populations using allele frequencies determined in pooled samples. The reduction in discrimination that is implicit in thresholding will tend to reduce the number of occasions where minor alleles are scored, and hence will skew estimates of heterogeneity.

A disadvantage of the application of variable thresholding is that it carries a requirement to review or repeat the thresholding process when new laboratories become contributors to the database or when existing laboratories substantially change their equipment. A set of control varieties, selected from among the thresholding set of varieties, should be included within each analytical batch and the data used to establish whether there has been a significant change in method performance. Where there has been a change in method performance or new laboratories become contributors to the database the thresholding exercise will need to be repeated at all laboratories. The review of thresholding may require that thresholding parameters be changed for historic data. Raw data, based on peak height data without thresholding, must be stored in order to facilitate this, which imposes a burden on the contributors to the database. Furthermore, the possibility that historic data may be changed if and when new thresholding parameters are applied must be explicitly stated to all stakeholders with an interest in the data.

The use of a unified database created in this way must acknowledge the fact that agreement between laboratories cannot always be perfect. If used to manage reference collections in the context of DUS testing, new varieties will be analysed and the database interrogated, to eliminate those varieties that were clearly sufficiently different, according to previously agreed criteria, and to produce a group of similar varieties against which the new variety would need to be compared in more detail. The data used for thresholding includes examples where laboratories produce different profiles for sub-samples of the same variety; the frequency of mis-matches in these data could be used to develop rules to be used when matching 'unknown' samples to the database. For example, if 'Variety X', a member of the thresholding set of 40 varieties, is genotyped and compared to the test set then, taking each marker in turn, the probabilities of obtaining 'Genotype X' (true positive), 'not Genotype X' (false negative), 'Genotype Y' (false positive) and 'not Genotype Y' (true negative) can all be calculated. Thus, when a candidate variety is compared to the full unified database, the cumulative probabilities for each outcome for all markers at each variety will allow calculation of likelihood for a match between the candidate variety and each of the database varieties. Therefore, database queries would be written to produce a group of 'most likely' matches from the database varieties that would allow candidate varieties to differ at one or more loci from group members yet still be considered as similar. An assessment of the risks of failing to group all 'similar' varieties in the grouping process would also be needed and the database would need to be modelled and tested using the known error rate. The likelihoods used in the database queries would be 'calibrated' against user requirements by this process. Development of such a procedure is beyond the scope of the reported work.

In conclusion, in this paper we have discussed an approach to unifying molecular marker data from collaborating laboratories, used to populate a centrally maintained database of variety profiles. Such a database could be used to assist in the management of the reference collections used in DUS testing of crop plants. However, the thresholding methods described are not limited to this highly specialised field, but could have wider application in any situation where data from collaborating laboratories are collated into a single database. The thresholding methods described will improve concordance between laboratories at the expense of discrimination power within the data set; success is achieved if the data set produced retains sufficient discrimination to meet the requirements of the end users.

References

Barendrecht CJ (1999) The concept of grouping in UPOV. Guidelines focused on ornamentals. Plant Var Seeds 12:143–148

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction length polymorphism. Am J Hum Genet 32:314–331

Bredemeijer GMM, Cooke RJ, Ganal MW, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor Appl Genet 105:1019–1026

Donini P, Cooke RJ, Reeves JC (2000) Molecular markers in variety and seed testing. In: Proceedings of the international symposium on plant genetic engineering, Havana, Cuba, 1999, pp 27–34

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe:

large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Ford R, Le Roux K, Itman C, Brouwer JB, Taylor PWJ (2002) Diversity analysis and genotyping in *Pisum* with sequence tagged microsatellite site (STMS) primers. Euphytica 124:397–405

Hao CY, Zhang XY, Wang LF, Dong YS, Shang XW, Jia JZ (2006) Genetic diversity and core collection evaluations in common wheat germplasm from the north-western spring wheat region in China. Mol Breed 17:69–77

Heckenberger M, Bohn M, Ziegle JS, Joe LK, Hauser JD, Hutton M, Melchinger AE (2002) Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol Breed 10:181–191

Hood GM (2005) PopTools version 2.7.1. http://www.cse.csiro.au/poptools

Jones H, Jarman RJ, Austin L, White J, Cooke RJ (2003) The management of variety reference collections in distinctness, uniformity and stability testing of wheat. Euphytica 132:175–184

Kresovich S, Szewc-McFadden 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

LeDuc C, Miller P, Lichter J, Parry P (1995) Batched analysis of genotypes. PCR Methods Appl 4:331–336

Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

Manly BFJ (1991) Randomization and Monte Carlo methods in biology. Chapman & Hall, London

Plieske J, Struss 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

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859

Research programme CPV5766 (2007) Management of winter oilseed rape reference collections. Supported by the Community Plant Variety Office Research and Development Section, 1 January 2005 to 31 December 2007

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Rogers JS (1972) Measures of genetic similarity and genetic distance. In: Wheeler MR (ed) Studies in genetics VII. University of Texas publication 7213. University of Texas Press, Austin, pp 145–173

Szewc-McFadden 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

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. Theor Appl Genet 106:1091–1101

UPOV Convention (1991) http://www.upov.int/en/publications/conventions/1991/content.htm

UPOV (1996) TG/36/6 guidelines for the conduct of tests for distinctness, uniformity and stability. International Union for the Protection of New Varieties of Plants, Geneva

Vosman B, Scott E, Spellerberg B, Barendrecht J, Tams S, Jones H, Esselink D (2006) A European reference collection of rose varieties. In: UPOV biochemical and molecular technologies meeting, Seoul, South Korea