ISSN 1479-2621

Plant Genetic Resources **4**(2); 117–124
DOI: 10.1079/PGR2005106

Redundancy and distinctness in flax germplasm as revealed by RAPD dissimilarity

Yong-Bi Fu*

Plant Gene Resources of Canada, Saskatoon Research Centre, Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N 0X2, Canada

Received 21 June 2005; Accepted 28 September 2005

Abstract

Molecular characterization of *ex situ* plant germplasm, although more attainable than before, has rarely been applied to a whole germplasm collection of 2000 accessions or larger. The benefits of screening large numbers of accessions have not been well recognized for germplasm management and utilization. Challenges also exist in identifying duplicated and genetically related accessions and in validating developed core subsets. Here we show how a new approach using an average marker-based dissimilarity of an accession in a collection can be applied to identify both redundancy and distinctness in a plant germplasm collection. Application of this dissimilarity measure to 2727 flax accessions genotyped by 149 randomly amplified polymorphic DNA (RAPD) markers revealed that up to 22% of accessions could be deemed to be redundant. Up to 500 of the most distinct flax accessions were identified and these can be directly screened for traits of interest to broaden the genetic base in a flax improvement programme. These results demonstrate that molecular screening of a large number of accessions with an informative diversity analysis can facilitate the management and utilization of *ex situ* plant germplasm.

Keywords: distinctness; flax; molecular characterization; RAPD; redundancy

Introduction

The characterization of *ex situ* plant germplasm using molecular marker techniques is more attainable than ever before (Karp, 2002), and thus has greatly increased over the last decade (Fu, 2003). This sort of characterization, however, has rarely been applied to large-sized collections (Fu, 2005), probably because the experimental effort required is too large and the characterization too expensive to undertake. However, bulking individual plants from one accession or group to form a representative sample has made such characterization practically feasible (Fu, 2003; Fu *et al.*, 2003a), but the motivation for these efforts still appears to be lacking (Hodgkin and Rao, 2002). Clearly, the benefits of screening a large collection for germplasm management and utilization have not been well documented and consequently less appreciated (Fu, 2005).

Duplicated and genetically related accessions are known to exist in most *ex situ* collections (Hintum and Knüpffer, 1995; Hintum and Visser, 1995), but little is known about their numbers (Dean *et al.*, 1999; McGregor *et al.*, 2002; Treuren *et al.*, 2004), particularly in collections with poor passport and pedigree information. Efforts to identify duplicated accessions have been made using DNA profiling (Waycott and Fort, 1994; Virk *et al.*, 1995; Phippen *et al.*, 1997; Zeven *et al.*, 1998; Lund *et al.*, 2003), but it is difficult to define how many loci need to be matched before declaring two accessions to be duplicated (Treuren and Hintum, 2001). It is even more difficult to identify

*Corresponding author. E-mail: fuy@agr.gc.ca

related, but non-identical accessions, when pedigree data are lacking and/or incomplete. Thus, rather than devoting a substantial effort to identify duplicates and related accessions, it is probably more productive and informative to attempt to assess the level of redundancy in a collection, the proportion of redundant accessions without which the overall genetic variation and/or structure of the collection remains unchanged. With molecular markers, it is possible to estimate the level of redundancy within a collection either by screening the whole collection or at least a large random sample (Dean *et al.*, 1999; Treuren *et al.*, 2004). An estimate of the extent of redundancy present would provide a useful guide to germplasm management (Engels and Visser, 2003).

Over the last two decades, many core subsets have been developed using passport and phenotypic data aiming to represent the range and structure of genetic variability in large collections (Brown and Spillane, 1999), but most have not been well characterized (Liu *et al.*, 2001; Karp, 2002; Ude *et al.*, 2003; Fu *et al.*, 2005). Little is known of the diversity and structure of core subsets (Fu *et al.*, 2005) and there is some concern surrounding the representation of the full collection in core subsets (Brown and Spillane, 1999). Thus few established core subsets have been widely screened for phenotypic traits and most large collections remain under-utilized. Screening a large number of accessions using molecular markers and identifying genetically distinct germplasm directly may encourage the use of *ex situ* plant germplasm, particularly in widening the genetic base of breeding materials.

Plant Gene Resources of Canada (PGRC; the Canadian national seed genebank) at Saskatoon maintains a substantial flax collection. To support its management and utilization, the collection was fingerprinted from 1999 to 2001 at 149 randomly amplified polymorphic DNA (RAPD) loci. The RAPD technique (Williams *et al.*, 1990), although now largely superseded, is nevertheless well suited for characterizing a large number of samples, as it is relatively simple, fast and cheap (Nybom, 2004). This large data set provides an opportunity to address not only issues associated with genetic diversity (Fu *et al.*, 2002), genetic diversity changes in improved gene pools (Fu *et al.*, 2003b) and the genetic structure of the collection (Fu, 2005), but also redundancy and distinctness in the collection.

The overall objectives of this study were to (i) assess the level of redundancy in the flax collection and (ii) identify distinct accessions. This was achieved by calculating the mean RAPD dissimilarity between each accession and the remaining accessions in the collection, and identifying those accessions, without (or with) which the original genetic structure in the collection was maintained.

Materials and methods

Plant collection

The PGRC collection of flax germplasm consists of 3158 accessions, representing the seventh largest collection of flax genetic resources in the world, but only 2813 of these are available for seed distribution (Diederichsen *et al.*, 2002). It represents for the most part a duplication of the collection maintained in the US National Plant Germplasm System. Newly acquired germplasm includes 94 accessions bred in Canada and 571 accessions of diverse origin from seed genebanks in Russia, Germany and the Czech Republic. During rejuvenation from 1999 to 2001, 238 additional accessions with defined phenotypic differences were newly selected from mixed samples, largely of landrace materials.

RAPD analysis

The entire active collection was RAPD fingerprinted using 16 informative primers, described by Fu (2005). The reproducibility of fingerprints has been mentioned elsewhere (Fu *et al.*, 2002, 2003a). Eighty-six accessions were not adequately characterized, and the remaining 2727 accessions consisted of 19 accessions of uncertain origin and 2708 originating from 63 countries in 12 major geographical regions as defined in Fu (2005). A total of 149 polymorphic bands were scored independently by two individuals as present (1), absent (0) or uncertain (9) for each accession.

A new approach for assessing redundancy and distinctness

In a collection of n accessions, a given accession can form $n - 1$ pairs with the remaining accessions of the collection. The RAPD similarity S_{ij} between accessions i and j is calculated as $(a + d)/(a + b + c + d)$, where a is the number of bands present in both i and j, b the number present in i and absent in j, c the number present in j and absent in i, and d the number absent from both i and j (equivalent to the simple matching coefficient of Sokal and Michener, 1958). The RAPD dissimilarity for such pair is then $1 - S_{ij}$. The mean RAPD dissimilarity for accession i is obtained by averaging all $n - 1$ RAPD dissimilarities. The higher the mean dissimilarity, the more genetically distinct the accession is in the collection. The lower the mean dissimilarity, the more genetically redundant the accession is in the collection. Thus, ranking the mean dissimilarities of all n accessions provides a means of identifying the most distinct and most redundant accessions. This ranking, however, does

https://doi.org/10.1079/PGR2005106 Published online by Cambridge University Press

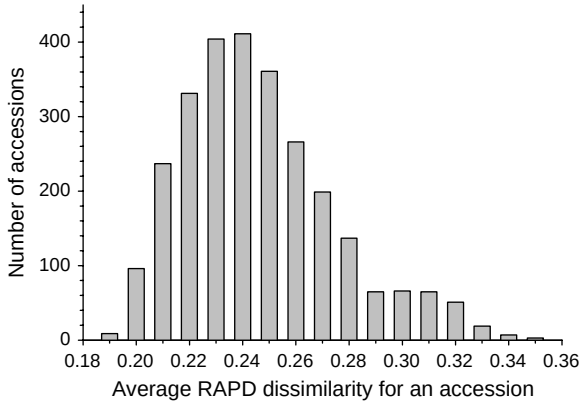

Fig. 1. Frequency distribution of mean RAPD dissimilarities for all 2727 flax accessions.

not provide any objective cut-off point to select a group of the most distinct or most redundant accessions in the collection.

A possible means to select a redundant group is to identify the most redundant accessions, without which the original genetic structure of the collection remains unchanged. The genetic structure in the collection and in a sample of selected accessions can be measured by AMOVA (analysis of molecular variance; Excoffier *et al.*, 1992). This is achieved by removing certain accessions with the lowest mean dissimilarities from the collection and assessing changes in molecular variance explained by countries or regions in the remaining accessions, until these variances approach the levels present in the original collection. Accessions whose removal has little impact on the original genetic structure of the collection are identified as the most redundant. The proportion of the most redundant accessions in the collection gives a rough estimate for the level of genetic redundancy in the collection. The same step-wise approach can be applied to identify a group of the most distinct accessions (called a distinct group) by selecting those accessions with the highest mean dissimilarity from the collection as a whole, and assessing the changes in molecular variance explained by countries and/or regions in the selected accessions, until these variances exceed those in the original collection. Such an approach, although computationally intensive for large numbers of accessions, provides an objective basis to address challenging issues such as redundancy and distinctness in a large collection.

Table 1. The proportions of RAPD variation explained by countries and regions in the flax collection without the genetically most redundant accessions (GMRAs) that were identified based on the least mean RAPD dissimilarities, as obtained by AMOVA

Number of GMRAs removed	Size of the remaining sample	Number of countries in the sample	Proportion of variance explained by countries	Number of regions in the sample	Proportion of variance explained by regions
0	2727	63	0.149	13	0.118
100	2627	63	0.150	13	0.117
200	2527	62	0.150	13	0.117
300	2427	62	0.152	13	0.117
400	2327	60	0.153	13	0.118
500	2227	59	0.156	13	0.120
600	2127	59	0.160	13	0.123
700	2027	59	0.163	13	0.124
800	1927	58	0.165	13	0.126
900	1827	57	0.171	13	0.128
1000	1727	56	0.172	13	0.129
1100	1627	55	0.176	13	0.131
1200	1527	55	0.182	13	0.135
1300	1427	54	0.186	13	0.136
1400	1327	54	0.192	13	0.139
1500	1227	53	0.198	13	0.142
1600	1127	50	0.205	13	0.146
1700	1027	47	0.213	13	0.153
1800	927	46	0.221	13	0.158
1900	827	43	0.232	12	0.165
2000	727	42	0.238	12	0.170
2100	627	39	0.244	12	0.172
2200	527	37	0.251	12	0.181
2300	427	36	0.249	12	0.176
2400	327	35	0.231	12	0.155
2500	227	32	0.203	12	0.133
2600	127	28	0.177	11	0.114

https://doi.org/10.1079/PGR2005106 Published online by Cambridge University Press

Analysis of RAPD data

Missing data were omitted in the dissimilarity calculation for each pair of accessions and the total number of RAPD bands may vary for different pairs of accessions. To visualize the variation of the mean dissimilarities obtained, the frequency distribution of the 2727 mean RAPD dissimilarities was plotted. To assess redundancy and distinctness in the collection, a specific SAS program, including AMOVA routines, was written in SAS IML (SAS Institute, 2004) and applied. This SAS program is available upon request.

To verify the redundant and distinct groups identified, a pairwise accession similarity matrix for all 2727 accessions was generated using the simple matching coefficient and converted to a Euclidean distance matrix for a principal coordinate analysis (NTSYS-pc; Rohlf, 1997). The first three principal coordinate scores were plotted to determine the genetic associations of the identified groups and the original collection. To assess genetic relationships between accessions showing the greatest distinctness, a pairwise accession similarity matrix was generated using the simple matching coefficient and clustered using NTSYS-pc with the algorithm of unweighted pair-group methods using arithmetic averages (Rohlf, 1997).

Results and discussion

The mean RAPD dissimilarities ranged from 0.197 to 0.371 with a mean of 0.251 and a standard deviation of 0.029. The frequency distribution of these dissimilarities is given in Fig. 1, and there were 342 accessions with mean dissimilarities less than 0.222 (=0.251 − 0.029) and 276 accessions greater than 0.28 (=0.251 + 0.029). Thus, given the mean dissimilarity value of a particular accession, it is possible to assess the overall genetic difference of the accession against the remaining accessions of the collection. If the value is much lower than 0.222 or much higher than 0.28, the accession could be considered as, respectively, redundant or distinct.

Genetic redundancy of flax germplasm

The AMOVA revealed that removing up to 600 accessions with the lowest mean dissimilarities did not greatly change the original proportion of variation explained by 63 countries (from 0.149 to 0.160; Table 1), with only 4.9% increase of the proportion of variance. Similarly, removing up to 700 accessions with the lowest mean dissimilarities did not greatly change the original proportion of the variance explained by 13 regions (from 0.118 to 0.124; Table 1), with 5.1% increase of the proportion of variance. Removing more accessions will magnify the departure from the original genetic structure. For example, removing 1400 accessions of the lowest mean dissimilarities resulted in changes of, respectively, 28.9% and 17.8% to the original

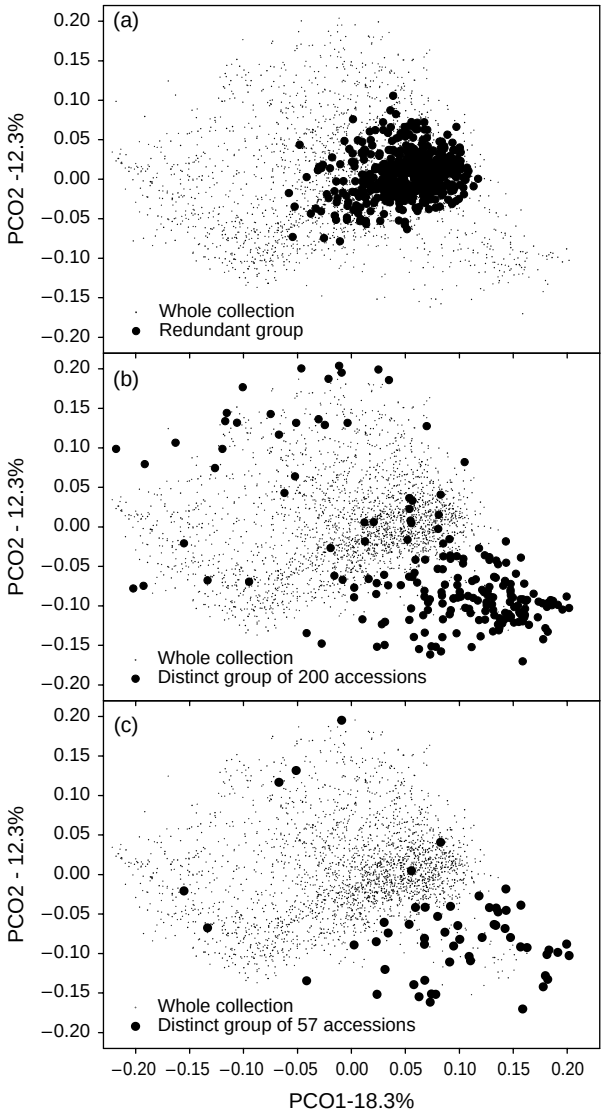

Fig. 2. Plot of the first two principal component scores based on the Euclidean distances converted from the simple matching coefficient matrix of 149 RAPD bands for 2727 flax accessions. These first two components accounted for 18.3% and 12.3% of the total variance, respectively. (a) Five hundred flax accessions with the lowest mean RAPD dissimilarities (redundant group) were identified against the whole collection; (b) 200 flax accessions with the highest mean RAPD dissimilarities (distinct group of 200 accessions) were identified against the whole collection; (c) 57 flax accessions with the highest mean RAPD dissimilarities (distinct group of 57 accessions) were identified against the whole collection.

https://doi.org/10.1079/PGR2005106 Published online by Cambridge University Press

country- and region-related variance. Raising this to 2200 accessions (i.e. about 500 accessions left) still results in an increase in the proportion of the variance explained either by countries or regions in the remaining collection (Table 1), indicating that the accessions removed still contributed to overall variance, although only to a small extent. With a nominal tolerance of 5% genetic structure change, the level of redundancy in the flax collection is around 22% with respect to country-related and 29% with respect to region-related genetic structure. Figure 2a shows the associations between the 500 accessions with the lowest mean dissimilarities and the remaining 2227 accessions. Clearly, removing these 500 most redundant accessions will not much affect the genetic structure of the collection.

Given the establishment history of the PGRC flax collection, it is not surprising to find such a level of redundancy. Extensive exchanges of germplasm over the last 70 years occurred between the USA and Canada and newly acquired germplasm, particularly obtained from Russia, Germany and the Czech Republic, includes likely duplicates as revealed by passport information. During rejuvenation from 1999 to 2001, splitting accessions from genetically nearly identical samples with only one or two trait differences may have generated more genetically redundant samples to the collection. However, the generality of this finding for other germplasm collections at PGRC or other countries remains to be determined. In spite of this, efforts have been made to validate the most redundant accessions and possible treatments of the validated accessions, including pooling duplicated accessions (Treuren *et al.*, 2001, 2004), are being considered with economic factors to make the management of the flax collection more cost-effective.

The mean dissimilarity approach developed here did not identify all true duplicates. For example, two identical accessions which are distinct from the remaining accessions would have the same mean dissimilarity, and this would be larger than the mean for the whole collection. Consequently, the duplicates would not be identified as

Table 2. The proportions of RAPD variation explained by countries and regions in various samples of the genetically most distinct accessions selected from the flax collection based on the largest mean RAPD dissimilarities, as obtained by AMOVA

Number of the most distinct accessions in a sample	Number of countries in the sample	Proportion of variance explained by countries	Number of regions in the sample	Proportion of variance explained by regions
50	19	0.135	9	0.086
57	21	0.148	9	0.090
70	22	0.167	10	0.094
90	24	0.166	10	0.110
110	27	0.191	10	0.128
130	28	0.173	11	0.116
150	29	0.169	11	0.108
170	30	0.174	12	0.108
190	30	0.178	12	0.114
210	32	0.185	12	0.120
230	32	0.198	12	0.130
250	32	0.213	12	0.150
270	33	0.222	12	0.156
290	33	0.222	12	0.158
310	34	0.231	12	0.161
330	35	0.232	12	0.154
350	35	0.236	12	0.156
370	35	0.245	12	0.165
390	35	0.248	12	0.171
410	36	0.252	12	0.177
430	36	0.249	12	0.175
450	37	0.251	12	0.175
470	37	0.256	12	0.181
490	37	0.257	12	0.185
510	37	0.257	12	0.186
530	37	0.251	12	0.181
550	37	0.248	12	0.180
570	37	0.248	12	0.179
590	37	0.246	12	0.177
610	38	0.247	12	0.175
Whole sample	63	0.149	13	0.118

https://doi.org/10.1079/PGR2005106 Published online by Cambridge University Press

redundant. Also, not all of the redundant accessions identified by this approach are necessarily duplicates. However, the approach generated a list of accessions most likely to be redundant, and these can be targeted for further validation if morphological identification or passport verification is unable to provide a guide. In particular, a pair of accessions with the same mean dissimilarity can be considered as likely duplicated and represent suitable candidates for validation. The obvious advantage of this approach lies in the rough estimate of the extent of redundancy in a collection, which is seldom available for collections of any size. As the estimate depends on the criteria used to measure the genetic structure, the resulting estimate may vary and is not as precise as hoped. A further study on the behaviour of this estimator is needed.

Genetic distinctnesss of flax germplasm

Separate AMOVAs show that about 50 accessions with the highest mean dissimilarities would approach the original proportion (0.149) of variance explained by 63 countries (Table 2). Specifically, there were 57 very distinct accessions originating from 21 countries, and displaying the same level of country-wise genetic structure observed in the original collection. However, there were about 210 accessions with the highest mean dissimilarities from 12 regions displaying a similar proportion of region-related genetic structure (Table 2). Figure 2b shows the associations between the 200 accessions with the highest mean dissimilarities with the remaining 2527 accessions. Clearly, these 200 most distinct accessions cover, although unevenly, the range of genetic structure observed in the original collection. When the number of selected accessions with the highest mean dissimilarities reached about 500, the variance explained either by countries or regions was maximized (0.257 and 0.186, respectively). This indicates that these accessions reflect the country- or region-related genetic structure of the collection.

To facilitate the utilization of the collection, the 57 most distinct accessions have been listed in Table 3, along with their Canadian National accession number, country of origin, original description and mean RAPD dissimilarity. These latter ranged from 0.324 to 0.371 with a mean of 0.333. This set showed a good, but uneven, coverage of the genetic structure of the whole collection (Fig. 2c). The genetic relationships of these 57 accessions are shown in Fig. 3, and five major clusters were observed at a similarity level of 0.71. Such a clustering does not match with geographical origin, but the two largest clusters were roughly consistent with the

Table 3. List of the 57 genetically most distinct flax accessions with their Canadian National (CN) accession number, country of origin, original description and mean RAPD dissimilarities (MRD)

Code	CN	Origin	Description	MRD
D1	100910	Portugal	Oil	0.371
D2	101378	Ukraine	Intermediate	0.355
D3	101308	India	Crown	0.354
D4	100885	Greece	Oil	0.347
D5	101268	Netherlands	Oil	0.346
D6	101367	Georgia	Crown	0.346
D7	101402	Russia	Intermediate	0.343
D8	101310	India	Crown	0.343
D9	101396	Russia	Fibre	0.342
D10	98415	India	Oil	0.342
D11	101386	Turkey	Intermediate	0.340
D12	101281	Canada	Oil	0.339
D13	98275	Hungary	Mediterranean	0.338
D14	101385	Turkey	Intermediate	0.337
D15	101237	Lithuania	Oil	0.337
D16	101403	Romania	Fibre	0.336
D17	101374	Russia	Crown	0.335
D18	101338	Afghanistan	Crown	0.334
D19	101407	Netherlands	Fibre	0.333
D20	98303	Hungary	Fibre	0.333
D21	101387	Netherlands	Fibre	0.333
D22	101392	France	Fibre	0.332
D23	101296	Russia	Oil	0.331
D24	101289	Russia	Oil	0.331
D25	101382	Turkey	Intermediate	0.331
D26	101397	Ukraine	Fibre	0.331
D27	101366	Georgia	Crown	0.331
D28	101373	Armenia	Crown	0.330
D29	101265	Great Britain	Oil	0.330
D30	101401	Russia	Fibre	0.330
D31	101405	Romania	Fibre	0.330
D32	101273	France	Oil	0.329
D33	101363	Russia	Fibre	0.329
D34	101375	Russia	Crown	0.329
D35	101394	Russia	Fibre	0.329
D36	101395	Russia	Fibre	0.329
D37	101240	Lithuania	Oil	0.328
D38	101331	Turkey	Crown	0.328
D39	101325	Greece	Crown	0.328
D40	101329	Egypt	Large-seeded	0.327
D41	101379	Ukraine	Intermediate	0.327
D42	101362	Russia	Fibre	0.327
D43	101230	China	Oil	0.327
D44	101274	Canada	Oil	0.326
D45	101307	Russia	Oil	0.326
D46	101406	Russia	Fibre	0.326
D47	101286	USA	Oil	0.325
D48	101327	Spain	Crown	0.325
D49	101299	Russia	Oil	0.325
D50	101301	Russia	Oil	0.325
D51	101298	Russia	Oil	0.325
D52	98969	India	Oil	0.325
D53	101332	Turkey	Crown	0.325
D54	101404	Romania	Fibre	0.325
D55	101348	Russia	Intermediate	0.324
D56	101364	Russia	Fibre	0.324
D57	97334	Argentina	Oil	0.324

https://doi.org/10.1079/PGR2005106 Published online by Cambridge University Press

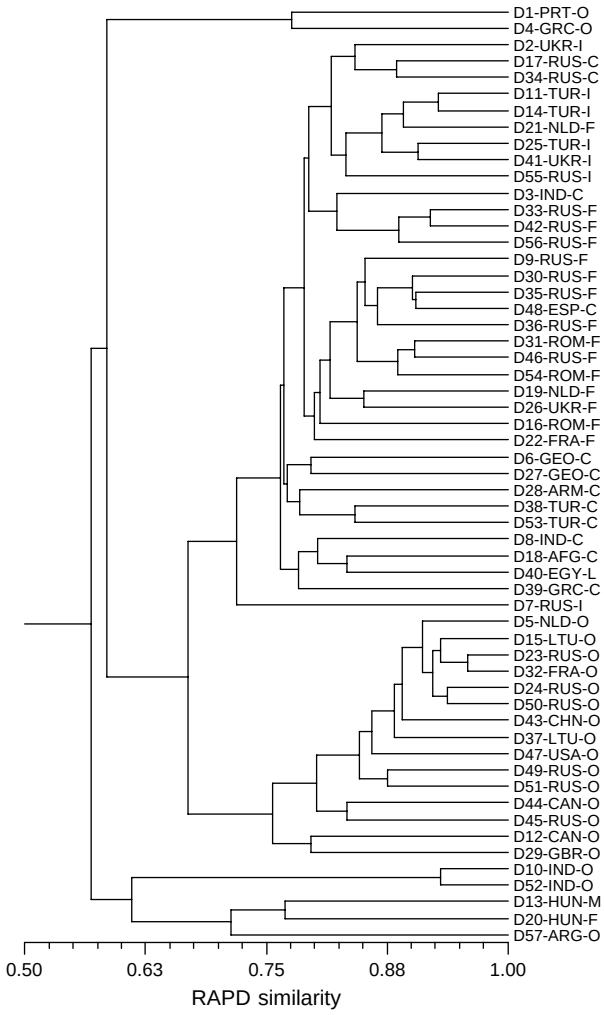

Fig. 3. Genetic associations of the 57 genetically most distinct flax accessions representing 21 countries and nine geographical regions, as revealed by the clustering of their RAPD similarities. Each accession was labelled with its code, country of origin and description (O, oil; I, intermediate; C, crown; F, fibre; L, large-seeded; M, Mediterranean; see Table 3).

types of flax (oil or fibre). The set of 57 includes many types of flax, including oil, fibre, crown and intermediate (Table 3). Interestingly, accessions D23 and D32 are genetically highly similar, but have different origins. To broaden the coverage of distinctness, the list given in Table 3 was expanded to the 500 most distinct flax accessions. This expanded set is available upon request and would be suitable for phenotypic screening aimed at a broadening of the genetic base of flax. The expanded set should be genetically more informative, although not necessarily more representative, than a core subset developed from passport or phenotypic data and requires no molecular validation of the genetic diversity captured in the core subset.

Concluding remarks

The mean dissimilarity approach introduced here is conceptually simple, but offers limited resolution in determining the extent of redundancy. A direct, precise estimator of genetic redundancy from a sample of randomly selected accessions is desirable. The most distinct accessions identified may be of value in broadening the genetic base of flax, but further screening of the distinct accessions for specific traits is needed before incorporating them into a breeding programme. We expect that these efforts will facilitate the management and utilization of *ex situ* plant germplasm generally, particularly in the context of large collections.

Acknowledgements

The author would like to thank Dallas Kessler for his assistance in searching the CN numbers and passport information on related flax accessions and Ken Richards, Axel Diederichsen, David Williams and two anonymous journal reviewers for their helpful comments on the early version of the manuscript.

References

Brown AHD and Spillane C (1999) Implementing core collections—principles, procedures, progress, problems and promise. In: Johnson RC and Hodgkin T (eds) *Collections for Today and Tomorrow*. Rome: International Plant Genetic Resources Institute, pp. 1–9.

Dean RE, Dahlberg JA, Hophins MS, Mitchell SE and Kresovich S (1999) Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. *Crop Science* 39: 1215–1221.

Diederichsen A, Raney JP, Fu YB and Richards KW (2002) Diversity in the flax collection at Plant Gene Resources of Canada. *Proceedings of the Flax Institute USA* 59: 138–143.

Engels JMM and Visser L (2003) *A Guide to Effective Management of Germplasm Collections*. IPGRI Handbooks for Genebanks No. 6. Rome: IPGRI.

Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

Fu YB (2003) Applications of bulking in molecular characterization of plant germplasm: a critical review. *Plant Genetic Resources* 1: 161–167.

Fu YB (2005) Geographic patterns of RAPD variation in cultivated flax. *Crop Science* 45: 1084–1091.

Fu YB, Diederichsen A, Richards KW and Peterson G (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linus usitatissimum* L.) as revealed by RAPDs. *Genetic Resources and Crop Evolution* 49: 167–174.

Fu YB, Guerin S, Peterson GW, Carlson JE and Richards KW (2003a) Assessment of bulking strategies for RAPD analyses

https://doi.org/10.1079/PGR2005106 Published online by Cambridge University Press

of flax germplasm. *Genetic Resources and Crop Evolution* 50: 743–746.

Fu YB, Rowland GG, Duguid SD and Richards KW (2003b) RAPD analysis of 54 North American flax cultivars. *Crop Science* 43: 1510–1515.

Fu YB, Peterson GW, Williams D, Richards KW and Mitchell Fetch J (2005) Patterns of AFLP variation in a core subset of cultivated hexaploid oat germplasm. *Theoretical and Applied Genetics* 111: 530–539.

Hintum ThJL van and Knüpffer H (1995) Duplication within and between germplasm collections. I. Identifying duplication and the basis of passport data. *Genetic Resources and Crop Evolution* 42: 127–133.

Hintum ThJL van and Visser DL (1995) Duplication within and between germplasm collections. II. Duplication in four European barley collections. *Genetic Resources and Crop Evolution* 42: 135–145.

Hodgkin T and Rao VR (2002) People, plants and DNA: perspectives on the scientific and technical aspects of conserving and using plant genetic resources. In: Engels JMM, Rao VR, Brown AHD and Jackson MT (eds) *Managing Plant Genetic Diversity*. Rome: International Plant Genetic Resources Institute, pp. 469–480.

Karp A (2002) The new genetic era: will it help us in managing genetic diversity? In: Engels JMM, Rao VR, Brown AHD and Jackson MT (eds) *Managing Plant Genetic Diversity*. Rome: International Plant Genetic Resources Institute, pp. 43–56.

Liu F, Sun GL, Salomon B and Bothmer von R (2001) Distribution of allozymic alleles and genetic diversity in the American barley core collection. *Theoretical and Applied Genetics* 102: 606–615.

Lund B, Ortiz R, Skovgaard IM, Waugh R and Andersen SB (2003) Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. *Theoretical and Applied Genetics* 106: 1129–1138.

McGregor CE, van Treuren R, Hoekstra R and van Hintum ThJL (2002) Analysis of the wild potato germplasm of the series of Acaulia with AFLPs: implications for *ex situ* conservation. *Theoretical and Applied Genetics* 104: 146–156.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* 13: 1143–1155.

Phippen WB, Kresovich S, Candelas FG and McFerson JR (1997) Molecular characterization can quantify and partition variation among genebank holdings: a case study with phenotypically similar accessions of *Brassica oleracea* var. capitata L. (cabbage) "Golden Acre". *Theoretical and Applied Genetics* 94: 227–234.

Rohlf FJ (1997) *NTSYS-pc 2.1. Numerical Taxonomy and Multivariate Analysis System*. Setauket, NY: Exeter Software.

SAS Institute (2004) *The SAS System for Windows V8.02*. Cary, NC: SAS Institute.

Sokal RR and Michener CD (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38: 1409–1438.

Treuren R van and Hintum ThJL van (2001) Identification of intra-accession genetic diversity in selfing crop using AFLP markers: implications for collection management. *Genetic Resources and Crop Evolution* 48: 287–295.

Treuren R van, Soest LJM van and Hintum ThJL van (2001) Marker-assisted rationalization of genetic resources collections: a case study in flax using AFLPs. *Theoretical and Applied Genetics* 103: 144–152.

Treuren R van, Magda A, Hoekstra R and Hintum ThJL van (2004) Genetic and economic aspects of marker-assisted reduction of redundancy from a wild potato germplasm collection. *Genetic Resources and Crop Evolution* 51: 277–290.

Ude G, Pillay M, Ogundiwin E and Tenkouano A (2003) Genetic diversity in an African plantain core collection using AFLP and RAPD markers. *Theoretical and Applied Genetics* 107: 248–255.

Virk PS, Newbury HJ, Jackson MT and Ford-Lloyd BV (1995) The identification of duplicate accessions with a rice germplasm collection using RAPD analysis. *Theoretical and Applied Genetics* 90: 1049–1055.

Waycott W and Fort SB (1994) Differentiation of nearly identical germplasm accessions by a combination of molecular and morphological analyses. *Genome* 37: 577–583.

Williams JKG, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531–6535.

Zeven AC, Dehmer JK, Gladis T, Hammer K and Lux H (1998) Are the duplicates of perennial kale (*Brassica oleracea* L. var. *ramosa* DC.) true duplicates as determined by RAPD analysis? *Genetic Resources and Crop Evolution* 45: 105–111.