Theor Appl Genet (2001) 103:144–152

R. van Treuren · L.J.M. van Soest · Th.J.L. van Hintum

Marker-assisted rationalisation of genetic resource collections: a case study in flax using AFLPs

Received: 25 June 2000 / Accepted: 27 October 2000

Abstract Removing redundant germplasm from collections is one of the options for genebanks to increase the efficiency of their genetic resource management. Molecular characterisation of germplasm is thereby becoming more and more important to verify suspected duplication. AFLPs were used to characterise 29 flax accessions of material derived from research activities (hereafter termed ''breeder's line"). Based on similar accession names, the breeder's lines could be classified into three series ('M 25', 'Ru' and 'Rm') that were expected to contain redundancies. In addition, 12 reference cultivars were analysed. A total number of 144 polymorphic bands (59.8%) were scored among the 164 individuals investigated. In general, relatively high levels of intra-accession variation were found, even for the cultivars examined. This finding was not in line with the low outcrossing rates reported for flax. A cluster analysis grouped the 'Ru' and 'Rm' series together, indicating their close genetic relationship. An analysis of molecular variance (AMOVA) showed a significant group effect (fibre/oil flax) only for 'M 25', explaining 34% of the variation observed within this series. For the cultivars 40.5% of the variation was distributed among accessions within groups and all pairwise comparisons were significantly different, except for one case. Both for the series of breeder's lines and the cultivars the major part of the variation was distributed among individuals within accessions. This component constituted 80.7% and 83.6% of the total variation for the 'Ru' and 'Rm' series, respectively. Pairwise comparisons of accessions were performed by AMOVA in order to identify redundant germplasm. Stepwise bulking of accessions until all remaining accessions were significantly different showed that the 29 accessions of breeder's lines could be reduced to 14. Only a small negative effect of this bulking approach on the among-population component of variance was observed, showing a reduction of 2.6%. Results are discussed in relation to improving the efficiency of collection management.

Keywords AFLP[1] · Collection management · Flax, Genetic resources · Rationalisation

Communicated by H.C. Becker

R. van Treuren (✉) · L.J.M. van Soest · Th.J.L. van Hintum
Centre for Genetic Resources, The Netherlands,
Plant Research International B.V., P.O. Box 16,
6700 AA Wageningen, The Netherlands
e-mail: r.vantreuren@plant.wag-ur.nl
Tel.: +31-317-477082, Fax. +31-317-418094

Introduction

Characterisation of the germplasm is one of the main aspects of genetic resource management (Bretting and Widrlechner 1995; Brown and Kresovich 1996; Clark et al. 1997). Assessment of the genetic relationships among *ex situ* conserved accessions allows genebanks to identify gaps and redundant germplasm in their collections, and to establish core collections (e.g. Hodgkin et al. 1995). Unwanted duplications within collections are a burden to genebanks because these do not contribute to the genetic diversity present within the collection, but do require capacity for storage and maintenance. Increasing collection size and decreasing financial resources have stimulated genebanks to identify and remove redundant germplasm in order to increase efficiency (Hintum and Visser 1995; Hintum et al. 1996).

Passport data may be used to identify probable duplicates (Hintum and Knüpffer 1995). For instance, passport data may show that accessions are conserved under identical or similar names, that samples have been collected at identical or nearby localities, or that pedigree data indicate a common genetic background. Obviously, additional information is needed to verify suspected redundancy in the above-mentioned cases. Such information can be obtained by morphological characterisation, but morphological variation is often found to be restrict-

[1] AFLP™ is a trademark of Keygene N.V. (Wageningen, The Netherlands)

ed and can be strongly influenced by the environment. Therefore, studies directed to the identification of redundant germplasm often combine morphological analyses with molecular characterisation (Waycott and Fort 1994; Virk et al. 1995; Oliveira et al. 1997; Zeven et al. 1998) or are carried out exclusively with molecular markers (Phippen et al. 1997; Cervera et al. 1998).

Absolute certainty that two samples are genetically identical can only be obtained by comparison of their complete genomes. However, samples do not necessarily have to be completely identical in order to be considered redundant. For example, two samples collected from the same outcrossing population will have a very small probability to be completely identical, yet they will share a similar genetic background. Therefore, we can consider redundancy by quantifying the genetic diversity between samples based on the screening of germplasm for a large number of polymorphic markers. Statistical tools, such as an analysis of molecular variance (AMOVA), may then be used to evaluate whether samples display sufficient genetic variation in order to consider them distinct (Excoffier et al. 1992).

To characterise germplasm at the DNA level, a variety of molecular techniques is currently available (Whitkus et al. 1994; Karp et al. 1996, 1997; Parker et al. 1998). Compared to alternative genetic marker systems, AFLPs are generally considered relatively powerful in germplasm analysis because of the high number of markers that can be generated per analysis (Powell et al. 1996; Milbourne et al. 1997; Russell et al. 1997). In the present study, AFLPs have been used to identify redundant germplasm in the flax collection of The Centre for Genetic Resources, The Netherlands (CGN).

In the past, CGN has included in this collection a substantial amount of material derived from research activities (material hereafter denoted as ‘‘breeder’s line”). Unfortunately, this material is rather poorly documented. However, crop type (fibre or oil flax) and an accession code are available in nearly all cases. A large part of these accessions appeared to have very similar names, suggesting that they may have originated from the same breeding programme and hence may share a common genetic background. Redundant germplasm might therefore be present among these accessions. Because this material is also rarely requested for utilisation, the present study was undertaken to investigate possibilities for the rationalisation of the collection through the use of AFLP analysis. Variance components within and among accessions were estimated from the molecular data and were used as criteria to constitute groups of genetically similar accessions. The results are discussed in relation to the value of this approach for the rationalisation of collections.

Material and methods

Germplasm studied

The flax (*Linum usitatissimum*) collection of the Centre for Genetic Resources, The Netherlands (CGN), currently comprises about 1000 accessions. Part of this collection consists of a considerable number of accessions that have been classified as breeder’s lines and for which only very limited passport data are available. Individual accessions can be distinguished based on a number, but can be clustered into different groups based on an identical group code (e.g. ‘M 25–341’, ‘M 25–343’, ‘M 25–355’, etc.; ‘526 Hi’–‘541 Hi’). Within the flax collection, 41 of such groups exist, consisting of 317 accessions and comprising around 30% of the entire collection. Out of the 41 groups, 26 have in common that they are denoted by a two-letter code (e.g. Ai, Cg, Hi, Rm, Sp, etc.) with all accessions having a unique preceding number. This suggested that redundancies might be encountered, not only within groups, but even between groups. Furthermore, screening of the flax database revealed that for all 26 groups an existing cultivar could be found of which the name could be related to the two-letter code (e.g. Ai: Aino, Rh: Rhodesia, Sv: Svetoc).

For the present study, 29 accessions from three different series of breeder’s lines (‘M 25’, ‘Ru’ and ‘Rm’) were selected. These series each comprised accessions characterised as either fibre flax or oil flax. In addition, 12 cultivars were included as references (see Table 1). The cultivars ‘Rusland I’, ‘Rusland II’, ‘Rembrandt’, ‘139 Rumanien’, ‘Russian Althanson’, ‘Roma’, ‘Italia Roma’ and ‘Roman Winter’ were selected because of the possible relationship with the ‘Ru’ and ‘Rm’ series based on their accession name and the group code of these two series of breeder’s lines. For the other four selected cultivars there was no reason to expect any relationship with the ‘M 25’, ‘Ru’ and ‘Rm’ series. All accessions used in the present study have been regenerated at least once in the past. During regeneration, the oil and fibre type of the accessions are routinely checked. Four plants per accession were analysed with AFLPs. In addition, two replicate samples from a single individual were included in the analyses to estimate the frequency of artefact bands.

Extraction of DNA

From each of the 41 accessions, seedlings were raised in a greenhouse from 20 sown seeds. After 13 days 50–100 mg of leaf tissue were obtained from four individuals per accession. Tissue samples were collected in 2-ml Eppendorf tubes and immediately frozen in liquid nitrogen. All samples were stored at –80°C upon return to the laboratory.

Samples were vacuum-dried overnight after which five glass pearls were added to each Eppendorf tube and tissue samples were ground mechanically into a fine powder. Subsequently, total genomic DNA was isolated using the microprep protocol of Fulton et al. (1995). Extracted DNA was re-suspended in 50–100 µl of 10 mM Tris/0.1 mM EDTA (pH 8.0), depending on the size of the pellet, and stored at 4°C. Two microliters of each sample were loaded onto an 0.8% agarose gel to estimate DNA concentrations, using 20, 40, 60, 80 and 100 ng of phage lambda DNA as references.

AFLP procedures

The AFLP protocol basically followed the procedures described by Vos et al. (1995). Briefly, five units of *Eco*RI and five units of *Mse*I were used to digest 300 ng of total genomic DNA to completion. AFLP adapters for both restriction enzymes were then ligated to the restriction fragments. Selection of the *Eco*RI–*Eco*RI and *Eco*RI–*Mse*I fragments through the use of biotinylated *Eco*RI adapter and magnetic beads was omitted. Subsequently, template DNA was pre-amplified using primer combinations based on the sequence of the adapters but 3′-extended with one selective nucleotide (*Eco*RI+A/*Mse*I+C). For each of the samples part of the PCR product was loaded onto a 2% agarose gel to check for successful amplification. Pre-amplification was followed by a second amplification reaction, using primer pairs further extended with one selective nucleotide at the 3′ end. All samples were analysed for two selective primer pairs, *Eco*RI+AC/*Mse*I+CC and

*Eco*RI+AG/*Mse*I+CC, which in a previous study had been identified as appropriate primer combinations for AFLP analyses in flax (Van Treuren, unpublished data). Prior to selective amplification, the *Eco*RI primer was radiolabelled with ^{33}P. PCR products were separated on 6% denaturing polyacrylamide gels (Biozym, Sequagel-6), followed by exposure to Kodak XOMAT AR film for several days. In all amplifications, Goldstar *Taq* DNA polymerase was used for PCR and all reactions were performed on a single block of a Perkin Elmer 9600 thermo cycler. The following touchdown thermal profile was used in all amplifications: one cycle of 30 s at 94°C, 30 s at 65°C and 60 s at 72°C; 12 cycles in which the initial annealing temperature of 65°C was lowered by 0.7°C each cycle; 23 cycles in which the annealing temperature was held constant at 56°C. Additional details about the AFLP protocol followed can be found in Arens et al. (1998).

Data analysis

Autoradiograms (approximate size range of the fragments: 50–500 bp) were scored manually and observed band positions were considered to represent different loci. Each observed locus was assumed to potentially have two allelic states (band presence or absence) and variations in band presence were recorded as polymorphisms. Band-sharing data were used to calculate genetic similarities between samples based on the simple matching coefficient (Sokal and Michener 1958). The similarity values were used to graphically represent genetic relationships between samples by principal co-ordinate plots (PCO). Similarity values and PCOs were generated by the Genstat 5 software package (release 4.1). Genetic relationships between accessions were calculated using Nei's unbiased estimate of standard genetic distance and were represented by a dendrogram using the UPGMA clustering algorithm (Nei 1987; Weir 1990). These calculations were carried out using the software package TFPGA (Miller 1997).

An analysis of molecular variance (AMOVA) on genetic distances based on the simple matching coefficient between all pairs of samples was used to compute molecular variance components at different hierarchical levels, i.e. within accessions, among accessions within groups and among groups of accessions. Variance components were tested for significance by a non-parametric resampling approach using 10000 permuted distance matrices (Excoffier et al. 1992). The AMOVA approach was also used in bulking procedures. Firstly, a distance matrix with the corresponding significance values was generated between pairs of accessions. In these comparisons a conservative significance level of 0.1 was used to identify genetically different accessions. Secondly, if not significantly different, the pair with the smallest genetic distance was bulked into one group. Thirdly, a new AMOVA was performed, after which the first two steps were repeated until all matrix elements were significantly different from each other. Analyses of molecular variance were carried out using version 1.55 of the software package WINAMOVA (see Excoffier et al. 1992). Monomorphic loci were included in all data analyses.

Results

Intra-accession variation

For the entire analysis, a total number of 241 bands was generated with the two AFLP primer combinations. Out of these bands, 97 (40.2%) were found to be monomorphic, whereas 144 (59.8%) were polymorphic. Polymorphic bands were distributed equally over the two primer pairs, which generated 57.6% and 62.1% polymorphic bands per primer set, respectively. For both primer combinations, the AFLP profiles of the replicate DNA samples were completely identical, indicating a low probability of artefact polymorphisms. Contrary to expectation based on the selfing nature reported for flax, substantial levels of intra-accession variation were observed, including the reference cultivars. In both accessions '75 Ru' and 'Moose Sel' 60 polymorphic bands were found, which means that these entries each contain 41.7% of the total number of polymorphisms detected. Mean similarity values within accessions ranged from 0.86 to 1.00 (Table 1). Genetic homogeneity was observed only for breeder's line 'M 25–221' and cultivar 'Rembrandt.' In general, lower similarity values were found in the oil flax accessions. On average, mean similarity was 0.905 for oil flax and 0.932 for fibre flax.

Comparison of banding profiles between adjacent rows of fragments suggested codominant inheritance in three cases. These were identified based on the absence of "double null scores" and a marked reduction in intensity of both bands in case both fragments were present. Mean observed heterozygosity in the total sample was 4.1% at these three loci. Heterozygosity at these positions was in nearly all cases accompanied by less intense fragments at other positions. Assuming that these fragments also indicated heterozygosity, up to 13 heterozygous loci per individual (in the accessions 'M 25–343' and 'Moose Sel') could be scored. Again, these results were not expected based on the predominant self-fertilisation of the species.

Inter-accession variation

A nested AMOVA showed a significant effect of crop type (fibre/oil flax) only for the 'M 25' series. For this series 34% of the variation detected could be ascribed to the difference between fibre and oil flax (Table 2). Surprisingly, the effect of crop type was found to be absent in the cultivars, explaining only 3% of the variation detected. No fixations for different alleles were observed when comparing fibre flax and oil flax. However, one fragment was found to be present in all individuals belonging to the fibre crop type, whereas in oil flax the fragment occurred in a frequency of 65.9%. Based on the estimated frequency of the fragment in the total sample, the probability of fixation in either fibre or oil flax due to chance alone was calculated to be 4.3×10^{-7}, suggesting a relationship between this fragment and crop type. For the 'M 25', 'Ru' and 'Rm' series, as well as for the group of cultivars, significant differences were observed among accessions within groups. For the series of breeder's lines this component of variance ranged from 13.3% ('Rm') to 19.5% ('Ru'), whereas it was found to be highest in cultivars (40.5%). However, differences among individuals within accessions accounted for the major part of the variation in all four groups. For the 'Ru' and 'Rm' series this component of variance explained 80.7% and 83.6%, respectively, of the total variation detected (Table 2).

An UPGMA cluster analysis partitioned the samples into separate groups of fibre and oil flax only for the

Table 1 Germplasm of flax investigated in the present study. Accessions that have not yet been registered with a definite CGN accession number are denoted by their receipt number (Recnr.). In the last column the mean intra-accession similarity between the samples based on the AFLP data is presented (*n*=4, SD=standard deviation)

Accession name	Accession number	Crop type	Country of origin	Mean similarity (SD)
M 25–64	CGN19481	Fibre	USA	0.99 (0.01)
M 25–221	CGN19482	Fibre	USA	1.00 (0.00)
M 25–245	CGN19483	Fibre	USA	0.90 (0.02)
M 25–330	Recnr. 964680	Oil	USA	0.90 (0.02)
M 25–341	Recnr. 964681	Oil	USA	0.93 (0.02)
M 25–343	Recnr. 964682	Oil	USA	0.89 (0.02)
M 25–355	Recnr. 964683	Oil	USA	0.88 (0.02)
M 25–362	Recnr. 964684	Oil	USA	0.96 (0.02)
M 25–410	CGN19484	Fibre	USA	0.92 (0.03)
75 Ru	Recnr. 964875	Oil	Unknown	0.86 (0.02)
76 Ru	CGN20424	Fibre	Unknown	0.92 (0.04)
341 Ru	CGN19509	Fibre	Unknown	0.90 (0.05)
342 Ru	CGN20524	Fibre	Unknown	0.92 (0.02)
343 Ru	CGN19510	Fibre	Unknown	0.91 (0.04)
345 Ru	Recnr. 964880	Oil	Unknown	0.93 (0.03)
349 Ru	Recnr. 964881	Oil	Unknown	0.93 (0.03)
350 Ru	CGN20431	Fibre	Unknown	0.93 (0.04)
351 Ru	Recnr. 964882	Oil	Unknown	0.92 (0.02)
352 Ru	Recnr. 964883	Oil	Unknown	0.89 (0.02)
346 Rm	CGN20411	Fibre	Unknown	0.87 (0.03)
347 Rm	Recnr. 964867	Oil	Unknown	0.90 (0.05)
348 Rm	Recnr. 964868	Oil	Unknown	0.93 (0.02)
386 Rm	Recnr. 964872	Oil	Unknown	0.91 (0.02)
387 Rm	CGN20416	Fibre	Unknown	0.90 (0.02)
442 Rm	CGN20417	Fibre	Unknown	0.91 (0.03)
443 Rm	CGN20418	Fibre	Unknown	0.91 (0.02)
444 Rm	Recnr. 964873	Oil	Unknown	0.92 (0.02)
469 Rm	CGN20419	Fibre	Unknown	0.91 (0.02)
471 Rm	Recnr. 964874	Oil	Unknown	0.89 (0.01)
Rusland I	CGN20432	Fibre	USSR	0.93 (0.02)
Rusland II	Recnr. 964377	Fibre	USSR	0.95 (0.03)
Rembrandt	Recnr. 964013	Fibre	The Netherlands	1.00 (0.00)
139 Rumanien	CGN19512	Fibre	Romania	0.99 (0.00)
Russian Althanson	CGN20358	Fibre	unknown	0.95 (0.03)
Roma	CGN19471	Fibre	Italy	0.94 (0.02)
Italia Roma	Recnr. 964621	Oil	Argentina	0.91 (0.05)
Roman Winter	Recnr. 964742	Oil	The Netherlands	0.91 (0.02)
Moose Sel	Recnr. 964935	Oil	Australia	0.86 (0.03)
Svetoc	Recnr. 964754	Oil	USSR	0.90 (0.04)
Dakota	Recnr. 964568	Oil	USA	0.91 (0.06)
India Type 55	Recnr. 964618	Oil	India	0.89 (0.02)

Table 2 Results of a nested AMOVA based on the AFLP data. Analyses were performed separately for the three series of breeder's lines ('M 25', 'Ru' and 'Rm') and the cultivars and were carried out at three hierarchical levels: among the two groups of crop type, among accessions within groups and within accessions. *P*-values were derived from permutation tests and denote the probability of observing larger variance components at random

Variance component	M 25	Ru	Rm	Cultivars
Among groups (fibre-oil)	2.07 (34.0%) *P*<0.001	–0.01 (–0.2%) *P*=0.469	0.16 (3.1%) *P*=0.117	0.18 (3.0%) *P*=0.143
Among accessions within groups	0.92 (15.1%) *P*<0.001	0.98 (19.5%) *P*<0.001	0.70 (13.3%) *P*<0.001	2.39 (40.5%) *P*<0.001
Within accessions	3.10 (50.9%) *P*<0.001	4.03 (80.7%) *P*<0.001	4.39 (83.6%) *P*<0.001	3.34 (56.5%) *P*<0.001

'M 25' series, again indicating that in general crop type is not a main discriminant between the accessions investigated (Fig. 1). The group of oil flax samples of the 'M 25' series was found to be more closely related to the oil flax accession 'Moose Sel' than to the group of fibre flax samples of the 'M 25' series (Fig. 1d), probably related to the fact that in this series the effect of crop type was found to be the most pronounced. The 'Ru' and 'Rm' series grouped into a single separate cluster, indicating their close genetic relationship. The cultivar that appeared least distantly related to this cluster of breeder's lines was 'Roma', which might suggest a relationship between the name of this cultivar and the 'Rm' code.

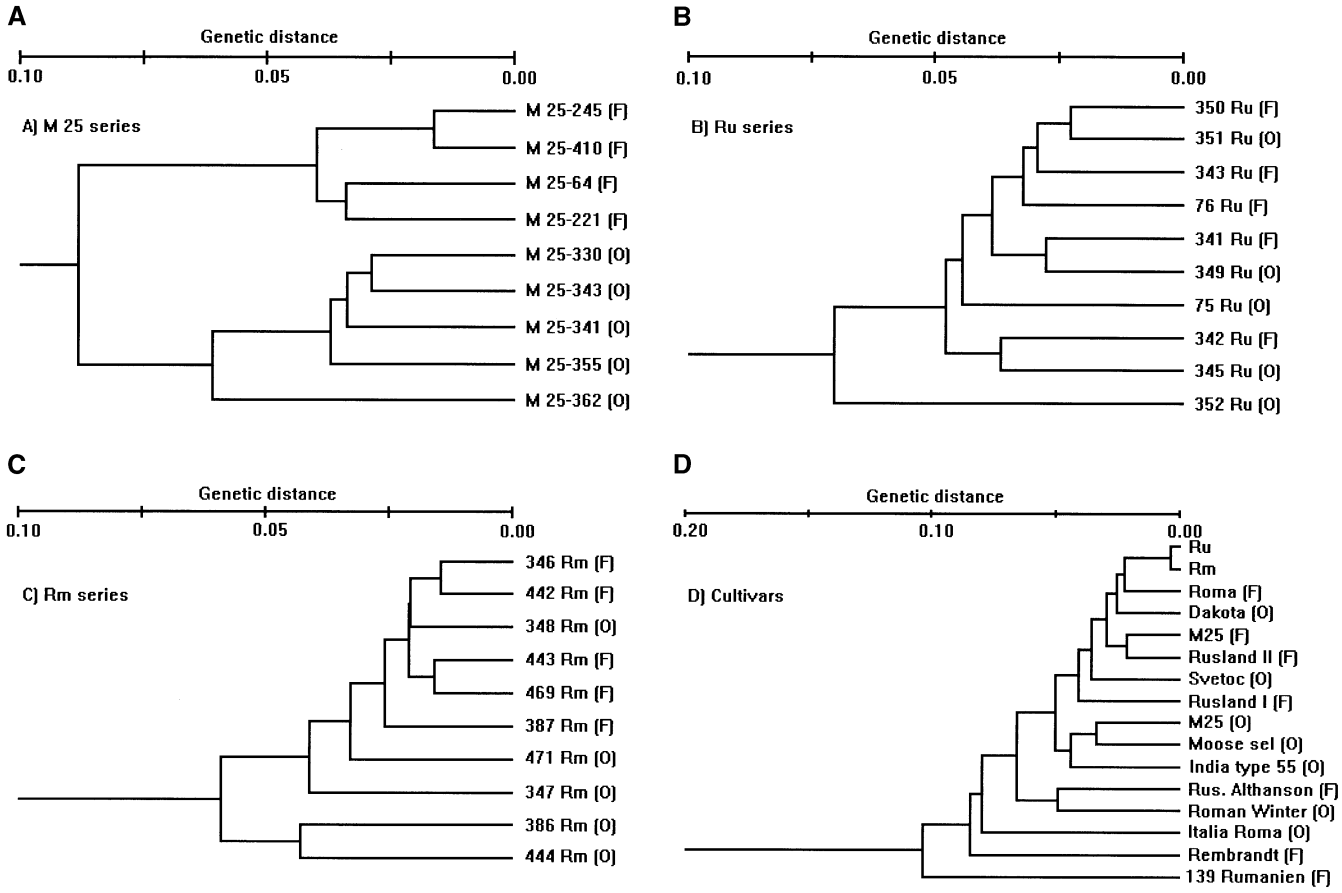

Fig. 1A–D UPGMA cluster analysis based on Nei's unbiased estimate of standard genetic distances within the three series of breeder's lines (**A**–**C**) and within the group of cultivars (**D**). In graph **D** individual samples from respectively the series 'Ru', the series 'Rm', the group of 'M 25 oil' and 'M 25 fibre' were lumped together and included for comparison. At the end of the branches the accession names are presented and the crop type is given between brackets (*F*=fibre, *O*=oil)

Rationalisation

Within the combined set of 29 accessions of breeder's lines, pairwise comparison of accessions by AMOVA revealed significant differences for all combinations consisting of an 'M 25' accession and an accession belonging to one of the two other series. These results indicated that the 'M 25' series is unrelated to the 'Ru' and 'Rm' series. Moreover, 'M 25' accessions of different crop type were always found to be significantly different, which was consistent with the results described in the previous section. Samples for which the among-accession component of variance was not found to be significantly larger than the within-accession component were stepwise bulked until all pairwise combinations of the matrix were significantly different from each other. Following this approach for the entire set of breeder's lines, accession 'M 25–245' could be bulked with 'M 25–410', and accession 'M 25–341' with 'M 25–343' and 'M 25–355.' Furthermore, the analysis resulted in five bulked groups consisting of accessions from the two other series: (1) '75 Ru', '349 Ru', '442 Rm', '443 Rm', '469 Rm' and '471 Rm', (2) '76 Ru', '343 Ru' and '387 Rm', (3) '341 Ru' and '342 Ru', (4) '346 Rm', '348 Rm', '350 Ru' and '351 Ru', (5) '347 Rm' and '352 Ru.' Based on the results of AMOVA, only accessions 'M 25–64', 'M 25–221', 'M 25–330', 'M 25–362', '345 Ru', '386 Rm' and '444 Rm' were maintained as separate entries. Thus, the entire set of 29 accessions of breeder's lines could be reduced to 14. Pairwise comparison of the reference accessions revealed significant differences between all cultivars, except for the combination 'Dakota'/'Moose Sel' (P=0.182).

As an illustration of the effects of the bulking procedures, a PCO plot of all individual samples of the 'M 25' series is presented in Fig. 2. The two principal axes shown in this two-dimensional plot explained 36.5% of the total variation. As might be expected when heterogeneous accessions are combined, the within-accession component of variance increased after bulking (Table 3). However, since only genetically similar samples were bulked, the increase was relatively small. As a direct consequence the among-accession component of variance was affected by bulking only to a small extent, showing a reduction of 2.6%. Thus, based on the material investigated in the present study a substantial reduction in the number of accessions can be realised in the flax collection of CGN, while at the same time maintaining similar levels of genetic variance.

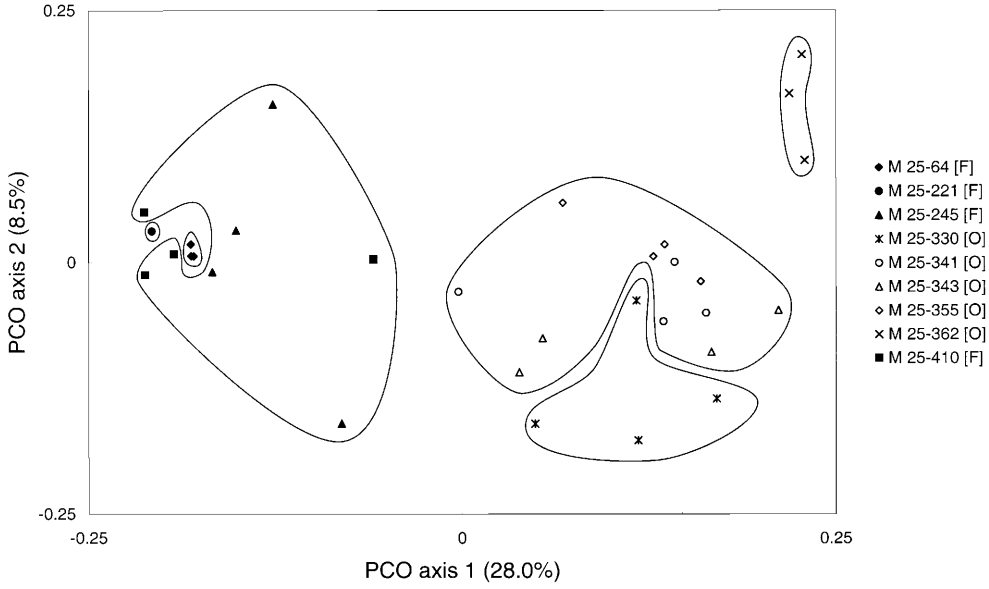

Fig. 2 Principal co-ordinate (PCO) plot of the individual flax samples of the M 25 series. The percentage variation explained by the axis is given between *brackets* in the axis legend. AMOVA was used to identify significantly different (groups of) accessions (*encircled* in the plot)

Table 3 Molecular variance components before bulking for the three separate series of breeder's lines. The effect of bulking on the variance among and within accessions is presented for the combined set of breeder's lines. The number of accessions before bulking and the number of elements after bulking is denoted by *n*

Germplasm	Variance component	Before bulking	After bulking
M 25 series		*n*=9	
	Among accessions	2.07 (40.0%)	
	Within accessions	3.10 (60.0%)	
Ru series		*n*=10	
	Among accessions	0.97 (19.4%)	
	Within accessions	4.03 (80.6%)	
Rm series		*n*=10	
	Among accessions	0.79 (15.2%)	
	Within accessions	4.39 (84.8%)	
M 25/Ru/Rm		*n*=29	*n*=14
combined	Among accessions	1.32 (26.4%)	1.21 (23.8%)
	Within accessions	3.68 (73.6%)	3.87 (76.2%)

Discussion

Knowledge about the distribution of genetic variation within and among populations plays a crucial role in conservation. For natural locations, these data can assist to determine where and how to sample populations in order to optimise the diversity captured for *ex situ* storage. For already existing collections, data on genetic structure may help to improve the composition of collections by identifying gaps within collections, removing redundant germplasm and establishing core collections. Ideally, the genetic diversity within a collection is distributed mainly between accessions, rather than within accessions. In the latter case, collection composition may be improved by combining accessions into bulked groups while ensuring minimum loss of the among-accession variation. This approach follows the rationale that in the absence of significant genetic differentiation between populations, samples can be considered as drawn from the same population, and hence can be combined. In the present study, this strategy was evaluated through analysis of accessions of a flax collection by estimating genetic variance components from AFLP data.

Various marker systems have been used for the identification of redundant germplasm, including isozymes (Hintum and Visser 1995; Hintum et al. 1996; Oliveira et al. 1997), RAPDs (Waycott and Fort 1994; Virk et al. 1995; Phippen et al. 1997; Zeven et al. 1998) and AFLPs (Cervera et al. 1998). AFLPs offer the advantage that the number of bands generated can be manipulated by modifying the number of selective nucleotides in PCR (Vos et al. 1995). This allows adjustment of the AFLP protocol to the genome size of the species involved. The genome of flax is relatively small, consisting of an estimated number of 280 Mbp per haploid genome (Cullis 1990). Therefore, only four selective nucleotides were used in PCR in the present study, generating a total number of 241 scorable bands by using only two primer combinations. This AFLP protocol provided an efficient strategy for germplasm analysis in flax since 144 of

these bands were found to be polymorphic, revealing sufficient resolving power in the present study. Completely identical genotypes were observed only within, and not between, accessions.

Apart from two cases, the flax accessions investigated appeared all but homogeneous. In addition, the AFLP profiles of some individuals seemed to indicate heterozygosity. For the 'M 25', 'Ru' and 'Rm' series this may be due to limited genetic purification following a cross between distinct parental lines or following a selection from a heterogeneous source. Mean observed heterozygosity for the three loci presumed codominant was estimated at 4.1%, which is in line with the value of 3.8% reported for predominantly selfing, dicotyledonous plant species on the basis of allozymes (Nevo et al. 1984). However, the reference cultivars were also found to be quite heterogeneous and to contain heterozygotes. Moreover, in case of observed heterozygosity at the loci presumed codominant, heterozygosity seemed to apply to multiple fragments. These results were not in line with predominant self-fertilisation, but rather suggest outcrossing between divergent populations. The rate of outcrossing in flax has been estimated at less than 3%, although values of 5–6% have been reported (Robinson 1937; Dillman 1938). However, since substantially higher values may be observed due to variation in flower morphology, safety measures to avoid outcrossing have been recommended by plant breeders (Williams 1988). Such precautions are not practised by genebanks during the regeneration of flax accessions. Therefore, the heterogeneity observed within the flax accessions studied may have possibly resulted from outcrossing during regeneration. This suggestion is supported by a high level of variation in flower colour observed within accessions during a recent field regeneration of 160 accessions in 1999. If genetic contamination is indeed common during the regeneration of flax, this could account for the lack of observed differentiation between fibre and oil flax cultivars. Because gene flow has a strong homogenising effect on populations, already contaminated accessions are potential candidates for bulking. This also means that the regeneration protocols of flax may need re-consideration in order to maintain the genetic integrity of accessions.

To determine whether observed differences between accessions were statistically significant, an analysis of molecular variance (AMOVA) was used. This procedure uses a non-parametric re-sampling approach to estimate the probability that the genetic distances of randomly generated groups exceed those of the observed data (Excoffier et al. 1992). Using this method, genetic differences between two homogeneous accessions will always be statistically significant, regardless of the number of polymorphisms, and closely related but homogeneous accessions will always be considered distinct. This is illustrated in Fig. 2 by the accessions 'M 25–64' and 'M 25–221' that despite their proximity or even encasement by accessions 'M 25–245' and 'M 25–410' remain as separate accessions. This is in line with the principle followed in cultivar registration, where varieties differing in just one characteristic are considered distinct provided that they are uniform and stable (UPOV 1991). Since AMOVA evaluates genetic differences between accessions in relation to variation within accessions, a reliable estimate of the latter component needs to be obtained. Because a high level of homogeneity was anticipated based on the high rate of self-fertilisation presumed, only four individuals per accession were analysed in the present study. However, the substantial levels of intra-accession variation encountered for the majority of samples suggests that sample sizes need to be increased in future analyses of flax.

To avoid unjustified bulking based on the results of AMOVA, a conservative level of significance of 0.1 was used to decide whether two accessions were to be considered different or not. Nevertheless, variance components will alter during each cycle when accessions are stepwise bulked. Eventually, this might result in bulked groups including accessions that in pairwise comparison differ significantly. However, this effect appeared to be restricted in the present study since it was only observed for the combination 'M 25–341'/'M 25–355' (P=0.081). Therefore, the conservative approach of bulking applied in the present study resulted in a reduction of only 2.6% of the initial among-accession component of variance. In case a larger reduction in the among-accession variance is accepted, a higher degree of rationalisation may be achieved by lowering the significance level in AMOVA to 0.05.

Within the series of breeder's lines, redundancies were to be expected based on the similarity of their accession names. However, redundancies were also identified between the 'Ru' and 'Rm' series. Moreover, these series clustered together into one single separate group indicating their close genetic relationship (Fig. 1d). The numbers that are part of the name of the accessions belonging to the two series, but also those from other series with comparable group codes (e.g. Ha, Hi, Pg, etc.), appeared to be unique. This suggests a common background of these groups and offers opportunities for substantial rationalisation. Potential rationalisation may apply not only to CGN's flax collection but also to other flax collections and other crops as genebanks often have included material from breeding programmes in their collections.

Implications for collection management

The results of the present study show how molecular data can assist genebank curators to identify redundant germplasm in order to increase efficiency in collection management. Obviously, accessions that are found identical or similar based on molecular genetic data may differ in just one important character. In practise, curators will therefore not decide for rationalisation only on the

basis of molecular data as long as additional information is available. However, in the absence of adequate passport and/or evaluation data, and when accessions are involved that are rarely requested by end-users, marker data can provide curators with the only available decision criterion. Rationalisation of a collection may be accomplished, e.g., by actually removing unwanted redundancies from the collection or by bulking genetically similar accessions. Removing accessions has the advantage that the genetic integrity of the remaining accessions is not violated, but also has the obvious disadvantage of the risk of losing unidentified but important diversity. Bulking has the advantage that in principle all the original material remains present in the collection and available for utilisation. However, bulking may increase the intra-accession diversity and hence may increase the threat of selection and drift if the usual genebank regeneration protocols are followed.

It was suggested that the three series of breeder's lines investigated could be reduced from 29 to 14 accessions. Assuming that these results can be extrapolated to the entire subset of possible redundancies, the total number of 317 accessions could be reduced to 153. Since the costs to maintain a flax accession at CGN are approximately €250 per generation, about €41,000 could be saved per generation following rationalisation. It is estimated that AFLP fingerprinting of four individuals from each of the 317 accessions using two primer pairs will cost approximately €35,000. Thus, the savings achieved through rationalisation of the collection will cover the expenses of the necessary molecular analyses within one generation.

Acknowledgements The authors thank Bert Visser and René Smulders for their comments on earlier drafts of the manuscript. The experiments described in this paper were carried out at Plant Research International B.V. (Wageningen, The Netherlands) and comply with current national legislation.

References

Arens P, Coops H, Jansen J, Vosman B (1998) Molecular genetic analysis of black poplar (*Populus nigra* L.) along Dutch rivers. Mol Ecol 7:11–18

Bretting PK, Widrlechner MP (1995) Genetic markers and plant genetic resource management. Plant Breed Rev 31:11–86

Brown SM, Kresovich S (1996) Molecular characterization for plant genetic resources conservation. In: Paterson H (ed) Genome mapping of plants. Academic Press, San Diego, pp 85–93

Cervera MT, Cabezas JA, Sancha JC, Martínez de Toda F, Martínez-Zapater JM (1998) Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theor Appl Genet 97:51–59

Clark RL, Shands HL, Bretting PK, Eberhart SA (1997) Germplasm regeneration: developments in population genetics and their implications. Managing large diverse germplasm collections. Crop Sci 37:1–6

Cullis CA (1990) DNA rearrangements in response to environmental stress. Adv Genet 28:73–97

Dillman AC (1938) Natural crossing in flax. J Amer Soc Agron 30:279–286

Excoffier L, Smouse PE, 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

Fulton TM, Chunwongse J, Tanksley SD (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209

Hintum ThJL van, Knüpffer H (1995) Duplication within and between germplasm collections. I. Identification of duplication on the basis of passport data. Genet Res Crop Evol 42: 127–133

Hintum ThJL van, Visser DL (1995) Duplication within and between germplasm collections. II. Duplication in four European barley collections. Genet Res Crop Evol 42:135–145

Hintum ThJL van, Boukema IW, Visser DL (1996) Reduction of duplication in a *Brassica oleracea* germplasm collection. Genet Res Crop Evol 43:343–349

Hodgkin T, Brown AHD, van Hintum ThJL, Morales EAV (1995) Core collections of plant genetic resources. John Wiley and Sons, Chichester, UK

Karp A, Seberg O, Buiatti M (1996) Molecular techniques in the assessment of botanical diversity. Ann Bot 78:143–149

Karp A, Edwards KJ, Bruford M, Funk S, Vosman B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D, Hewitt GM (1997) Molecular technologies for biodiversity evaluation: opportunities and challenges. Nature Biotechnol 15:625–628

Milbourne D, Meye R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Mol Breed 3:127–136

Miller MP (1997). Tools for population genetic analyses (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nevo E, Beiles A, Ben-Shlomo R (1984) The evolutionary significance of genetic diversity: ecological, demographic and life history correlates. In: Mani GS (ed) Evolutionary dynamics of genetic diversity. Springer, Berlin, pp 13–213

Oliveira JA, Lindner R, Bregu R, García A, González A (1997) Genetic diversity of westerwold ryegrass landraces collected in Northwest Spain. Genet Res Crop Evol 44:479–487

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998) What molecules can tell us about populations: choosing and using a molecular marker. Ecology 79:361–382

Phippen WB, Kresovich S, Candelas FG, 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.' Theor Appl Genet 94:227–234

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Robinson BB (1937) Natural cross-pollination studies in fibre flax. J Am Soc Agron 29:644–649

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438

UPOV (1991) International convention for the protection of new varieties of plants. International Union for the Protection of New Varieties of Plants, Geneva

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd, BV (1995) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90:1049–1055

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

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

Weir BS (1990) Genetic data analysis. Sinauer Associates, Sunderland, Massachusetts, USA

Whitkus R, Doebley J, Wendel JF (1994) Nuclear DNA markers in systematics and evolution. In: Phillips RL, Vasil IK (eds) DNA-based markers in plants. Adv Cell Mol Biol Plants, vol. 1, pp 116–141

Williams JH (1988) The pollination of linseed and flax. Bee World 69:145–151

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