Plant Breeding **124**, 446—453 (2005)
© 2005 Blackwell Verlag, Berlin

Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism

J. Vollmann[1], H. Grausgruber[1], G. Stift[2], V. Dryzhyruk[3] and T. Lelley[2]

[1]Department of Applied Plant Sciences and Plant Biotechnology, BOKU – University of Natural Resources and Applied Life Sciences Vienna, Gregor Mendel Str. 33, A-1180 Vienna, Austria, E-mail: johann.vollmann@boku.ac.at; [2]Department for Agrobiotechnology, IFA Tulln, BOKU – University of Natural Resources and Applied Life Sciences Vienna, Konrad Lorenz Str. 20, A-3430 Tulln, Austria; [3]Poltava Agricultural Academy, Skovoroda Str. 1/3, Poltava 36003, Ukraine

With 3 figures and 7 tables

Received November 6, 2004/Accepted March 23, 2005
Communicated by G. Röbbelen

Abstract

Camelina is an alternative oilseed crop species with limited information about the origin and diversity of available germplasm. Therefore, a set of 130 camelina accessions from a world collection was evaluated for oil content, protein content and 1000-seed weight in field experiments grown in three macro-environments in Austria. Based on phenotypic data, accessions were categorized into four groups with different seed characteristics using *k*-means cluster analysis or principal component extraction. Subsequently, a representative set of 41 accessions was subjected to random amplified polymorphic DNA (RAPD) analysis. Of 24 primers, 15 were polymorphic producing a total of 30 marker loci. Genetic distance estimates between the 41 accessions were calculated, based both on RAPD polymorphism and on seed quality characteristics, and dendrograms were generated for comparison. Similarities were found between the two different clustering approaches, and grouping was partly in agreement with pedigree information or geographic origin. However, as the two estimates of diversity sampled different segments of the genome, i.e. regions coding for seed characteristics or phenotypically neutral genomic regions highlighted by discrete markers, the correlation between the two distance matrices was low.

Key words: *Camelina sativa* — false flax — genetic distance — neutral genetic markers — oil content

Camelina (gold-of-pleasure, false flax) is an underexploited crop species from the Brassicaceae family, which is occasionally grown in Europe and in North America. Interest in camelina as an oilseed crop is due to its low-input features (Putnam et al. 1993, Vollmann et al. 1996) and a unique fatty acid profile, made up by a high content of α-linolenic acid and a low content of erucic acid (Zubr and Matthäus 2002). Similar to olive and rapeseed oils, diets of camelina oil can – apart from other effects – significantly reduce serum cholesterol concentration in human nutrition (Karvonen et al. 2002). Because of the high amount of ω-3 fatty acids present, the oxidative stability of camelina oil is relatively low, reducing shelf life when marketed, although it has proved to be more stable than fish oil or linseed oil (Ni Eidhin et al. 2003). In non-food applications, camelina oil may be utilized for environmentally safe paintings and coatings, in cosmetics and dermatological products, as well as a fuel for diesel engines (Bonjean and Le Goffic 1998, Bernardo et al. 2003).

Breeding of camelina, which is a strictly self-pollinating species, has not been carried out on an intensive level in the past. While its agronomic features have been considered to be quite acceptable, the oil content has been described as being in the low range of 320–440 g/kg seed dry matter (Seehuber 1984, Marquard and Kuhlmann 1986, Vollmann et al. 1996, Zubr 2003). Therefore, improvement of seed oil content would undoubtedly enhance the competitiveness of camelina in comparison with other oilseed crops. Because of its short breeding history, the availability of germplasm is limited. Most camelina genotypes accessible to date have originated from germplasm collections, botanical gardens and a first cycle of hybridization and selection (Seehuber et al. 1987). Genetic and other background information about the origin of particular genotypes would be essential for initiating well-focused breeding programmes (Mohammadi and Prasanna 2003), but is not available for most of the camelina accessions. In other oilseed crop species, molecular markers have frequently been used for categorizing germplasm, estimating genetic diversity, monitoring genetic variation in breeding processes or predicting heterosis, e.g. in oilseed rape (Becker et al. 1995, Riaz et al. 2001), sunflower (Köhler and Friedt 1999), linseed (Fu et al. 2003), soybean (Powell et al. 1996, Xu and Gai 2003) or oil pumpkin (Baranek et al. 2000, Paris et al. 2003). Apart from neutral molecular markers, highly heritable seed traits such as 1000-seed weight, seed protein and oil content or fatty acid patterns, which are the characters of primary interest in oilseed crops, have also been used sporadically for investigating the genetic diversity of germplasm collections and for grouping breeding material, e.g. in linseed (Diederichsen 2001) or soybean (Cui et al. 2001).

The aim of the present study was to provide an overview of the genetic diversity present within a camelina germplasm collection. In an initial approach, seed quality characteristics from a total of 130 accessions were used for categorizing genotypes. Subsequently, a reduced core set of accessions was subjected to random amplified polymorphic DNA (RAPD) analysis, and the level of similarity between the two estimates of genetic diversity was evaluated.

Materials and Methods

Plant materials: A total of 130 germplasm accessions of camelina, *Camelina sativa* [L.] Crtz., were evaluated for seed quality characters in three macro-environments as described below; the names of all genotypes are listed in Table 1. Entries designated with BGRC codes were developed during a pre-breeding project at the Federal

Table 1: List of 130 camelina accessions investigated and their grouping according to a *k*-means cluster analysis based on seed quality characters from three macro-environments

Number	Group 1	Group 2	Group 3	Group 4
1	BGRC 30019	BGRC 51562	258366 USSR	304270 Sweden
2	BGRC 51556	BGRC 51564	258367 USSR	304271 Sweden
3	BGRC 51557	BGRC 51566	304268 Sweden	BGRC 28347
4	BGRC 51559	BGRC 51569	304269 Sweden	BGRC 28347 selection
5	BGRC 51560	BGRC 51578	311735 Poland	BGRC 51572
6	BGRC 51563	Borowski Ikar	BGRC 30021 (Sv. Came)	BGRC 51575
7	BGRC 51565	Bot. Gard. Univ. Halle/Saale	BGRC 30021 (Sv. Came) selection	BGRC 51580
8	BGRC 51568	Kirgizskij	BGRC 51558	BGRC 51581
9	BGRC 51576	Land variety Poland 1978 No. 85	BGRC 51561	BGRC 51590
10	BGRC 51577		BGRC 51567	BGRC 51592
11	BGRC 51579		BGRC 51570	BGRC 51593
12	BGRC 51585		BGRC 51571	BGRC 51598
13	BGRC 51591		BGRC 51573	BGRC 51599
14	BGRC 51594		BGRC 51574	BGRC 51602
15	BGRC 51595		BGRC 51582	BGRC 51604
16	BGRC 51597		BGRC 51583	BGRC 51606
17	BGRC 51605		BGRC 51584	BGRC 51607
18	BGRC 51609		BGRC 51586	BGRC 51608
19	BGRC 51611		BGRC 51587	BGRC 51612
20	BGRC 51613		BGRC 51588	BGRC 51614
21	BGRC 51616		BGRC 51589	BGRC 51615
22	Bot. Gard. Univ. Dnepropetrovsk, USSR		BGRC 51596	BGRC 51617
23	Epona		BGRC 51600	Branisovicka
24	Inst. Agrobiol. Greifswald		BGRC 51601	Braunschweig C043
25	Gomholka		BGRC 51603	Brzybrodzka II
26	I.N.I.A. (2073-72)		BGRC 51610	Bot. Gard. Berlin-Dahlem
27	Land variety from the region Malopolska		Bronowska	LD 303
28	LD 2861		BOKU CA13X1S-15	LD-14
29	Minnesota		BOKU CA13X1S-23	LD-D
30	Oil plant assortment Giessen C038		Calena CA13X1S-9	False flax C115
31	Oil plant assortment Giessen (Polish origin)		Calinka CA2X-10	Korneuburg
32	Omskij		Celine	No. 637
33	Omskij Local		Czenstochowska	Oil plant assortment Giessen C039
34	Omskij Mestnyj		Irkutskij Local	Omskij Mestnyj
35	Pernice		Iwan	Rollsdorf
36	Type Klagenfurt		Lindo	VIR 4144
37	Volynskaja		Soledo	
38	Yamaji		Sortandinskij	
39	Zarja Socializma		Sortandinskij selection	
40	Zavolzskij		Sv. Came	
41			Ukrajinskij	
42			Voronezskij	
43			Voronezskij 339	
44			Wiledo	
45			Wroctawska	

Agricultural Research Centre (FAL) Braunschweig, Germany, during the 1980s (Seehuber et al. 1987) and are now maintained as a base population of camelina at the Institute of Plant Genetics and Crop Plant Research (IPK) at Gatersleben, Germany. Other accessions were received from different germplasm collections and institutions in Austria, Bulgaria, Czech Republic, France, Germany, Hungary, Poland, Romania, Russia, Spain, Sweden and the USA. For RAPD analysis, a representative set of 41 genotypes was selected out of the 130 accessions mentioned above; accession names, origin of samples, pedigree or other information available are given in Table 2. This selection of a subset of camelina accessions for RAPD analysis aimed at covering the whole geographic range of genotypes present in the larger set and was only based on geographic origin information, as pedigree or genetic information is available in a very few cases only.

Field experiments: Field experiments were carried out at Gross Enzersdorf in 1998 (GE 1998) and at Raasdorf in 1999 (RA 1999) and 2001 (RA 2001). Both sites are located approximately 25 km east of Vienna, Austria; location by year combinations were considered as three macro-environments. The set of 130 camelina accessions was planted in single-row plots of 2.5 m length and 0.3 m row spacing during the last week of March in each year. All accessions were sown in two replications using a 13×10 generalized lattice design. Prior to planting, nitrogen fertilization was applied at a rate of 50 kg/ha nitrogen. Single-row plots were harvested at the full maturity stage of the experiment between July 25 and July 31 in each season. From the harvest, about 15 g of dry seeds were analysed for oil and protein content by near-infrared reflectance spectroscopy (NIRS) using an InfraAlyzer model 450 spectrometer and IDAS CALIBRATION software (Bran & Luebbe, Norderstedt, Germany). NIRS calibrations with

Table 2: Name, origin, pedigree and other information of a core set of 41 camelina accessions subjected to random amplified polymorphic DNA (RAPD) analysis

Number	Name	Pedigree/information	Origin
1	Czenstochowska		Prague, Czech Republic
2	Bronowska	From Poland	Prague, Czech Republic
3	Brzybrodzka II		Poznan, Poland
4	Omskij		Prague, Czech Republic
5	Volynskaja		Prague, Czech Republic
6	Voronezskij		Cernovic, Russia
7	Sortandinskij		Kazakhstan
8	Kirgizskij		Kyrgyzstan
9	Zavolzskij		Prague, Czech Republic
10	Ukrajinskij	From Ukraine	Prague, Czech Republic
11	Wroctawska	From Poland	Prague, Czech Republic
12	Borowski Ikar		Sadovo, Bulgaria
13	Branisovicka		Sadovo, Bulgaria
14	Sv. Came		Alnarp, Sweden
15	Voronezskij 339		St Petersburg, Russia
16	Omskij Local		St Petersburg, Russia
17	Irkutskij Local		St Petersburg, Russia
18	Pernice	From Slovenia	Wies, Austria
19	Type Klagenfurt	From Carinthia (Austria)	Wies, Austria
20	Korneuburg		Korneuburg, Austria
21	Rollsdorf	Collected as a weed in a cabbage field in Styria (Austria)	Rollsdorf, Austria
22	BGRC 51558	Omskaja/Came	Braunschweig, Germany
23	BGRC 30021		Braunschweig, Germany
24	BGRC 51572	Sortandinskij/BGRC 28345	Braunschweig, Germany
25	BGRC 51586	Came/BGRC 28329	Braunschweig, Germany
26	BGRC 51587	Came/BGRC 28329	Braunschweig, Germany
27	BGRC 51592	Came/BGRC 28329	Braunschweig, Germany
28	BGRC 51593	Came/BGRC 28329	Braunschweig, Germany
29	Gomholka		Gödöllö, Hungary
30	I.N.I.A. (2073-72)		Madrid, Spain
31	Minnesota	'1980 crop' from Dr Robinson	St Paul, USA
32	Lindo		Lippstadt, Germany
33	Wiledo	Winter type	Braunschweig, Germany
34	Soledo		Braunschweig, Germany
35	Epona	Winter type	Paris, France
36	Celine		Paris, France
37	Yamaji	Landrace	Xingjiand, China
38	Calinka CA2X-10	Voronezskij//Omskaja/Came	Vienna, Austria
39	Calena CA13X1S-9	Omskaja/Came//Sortandinskij/BGRC 28345	Vienna, Austria
40	BOKU CA13X1S-23	Omskaja/Came//Sortandinskij/BGRC 28345	Vienna, Austria
41	BOKU CA13X1S-15	Omskaja/Came//Sortandinskij/BGRC 28345	Vienna, Austria

Table 3: Characteristics of random amplified polymorphic DNA (RAPD) primers exhibiting a polymorphic banding pattern in the set of 41 camelina accessions

Primer number	Designation	Polymorphic loci	Nucleotide sequence
1	88	2	5′-AAGACGACGG-3′
2	89	2	5′-AATCCGCTGG-3′
3	90	3	5′-AGTCGGCCCA-3′
4	91	2	5′-AACAGGGCAG-3′
5	92	1	5′-TGGAAGCACC-3′
6	93	1	5′-AGGCAGCCTG-3′
7	94	3	5′-GATGCGACGG-3′
8	96	1	5′-CCAGATGGGG-3′
9	98	2	5′-GGGGGCTTGA-3′
10	99	1	5′-TCCCGGTCTC-3′
11	102	2	5′-GAAGGCTCCC-3′
12	105	3	5′-GGGAGCGCTT-3′
13	106	2	5′-GTTCTCGGAC-3′
14	108	2	5′-AACTGGCCCC-3′
15	109	3	5′-TGGTCATCCC-3′

validation r^2 values of 0.94 and 0.90 for protein and oil content, respectively, had earlier been developed with reference samples from eight different environments. All seed oil and protein content data were expressed in g/kg oven-dried seed. The 1000-seed weight of samples was also calculated from the NIR scans as described by Vollmann et al. (1997). For this character, calibrations (validation r^2 from 0.86 to 0.90) were developed separately for each sampling season.

RAPD polymorphism analysis: Leaf material was bulked from 10 single plants for each of the 41 camelina genotypes described in Table 2. DNA extraction was carried out with a Promega Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA).

A set of 24 10-mer RAPD primers (Roth, Karlsruhe, Germany) was used for DNA amplification. The 15 µl amplification reaction contained 40 ng of genomic DNA, 0.3 µM of the 10-mer random primer, 1X reaction buffer, 1.5 mM $MgCl_2$, 200 µM dNTP and 0.5 U *Taq* polymerase. For amplification a MWG Primus 96 plus thermocycler (MWG AG, Ebersberg, Germany) was used. Nucleotide sequences of polymorphic primers together with their respective number of loci are given in Table 3.

The temperature regime was set to an initial denaturation of 60 s at 94°C, followed by 34 cycles of 60 s at 94°C, 45 s at 36°C, 30 s at 72°C, and finishing with an extension step of 5 min at 72°C.

Non-denaturing 10% polyacrylamide gels were prepared according to standard protocols. Fragment separation was done in a vertical electrophoresis unit (C.B.S. Scientific Company, Del Mar, CA, USA). Silver staining was carried out according to a protocol of Bassam and Caetano-Anollés (1993) with modifications described by Stift et al. (2003). For statistical evaluation of polymorphism, the presence of

bands was recorded visually, and data were converted to a binary (1–0) scale.

Statistical analysis of data: For each of the three field experiments, lattice-adjusted plot values and genotype mean values were generated for the seed quality characters using the appropriate lattice design ANOVA model and PLABSTAT software (Utz 1996). Subsequently, lattice-adjusted plot values were used in a combined analysis of variance across the three macro-environments, and the significance of genotype, macro-environment and genotype by macro-environment interaction effects was evaluated using *F*-tests. For each macro-environment, genotype mean values for oil content, protein content and 1000-seed weight were subjected to *z*-transformation (standardization with mean = 0 and SD = 1, respectively), and the nine variables from the 130 genotypes were entered to a *k*-means cluster analysis using SPSS version 8 software (SPSS Inc. 1997). The number of clusters was set to 4, cluster membership was recorded for each genotype, and the quality of the grouping obtained was evaluated in each variable by ANOVA for differences among cases in different clusters. For data reduction and visualization, the nine variables were also subjected to factor analysis using principal components (PCs) as an extraction method. Varimax rotation was applied to improve the interpretability of the factor solution and factor scores were saved as new variables (SPSS Inc. 1997). The RAPD polymorphism matrix of 41 genotypes and 30 marker loci was used to calculate Jaccard genetic distances between genotypes using RAPDISTANCE version 1.04 software (Armstrong et al. 1994). The genetic distance matrix was then subjected to UPGMA (average linkage) cluster analysis using NTSYS-PC version 1.8 software (Rohlf 1993). For assessing both RAPD polymorphism and seed quality-based estimates of diversity, a hierarchical cluster analysis was also carried out using the standardized seed quality data (Euclidean distance measure, average linkage) from the same 41 genotypes as selected for RAPD analysis, and dendrograms generated from both types of analysis were compared. In addition, the Mantel *Z*-statistic (Mantel 1967) was used to evaluate the correspondence between the RAPD-based distance matrix and the distance matrix calculated from seed quality data; the significance of *Z* was determined by comparing the observed *Z* value to a permutational distribution of *Z* values obtained by calculating *Z* values between one matrix and 2000 random permutations of the second matrix.

Results

Variation in seed quality characters

An overview of variation for seed quality characters across all 130 accessions of camelina is presented in Table 4 for each of the three macro-environments. An oil content of up to 460 g/kg was found in particular accessions grown at GE 1998 as well as RA 1999, whereas oil content was clearly lower at RA 2001; protein content was highest at GE 1998 and 1000-seed weight was largest at RA 1999. In a combined ANOVA (details not shown), differences between environments as well as between genotypes were found to be highly significant (*F*-test, $P < 0.001$) for each character. Estimates of heritability calculated from components of variance of a random effects model were 0.74, 0.68 and 0.80 for oil content, protein content and 1000-seed weight, respectively. Moreover, genotype by environment interaction was also highly significant, and the interaction component of variance was of the same magnitude as the genotype component for each of the characters, suggesting a crossover interaction. For this reason, oil content, protein content and 1000-seed weight were considered as separate variables in each of the three macro-environments in subsequent analyses of diversity. As a further consequence of the significant genotype by environment interaction, most accessions were ranked differently between the environments for a given character; only a few 'exceptional' genotypes were identified, e.g. *BOKU CA13X1S-15*, which was ranked within the top three of 130 genotypes for oil content in each environment. Other accessions with high oil content across environments were *LD-14*, *BGRC 28347*, *BGRC 51572*, *BGRC 51612* (from group 4 in Table 1) and *Celine* (from group 3 in Table 1). In addition, accessions *BOKU CA13X1S-15* and *Celine* were also high in 1000-seed weight. The correlation between oil and protein content was clearly negative in each experiment, whereas correlations between each of the two characters and 1000-seed weight were of minor importance (Table 5).

Germplasm diversity in seed quality characters

Based on seed quality data, the 130 camelina accessions were divided into four distinct groups by a *k*-means cluster analysis; group membership of particular genotypes is given in Table 1. Differences between groups were highly significant ($P < 0.001$) for each of the nine variables involved, thus demonstrating the meaningfulness of the grouping established. Using final cluster centres, groups may be characterized as follows. Group 1: genotypes low in seed weight and below average in oil content; group 2: lowest oil content, highest protein content and high seed weight; group 3: accessions with above average to high oil content, lowest protein content and medium seed weight; group 4: highest oil content, low seed weight. As no selection for seed quality characters had previously been carried out, recently developed breeding lines (BGRC codes) and other accessions were evenly distributed among all four groups (Table 1).

Factor analysis was applied as another approach of visualizing diversity in seed quality. Three PCs with an eigenvalue > 1 were extracted from the input data, explaining over 75% of the total variation present in the nine variables (Table 6). In a

Table 4: Overview of variation for oil content (g/kg), protein content (g/kg) and 1000-seed weight (g) in 130 camelina accessions for each of three macro-environments

	GE 1998			RA 1999			RA 2001		
	Oil content	Protein content	1000-seed weight	Oil content	Protein content	1000-seed weight	Oil content	Protein content	1000-seed weight
Mean	417.1	292.1	0.77	425.6	264.0	1.24	389.0	290.5	1.12
Minimum	372.5	266.1	0.56	380.9	235.3	0.97	357.8	272.7	0.74
Maximum	462.6	325.4	1.13	460.7	292.6	1.56	417.3	307.8	1.46
CV (%)	4.1	3.8	18.6	3.1	4.2	9.2	2.8	2.6	12.5

GE, Gross Enzersdorf; RA, Raasdorf; CV, coefficient of variation.

Table 5: Phenotypic coefficients of correlation between seed quality characters of camelina within as well as across environments

	GE 1998			RA 1999			RA 2001		
	Oil content	Protein content	1000-seed weight	Oil content	Protein content	1000-seed weight	Oil content	Protein content	1000-seed weight
Oil GE 1998	–								
Protein GE 1998	−0.786**	–							
1000-sw GE 1998	−0.158	0.011	–						
Oil RA 1999	0.457**	−0.35**	−0.192*	–					
Protein RA 1999	−0.080	0.333**	−0.133	−0.610**	–				
1000-sw RA 1999	0.150	−0.174*	0.722**	−0.116	0.062	–			
Oil RA 2001	0.555**	−0.395**	−0.164	0.571**	−0.071	−0.025	–		
Protein RA 2001	−0.160	0.480**	−0.189*	−0.242**	0.521**	−0.054	−0.306**	–	
1000-sw RA 2001	0.356**	−0.318**	0.435**	0.192*	−0.049	0.598**	0.525**	−0.179*	–

*,**, significant at P = 0.05 and P = 0.01, respectively.
GE, Gross Enzersdorf; RA, Raasdorf; sw, seed weight.

subsequent analysis, factor scores for the first and second PCs were calculated, and individual genotypes were plotted using both the respective factor scores and group membership information from the cluster analysis described above (Fig. 1). The figure reveals that grouping of genotypes based on cluster analysis of seed quality data proved to be consistent. Groups of camelina genotypes differing in oil content, protein content and 1000-seed weight were also separated in principal component analysis (PCA).

Table 6: Summary results of a principal component extraction from nine seed quality variables (components with eigenvalue > 1)

		Percentage of variance explained	
Component	Eigenvalue	Individual component	Cumulative
1	3.258	36.2	36.2
2	2.138	23.8	60.0
3	1.425	15.8	75.8

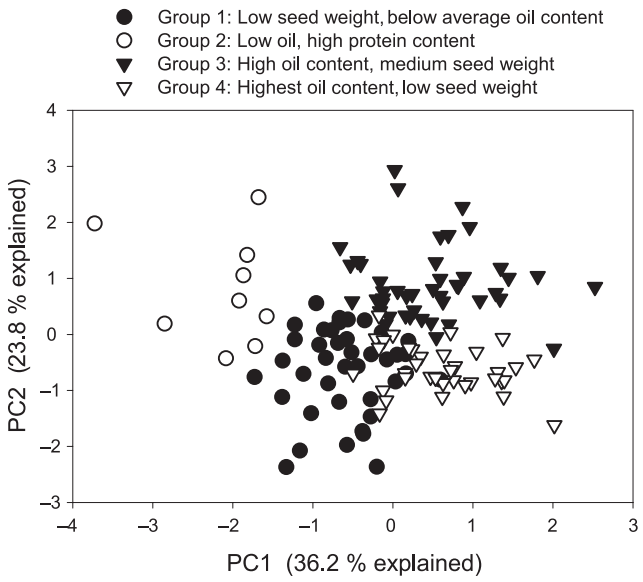

Fig. 1: Dispersion of 130 genotypes of camelina according to first and second principal components of seed quality data; symbols refer to group membership of genotypes according to *k*-means cluster analysis (Table 1)

RAPD-based analysis of diversity

From 24 primers applied in the 41 camelina accessions, 15 (63%) of them yielded polymorphic bands producing a total of 30 marker loci (Table 3). The remaining nine primers produced either no polymorphic band or were excluded because of unclear banding patterns. Based on the RAPD polymorphisms, pairwise genetic distances were calculated and the genetic diversity among accessions was visualized in the dendrogram given in Fig. 2. Two distinct groups of genotypes can be recognized, containing 11 and 28 of the 41 entries, respectively. In the smaller group of 11 genotypes, accessions *Czenstochowska*, *Bronowska* and *Celine* appeared to be genetically very similar (no polymorphism detected). Other members of that group are closely related because of their similar pedigrees (*BGRC 51587*, *BGRC 51558*, *Calinka*, *Calena*, *BOKU CA13X1S-23* and *BOKU CA13X1S-15*, see Table 2 for the respective pedigree information). All members of that smaller group were clustered together in group 3 (high oil content, medium seed weight) according to their seed quality data (see Table 1, Fig. 1). The larger group of 28 genotypes (Fig. 2) consists of a number of subgroups partly corresponding to the grouping based on the phenotypic data, subgroups because of geographic origin (e.g. *Volynskaja*, *Voronezskij*, *Voronezskij 399* or *Omskij*, *Kirgizskij*) and other subgroups previously not explainable due to a lack of pedigree information [e.g. the group made up by *Brzybrodzka II*, *Branisovicka*, *BGRC 51593*, *I.N.I.A. (2073-72)*, *Epona* and *Yamaji*].

Comparative evaluation of diversity estimates

In an attempt to evaluate the meaningfulness and similarity of the two diversity patterns obtained from phenotypic data (seed quality characters) vs. genetic distances (RAPD polymorphisms), the phenotypic classification of genotypes (cluster membership of Table 1) was used as a grouping variable to compare RAPD-based genetic distances within- and between-groups. In Table 7, average pairwise genetic distances are listed within and between groups 1, 3 and 4 (group 2 was excluded, because it contained only two entries in the core set of accessions used for RAPD analysis); distances within groups are smaller than between groups, except for the

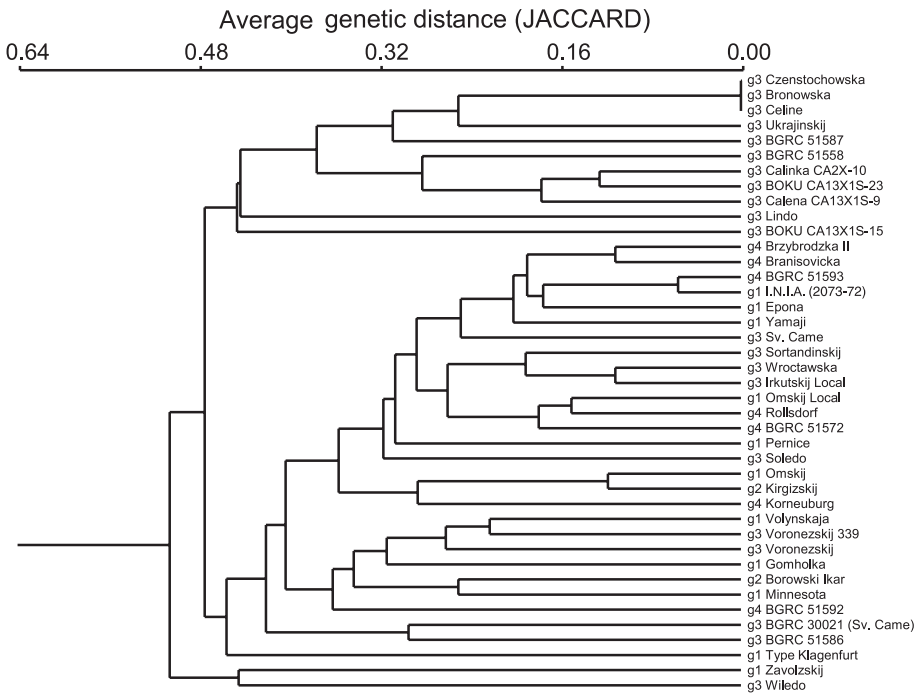

Fig. 2: Relationship between 41 accessions of camelina (core set) as revealed by hierarchical cluster analysis of random amplified polymorphic DNA (RAPD)-based genetic distance estimates; prefixes g1, g2, g3 or g4 in genotype designations indicate the group membership of genotypes according to the *k*-means cluster analysis (see Table 1)

Table 7: Average pairwise genetic distances within and between groups 1, 3 and 4 (grouping according to cluster analysis of seed quality characters, see Table 1) based on the Jaccard distances measure of random amplified polymorphic DNA (RAPD) data (741 pairwise distances) from the camelina core set

Group	1	3	4
1	0.416		
3	0.442	0.419	
4	0.358	0.432	0.307

distance between groups 1 and 4. Although these two groups are clearly different in oil content, the members of the two groups selected for RAPD analysis are apparently genetically rather similar to each other; therefore, group 4 has to be understood as a high oil content subgroup of group 1 in Fig. 1.

Moreover, a hierarchical cluster analysis was carried out using seed quality data from the reduced set of 41 genotypes. For comparison with the RAPD-based dendrogram (Fig. 2), a dendrogram based on phenotypic dissimilarity was generated (Fig. 3) revealing two distinct groups of entries and one separated individual (accession *Minnesota*). The upper group in Fig. 3 (from *Czenstochowska* to *BOKU CA13X1S-15*) contains all entries from the smaller group in Fig. 2, except for accession *BGRC 51587*; all entries in the upper group of Fig. 3, except for *Rollsdorf* and *BGRC 51572*, are from the high oil content/medium seed weight group of accessions (group 3 in Table 1 and Fig. 1, respectively). The second group in Fig. 3 consists of a few subgroups, which were also partly found in the previous analysis (e.g. *Brzybrodzka II*, *Branisovicka* and *BGRC 51593*), but are partly different from the RAPD-based associations. Although there are apparent similarities between the groupings of particular genotypes, the overall correspondence between the two distance matrices appears to be rather low. A correlation between distance estimates of $r = 0.101$ indicates a poor fit. The Mantel Z-statistic evaluating the correspondence between the two distance matrices is significant at a level of $P = 0.076$ only.

Discussion

Phenotypic diversity in seed quality characters such as oil content is crucial for genetic improvement and establishment of camelina as an oilseed crop. Therefore, an overview of variation within the available germplasm can immediately be utilized in breeding approaches, e.g. for developing populations combining larger seed size and high oil content using genotypes of different groups listed in Table 1. In camelina germplasm, the presence of significant genetic variation in seed quality parameters such as oil content, fatty acid profile or glucosinolate content has repeatedly been described (e.g. Budin et al. 1995, Schuster and Friedt 1998, Zubr and Matthäus 2002, Zubr 2003). In the present study, genetic variation in 1000-seed weight, oil and protein content allowed an informative grouping of the 130 germplasm accessions investigated (Fig. 1).

Phenotypic variation in a set of characters reflects genetic distinctiveness between genotypes. For this reason, similarity in diversity patterns obtained from seed quality (Fig. 2) and molecular markers (Fig. 3) has been found in the set of camelina germplasm investigated. However, the classification of particular genotypes was different for the two systems, and the correlation between genetic distances derived from each of the two approaches was low, raising the question of whether phenotypic or DNA data would provide more accurate estimates of diversity. Neutral DNA markers such as RAPD are discretely highlighting polymorphism in anonymous regions randomly distributed over the whole genome, thus depicting overall variation at the DNA level independent of selection. Diversity estimates from phenotypic data are based on differential gene expression of functional genomic regions coding for the phenotypic characters evaluated, thus reflecting polymorphism in certain regions of the genome, which may be

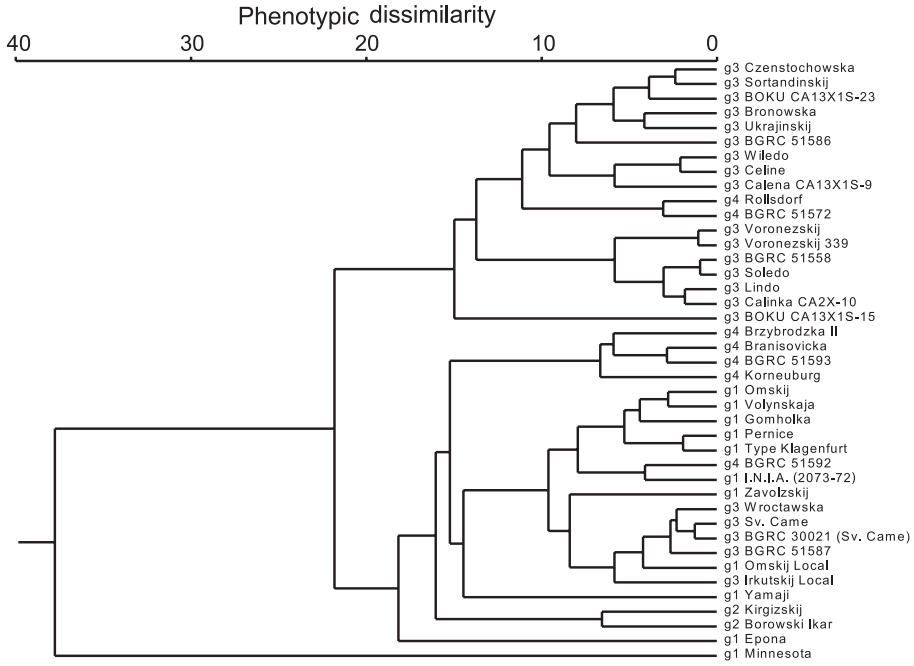

Fig. 3: Relationship between 41 accessions of camelina (core set) according to hierarchical cluster analysis of phenotypic dissimilarity in seed quality characters; prefixes g1, g2, g3 or g4 in genotype designations indicate the group membership of genotypes according to the *k*-means cluster analysis (see Table 1)

subject to selection. For this reason, the two different approaches will not necessarily yield closely matching results. A diversity estimate from phenotypic data – although covering only a small fraction of the overall diversity – would be most valuable for plant breeding or germplasm conservation applications. In the current study, the low correlation between phenotypic and DNA marker-based estimates of diversity might – apart from several other reasons – be due to low genetic variation present. Only 63% of markers were polymorphic yielding only one to three loci per marker, which suggests a low degree of genetic diversity within this possibly polyploid species when compared with other self-pollinated crops. Thus, a much higher number of markers and/or more phenotypic data would be required for a closer matching of results. In such a case, the merit of using both sources of information can be demonstrated. RAPD distance between clusters 1 and 4 was low identifying them as one group (Table 7), whereas the two clusters were clearly separated by their difference in oil content (Fig. 1).

Comparisons between phenotypic and marker-based estimates of diversity have rarely been published in plant breeding research (Mohammadi and Prasanna 2003). In core collections of sorghum, similarities of a RAPD-based grouping and various clusterings according to seed morphological traits were low and differed from race to race and for different clustering techniques (Dahlberg et al. 2002). In soybean, Chen and Nelson (2004) reported that RAPD polymorphisms and phenotypic traits (agronomic character scores, seed traits, fatty acid profile) were in good agreement to discriminate between accessions of cultivated soybean (*Glycine max* [L.] Merr.), wild soybean (*G. soja* Sieb & Zucc.) and semiwild intermediate types; however, similarities in clustering within each of the three groups were not evaluated. In evolutionary studies, the correlation between quantitative characters and diversity in DNA or allozyme markers has more frequently been investigated, but results were inconsistent in meta-analyses including a large number of individual studies (Merilä and Crnokrak 2001, Reed and Frankham 2001). In order to achieve estimates of genetic diversity more relevant to plant breeding or germplasm conservation, Van Tienderen et al. (2002) have suggested exploiting polymorphisms of genetic markers in functional regions of the genome, thus targeting genes rather than random markers. Apart from diversity estimates, functional markers would also improve quantitative trait loci (QTL) mapping and could allow an easier transfer of QTL markers between different populations (Andersen and Lübberstedt 2003). The effectiveness of non-random markers in a plant breeding application has been demonstrated by Li and Quiros (2001), who developed a set of sequence-related amplified polymorphisms (SRAP) targeting at coding sequences in oilseed rape. Their SRAP markers were used to determine genetic distances between a set of oilseed rape lines, and genetic distances were highly correlated with grain yield of the respective hybrids (Riaz et al. 2001).

In the light of the above discussion and with an emphasis on camelina breeding, both the diversity estimates based on seed quality data and on RAPD polymorphisms might be equally relevant for characterization of the germplasm accessions investigated in this study, as they revealed informative groupings of the material available. Apparently, more research is needed for combining and jointly interpreting phenotypic and genetic marker data in diversity analysis for plant breeding and germplasm management.

References

Andersen, J. R., and T. Lübberstedt, 2003: Functional markers in plants. Trends Plant Sci. **8,** 554—560.

Armstrong, J. S., A. J. Gibbs, R. Peakall, and G. Weiller, 1994: The RAPDistance Package. Available at: http://life.anu.edu.au/molecular/software/RAPDistance/, accessed on: 21 October 2004.

Baranek, M., G. Stift, J. Vollmann, and T. Lelley, 2000: Genetic diversity within and between the species *Cucurbita pepo*, *C. moschata* and *C. maxima* as revealed by RAPD markers. Cucurbit Genet. Coop. Rep. **23,** 73—77.

Bassam, B. J., and G. Caetano-Anollés, 1993: Silver staining of DNA in polyacrylamide gels. Appl. Biochem. Biotechnol. **42,** 181—188.

Becker, H. C., G. M. Engqvist, and B. Karlsson, 1995: Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. Theor. Appl. Genet. **91,** 62—67.

Bernardo, A., R. Howard-Hildige, A. O'Connell, R. Nichol, J. Ryan, B. Rice, E. Roche, and J. J. Leahy, 2003: Camelina oil as a fuel for diesel transport engines. Ind. Crops Prod. **17,** 191—197.

Bonjean, A., and F. Le Goffic, 1998: La cameline – *Camelina sativa* (L.) Crantz: une opportunité pour l'agriculture et l'industrie européenne. OCL – Oleagineux Crops Gras Lipides **5,** 20—25.

Budin, J. T., W. M. Breene, and D. H. Putnam, 1995: Some compositional properties of camelina (*Camelina sativa* L. Crantz) seeds and oils. J. Am. Oil Chem. Soc. **72,** 309—315.

Chen, Y., and R. L. Nelson, 2004: Genetic variation and relationships among cultivated, wild, and semiwild soybean. Crop Sci. **44,** 316—325.

Cui, Z., T. E. Carter Jr, J. W. Burton, and R. Wells, 2001: Phenotypic diversity of modern Chinese and North American soybean cultivars. Crop Sci. **41,** 1954—1967.

Dahlberg, J. A., X. Zhang, G. E. Hart, and J. E. Mullet, 2002: Comparative assessment of variation among sorghum germplasm accessions using seed morphology and RAPD measurements. Crop Sci. **42,** 291—296.

Diederichsen, A., 2001: Comparison of genetic diversity of flax (*Linum usitatissimum* L.) between Canadian cultivars and a world collection. Plant Breeding **120,** 360—362.

Fu, Y.-B., G. G. Rowland, S. D. Duguid, and K. W. Richards, 2003: RAPD analysis of 54 North American flax cultivars. Crop Sci. **43,** 1510—1515.

Karvonen, H. M., A. Aro, N. S. Tapola, I. Salminen, M. I. J. Uusitupa, and E. S. Sarkkinen, 2002: Effect of alpha-linolenic acid-rich *Camelina sativa* oil on serum fatty acid composition and serum lipids in hypercholesterolemic subjects. Metabolism **51,** 1253—1260.

Köhler, H., and W. Friedt, 1999: Genetic variability as identified by AP-PCR and reaction to mid-stem infection of *Sclerotinia sclerotiorum* among interspecific sunflower (*Helianthus annuus* L.) hybrid progenies. Crop Sci. **39,** 1456—1463.

Li, G., and C. F. Quiros, 2001: Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theor. Appl. Genet. **103,** 455—461.

Mantel, N. A., 1967: The detection of disease clustering and a generalized regression approach. Cancer Res. **27,** 209—220.

Marquard, R., and H. Kuhlmann, 1986: Untersuchungen über Ertragsleistung und Samenqualität von Leindotter (*Camelina sativa* Crtz.). Fette - Seifen - Anstrichmittel **88,** 245—249.

Merilä, J., and P. Crnokrak, 2001: Comparison of genetic differentiation at marker loci and quantitative traits. J. Evol. Biol. **14,** 892—903.

Mohammadi, S. A., and B. M. Prasanna, 2003: Analysis of genetic diversity in crop plants – salient statistical tools and considerations. Crop Sci. **43,** 1235—1248.

Ni Eidhin, D., J. Burke, and D. O'Beirne, 2003: Oxidative stability of omega-3-rich camelina oil and camelina oil-based spread compared with plant and fish oils and sunflower spread. J. Food Sci. **68,** 345—353.

Paris, H. S., N. Yonash, V. Portnoy, N. Mozes-Daube, G. Tzuri, and N. Katzir, 2003: Assessment of genetic relationships in *Cucurbita pepo* (Cucurbitaceae) using DNA markers. Theor. Appl. Genet. **106,** 971—978.

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

Putnam, D. H., J. T. Budin, L. A. Field, and W. M. Breene, 1993: Camelina: a promising low-input oilseed. In: J. Janick, and J. E. Simon (eds), New Crops, Exploration, Research and Commercialization, 314—322. John Wiley and Sons, Inc., New York, USA.

Reed, D. H., and R. Frankham, 2001: How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution **55,** 1095—1103.

Riaz, A., G. Li, Z. Quresh, M. S. Swati, and C. F. Quiros, 2001: Genetic diversity of oilseed *Brassica napus* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. Plant Breeding **120,** 411—415.

Rohlf, F. J., 1993: NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System, Vers. 1.80. Applied Biostatistics Inc., New York, USA.

Schuster, A., and W. Friedt, 1998: Glucosinolate content and composition as parameters of quality of *Camelina* seed. Ind. Crops Prod. **7,** 297—302.

Seehuber, R., 1984: Genotypische Variabilität in Ertrags- und Qualitätsmerkmalen bei Mohn und Leindotter. Fette - Seifen - Anstrichmittel **86,** 177—180.

Seehuber, R., J. Vollmann, and M. Dambroth, 1987: Anwendung der Single-Seed-Descent Methode bei Leindotter (*Camelina sativa* (L.) Crantz) zur Erhöhung des Ertragsniveaus. Landbauforschung Völkenrode **37,** 132—136.

SPSS Inc., 1997: SPSS for Windows, Release 8.0.0. SPSS Inc., Chicago, IL, USA.

Stift, G., M. Pachner, and T. Lelley, 2003: Comparison of RAPD fragment separation in agarose and polyacrylamide gel by studying *Cucurbita* species. Cucurbit Genet. Coop. Rep. **26,** 62—65.

Utz, H. F., 1996: PLABSTAT – Plant Breeding Statistical Program, Vers. 2N. Institut für Pflanzenzüchtung, Saatgutforschung und Populationsgenetik, Universität Hohenheim, Stuttgart, Germany.

Van Tienderen, P. H., A. A. de Haan, C. G. van der Linden, and B. Vosman, 2002: Biodiversity assessment using markers for ecologically important traits. Trends Ecol. Evol. **17,** 577—582.

Vollmann, J., A. Damboeck, A. Eckl, H. Schrems, and P. Ruckenbauer, 1996: Improvement of *Camelina sativa*, an underexploited oilseed. In: J. Janick (ed.), Progress in New Crops, 357—362. American Society for Horticultural Science, Alexandria, VA, USA.

Vollmann, J., A. Damboeck, S. J. H. Kuyt, and P. Ruckenbauer, 1997: Determination of camelina seed weight using near-infrared reflectance spectroscopy. Plant Varieties Seeds **10,** 95—101.

Xu, D. H., and J. Y. Gai, 2003: Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. Plant Breeding **122,** 503—506.

Zubr, J., 2003: Qualitative variation of *Camelina sativa* seed from different locations. Ind. Crops Prod. **17,** 161—169.

Zubr, J., and B. Matthäus, 2002: Effects of growth conditions on fatty acids and tocopherols in *Camelina sativa* oil. Ind. Crops Prod. **15,** 155—162.