Breeding Science 59: 447–452 (2009)

Note

Assessment of the congruity of genetic relationships and variation revealed by individual- and bulked-samples-based approaches using RAPD and ISSR markers in Japanese turnip (*Brassica rapa* ssp. *rapa*) cultivars

Kenji Wakui*[1], Hiroyoshi Iwata[2], Yu Takahashi[3], Yoshihito Takahata[3] and Junzo Fujigaki[1]

[1] *Department of Bioproduction Technology, Junior College of Tokyo University of Agriculture*, 1-1-1 Sakuragaoka, Setagaya, Tokyo 156-8502, Japan

[2] *Department of Information Science and Technology, National Agricultural Research Center*, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8666, Japan

[3] *Faculty of Agriculture, Iwate University*, 3-18-8 Ueda, Morioka, Iwate 020-8550, Japan

To find out an efficient and accurate way to estimate the genetic relationship among cultivars having within-variety genetic variation by using different types of approaches, we compared among-varieties genetic similarity estimated based on individual samples and also on bulked population samples of six Japanese turnip (*Brassica rapa* ssp. *rapa*) cultivars using RAPD and ISSR markers. In both individual- and bulked-samples-based approaches, 37 individuals were sampled from each cultivar, amplified individually and also as bulked DNA, respectively. Genetic similarity matrices among six cultivars were estimated based on four similarity measures estimated with 125 or 124 polymorphic bands scored in the individual- or bulked-samples-based approach, respectively, and the correlations between the similarity matrices were calculated to evaluate the degree of agreement between matrices. Correlations between similarity matrices calculated based on individuals on the one hand and those calculated based on bulked samples on the other were also high and significant in all combinations of the four similarity measures. Use of computational analysis indicated that 6 to 8 individuals per cultivar were enough to accurately estimate similarity with the individual-based approach.

Key Words: *Brassica rapa* ssp. *rapa*, Random amplified polymorphic DNA (RAPD), Inter-simple sequence repeat (ISSR), Individual-based approach, Bulked-samples-based approach.

Introduction

RAPD and ISSR markers are widely employed to reveal genetic variation in many plant species because they do not require knowledge of genomic sequences and also because these protocols are relatively simple, rapid and cost effective (Bussell *et al.* 2005, Semagn *et al.* 2006, Agarwal *et al.* 2008). When evaluating the genetic relationship between varieties which have genetic heterogeneity in each variety using such technology, heterogeneity should be taken into consideration by sampling multiple individuals from each variety (Yu and Pauls 1993, Segovia-Lerma *et al.* 2003). It is also desirable to use a number of individuals that may contain all the alleles that constitute each population and variety. That is, a number of individuals that include the variability within each variety will allow for the determination and suitable evaluation of genetic variability. Moreover, analysis of multiple individual genotypes per variety provides highly informative data about a population's genetic structure (Yu and Pauls 1993, Ghérardi *et al.* 1998). However, these practical applications are still limited in cases where a large number of populations need to be examined. One approach to overcome this limitation is to bulk genomic DNA samples. Bulked DNA drastically reduces the number of samples that need to be processed, which allows the generation of genetic information in a shorter period at a cheaper cost (Fu 2000). Bulked DNA has been used to investigate genetic variation among cultivars of several species including alfalfa (Segovia-Lerma *et al.* 2003), flax (Fu *et al.* 2003a), radish (Muminović *et al.* 2005) and *Brassica napus* (Mailer and May 1999). In some species, although suitable bulk-size and the congruity between genetic relationships estimated from individual samples on one hand and from bulked population samples on the other have been investigated, the species and variety used were not uniform.

B. rapa, including turnip (ssp. *rapa*), is an allogamous species owing to its self-incompatible system, thus turnip cultivars are expected to be genetically heterogeneous, especially

Communicated by T. Hayashi
Received July 22, 2009. Accepted October 30, 2009.
*Corresponding author (e-mail: wakuik@nodai.ac.jp)

when they are indigenous cultivars. Turnip differentiated into over 100 local varieties in each district after it was introduced in Japan (Yoshikawa and Yui 1991, Shinohara 1984). To assess the genetic relationships among such local varieties of turnip, the number of individuals should be so large that they can sufficiently capture the heterogeneous variation within a cultivar. When trying to use bulked population samples of local turnip varieties in the future, the congruity between estimated genetic relationships based on individual samples and on bulked population samples is also indispensable. From this point of view, we conducted a study to compare variation between individual and bulked population samples of Japanese turnip (*Brassica rapa* ssp. *rapa*) cultivars using RAPD and ISSR markers. In this study, we also determined the number of individuals per cultivar that is necessary to capture the genetic variation within a cultivar with a simple computational approach.

Materials and Methods

Plant materials and DNA preparation

Six Japanese commercial varieties of turnip (*Brassica rapa* ssp. *rapa*), Kamo-sugukina, Hinona-kabu, Tennoji-kabu, Kuretsubo-kabu, Nozawana and Hakuba, provided by Noguchi Co. (Saitama, Japan), were used. Thirty-seven seedlings were randomly collected for each cultivar, and then DNA of individual samples was extracted from the first leaf of seedlings by the CTAB method (Rogers and Bendich 1988). In the individual-based approach, DNA of 37 individuals was individually used as template for PCR. In the bulked-samples-based approach, 500 ng of DNA from each individual plant was mixed, and the mixed DNA solution was analyzed in the same way as the individual-based approach.

RAPD and ISSR amplification

Primers were chosen by prior analysis. RAPD analysis was performed using 10 random 10-mer primers (Operon Co.) (Table 1). RAPD PCR was carried out in a 10 µl volume containing 5–20 ng genomic DNA, 1X PCR (2 mM Tris-HCl pH 8.0, 10 mM KCl, 2 mM $MgCl_2$, 10 µM EDTA, 0.1 mM DTT, 0.05% Tween 20 and 0.05% Nonidet P-40), 200 µM dNTPs, 0.5 µM primer and 0.25 U *Taq* polymerase HS (Takara Bio Inc.). Amplification was performed with a PTC-100 thermal cycler (MJ Research Inc. USA) programmed for 3 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 38°C and 1 min at 72°C. The reaction was terminated with a final step of 5 min at 72°C, followed by soaking at 4°C until recovery. For ISSR analysis, 10 ISSR primers from primer set No. 9 (Biotechnology Laboratory of University of British Columbia, Vancouver, B.C., Canada) were used for PCR (Table 1). These primers were mostly 17- to 18-mers. Amplification was carried out as for RAPD PCR except for primer concentration, which was 0.2 µM in the reaction mixture. Amplification was programmed for 3 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 1 min at 53°C and 2 min at 72°C and a final 10-min extension at 72°C, and by holding at 4°C. Amplification products were separated by electrophoresis through a 1.5% (w/v) agarose gel in 1X TAE. DNA bands were stained with ethidium bromide,

Table 1. Sequence of primers employed and the number of scored total and polymorphic bands based on RAPD and ISSR analysis

Name of primers	Sequences (5′-3′)	Individual-based approach		Bulked-samples-based approach		Size range (bps)
		Total no. of bands	No. of polymorphic of bands (P%)[a]	Total no. of bands	No. of polymorphic of bands (P%)[a]	
OPA-1	CAGGCCCTTC	7	7	7	3	600–3300
OPA-2	TGCCGAGCTG	4	4	9	4	700–2700
OPA-3	AGTCAGCCAC	4	2	3	1	700–1700
OPA-4	AATCGGGCTG	4	3	9	7	800–4000
OPA-7	GAAACGGGTG	5	5	8	5	<500–4000
OPA-8	GTGACGTAGG	3	3	4	2	600–2300
OPA-9	GGGTAACGCC	3	3	5	5	700–3000
OPA-10	GTGATCGCAG	6	6	10	8	600–3800
OPA-11	CAATCGCCGT	9	9	9	7	500–3500
OPA-20	GTTGCGATCC	5	4	8	5	600–3500
UBC-807	$(AG)_8T$	17	17	22	19	600–4300
UBC-808	$(AG)_8C$	4	3	16	14	800–4500
UBC-810	$(GA)_8T$	8	8	13	11	700–4500
UBC-811	$(GA)_8C$	13	12	10	5	650–5000
UBC-812	$(GA)_8A$	12	12	10	7	800–5000
UBC-813	$(CT)_8T$	4	4	7	6	2000–>5000
UBC-814	$(CT)_8A$	5	5	4	1	2000–5000
UBC-815	$(CT)_8G$	7	7	9	7	950–5000
UBC-822	$(TC)_8A$	6	6	6	2	1800–>5000
UBC-823	$(TC)_8C$	5	5	7	5	600–5000
Total		131	125 (95%)	176	124 (70%)	

[a] polymorphism rate.

visualized under UV light and photographed. RAPD and ISSR bands were scored manually as present (1) or absent (0). Only distinct major bands were chosen for this study.

Data analysis

Genetic similarity among the six cultivars was calculated using four similarity measures. Nei's index for genetic similarity (Nei 1972) and Wright's Fst based similarity (1-Fst; Wright 1951) were calculated for individual data. Jaccard's similarity coefficient (Jaccard 1901) and simple matching (SM) similarity index (Sneath and Sokal 1973) were calculated for bulked-samples data. Correlation coefficients between the four different similarity matrices were calculated and their significance was tested by using the Mantel matrix correspondence test using 10,000 random permutations (Mantel 1967). Genetic relationships among cultivars were investigated by cluster analysis using the unweighted pair-group method with arithmetic mean (UPGMA) based on genetic distance matrices for individual data (i.e., Nei's genetic distance and Fst) calculated from the similarity indices or the bulked-samples data (i.e., 1-[Jaccard coefficients] and 1-[Simple matching index].

To determine the number of samples that are necessary to accurately estimate similarity among cultivars, we conducted a simple computational analysis. The correlation between genetic similarity calculated from data with a complete dataset (i.e., from 37 individuals) based on a genetic distance (Nei 1972) and the genetic similarity based on a partial dataset (i.e., dataset with m (m=2, 4, 6, 8, 12, 14, 16, 18) individuals selected randomly from the 37 individuals) were calculated. We replicated the above calculation 1000 times, and calculated the minimum, maximum, and mean of correlation coefficients for each m.

Results

After preliminary screening of 20 RAPD and 100 ISSR primers, 10 primer combinations for each marker system (i.e., RAPD and ISSR) produced good amplification products and were chosen for RAPD and ISSR analysis. Twenty primer combinations used in this study amplified 125 and 124 polymorphic bands in the individual- and bulked-samples-based approaches, respectively (Table 1). In the individual-based approach, amplification of 37 individual DNA per primer detected clear variations within each variety (Fig. 1). Although the band pattern was the sum of variation within the variety most detected by the individual method in bulk, there was also a band which was not detected.

Similarity matrices among six cultivars were calculated with four different similarity measures, and correlation coefficients between matrices were estimated (Table 2). Correlation coefficients between two similarity measures based on individual samples (i.e., Nei vs. 1-Fst) and one between two similarity measures for bulked population samples (i.e., Jaccard vs. SM) both exceeded 0.99. Correlation coefficients between a similarity measure calculated on individual samples on one hand and those calculated on bulked population samples on the other were also high and highly significant ($P < 0.01$). Of them, the correlation between similarity matrices based on individual and bulked samples was highest (0.873) between Nei and Jaccard, suggesting that Jaccard's similarity coefficient is a better similarity measure than SM for bulked population samples, assuming that the similarity estimate based on individual samples is closer to true genetic similarity than one based on bulked population samples. UPGMA trees constructed on the basis of individual- and bulked-samples data had the same topology among different similarity (or distance) measures (Fig. 2), indicating that similarity estimates based on bulked population samples were so close to the those based on individual samples that we could estimate a similar pattern of genetic diversity among varieties. The branch length of Hakuba, however, was longer in trees based on individual data (Fig. 2A and 2B) than in trees based on bulked-samples data (Fig. 2C and 2D), indicating that the estimates of genetic similarity might be biased at some level by using bulked population samples instead of population samples.

Using computational analysis to determine the number of individuals that were necessary for the individual-based approach, we obtained the minimum, maximum, and mean correlation between genetic similarity calculated from data with the complete dataset (i.e., dataset with 37 individuals) on one hand and the genetic similarity calculated on a partial dataset (i.e., dataset with m (m=2, 4, 6, 8, 12, 14, 16, 18) individuals) on the other (Fig. 3). As the number of individuals per cultivar increased so too did the correlation with genetic

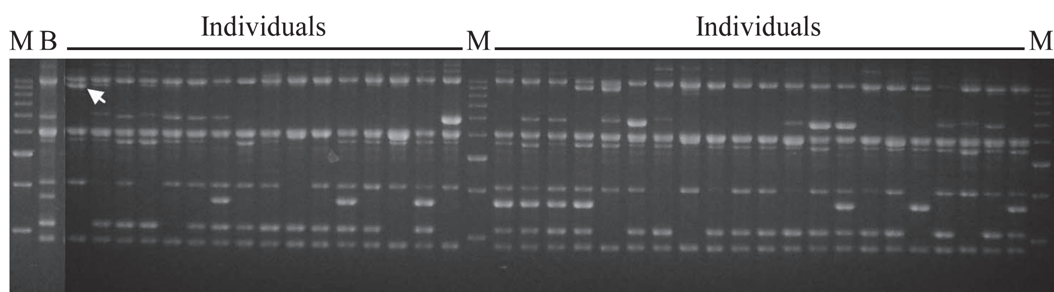

Fig. 1. ISSR banding patterns produced by primer UBC-815 primer with bulked sample (lane; B) and 37 individual samples of *B. rapa* ssp. *rapa* cv. Hinona-kabu. The arrow shows a rare band detected in the individual-based approach. M is a 500 bp ladder DNA.

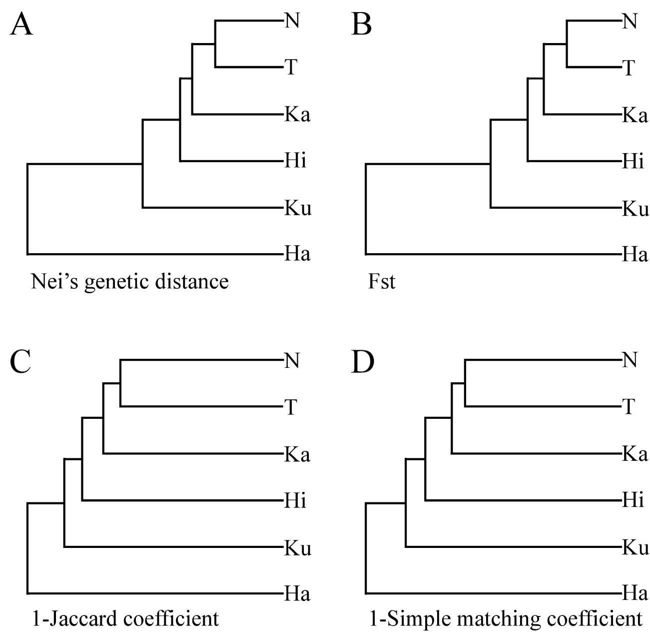

Fig. 2. UPGMA trees based on genetic distance estimated by using RAPD and ISSR among six cultivars. The trees were constructed on the basis of Nei's genetic distance (A), Fst (B), 1-Jaccard coefficient (C) and 1-Simple matching coefficient (D). N: Nozawana, T: Tennoji-kabu, Ka: Kamo-sugukina, Hi: Hinona-kabu, Ku: Kuretsubo-kabu, Ha: Hakuba.

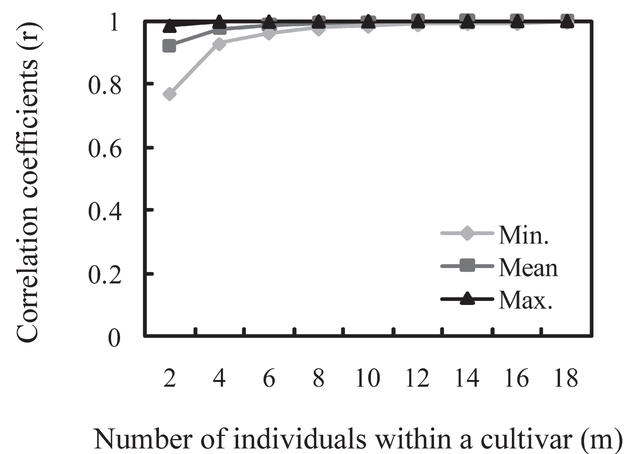

Fig. 3. Correlation coefficients between the genetic identities based on all data vs. each m revealed by 1000 times simulation. The values of minimum, maximum and mean for each correlation are shown.

similarity. However, the degree of increase in the correlation was larger when the number of individuals was smaller. The correlation between datasets based on four individuals (m = 4) and the complete dataset was already 0.9275, which was higher than 0.8737 (Nei vs. Jaccard), shown in Table 2. The mean correlation was 0.987 at m = 6 (Fig. 4A) and 0.992 at m = 8 (Fig. 4B), respectively.

Discussion

In RAPD or AFLP analysis, when evaluating the consistency of an individual method and a bulk method from the degree of coincidence of the band pattern detected, in many cases, it is inharmonious (Fu *et al.* 2003b, Mailer and May 1999). Michelmore *et al.* (1991) found that rare RAPD markers could not be detected in bulks when the allele was contained in the DNA mixture at less than 10%. Lower-frequency or unique alleles present in only a few individuals may have been diluted to such an extent that they were not amplified efficiently (Segovia-Lerma *et al.* 2003). The rare band detected in the individual-based approach disappeared in the bulked-sampled-based approach in this study (Fig. 1). A band pattern is simplified in a bulk method and is effective for characterization of a variety (Kölliker *et al.* 2001). Although the band pattern of individual and bulk methods is not necessarily coincident, the major difference in the genetic relationship between varieties is not found in some cases (Fu *et al.* 2003a, Kölliker *et al.* 2001). Divaret *et al.* (1999) reported that scoring the most representative bands of one population is not necessarily equivalent to scoring all the bands existing in this population. The more frequent bands as well as the intense bands were characteristics of one population. In this study, scoring only of the clear band of the method according to individual and each bulk method was carried out independently, and high coincidence was obtained between genetic-relatedness. However, in the bulk method, use of a small sample size produces distortion in the detection of rare alleles (Divaret *et al.* 1999, Fu *et al.* 2003b, Mailer and May 1999). On the other hand, if sample size of bulk becomes large, a large difference in banding pattern will no longer be found (Divaret *et al.* 1999). This shows the importance of evaluating the number of individuals which include variance of the variety used for analysis (Yu and Pauls 1993). Further, determining the minimum number of individuals by which the variety within a species is characterized increases the efficiency of analysis, and it provides reliability to a result.

Fujimoto and Yamagishi (1996) used 12 individuals were per cultivar to investigate the phylogenetic relationships of Sugukina and turnip by RAPD analysis. In this study, we used 37 individuals for each cultivar to determine the number of individuals necessary for capturing the genetic variation within a cultivar. We estimated genetic similarity based on data with a different number of samples, and found that the correlation between genetic similarity calculated on the complete dataset (with 37 individuals) and similarity on a partial dataset (with m individuals) was close to 1 even when m was small (Fig. 3). This result indicates that 37 individuals were sufficient to evaluate the genetic relationship among cultivars.

Our findings have several implications for RAPD and ISSR analysis of genetic relationships in Japanese turnip. First, high correlations between different similarity measures for individual samples and bulked population samples indicate that genetic similarity among cultivars is stable (i.e., highly congruent) between different measures and also

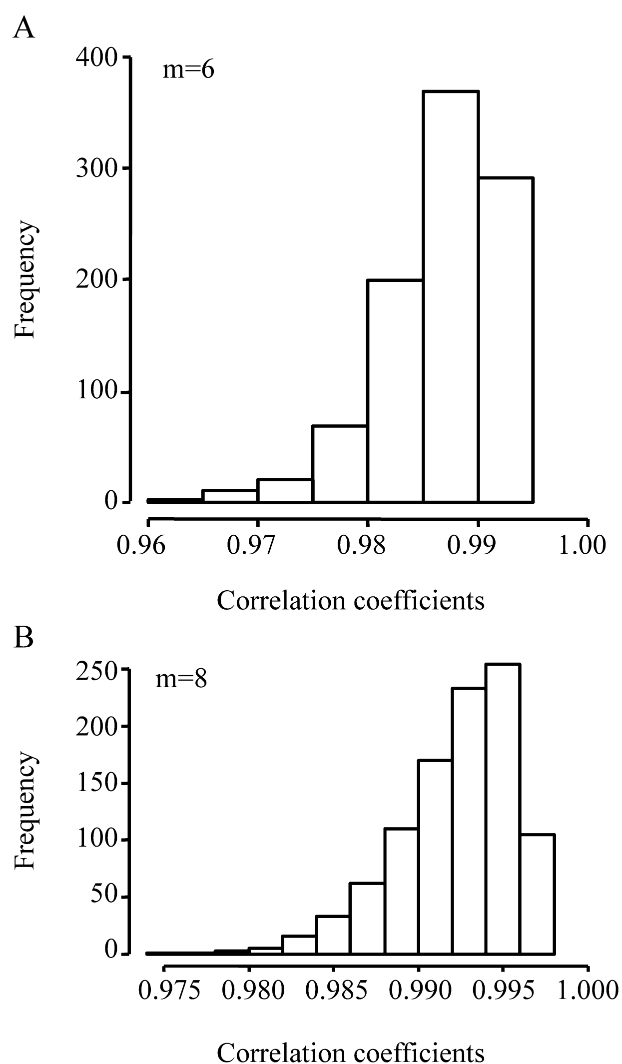

Fig. 4. Correlation coefficients between genetic identities based on all data vs. each m = 6 (A) and m = 8 (B) revealed by 1000 times simulation. The values of minimum, maximum and mean for each correlation are shown.

between different approaches (individual- and bulked-samples-based), suggesting that the latter approach can detect genetic relationships that are quite close to the former approach. Second, by simple computational analysis, the rapid increase of correlation coefficients indicates that the minimum number of individuals to assess genetic variation was 4. Since the correlation coefficients among genetic similarities were close to 1 for datasets with 6 and 8 individuals per cultivar, 6 to 8 individuals are desirable to be used as samples per cultivar to capture the genetic variation profiling a cultivar. This result can serve as a basis of determination of the number of individuals in a bulk method, when evaluating genetic similarity among Japanese cultivars of *B. rapa* ssp. *rapa* broadly from now on.

Acknowledgements

We thank Dr. H. Fujimaki, Tokyo University of Agriculture, for constructive comments on an earlier version of the manuscript.

Literature Cited

Agarwal, M., N. Shrivastava and H. Padh (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. 27: 617–631.

Bussell, J.D., M. Waycott and J.A. Chappill (2005) Arbitrarily amplified DNA markers as characters for phylogenetic inference. Perspect. Plant Ecol. Evol. System 7: 3–26.

Divaret, I., E. Margalé and G. Thomas (1999) RAPD markers on seed bulks efficiently assess the genetic diversity of a *Brassica oleracea* L. collection. Theor. Appl. Genet. 98: 1029–1035.

Fu, Y.B. (2000) Effectiveness of bulking procedures in measuring population-pairwise similarity with dominant and co-dominant genetic markers. Theor. Appl. Genet. 100: 1284–1289.

Fu, Y.B., S. Guerin, G.W. Peterson, J.E. Carlson and K.W. Richards (2003a) Assessment of bulking strategies for RAPD analyses of flax germplasm. Genet. Res. Crop Evol. 50: 743–746.

Fu, Y.B., S. Whitwill, Y.S.N. Ferdinandez, B.E. Coulman and K.W. Richardes (2003b) Bulking generated considerable bias in detection of amplified fragment length polymorphism variations in oat, fringed brome and smooth bromegrass. Mol. Ecol. Notes 3: 476–478.

Fujimoto, T. and H. Yamagishi (1996) Investigation on the phylogenetic relationship between 'Sugukina' and turnip by RAPD analysis. J. Japan. Soc. Hort. Sci. 65: 609–614.

Ghérardi, M., B. Mangin, B. Goffinet, D. Bonnet and T. Huguet (1998) A method to measure genetic distance between allogamous populations of alfalfa (*Medicago sativa*) using RAPD molecular markers. Theor. Appl. Genet. 96: 406–412.

Jaccard, P. (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles 37: 547–579.

Kölliker, R., E.S. Jones, M.Z.Z. Jahufer and J.W. Forster (2001) Bulked AFLP analysis for the assessment of genetic diversity in white

Table 2. Correlations between similarity matrixes calculated with each matching coefficient among six cultivars in *B. rapa* ssp. *rapa* (above diagonal). The significance of the correlation was tested by 1000 random permutations (** significant at the 0.1% probability level) (below diagonal)

Matching coefficient	Nei's genetic identity[a]	1-Fst[a]	Jaccard coefficient[b]	Simple matching coefficient[b]
Nei's genetic identity		0.9985	0.8737	0.8529
1-Fst	0.0013**		0.8572	0.8364
Jaccard coefficient	0.0014**	0.0016**		0.9967
Simple matching coefficient	0.0021**	0.0020**	0.0014**	

[a] Matching coefficient of individual-based approach.
[b] Matching coefficient of bulked-samples-based approach.

clover (*Trifolium repens* L.). Euphytica 121: 305–315.

Mailer, R.J. and C.E. May (1999) Heterogeneity of random amplified polymorphic DNA sequences in individual seedlings and bulked samples of four cultivars of *Brassica napus*. Plant Breed. 118: 465–470.

Mantel, N. (1967) The detection of desease clustering and generalized regression approach. Cancer 27: 209–220.

Michelmore, R.W., I. Paran and R.V. Kesseli (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. 88: 9828–9832.

Muminović, J., A. Merz and A.E. Melchinger (2005) Genetic structure and diversity among radish varieties as inferred from AFLP and ISSR analyses. J. Amer. Soc. Hort. Sci. 130: 79–87.

Nei, M. (1972) Genetic distance between populations. Am. Nat. 106: 283–292.

Rogers, S.O. and A.J. Bendich (1988) Extraction of DNA from plant tissues. *In*: Galvin, S.B. and R.A. Schilperoort (eds.) Plant Molecular Biology Manual, Kluwer Academic Publishers, Dordrecht, The Netherlands, A6: 1–10.

Segovia-Lerma, A., R.G. Cantrell, J.M. Conway and I.M. Ray (2003) AFLP-based assessment of genetic diversity among nine alfalfa germplasms using bulk DNA templates. Genome 46: 51–58.

Semagn, K., Å. Bjørnstad and M.N. Ndjiondjop (2006) An overview of molecular marker methods for plants. Afr. J. Biotech. 5: 2540–2568.

Shinohara, S. (1984) Turnip. *In*: Vegetable Seed Production Technology of Japan, Particulars. 1, pp. 239–268.

Sneath, P.H.A. and R.R. Sokal (1973) Numerical Taxonomy, W.H. Freeman and Company, San Francisco, CA.

Wright, S. (1951) The genetical structure of populations. Ann. Eugen. 15: 323–354.

Yoshikawa, H. and S. Yui (1991) Classification and identification of crucifer vegetables in Japan. JICA Ref. Series 4: 97–114.

Yu, K. and K.P. Pauls (1993) Rapid estimation of genetic relatedness among heterogeneous populations of alfalfa by random amplification of bulked genomic DNA samples. Theor. Appl. Genet. 86: 788–794.