Theor Appl Genet (2001) 102:689–694

J. Plieske · D. Struss

Microsatellite markers for genome analysis in *Brassica*. I. development in *Brassica napus* and abundance in *Brassicaceae* species

Received: 14 April 2000 / Accepted: 3 July 2000

Abstract One hundred and twenty one microsatellites were identified by screening a λ phage library of *Brassica napus*. The distribution of these microsatellites within *Brassicaceae* species was estimated using 81 locus-specific primer pairs. Most of them (83%) amplified fragments either from *Brassica oleracea* or *Brassica campestris*, or from both species, whereas less than 30% detected loci in *Brassica nigra*. The same was true (30–35%) for more-distantly related crucifer species such as *Diplotaxis* ssp., *Brassica tournefortii*, *Sinapis alba*, *Raphanus sativus* and *Eruca sativa*. Only 16 microsatellite-specific primer pairs (19.8%) amplified fragments from *Arabidopsis thaliana*. Moreover, 61 of the primer pairs detecting 198 polymorphisms were used to estimate the extent of genetic diversity among 32 *Brassica napus* varieties and breeding lines. On average, four alleles per locus were observed. The spring and winter types of oilseed rape could be clearly distinguished by using the microsatellite markers in a cluster analysis. The results demonstrated the high efficiency of these markers for monitoring genetic diversity.

Keywords *Brassica napus* · Microsatellite marker · Genetic diversity

Introduction

Various molecular markers have been developed for many crops and applied both to study genetic diversity and to discriminate between genotypes. In *Brassica*, RFLPs and RAPDs have been extensively used for phylogenetic studies and genetic mapping (Song et al. 1990, 1991; Hu and Quiros 1991; Chyi et al. 1992; Uzunova et al. 1995; Struss et al. 1996; Chèvre et al. 1997; Cloutier et al. 1997). However, the utilization of RAPDs and RFLPs for genome analysis seems to be restricted because of the dominant character and the low specificity of RAPDs, and the cumbersome technique of RFLPs.

Microsatellites, defined as simple sequence repeats (SSRs), are randomly interspersed within eukaryotic genomes. They are highly variable with regard to repeat number, show co-dominant inheritance and are highly efficient in the fingerprinting and pedigree analysis of different crops (Plaschke et al. 1995; Rongwen et al. 1995; Guilford et al. 1997; Struss and Plieske 1998). The RFLP technique was one of the first methods used for genome analysis but it is laborious and less polymorphic than microsatellites (Akkaya et al. 1992; Bell and Ecker 1994). For microsatellite analyses only a small amount of DNA is needed and, since they are a PCR-based marker system, they are amenable to automation.

Microsatellites of *Brassica napus* have been previously studied by Lagercrantz et al. (1993), Kresovich et al. (1995), Szewc-McFadden et al. (1996) and Uzunova and Ecke (1999). However, the limited number of markers seems to be insufficient for detailed genetic diversity studies, particularly in advanced breeding material and elite cultivars.

In this paper we report on the identification of more SSR loci in *B. napus* and their distribution within *Brassicaceae* species. The microsatellites were used to estimate the genetic distances between rapeseed varieties and breeding lines.

Communicated by H.C. Becker

J. Plieske (✉) · D. Struss
Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), 06466 Gatersleben, Germany

Present address:
D. Struss, Department of Pomology, University of California, Davis, CA 95616, U.S.A.

Materials and methods

Plant material

Thirty two rapeseed varieties and breeding lines from different European breeders and countries were surveyed in this study (Ta-

Table 1 Rapeseed varieties and breeding lines surveyed with SSRs

Number of lines	Variety	Source of material Breeder	Country	Type	Canola quality[a]
1	Breeding line	Serasem	France	Winter	00
2	Bienvenu	Serasem	France	Winter	0+
3	Breeding line	SLM	Germany	Winter	00
4	Breeding line	SLM	Germany	Winter	00
5	Breeding line	SLM	Germany	Winter	00
6	Breeding line	SLM	Germany	Winter	00
7	Breeding line	SLM	Germany	Winter	00
8	DCH 10/87	DSV/Cargill	Germany/France	Winter	00
9	Vivol	DSV/Cargill	Germany/France	Winter	00
10	Falcon	NPZ	Germany	Winter	00
11	Breeding line	NPZ	Germany	Winter	00
12	Breeding line	NPZ	Germany	Winter	00
13	Idol	DSV/Cargill	Germany/France	Winter	00
14	Breeding line	DSV	Germany	Winter	00
15	Legend	Svalöf-Weibull	Sweden	Spring	00
16	Lenora	SLM	Germany	Winter	++
17	Liberator	DSV	Germany	Winter	00
18	Lirajet	DSV	Germany	Winter	00
19	Madora	SLM	Germany	Winter	00
20	Bristol	DSV/Cargill	Germany/France	Winter	00
21	Breeding line	NPZ	Germany	Winter	00
22	Breeding line	NPZ	Germany	Winter	00
23	Breeding line	NPZ	Germany	Winter	00
24	Breeding line	NPZ	Germany	Winter	00
25	PF 044	PajbjergFonden	Denmark	Spring	00
26	Profit	AgCd	Canada	Spring	00
27	Sollux	SAS	Germany	Winter	++
28	Wotan	NPZ	Germany	Winter	00
29	Breeding line	DSV	Germany	Winter	00
30	Norde	Svalöf-Weibull	Sweden	Winter	++
31	Gorczanski	ZNRIO	Poland	Winter	++
32	Score	CPB	United Kingdom	Winter	00

[a] 00 = erucic acid free, low in glucosinolates
0+ = erucic acid free, high in glucosinolates
++ = contains erucic acid, high in glucosinolates

ble 1). The material consisted of 29 winter and three spring types including old land races such as ''Leonora'' as well as new high yielding varieties of canola quality like ''Falcon'' and ''Idol''. Moreover, a series of various crucifers including the *Brassica* species forming the U triangle (U 1935) were investigated, i.e. *Brassica campestris*, *Brassica nigra*, *Brassica oleracea*, *B. napus*, *Brassica juncea*, *B. carinata*, *Diplotaxis* ssp., *Brassica tournefortii*, *Sinapis alba*, *Raphanus sativus*, *Eruca sativa* and *Arabidopsis thaliana*. These were obtained from the germplasm collection of the IPK, Gatersleben, or from the *Brassica* collection of the Institute of Agronomy and Plant Breeding, University of Göttingen, respectively. To avoid problems resulting from heterogeneity and heterozygosity of the analyzed plant material, each species was represented by a single plant and the varieties and breeding lines, additionally, by their selfed progenies.

Identification of microsatellites

Identification of microsatellites was performed according to Struss and Plieske (1998). In detail, a phage library was constructed from genomic DNA of the rapeseed variety ''Vivol'' by digestion of the DNA with the methylation-sensitive restriction enzyme *Pst*I. The low-molecular-weight fraction was isolated from an agarose gel, digested with *Mbo*I and cloned into the *Bam*HI site of the λ phage vector ZAP-Express (Stratagene). The library was screened for the presence of repeats by hybridization with $(GA)_n$ and $(GT)_n$ probes, labeled by random priming with ^{32}P-αdCTP (Feinberg and Vogelstein 1983). Positive plaques were isolated and used to transfect *Escherichia coli* strain XL 1 blue. The colonies were again hybridized with $(GA)_n$ and $(GT)_n$ probes, and positive clones were sequenced on A.L.F. (Pharmacia) or ABI 377 (Applied Biosystems) automated sequencers.

Primer design and PCR

Primers flanking microsatellite loci were designed using the computer program Primer 0.5 (E. Lander, Cambridge, Mass., USA), the parameters were chosen as recommended in the program. The forward primers were labelled with Fluorescein or FAM. For PCR 50–150 ng of template DNA, 250 nmol of each primer, 200 µmol of dNTPs, 1 U of *Taq* polymerase, 1.5 mM of $MgCl_2$ were mixed in a total volume of 25 µl. Forty five cycles of amplification (denaturing at 94°C for 1 min, annealing at 50°C, 55°C or 60°C, depending on the primer pair, for 1 min, extension at 72°C for 2 min) were followed by a final extension step at 72°C for 1 h. The amplification products were detected on denaturing polyacrylamide gels using A.L.F. or ABI 377 automated sequencers following the manufacturer's protocol.

Data analysis

The fragments amplified by microsatellite primers were scored as present (1) or absent (0). Genetic distances (GD) were estimated according to Nei and Li (1979) based on the probability that the amplified fragment from one genotype would be present in another one, i.e.

$$GD = 1–[2a/(2a + b + c)]$$

, where a is the number of shared fragments, b the number of fragments in line A and c the number of fragments in line B. Based on an average linkage algorithm (UPGMA, unweighted pair-group method with an arithmetic average), the computer program package Syntax 5.02 (Podani 1995) was used for cluster analysis. Gene diversity (heterozygosity) was calculated according to Weir (1990). Thus,

Table 2 Characterization of *B. napus* microsatellites

Motif	Number of repeats		Composition of repeats		Range of product size (bp)
	Average	Range	Perfect	Imperfect	
GA	21	8–55	84	24	80–220
GT	18	8–35	6	–	90–161
AT	13	12–14	2	–	158
TTC	8	8	1	–	109
GAGCG	12	12	1	–	132
AGAGGG	7	7	1	–	163

gene diversity = $1 - \sum P_{ij}^2$,

where P_{ij} is the frequency of the jth pattern for SSR marker i and is summed across n patterns. The gene diversity was considered as the polymorphism information content (PIC) according to Anderson et al. (1993).

Results

Detection of microsatellite loci from rapeseed

Approximately 500 positive clones were isolated by screening the phage library of rapeseed with $(GA)_n$ and $(GT)_n$ probes, and were sequenced. About 25% of them resulted in 121 *Brassica* SSR-specific primer pairs. The majority of the microsatellites showed GA motifs (108), followed by GT (6) and AT (2) as well as two composite dinucleotides (AT/GA, AT/GT). Moreover, one tri-, one penta- and one hexa-nucleotide microsatellite were found by chance. The number of repeats ranged from seven (the hexanucleotide) to 55 (an average of 21) and the size of amplification products from 80 to 220 base pairs (bp). On average, dinucleotide GA repeats were longer than GT repeats (Table 2). While 29% of the GA sequences were imperfect repeats, all the other microsatellite loci showed perfect repeats.

Distribution within species of the *Brassicaceae* family

Eighty one microsatellite-specific primer pairs were analyzed in order to estimate the distribution of the microsatellites among the different *Brassica* species and the other crucifers. The amplified fragments were classified as microsatellite-specific amplification products if the size of the fragments was similar to the expected size amplified from *B. napus* variety ''Vivol''. For example, the size range of specific amplifications of a microsatellite-specific primer pair, with an SSR motif length of 30 bp (15 repeats of a dinucleotide) and an expected fragment size of 130 bp, was between 100 and 210 bp. The upper size limit of a microsatellite motif was fixed at 55 repeats, which was the observed maximum number (Table 2).

Most of the microsatellites were amplified either from *B. campestris* (6) or *B. oleracea* (26), or from both species (35). Fourteen microsatellites could not be assigned to the parental genomes of *B. napus* because specific fragments were only amplified from *B. napus* variety ''Vivol''. As a result of amplification from both parental genomes (*B. campestris* = AA and *B. oleracea* = CC*)* and of intragenomic duplications, the microsatellite-specific primer pairs detected predominantly two or more loci in rapeseed (70%). From the B genome of *Brassica*, represented by *B. nigra*, 24, and from the other two amphidiploid species of the triangle of U (U 1935), *B. juncea* and *B. carinata*, 51 and 62, primer pairs amplified specific fragments, respectively. Between 24 and 28 primer pairs amplified specific fragments from the more-distantly related *Brassicaceae* species, i.e. *Diplotaxis* ssp., *B. tournefortii*, *S. alba*, *R. sativus* and *E. sativa*. Only 16 microsatellite primer pairs amplified a specific fragment from *A. thaliana* (Fig. 1). As an example, the distribution of *Brassica* microsatellite loci BMS 344/2 among the tested species is shown in Fig. 2. Except from *E. sativa*, all amplified fragments were within the size range of specific amplification. Nevertheless, direct sequencing of specific fragments obtained from the more-distantly related crucifer species revealed that, even when the expected size was amplified, a microsatellite motif was not necessarily present. In *B. napus*, however, even if three products were amplified, all fragments contained the expected microsatellite motif and the length polymorphism of the fragments was indeed due to the number of repeats of the microsatellite.

Genetic diversity

Sixty one microsatellite-specific primer pairs were tested for their efficiency in detecting polymorphisms among 32 rapeseed varieties and breeding lines. A total of 198 polymorphisms were observed among the varieties and lines. The number of alleles per microsatellite locus varied from 2 to 7. Most of the primer pairs detected two alleles (39%), 21% three, 21% four and 14% five alleles. Six alleles were detected by 3% and seven by 2% of them. There was no significant difference in polymorphism between GA and GT microsatellites, and no correlation between the number of repeats of a microsatellite and the number of alleles detected within the 32 rapeseed varieties and breeding lines ($r = 0.28$). As an example, amplification products of BMS 25 are shown

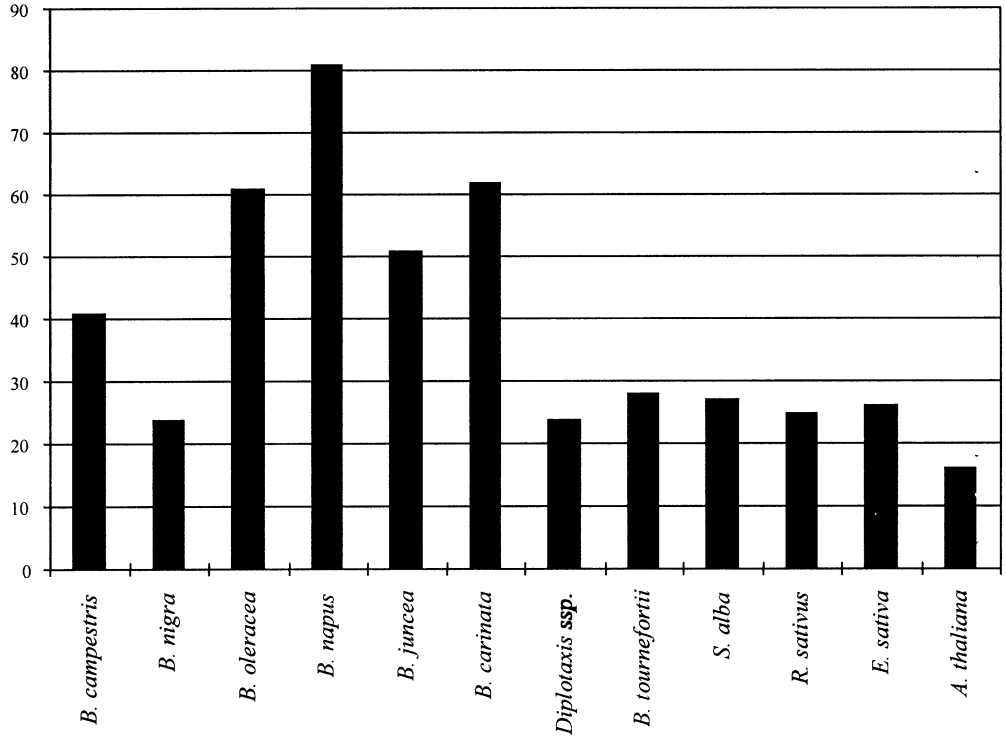

Fig. 1 Number of primer pairs from a total of 81 used which amplified specific products from different *Brassicaceae* species

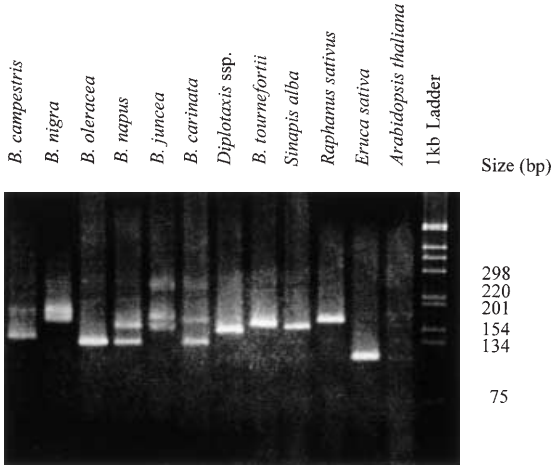

Fig. 2 Amplification products of microsatellite-specific primer pair BMS 344/2 obtained from different *Brassicaceae* species. The expected fragment size in *B. napus* was 162 bp, the motif length was 30 bp. The fragments were separated in 4% Metaphor agarose

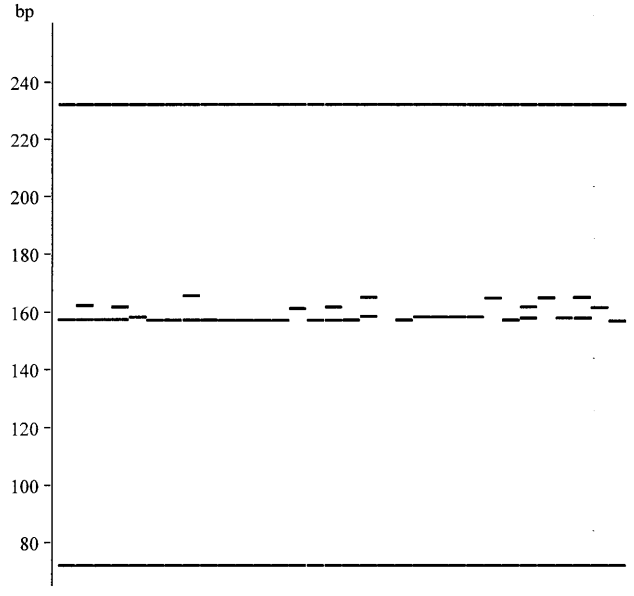

Fig. 3 A.L.F. sequencer fluorogram of amplification products of the microsatellite-specific primer pair BMS 25 from 32 *B. napus* varieties and breeding lines. Character of the microsatellite motif: dinucleotide repeats. Fragment sizes of alleles: 158, 162, 166 bp. Fragments of 231 and 71 bp are size standards

in Fig. 3. The size of the alleles ranged from 158 bp to 166 bp.

The gene diversity of microsatellite markers varied from 0.04 to 0.80 with an average of 0.50. Using 198 polymorphisms the estimated genetic distances revealed by the markers were, on average, 0.41 among the 32 rapeseed varieties. All 32 varieties were discriminated in a dendrogram (Fig. 4) constructed on the basis of shared fragments. The 32 genotypes were classified into two groups, representing winter and spring types. The winter-type varieties were further divided into seven sub-groups.

Discussion

The higher frequency of GA repeats observed in the *B. napus* genome confirmed the data of earlier studies on rapeseed (Lagercrantz et al. 1993; Szewc-Mcfadden et al. 1996; Uzunova and Ecke 1999), as well as on rice

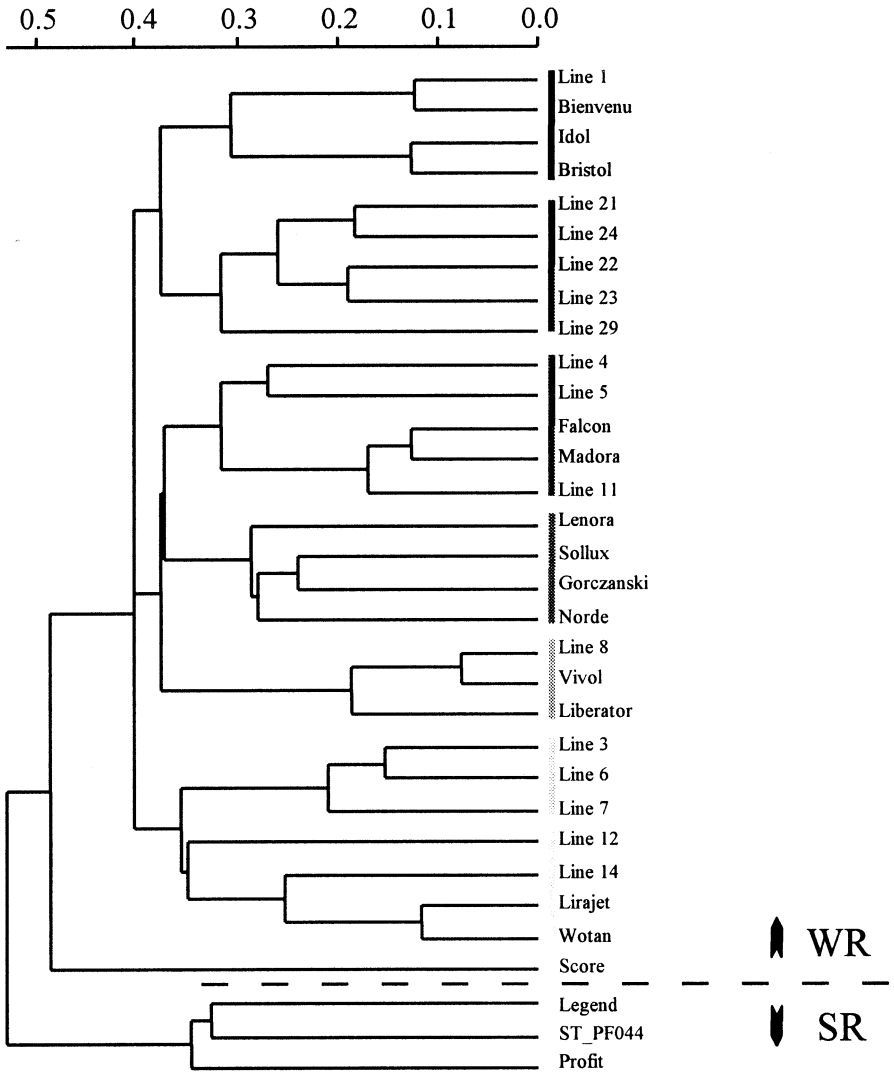

Fig. 4 Cluster analysis of 32 rapeseed varieties and breeding lines based on the data of 61 microsatellites. *WR* = winter types, *SR* = spring types

(Wu and Tanksley 1993), apple (Guilford et al. 1997) and barley (Struss and Plieske 1998). In contrast to the weak correlation between the number of alleles and the number of repeats observed in wheat (Plaschke et al. 1995), and the stronger one in barley (Saghai Maroof et al. 1994), no correlation was detected for rapeseed microsatellite markers in our investigations, confirming the results of Szewc-McFadden et al. (1996).

The distribution of the microsatellite loci among different species of the *Brassicaceae* family showed a high variation potential, especially among the species of the triangle of U. However, the transferability of microsatellite markers into distantly related *Brassicaceae* species seems to be limited. The same was observed by Röder et al. (1995**)** where, of few amplifying wheat microsatellite primer pairs in rye and barley, only one amplification product contained a microsatellite motif. Since most of the amplifying primer pairs in our study gave strong bands, and amplification products were polymorphic between the species, they are suitable as STS markers in genetic analyses. However, the variability of these loci is expected to be lower.

The rapeseed microsatellite markers distinguished clearly between the 32 assayed varieties and breeding lines. They monitored a large amount of genetic variation in closely related advanced genetically material. The high values for the gene diversity obtained demonstrates the high efficiency of these markers for genome analysis in *B. napus*. One of the most important concerns of hybrid breeders is the selection of appropriate donors and the prediction of hybrid performance. Enhancement of genetically diverse gene pools is an essential requirement in hybrid breeding. The importance of the genetic diversity of the parents for the expected heterosis of their hybrids is well known. Molecular markers have been previously used for estimation of the genetic diversity and the prediction of hybrid performance and heterosis (Melchinger 1993; Becker and Engqvist 1995; Xiao et al. 1998). Whereas Becker and Engqvist (1995) found a correlation between the genetic distance based on RAPD markers and heterosis for leaf dry matter in rapeseed, Xiao et al. (1998) could not find a significant correlation between microsatellite heterogeneity and the prediction of hetero-

sis in rice. Nevertheless, they found that the genetic diversity among the parental lines is related to heterosis. In our study, microsatellite markers were able to detect a high level of polymorphism among rapeseed inbred lines. Thus the results demonstrate the high efficiency of microsatellite markers for monitoring genetic diversity and suggest that they provide a reliable and effective means for predicting heterosis in *Brassica*.

Acknowledgments We thank C. Knaak for providing the rapeseed varieties. We also thank C.F. Quiros, K. Glozer, I. Schubert and B. Saal for reviewing the manuscript. Excellent technical assistance is acknowledged to M. Hantschmann and B. Schmidt. This project was funded by Deutsche Saatveredelung (DSV) and Semences Cargill (France). Primers will be provided on request for research purposes only, correspondence to D. Struss.

References

Akkaya SM, Bhagwat AA, Cregan PB (1992) Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132: 1131–1139

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36: 181–186

Becker H, Engqvist GM (1995) The potential of re-synthesized rapeseed for hybrid breeding. Proc 9th Int Rapeseed Congr, GCIRC, Cambridge, England, pp 113–115

Bell CJ, Ecker JH (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19: 137–144

Chèvre AM, Barret P, Eber F, Dupuy P, Brun H, Tanguy X, Renard M (1997) Selection of stable *Brassica napus-B. juncea* recombinant lines resistant to blackleg (*Leptosphaeria maculans*). 1. Identification of molecular markers, chromosomal and genomic origin of the introgression. Theor Appl Genet 95: 1104–1111

Cloutier S, Cappadocia M, Landry BS (1997) Analysis of RFLP mapping inaccuracy in *Brassica napus* L. Theor Appl Genet 95: 83–91

Chyi YS, Hoenecke ME, Sernyk JL (1992) A genetic linkage map or restriction fragment length polymorphism loci for *Brassica rapa* (syn. *campestris*). Genome 35: 746–757

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction fragments to high specific activity. Anal Biochem 132: 6–13

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in *Malus* × *domestica* (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94: 249–245

Hu J, Quiros CF (1991) Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep 10: 505–511

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple sequence repeats (SSR) loci isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor Appl Genet 91: 206–211

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21: 1111–1115

Melchinger AE (1993) Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. In International Crop Science I. Crop Science Society of America, Madison, Wisconsin, pp 621–628

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91: 1001–1007

Podani J (1995) SYN_TAX-pc-supplement 3: Macintosh version. Scientia Publishing, Budapest, Hungary

Röder MS, Plaschke J, König SU, Börner A, Sorrells E, Tanksley SD, Ganal M (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246: 327–333

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Gregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90: 43–48

Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91: 5466–5470

Song KM, Osborn TC, Williams PH (1990) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 3. Genome relationships in *Brassica* and related genome origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor Appl Genet 79: 497–506

Song KM, Suzuki JY, Slocum MK, Williams PH, Osborn TC (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. Theor Appl Genet 82: 296–304

Struss D, Plieske J (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet 97: 308–315

Struss D, Quiros CF, Plieske J, Röbbelen G (1996) Construction of *Brassica* B genome synteny groups based on chromosomes extracted from three different sources by phenotypic, isozyme and molecular markers. Theor Appl Genet 93: 1026–1032

Szewc-McFadden AK, Kresovich S, Bliek SM, Mitchell SE, McFerson JR (1996) Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. Theor Appl Genet 93: 534–538

U N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn J Genet 7: 784–794

Uzunova M, Ecke W, Weissleder K, Röbbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90: 194–204

Uzunova MI, Ecke W (1999) Abundance, polymorphism and genetic mapping of microsatellites in oilseed rape (*Brassica napus* L.). Plant Breed 118: 323–326

Weir BS (1990) Genetic data analysis methods for discrete genetic data. Sinauer Assoc Inc Sunderland, Massachusetts, USA

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241: 225–235

Xiao C, Wu L, Wu JL (1998) SSR heterogenic patterns of parents for marking and predicting heterosis in rice breeding. Molecular Breeding 4: 263—268