Theor Appl Genet (2004) 108:1103–1112
DOI 10.1007/s00122-003-1522-7

ORIGINAL PAPER

A. J. Lowe · C. Moule · M. Trick · K. J. Edwards

Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species

Received: 10 September 2003 / Accepted: 29 October 2003 / Published online: 5 December 2003

Abstract A set of 398 simple sequence repeat markers (SSRs) have been developed and characterised for use with genetic studies of *Brassica* species. Small-insert (250–900 bp) genomic libraries from *Brassica rapa*, *B. nigra*, *B. oleracea* and *B. napus*, highly enriched for dinucleotide and trinucleotide SSR motifs, were constructed. Screening the clones with a mixture of oligonucleotide repeat probes revealed positive hybridisation to between 75% and 90% of the clones. Of these, 1,230 were sequenced. Primer pairs were designed for 398 SSR clones, and of these, 270 (67.8%) amplified a PCR product of the expected size in their focal and/or closely related species. A further screen of 138 primers pairs that produced a PCR product in *B. napus* germplasm found that 86 (62.3%) revealed length polymorphisms within at least one line of a test array representing the four *Brassica* species. The results of this screen were used to identify 56 SSRs and were combined with 41 SSRs that had previously shown polymorphism between the parents of a *B. napus* mapping population. These 97 SSR markers were mapped relative to a framework of RFLP markers and detected 136 loci over all 19 linkage groups of the oilseed rape genome.

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-003-1522-7

Communicated by O. Savolainen

A. J. Lowe · C. Moule · K. J. Edwards (✉)
IACR Long Ashton Research Station,
Long Ashton, Bristol, BS41 9AF, UK
e-mail: K.J.Edwards@bristol.ac.uk
Tel.: +44-0-117-3317079

M. Trick
John Innes Centre, Norwich Research Park,
Norwich, NR4 7UH, UK

Present address:
K. J. Edwards, Functional Genomics Unit,
School of Biological Sciences, University of Bristol,
Bristol, BS8 1UG, UK

Present address:
C. Moule, Cambridge Institute for Medical Research,
Hills Rd, Cambridge, CB2 2XY, UK

Present address:
A. J. Lowe, School of Life Sciences,
University of Queensland, QLD 4072 Brisbane, Australia

Introduction

The Brassicas are a diverse range of species of considerable global economic importance, having been cultivated for use as a variety of oilseed, vegetable and fodder crops. The major crop types are derived from three species, *Brassica rapa* (turnips, swede and Chinese cabbage), *B. nigra* (mustards) and *B. oleracea* (cabbages, brussel sprouts, kale) and also the amphidiploid combinations thereof whose provenance was originally deduced cytogenetically (U 1935). For instance, the widely grown crop *B. napus* (oilseed rape or canola) is the AACC amphidiploid comprising the *B. rapa* (A) and *B. oleracea* (C) genomes. The Brassicas are characterised by an extraordinary level of intraspecific morphological phenotypic variation, fixed by artificial selection to form various crop types. This feature seems to be accompanied by a high degree of genetic variation pervasive throughout the entire genome, which in turn, has enabled a number of dense RFLP-based genetic linkage maps to be assembled for most of the *Brassica* species (Howell et al. 2002; Kowalski et al. 1994; Lionneton et al. 2002; Lombard et al. 2001; Lydiate et al. 1993; Parkin et al. 1995; Saal et al. 2001). Where common sets of RFLP probes and parental genotypes have been used, definitive allele matching is possible, which allows the formulation of integrated or comparative maps (Parkin et al. 1995; Sebastian et al. 2000).

Although the use of RFLP markers has been instrumental in developing an understanding of genome structure and evolution in the Brassicas (Cavell et al.

1998; Lagercrantz and Lydiate 1996; Soltis and Soltis 1995; Song et al. 1995) and, comparative evolution in the related model crucifer, *Arabidopsis thaliana* (Kowalski et al. 1994), these types of marker have some shortcomings when applied to marker-assisted selection. Here the aim is to accelerate plant breeding programmes based on the introgression of desirable traits into elite backgrounds through the counter-selection of unwanted segments of linked donor genome. High throughput potential, ease of data interpretation and application of simple and transferable technology are all crucial to success and RFLP markers can be deficient by each criterion to varying degrees. Over the past few years a number of new PCR-based marker technologies such as RAPDs, AFLPs and microsatellites have been developed and applied to crop plants. Of these, microsatellites or simple sequence repeats (SSRs) have a great deal of potential. SSR markers are based on unique DNA sequences flanking short repetitive tracts of simple sequence motifs such as di-or trinucleotides. The extraordinary level of informative polymorphism at a given SSR locus stems from the apparent propensity for expansion or contraction of repeat number through slippage replication or unequal crossing over events during meiosis (Goldstein and Schlötterer 1999). Advances in techniques for enrichment and selection of SSR-containing clone libraries have resulted in the development of large numbers of SSR markers in a number of plant species such as maize (Sharopova et al. 2002), wheat (Khlestkina et al. 2002), sunflower (Tang et al. 2002), and sugarbeet (Rae et al. 2002). However, this is not to say that such large-scale development of SSR markers is not without problems (Zane et al. 2002). When large numbers of SSR markers are required, difficulties and wastage due to clone duplication, chimera formation, lack of flanking sequence and poor amplification of PCR primers are all encountered, and can lead to massive attrition rates relative to the initial numbers of clones sequenced (up to 90%, Squirrell et al. 2003). However, with the accurate reporting of attrition rates at each step, the SSR development process can be further refined and improved to give greater efficiency of marker production.

Hitherto, the number of SSRs developed from *Brassica* has been small (Lowe et al. 2002; Plieske and Struss 2001; Saal et al. 2001; Suwabe et al. 2002; SzewcMcFadden et al. 1996; Westman and Kresovich 1999) and includes some originally transferred from *A. thaliana* (Westman and Kresovich 1998). This report describes the development and characteristics of 398 SSR markers derived from *B. rapa*, *B. nigra*, *B. oleracea* and *B. napus*, and includes details of the attrition rates encountered at each stage of development, together with surveys of genome specificity and polymorphism for successfully designed primer pairs. We also present a genetic linkage map based on their segregation in a *B. napus* population that is integrated with 90 RFLP loci previously described (Parkin et al. 1995). All the primer sequences (supplied with the Electronic Supplementary Material), together with polymorphism information and in silico analysis, are publicly available through the BrassicaDB database at http://ukcrop.net. It is anticipated that these should supply a useful resource to the Brassica research community.

Materials and methods

Brassica germplasm and DNA extraction

Genomic DNA libraries were constructed from four species of *Brassica* encompassing the three diploid genomes described by U (1935), using the following doubled-haploid or inbred lines; *Brassica napus* N-o-9 (Sharpe et al. 1995), *B. rapa* ssp. *chinensis* R-c-50 (Axelsson et al. 2000; D. Lydiate, AAFC Saskatoon Research Centre, Canada), *B. oleracea* spp. *alboglabra* A12DHd (Bohuon et al. 1996) and *B. nigra* N1 (Lagercrantz and Lydiate 1995). These lines were also used in an initial screen of candidate SSR primers to establish if a product of expected size could be amplified.

Germplasm accessions used for polymorphism screening comprised the following: *B. napus* N-o-9, *B. napus* SYN1 (Parkin et al. 1995), *B. napus* cv. Westar, *B. napus* cv. Apache, *B. napus* cv. Tapidor, *B. nigra* N1 and *B. nigra* A1 (Lagercrantz and Lydiate 1995), *B. rapa* ssp. *oleifera* RM29 (R. Mithen, IFR Norwich UK), *B. rapa* spp. *chinensis* R-c-50, *B. rapa* spp. *trilocularis* r-o-18 (D. Lydiate, AAFC Saskatoon Research Centre, Canada) and the *B. oleracea* F1 hybrids GK97370, GK97377, DJ7560, DJ7773 and CR96004 (G. King, Horticulture Research International, Wellesbourne UK).

DNA was extracted from plant leaf material using either the Nucleon Phytopure kit (Amersham) or by the CTAB technique (Doyle and Doyle 1987).

Genomic libraries highly enriched for SSRs

Enriched genomic DNA libraries were prepared according to Edwards et al. (1996). Briefly, 1 μg of genomic DNA was digested with *Ssp*I or *Rsa*I. Oligonucleotides of CA, CT, CAA, CATA, and GATA microsatellite repeat motifs were adhered to separate filter strips and were used to select for genomic fragments containing these SSRs. To do this the restriction-digested DNA was independently incubated, and allowed to anneal, with each of the oligonucleotide filters, and this was done separately by filter type and species. Washing of DNA fragments from filters (presumably enriched for SSR repeats) combined the different oligonucleotide adhered filters for each species. Microsatellite enriched fragments were then amplified by PCR and ligated into the modified pUC19 vector pJV1 (Edwards et al. 1996). Plasmids were transformed into DH10B (Life Technologies) and plated on to L-agar plates containing 100 μg ml^{-1} ampicillin. After overnight incubation at 37°C, single colonies were transferred to microtitre plates. Following an overnight incubation at 37°C, glycerol was added to a final concentration of 25% for long-term storage at –70°C.

Hybridisation screening of enriched libraries

The genomic libraries were screened by transferring clones onto nylon membranes and probing with radiolabelled oligonucleotides (Sambrook et al. 1989). A Beckmann Biomek 2000 was used to replicate 1,536 colonies onto an 8×12 cm membrane (Hybond N, Amersham), which was placed on L-agar containing 100 μg ml^{-1} ampicillin, and incubated for 16 h at 37°C. The membranes were transferred onto Whatman 3MM paper, soaked in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 5 min, and then twice onto paper soaked in neutralisation buffer (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.5) for 3 min. The membranes were then washed vigorously in 2xSSC (0.3 M NaCl, 0.03 M trisodium citrate) for 5 min. After air-drying the filters, DNA was fixed by crosslinking to the membrane using the optimal crosslink function of a Spectrolinker XL-1000 UV Crosslinker (Spectronics Corporation).

Filters were screened using each of the following oligonucleotides probes: $[CA]_{15}$, $[CT]_{15}$, $[CATA]_{10}$, $[CAA]_{10}$ and $[GATA]_{10}$, each end-labelled using [γ-^{32}P]-ATP (DuPont-NEN) and T4 polynucleotide kinase (Pharmacia Biotech). Membranes were prewashed with hybridisation buffer (6× SSC, 0.25% dried milk powder and 0.01% SDS) and incubated in an OV5 (Biometra) rotary oven at 50°C for 1 h. The wash buffer was replaced by 25 ml hybridisation buffer containing 100 ng of the radiolabelled oligonucleotide and incubated at 50°C for 16 h. Membranes were washed at 50°C, 4 times for 5 min, with 200 ml of 2× SSC containing 0.1% SDS, then air dried and exposed to X-ray film overnight.

Plasmid purification

Plasmid DNA was extracted using Wizard Minipreps D Purification System (Promega) from 5 ml overnight cultures in L-broth (1% w/v NaCl, 5% w/v yeast extract, 1% w/v bacteriological peptone) containing 50 μg ml^{-1} ampicillin.

Sequencing

Plasmid DNA was added to the Dye Terminator Cycle Sequencing Ready Reaction mix (Perkin-Elmer) together with M13 universal primers and cycle sequenced using the following conditions: 25 cycles of 96°C for 10 s, 50°C for 5 s, 60°C for 2 min, and then held at 25°C. The amplified product was precipitated with ethanol and sequenced using an ABI PRISM 377 DNA Sequencer (Perkin Elmer) according to the manufacturer's instructions.

Primer design

Before primers were designed, each sequence was examined using the program Autoassemble v1.4 (Applied Biosystems, Perkin Elmer), for sequence similarity to other clones. For those clones containing three or more repeats of an SSR element, pairs of primers 18–21 nucleotides in length were designed from the regions flanking the repeat using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi/primer3_www.cgi). Oligonucleotide primers were synthesised by Genosys.

Microsatellite primer characterisation

All primer pairs were initially tested for amplification using a panel of *Brassica* DNAs of four species; *B. nigra*, *B. rapa*, *B. oleracea* and *B. napus*. Amplification of products was achieved by using 25 μl reactions containing: 50 ng DNA template, 10× PCR buffer (Perkin Elmer), 1 U Amplitaq GOLD *Taq* DNA polymerase, 200 mM of each dNTP and 250 ng of both primers. PCR was performed using a Perkin Elmer 9600, using conditions similar to those of Lowe et al. (2002). Products were separated by electrophoresis in 2% agarose gel and visualised under a UV lamp after staining with ethidium bromide. SSR primer pairs were scored for their ability to amplify a product in at least one of these test species.

To screen for polymorphism, a batch of 138 SSR primer pairs was selected from those that produced a PCR product in the *B. napus* accessions and comprised markers originating from all 4 enrichment libraries (43 SSRs from *B. napus*, 31 from *B. nigra*, 27 from *B. oleracea* and 37 from *B. rapa*). The forward primer of each pair was end-labelled with 20 nCi [γ-^{33}P]-ATP (DuPont-NEN) per 30 pmol of primer. The SSR loci were then amplified across a test range of the four *Brassica* species and included parents of the N-o-61-9 *B. napus* doubled haploid (DH) mapping population (Parkin et al. 1995). For PCR, 25 μl reactions were used, each containing: 50 ng DNA template, 1× PCR buffer (GibcoBRL), 1 U *Taq* DNA polymerase (Perkin Elmer), and 200 mM of each dNTP, 45 pmol of both primers. Thermocycling was performed using a Perkin Elmer 9600, using 35 cycles of 94°C for 30 s, 56°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 10 min before holding at 25°C. Prior to electrophoresis on 6% denaturing polyacrylamide gels (19:1 acrylamide:bisacrylamide - EASIgel, Scotlab), an equal volume of formamide (0.4% w/v bromophenol blue and 0.25% w/v xylene cyanol FF), was added to each reaction. The gels contained 7 M urea and 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.4). Electrophoresis was at 55 W constant power for between 3–4 h (Sambrook et al. 1989). The gels were dried onto Whatman 3MM paper and exposed to X-ray films. Gels were scored for possible polymorphism (i.e. more than one amplification product detected) across the different species used in the screen.

Linkage analysis and map construction

All primers that exhibited polymorphism between the *B. napus* lines N-o-9 and SYN1 from the above described screen and a previous screen (C. Moule, unpublished data) were selected for further genetic analysis. SSR alleles were scored and genetically mapped in a maximally informative subset of 35 lines selected from the N-fo-61-9 *B. napus* DH population (from a cross between a natural winter type oilseed rape, N-o-9 and a resynthesised *B. napus* line, SYN1, Parkin et al. 1995), with integration into the pre-existing RFLP marker segregation data (proprietary to John Innes Centre and industrial sponsors SES and CPB Twyford). The Mapmaker v3.0 program (Lincoln et al. 1987) was used to recalculate map intervals using a LOD value of 4.0. A framework of 90 RFLP loci was selected on which to present the SSR loci, 5 per linkage group except N16, which cannot be mapped in this population due to a non-disjunction event in meiosis of the resynthesised parent of the cross. However, the finished linkage maps show inferred genetic intervals calculated from the full dataset. The segregation of all SSR markers was checked by a g^2 test. Markers, showing significant distortion ($P<0.05$) from the 1:1 segregation expected in a DH population, and including those possessing null alleles, were excluded, and loci from a remaining set of 97 primer pairs were mapped. A set of 7 loci were inferred to be on N16, all individuals showing the synthetic parental genotype.

Results

Clone enrichment, stage attrition and efficiency of microsatellite primer design

Initial screening of the library confirmed that the enrichment procedure had been very successful and approximately 75–90% of clones contained a microsatellite locus. Such was the level of microsatellite enrichment that it was decided to proceed straight to sequencing of a sample of the clones without a time-consuming selection phase.

A total of 1,230 clones were sequenced from across the microsatellite libraries constructed from *B. napus, B. rapa*, *B. nigra* and *B. oleracea*. Out of these, 189 sequencing reactions failed (15.4%), and were not repeated as enough sequenced clones had been obtained to proceed with the primer development procedure, thus 1,041 clone sequences were available. Duplicates (126) and partially sequenced SSR motifs (157) were also removed to leave a total of 758 sequences. From these data, the distribution of microsatellite classes found is presented in Table 1. Dinucleotide repeat motifs were most common (67.9%), with a particular abundance of the CT motif type (44.1%). However, there were also differences between the different species libraries, for example the *B. rapa* (A genome) library had an increased

Table 1 Distribution of SSR repeat motif type for 758 successfully sequenced clones (removing duplications and most sequences for which incomplete repeat motif information was available) from enriched libraries for four *Brassica* species

Species/motif	GA/CT	GT/CA	GGC/CCG	Other	Compound/ disrupted	None	Total
B. nigra	111	32	8	2	7	37	197
B. rapa	39	109	12	3	12	14	189
B. oleracea	100	26	25	3	2	56	212
B. napus	84	13	15	2	9	37	160
Total%	44.1	23.7	7.9	1.3	4.0	19.0	758

Table 2 Details of attrition for successfully sequenced clones from each of the four enriched libraries, and globally. The number of clone sequences lost and the percentage of the total number of sequences are shown for each stage down to successful primer design

	B. nigra	*B. rapa*	*B. oleracea*	*B. napus*	Total
Clones successfully sequenced	252	251	329	209	1041
No SSR	37 (14.7)	15 (6.0)	57 (17.3)	39 (18.7)	14.2%
Couldn't sequence past SSR	34 (13.5)	59 (23.5)	79 (24.0)	16 (7.7)	17.2%
SSR too close to end of flanking sequence	29 (11.5)	48 (19.1)	39 (11.9)	18 (8.6)	12.8%
Duplicate	34 (13.5)	18 (7.2)	41 (12.5)	33 (15.8)	12.2%
Problems with primer design	4 (1.6)	16 (6.4)	16 (4.9)	3 (1.4)	3.6%
Chimera, but primers designed	1 (0.4)	1 (0.4)	8 (2.4)	1 (0.5)	0.9%
No PCR product	30 (11.9)	25 (9.9)	29 (8.8)	38 (18.2)	11.7%
PCR primers amplify correct sized product	83 (32.9)	64 (25.5)	66 (20.1)	61 (29.2)	26.3%

representation of CA repeats (n=109) over CT type (n=39). This bias in dinucleotide repeat type was found to be statistically significant across the range of the focal species libraries (chi-squared test, df=3, P<0.001).

For the 1,041 sequenced clones, character data were screened using search and matching algorithms for the suitability of primer design sites. Clones were excluded if they contained no SSR, were duplicated (although chimeras were included), there was no or not enough unique sequence flanking the region from which to design primers (due to sequencing failures past the SSR or the location of clone insertion sites), or if no suitable primers could be designed/located. All primer sequences are presented in the ESM. These primer pairs were screened against a test array of the various library genomes to see if a product of expected size could be amplified. The overall primer design efficiency from the four clone libraries and attrition rates at the various stages of processing up to successful primer design are presented in Table 2.

The highest rates of attrition were due to problems associated with lack of available flanking region with which to design primers (sequencing difficulties; 17.2%, location of clone too close to insert site; 12.8%), and lack of microsatellite insert (14.2%). In addition, there appears to be a relationship between clone insert size and some of the attrition factors (Fig. 1). For example, if only those clones with an insert size of greater than 200 bp were taken, then the number of clones lost due to the lack of a microsatellite insert and the lack of sufficient flanking region would be approximately halved (i.e. reduced from 14.2% to 7.6% and 12.8% to 6.6% respectively). As the insert size increases, there is also a notable increase in the percentage of usable primer pairs produced per clone successfully sequenced (from 14% for insert sizes between 100 and 150 bp up to 57% for those inserts between 600 and 700 bp). However, the absolute number of successful primer pairs obtained falls off once insert size exceeds 250 bp (Fig. 1). The full list of primer pairs

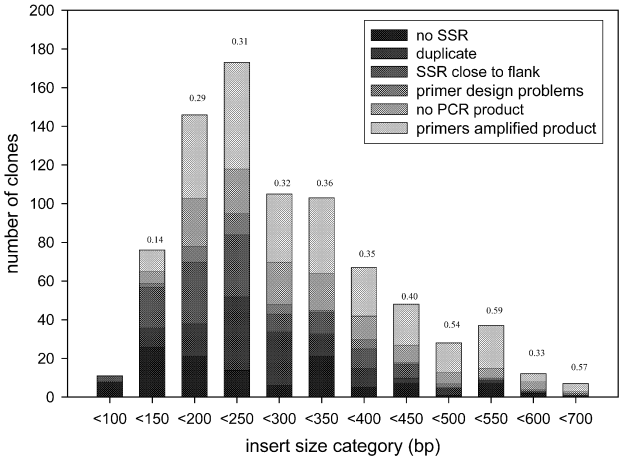


Fig. 1 Frequency of causes of attrition and eventual successful primer design for different size classes of library clone inserts

Table 3 Percentage of microsatellites (and absolute numbers in parentheses) that amplify a product within each of the test species. The global total is broken down by the focal species of the enriched library

Specificity test screen/library focal species	*B. nigra*	*B. rapa*	*B. oleracea*	*B. napus*
B. nigra	85.5% (71)	65.1% (54)	54.2% (45)	61.4% (51)
B. rapa	51.6% (33)	81.2% (52)	64.1% (41)	71.9% (46)
B. oleracea	60.6% (40)	65.1% (43)	63.6% (42)	72.7% (48)
B. napus	45.9% (28)	65.6% (40)	60.6% (37)	83.6% (51)
Total	62.8%	68.9%	60.2%	71.5%

Table 4 Results of a polymorphism screen for 138 microsatellites derived from enriched libraries of 4 *Brassica* species. In addition to giving global totals for the number and percentage of SSRs that were polymorphic for at least one of the screening lines, results are broken down by species used within the screening population for each library. For each library these more detailed results are presented as the number of SSRs which were polymorphic within a particular screening line and the proportion which were polymorphic within that screening line compared to the total number of SSRs which were polymorphic from that library, where *Ni* is *B. nigra*, *Ol* is *B. oleracea*, *Ra* is *B. rapa* and *Na* is *B. napus*. Note that the global totals are not a simple additive total of the individual species totals, as some primers were polymorphic in more than one species

Library	*Ni*	*Ol*	*Ra*	*Na*	Total
Total number of SSRs screened	31	27	37	43	138
No. polymorphic SSRs	13	21	19	33	86
% of polymorphic loci	41.9	77.8	51.3	76.7	62.3
No. polymorphic in *Ni*	2	3	0	2	7
Proportion of polymorphic loci	0.15	0.14	0	0.06	0.08
No. polymorphic in *Ol*	4	15	13	19	51
Proportion of polymorphic loci	0.31	0.71	0.68	0.58	0.59
No. polymorphic in *Ra*	3	5	10	16	34
Proportion of polymorphic loci	0.23	0.24	0.53	0.48	0.39
No. polymorphic in *Na*	12	21	19	31	83
Proportion of polymorphic loci	0.92	1	1	0.94	0.96

and associated descriptions are provided in the Electronic Supplementary Material.

Genome specificity, species transferability and diversity of microsatellite loci

Following screening of primer pairs against the panel of *Brassica* species, genome specificity was also assessed for those that produced a product of expected size (see Electronic Supplementary Material). A summary of the genome specificity of primer pairs is presented in Table 3 and is subdivided by library focal species. Primers tended to amplify best in the screening accession of the library focal species. However, there were also primers that failed to work in accessions screened for their focal species but that worked in one of the other species in the test array. This was most evident for *B. oleracea*, for which only 63.6% (n=42) of the primers designed from the *B. oleracea* library which amplified a product of expected size worked in the *B. oleracea* screening accession, although 72.7% of these primers (n=48) amplified a product in *B. napus* screening accession. Overall, the amount of transferability is higher between A and C genome species (*B. oleracea*, *B. rapa* and *B. napus*) than between this group and the B genome species *B. nigra*.

Polymorphism, measured as the presence of more than one amplification product (probably equating to alleles) in at least one accession of a given species, was assessed for a test sample of 138 SSR primers. Overall the level of polymorphism was high, i.e. 62.3% of all primers tested. The library from which primers were designed appeared to influence the level of polymorphism, where SSRs from *B. nigra* and *B. rapa* libraries exhibited lower polymorphism (41.9% and 51.3% respectively) and SSRs from *B. oleracea* and *B. napus* libraries exhibited slightly higher polymorphism (77.8% and 76.7% respectively). In addition, a higher proportion of primers tended to be polymorphic when screened against their focal species than when tested in transferred species (Table 4). However, this was not the case for *B. napus*, where the highest levels of polymorphism (100%) were detected by both *B. rapa* and *B. oleracea* SSRs. Overall the level of polymorphism was higher for SSRs transferred amongst A and C genome species (*B. rapa*, *B. oleracea* and *B. napus*) than between this group and the B genome species, *B. nigra* (Table 4). Such observations may be complicated by the diversity of the test species; the *B. napus* lines used in the polymorphism screen were probably the most diverse and the *B. nigra* lines the least diverse.

Another observation from this screen was that GA/CT motif types tended to have higher numbers of repeat units (31.5) than GT/CA (25.1) and a higher proportion were polymorphic [53 out of 78 (67.9%), compared with 17 out of 37 (45.9%), respectively].

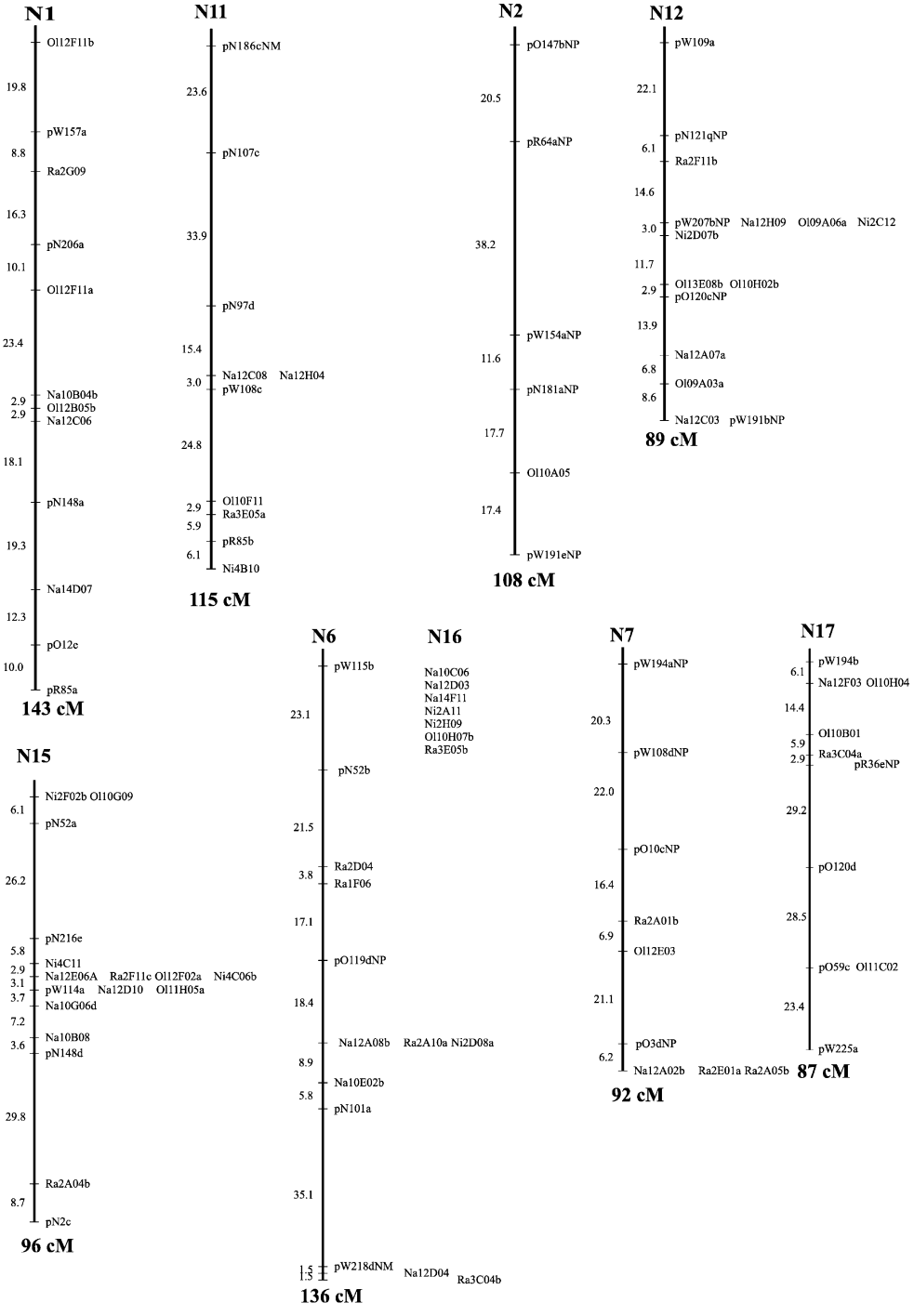
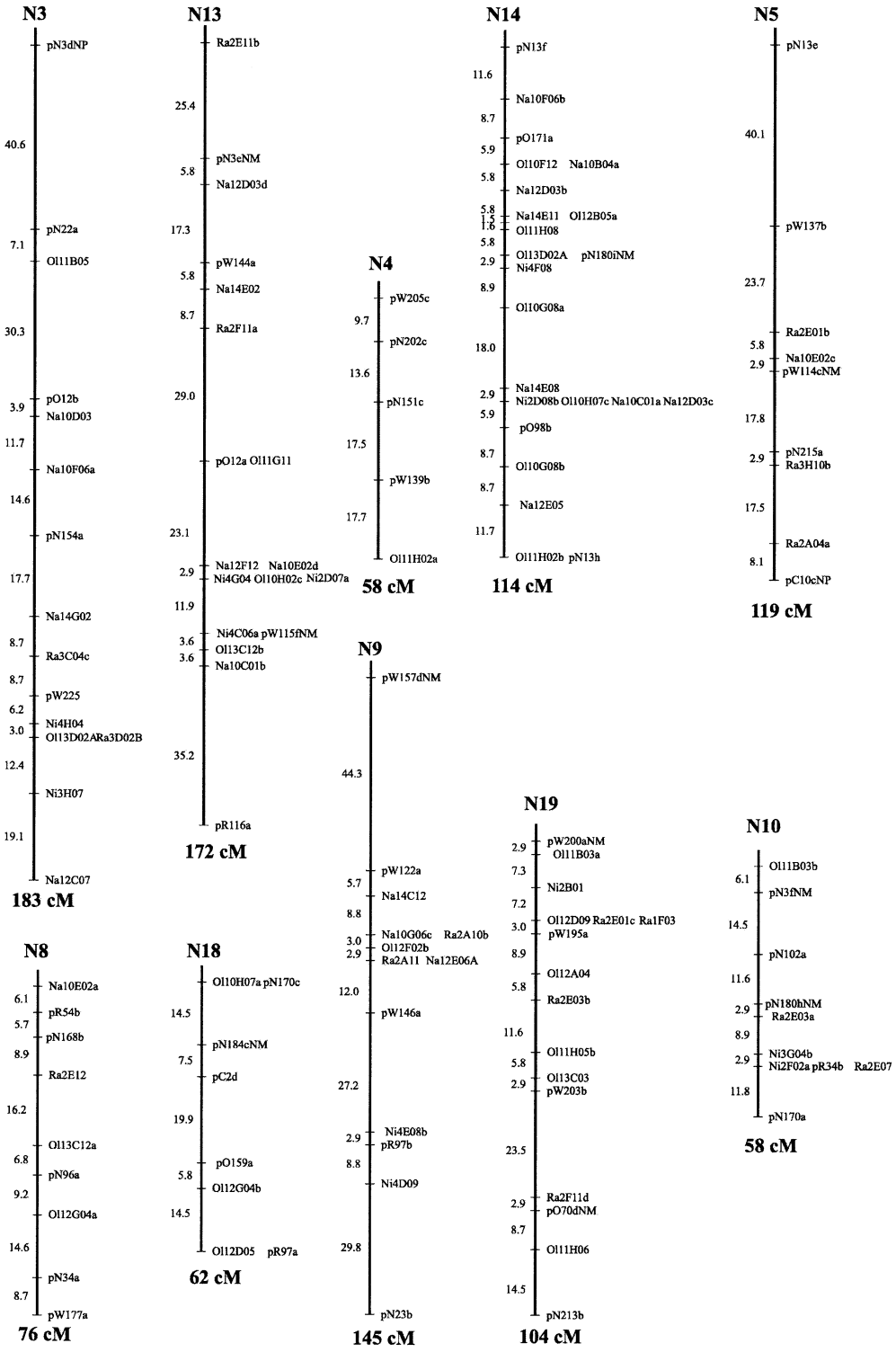


Fig. 2 Distribution of 97 SSRs detected at 136 loci within a genetic map of the N-o-61-9 *Brassica napus* doubled haploid population, with integration into the framework of 90 pre-existing RFLP marker loci (proprietary to John Innes Centre and industrial sponsors SES and CPB Twyford). Five RFLPs per linkage group are present except N16, which cannot be mapped in this population due to a non-disjunction event in meiosis of the resynthesised parent of the cross. Multiple SSR loci are labelled with subscripts

Genetic mapping of microsatellite loci and linkage group analysis

A total of 97 SSRs were successfully mapped to linkage groups established in the *B. napus* mapping population and their positions are presented relative to five RFLP markers spaced along each linkage group (see Fig. 2 and Electronic Supplementary Material. Most SSR primer pairs detected a single locus (*n*=68, 69.4%, Table 5) but a sizeable proportion detected two loci (*n*=25, 25.5%), and a smaller proportion detected three or four loci (*n*=7, 7.1%). Of the 32 SSRs which detected more than one locus, 13 detected multiple loci occurring within the same genome (11 within the C genome and 2 from the A genome), and for two of these SSRs both loci occurred on the same linkage group (Ol12F11 and Ol10G08). However, for the majority (*n*=19), loci were found in both the A and C genome linkage groups, and for six of these SSRs, loci were located on corresponding homologous portions of linkage groups (for example, SSR Ra2A04

Fig. 2 (continued)



detects two loci, one on linkage group N5 the other on N15).

The distribution of the location of SSR loci according to origin and type showed some skewness (Table 6). SSRs developed from *B. rapa* tended to be more commonly distributed in the A genome component (i.e. linkage groups N1 to N10; 18 out of 30, 60%) and *B. oleracea* SSRs tended to be more commonly found in the C genome component of the map (i.e. linkage groups 11 to 19; 32 out of 44, 72.1%), although this skew was not significant. However, the skewed distribution of repeat motif types was significant. GT/CA repeat types were located significantly more often in the A genome component (13 out of 21, 61.9%) and GA/CT repeats

Table 5 Total number of SSRs mapped onto the *B. napus* genetic map showing their library origin and the number of loci detected

Library focus	No. SSRs mapped	No. loci				Total
		1	2	3	4	
B. napus	31	24	5	0	2	42
B. nigra	16	12	4	0	0	20
B. rapa	19	12	4	2	1	30
B. oleracea	31	19	11	1	0	44
Total	97	67	24	3	3	136

Table 6 Distribution of SSRs derived from the different enrichment libraries and by repeat type in the A (linkage group N1-N10) and C genomic component (Linkage group N11-N19) of the *B. napus* genetic map

Genome detection	A genome	C genome
Library focal species		
B. napus	16	26
B. nigra	7	13
B. rapa	18	12
B. oleracea	12	32
Repeat motif type		
A/T	0	2
GT/CA	13	8
GA/CT	31	61
Tris	8	12
Compound	0	1

more often in the C genome component (61 out of 92, 66.3%) of the *B. napus* map than vice versa (chi-squared test, df=1, P=0.017; Table 6). In addition, the C genome contained a higher proportion of the tri-nucleotide repeats (60%) than the A genome, and was the location of both A/T repeats and the single compound SSR surveyed. Finally there was no apparent relation between repeat type and position along a linkage group (i.e. central or terminal regions).

Discussion

Efficiency of microsatellite development

The efficiency of developing a large set of microsatellites largely depends on the library enrichment techniques (Zane et al. 2002). The method used in this study (Edwards et al. 1996) provided a good enrichment of microsatellites within all the libraries used. However, the nature of the enrichment could lead to some bias in the type of microsatellite selected for subsequent mapping. For example there were significantly more GT/CA repeat motif type SSRs identified in the *B. rapa* clone library than GA/CT types, whereas significantly more of the latter type were isolated from the libraries of *B. nigra*, *B. oleracea* and *B. napus*. This bias is probably due differences in the abundance of motif types in these species. Different motif type frequency biases have been reported for a range of plant species genomes and major differences occur between plant and animal taxa (Zane et al. 2002). To ensure that the innate frequency of different microsatellite types within the genome of a focal species does not impact on the total efficiency of SSR isolation, the use of a range of mono-nucleotide, di-nucleotide and tri-nucleotide selective SSR motifs appears to be necessary. This difference may also have been due to binding differences during library preparation. Whilst experimental protocols were standardised and carried out in parallel for the different species to reduce such bias, this explanation cannot be ruled out and further genome probing experiments would establish more rigorously the relative proportions of different repeat motif types.

Examination of the attrition rates we experienced at the various stages of SSR development is also informative. Due to the relatively high enrichment of libraries (an average of 85.8% across the libraries), attrition due to the lack of an SSR repeat was relatively low (14.2%). Implementing a screening stage to determine which clones contain SSR inserts could reduce redundancy, but such a process is time consuming. The library enrichment process could also be improved, although few enrichment techniques are able to realise the proportions achieved in this study (Zane et al. 2002), and it is likely that this process has already reached an optimum.

The highest rate of attrition was due to an inability to sequence beyond the SSR repeat within a clone insert (17.2%). Improvements in PCR, sequencing and dye chemistry should help with this problem and it may be alleviated by future sequencing developments. An alternative is to sequence the reverse strand of a refractory insert, but this process requires initial screening and/or additional sequencing. A bi-directional sequencing tactic may become standard as sequencing costs reduce in the future.

Another significant cause of attrition was location of the SSR repeat tract too close to the clone insertion site, thus offering little or no flanking region from which to design primers (12.8%). A plot of insert size against attrition factors indicates (Fig. 1) that inclusion of a size selection step during library preparation may help reduce attrition due to lack of flanking region. For example, 28.4% of clones have an insert size less that 200 bp but 44.8% of all clones lost due to lack of flanking region fall in this size class. Thus a clone size selection step that excludes inserts of below 200 bp in size would significantly improve output efficiency. Furthermore, the set of clones carrying inserts of up to 200 bp also contained 46.2% of all clones lost due to the lack of an SSR insert, and so a size exclusion step may further improve efficiency due to this factor.

The third main reason for insert exclusion was duplication (12.2% of inserts), and is relatively high, although it is probably correlated with the large total number of inserts screened. Smaller scale SSR production projects are unlikely to encounter such high attrition due to this factor. The number of chimeras found was low, and most went on to produce working primer pairs, and so such inserts should not be discarded out of hand.

Following primer design, a further 11.7% of clones were excluded as no product of expected size was observed in any of the test accessions. Overall, SSR primers that produced the expected amplification product could be designed for only 26.3% of sequenced clones. This rate of attrition is within the range of rates encountered during a review of the development of SSRs for 71 plants species. Squirrell and coworkers (2003) found that, on average, 34.5% of sequenced clones successfully produced primers, but that, on average, half of these were excluded later due to lack of amplification product or production of multi locus or monomorphic profiles.

Microsatellite polymorphism and genome specificity

Screening the microsatellites against test accessions of *Brassica* indicated that for three of the four *Brassica* species examined, SSRs more often amplified a product in the focal species (except SSRs developed from *B. oleracea* DNA which amplified more successfully in *B. napus*, Table 3). In addition transferability is higher between A and C genome species (*B. oleracea*, *B. rapa* and *B. napus*) than between this group and the B genome species *B. nigra*. Such observations confirm the known evolutionary relationship between these taxa (Warwick et al. 1992).

For three of the four *Brassica* species examined here, polymorphism (as measured by the presence of more than one amplification product) was higher within material of the focal species than transferred species (except SSRs developed from *B. napus*, which were more polymorphic in *B. oleracea* and *B. rapa* lines than the *B. napus* material examined, Table 4). However, the results of the polymorphism screen are likely to be confounded by the level of diversity within screened lines, and a further more rigorous screening of candidate loci is recommended. Overall the polymorphism screen, conducted on 138 SSRs, found that 62.3% (n=86) were polymorphic in at least one accession. If this polymorphism rate is transferred across to all 274 SSRs that successfully amplified a product, then 171 of the loci described in this report (62.3%) would be expected to be polymorphic. In total, therefore approximately 16.4% of clones produced primers that amplified a single-locus polymorphic product and compares very well with the average total redundancy rate for much smaller numbers of SSRs developed [to the same stage for 71 plant species (17.2%), Squirrell et al. 2003].

In addition, it was possible to locate 97 of the SSRs on the *B. napus* map. Some loci may not be incorporated onto the map due to low LOD scores or because they are only encoded in the B genome (i.e. from *B. nigra*). There was a significant difference in the distribution of motif repeat types across the *B. napus* genome. A genome linkage groups contained significantly more GT/CA repeats than C genome linkage groups, and C genome linkage groups contained a higher frequency of GA/CT repeats than A genome linkage groups (Table 6). This pattern is similar to the proportion of repeat motif types isolated from the *B. rapa* and *B. oleracea* (more GT/CA and GA/CT repeat types respectively, Table 1), and is it is likely to be directly linked.

Most of the mapped SSRs (69.4%) detected a single locus, although a sizeable proportion (25.5%) detected two loci. The encoding of SSRs at multiple loci within the amphidiploid *B. napus* genome is expected due to the high rate of transferability of microsatellites across the diploid species (Table 3; e.g. Lowe et al. 2002). Indeed for the 32 SSRs which detected two or more genomic loci, the majority (59%) were detected in both A and C genomes, although only six of these detected corresponding homoeologous portions of linkage groups. Apart from these observations, there appeared to be little observable pattern to the location of SSRs across the linkage groups. Indeed all linkage groups were covered by at least one SSR locus and on average most were covered by seven SSR loci (a maximum of 18 SSR loci were located on linkage group 14), and there appeared to be little bias in the placement of SSRs along linkage groups (i.e. distal vs proximal).

Such observations bode well for the use of SSRs in constructing genetic maps and for marker assisted breeding, as it should be possible to construct dense SSR maps for *Brassica* species (of which the ones listed here are a start) for a range of agronomic purposes.

Acknowledgements We would like to thank staff at IACR Long Ashton and the John Innes Centre who provided support and help during laboratory work. This work was primarily funded by a grant (ref D08078) from the Biotechnology and Biological Sciences Research Council. Supporting finance was provided by IACR Long Ashton, John Innes Centre, BBSRC and the Natural Environment Research Council.

References

Axelsson T, Bowman CM, Sharpe AG, Lydiate DJ, Lagercrantz U (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43:679–688

Bohuon EJR, Keith DJ, Parkin IAP, Sharpe AG, Lydiate DJ (1996) Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor Appl Genet 93:833–839

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M (1998) Collinearity between a 30 centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome 41:62–69

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20:759–760

Goldstein DB, Schlötterer C (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford

Howell EC, Barker GC, Jones GH, Kearsey MJ, King GJ, Kop EP, Ryder CD, Teakle GR, Vicente JG, Armstrong SJ (2002) Integration of the cytogenetic and genetic linkage maps of *Brassica oleracea*. Genetics 161:1225–1234

Khlestkina EK, Pestsova EG, Salina E, Roder MS, Arbuzova VS, Koval SF, Borner A (2002) Genetic mapping and tagging of wheat genes using RAPD, STS and SSR markers. Cell Mol Biol Lett 7:795–802

Kowalski SP, Lan T-H, Feldmann KA, Paterson AH (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics 138:1–12

Lagercrantz U, Lydiate DJ (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. Genome 38:255–264

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in *Brassica.* Genetics 144:1903–1910

Lincoln SE, Daly MJ, Lander ES (1987) MAPMAKER v3.0. A computer programme. Whitehead Institute, Cambridge, Mass.

Lionneton E, Ravera S, Sanchez L, Aubert G, Delourme R, Ochatt S (2002) Development of an AFLP-based linkage map and localization of QTLs for seed fatty acid content in condiment mustard (*Brassica juncea*) Genome 45:1203–1215

Lombard V, Delourme R (2001) A consensus linkage map for rapeseed (*Brassica napus* L.): construction and integration of three individual maps from DH populations. Theor Appl Genet 103:491–507

Lowe AJ, Jones AE, Raybould AF, Trick M, Moule CJ, Edwards KJ (2002) Transferability and genome specificity of a new set of microsatellite primers among *Brassica* species of the U triangle. Mol Ecol Notes 2:7–11

Lydiate D, Sharpe A, Largercrantz U, Parkin I (1993) Mapping the *Brassica* genome. Outl Agric 22:85–92

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome 38:1122–1131

Plieske J, Struss D (2001) Microsatellite markers for genome analysis in *Brassica*. I. Development in *Brassica napus* and abundance in Brassicaceae species. Theor Appl Genet 102:689–694

Rae SJ, Aldam C, Dominguez I, Hoebrechts M, Barnes SR, Edwards KJ (2002) Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp.) Theor Appl Genet 100:1240–1248

Saal B, Plieske J, Hu J, Quiros CF, Struss D (2001) Microsatellite markers for genome analysis in Brassica. II. Assignment of rapeseed microsatellites to the A and C genomes and genetic mapping in *Brassica oleracea* L. Theor Appl Genet 102:695–699

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sebastian RL, Howell EC, King GJ, Marshall DF, Kearsey MJ (2000) An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. Theor Appl Genet 100:75–81

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48: 463–481

Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ (1995) Frequent non-reciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38:1112–1121

Soltis DE, Soltis PS (1995) The dynamic nature of polyploid genomes. Proc Natl Acad Sci USA 92:8089–8091

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci USA 92:7719–7723

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348

Suwabe K, Iketani H, Nunome T, Kage T, Hirai M (2002) Isolation and characterization of microsatellites in *Brassica rapa* L. Theor Appl Genet 104:1092–1098

SzewcMcFadden AKS, 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

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

U N (1935) Genomic analysis of *Brassica* with special reference to the experimental formation of *B. napus* and its peculiar mode of fertilization. Jpn J Bot 7:389–452

Warwick SI, Black LD, Aguinagalde I (1992) Molecular systematics of *Brassica* and allied genera (Subtribe Brassicinae, Brassiceae) - chloroplast DNA variation in the genus *Diplotaxis*. Theor Appl Genet 83:839–850

Westman AL, Kresovich S (1998) The potential for cross-taxa simple-sequence repeat (SSR) amplification between *Arabidopsis thaliana* L. and crop brassicas. Theor Appl Genet 96:272–281

Westman AL, Kresovich S (1999) Simple sequence repeat (SSR)-based marker variation in *Brassica nigra* genebank accessions and weed populations. Euphytica 109:85–92

Zane L, Bargelloni L, Patarnello T (2002) Statregies for microsatellite isolation: a review. Mol Ecol 11:1–16