Theor Appl Genet (2008) 117:977–985
DOI 10.1007/s00122-008-0837-9

ORIGINAL PAPER

Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype

Federico L. Iniguez-Luy · Amy V. Voort · Thomas C. Osborn

Received: 19 November 2007 / Accepted: 21 June 2008 / Published online: 24 July 2008

Abstract The traditional development of simple sequence repeat (SSR) or microsatellite markers by probe hybridization can be time-consuming and requires the use of specialized laboratory equipment. In this study, probe hybridization was circumvented by using sequence information on 3,500 genomic clones mainly from *Brassica oleracea* to identify di, tri, tetra and penta-nucleotide repeats. A total of 587 primer pairs flanking SSR were developed using this approach. From these, 420 SSR markers amplified DNA in two parental lines of *B. rapa* (26% were polymorphic) and 523 in two parental lines of *B. oleracea* (32% were polymorphic). A diverse array of motif types was identified, characterized and compared with traditional SSR detection methods. The most abundant motifs found were di- (38%) and trinucleotides (33%) followed by penta- (16%) and tetranucleotide (13%) motifs. The type of motif class, motif length and repeat were not indicative of polymorphisms. The frequency of *B. oleracea* SSRs in genomic shotgun sequence was estimated to be 1 every 4 Kb. In general, the average motif length and repeat numbers were shorter than those obtained previously by probe hybridization, and they contained a more balanced representation of SSR motif types in the genome by identifying those that do not hybridize well to DNA probes. *Brassica* genomic DNA sequence information is a promising resource for developing a large number of SSR molecular markers in *Brassica* species.

Communicated by M. Xu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0837-9) contains supplementary material, which is available to authorized users.

F. L. Iniguez-Luy · A. V. Voort
Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA

T. C. Osborn
Seminis Vegetables Seeds, 36437 State Highway 16, Woodland, CA 95695, USA

F. L. Iniguez-Luy (✉)
Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA
e-mail: fliniguezluy@wisc.edu

Introduction

Identification of simple sequence repeats (SSR) or microsatellites for the development of molecular markers in *Brassica* species, has traditionally been based on probe hybridization (containing repeated motifs) against genomic or cDNA libraries followed by DNA sequencing (Kresovich et al. 1995; Lowe et al. 2004; Plieske and Struss 2001; Suwabe et al. 2002; Szewc-McFadden et al. 1996; Uzunova and Ecke 1999; Varghese et al. 2000). The recent large increase in *Brassica* DNA sequence information in Genebank databases (Ayele et al. 2005; Katari et al. 2005) provides researchers with a vast resource to survey for SSR motifs. This approach has proven to be useful in the development of SSR markers in expressed sequence tags (ESTs) of cotton (Park et al. 2005), barley, maize, rice and wheat (Peng et al. 2005; Ramesh et al. 2002; Yu et al. 2004); and also genomic sequence derived SSRs in tomato (He et al. 2003) and sorghum (Schloss et al. 2002) among other crops. Tonguc and Griffiths (2004) reported the use of 13 SSRs developed from *B. oleracea* sequence information for the inspection of genetic relationships within the clade. However, the polymorphisms detected by this number of markers may not be representative of the entire *Brassica*

genome, nor is this number sufficient to carry out genetic studies in *B. oleracea* and related species. The utilization of publicly available *Brassica* sequence information to detect SSR motifs provides a promising methodology for the development of a large number of useful molecular markers.

The increasing availability of genomic DNA sequence from different species has enabled the study of genome distribution, putative function and mutational mechanisms of SSRs evolution (reviewed in Li et al. 2002). Extensive surveys of DNA sequences in mammal and plant genomes have found that SSRs are more abundant than previously believed and their distributions differ not only among species but also between transcribed and non-transcribed regions (Morgante et al. 2002; Toth et al. 2000). Li et al. (2004) noted a recent increase in the number of scientific articles reporting the location of SSRs in transcribed regions of genomes, including protein-coding genes and ESTs. In addition, studies of motif types showed that plant SSR motifs are A/T rich (e.g., the most common dinucleotide repeat is AT/TA), while mammalian SSR motifs are G/C rich (Cardle et al. 2000; Fujimori et al. 2003; Morgante et al. 2002;Toth et al. 2000). Sequence-based SSR detection, thus, enhances the ability to identify motifs that may otherwise be missed due to technical issues (e.g., AT probe hybridization), especially in plant species (Fujimori et al. 2003; Morgante et al. 2002).

The objectives of this research were to use sequence information from the rapid cycling *B. oleracea* TO1000DH3 line to identify mono, di, tri, tetra and pentanucleotide repeats; to generate primer pairs containing SSR motifs; to screen the SSR markers for useful polymorphisms between parents of B. *rapa* and *B. oleracea* mapping populations; and to describe the nature of these SSR markers in comparison to SSR markers previously reported in the literature.

Materials and methods

SSR marker development

Information from 3,500 genomic shotgun sequences (GSS) obtained directly from GeneBank (http://ukcrop.net/perl/ace/search/BrassicaDB) were used to identify di-, tri-, tetra- and pentanucleotide repeats. *Brassica* GSS were mainly sampled from the rapid cycling *Brassica oleracea* TO1000DH3 (Ayele et al. 2005; Katari et al. 2005). There were also few other *Brassica* GSS sampled from other sources rather than TO1000DH3 (*B. rapa* and *B. napus*, clone identification and further information can be obtained in Electronic Supplementary Material 1 column 2). The identification of SSR motifs was achieved using PCB-SSR Discovery, a publicly available, web-based software package (Robinson et al. 2004) The software selects for genomic sites containing SSR motifs. PCB-SSR uses sequences flanking the repeats to design primer pairs for PCR using the "Primer 3 Workbench Biology" (http://workbench.sdsc.edu) software technology.

The following set of parameters were used in designing SSR primer pairs: spacing to amplify fragments of 100–400 bp (potential for multiplexing); high and constant annealing temperature ($\sim$50–60°C, for high specificity and automation); primers longer than 20 nucleotides (nt) and of at least 50% GC content (for stability); variation in the number, types (tri, tetra, penta) and length of SSR motif; and presence in both coding and non-coding regions.

The SSR markers developed in this study were named with the prefix FITO (a laboratory name used hereon to refer to SSR markers developed in this study using sequence information) followed by an ascending number (e.g., fito001, fito002, fito003, etc.)

Plant DNA sources and PCR amplification

Four parent lines of two mapping populations, one *Brassica rapa* L. (parental lines IMB211 rapid cycling and R500 yellow sarson) and one *B. oleracea* L. (parental lines TO1000DH3 and "Early Big" broccoli) were used to screen SSR markers for useful polymorphisms. The parental lines were grown under fluorescent lights (150 µmol) with a 12-h photoperiod. In addition, DNA from 19 *B. oleracea* varieties were used to assess the usefulness of a set of 50 FITO SSR markers developed in this study. These DNA samples include *B. oleracea* var. *medullosa* Thell. (1), *B. oleracea* var. *costata* DC. (1), *B. oleracea* var. *alboglabra* L. H. Bailey (1), *B. oleracea* ssp. *acephala* (DC) Schubler (1), *B. oleracea* var. *botrytis* L. (4), *B. oleracea* var. *gongylodes* L. (1), *B. oleracea* var. *capitata* L. (5), *B. oleracea* var. *gemmifera* Zenker, (2), and *B. oleracea* var. *italica* Plenck. (3). DNA isolations were conducted using 0.8–1 g of lyophilized seedling tissue following the CTAB procedure described by Kidwell and Osborn (1992).

PCR amplifications were conducted in a 20 µl final volume reaction. The reaction mixture contained 50 ng of template DNA, 1 µM forward and reverse primers, 2.5 mM $MgCl_2$, 2 mM each of dNTP, 5 U of Taq Polymerase and 1X PCR buffer (Promega, Madison WI). PCR products were analyzed using 2.5% high grade/low melting agarose (sensitive to 5–10 bp differences) and/or 6–8% acrylamide in 1× TBE buffer gel electrophoresis (sensitive to 2–5 bp differences).

SSR marker screening

A total of 587 FITO SSR markers were screened for useful polymorphisms against the four parental lines described

above. Number of loci detected by each SSR primer pair was estimated by following the segregation pattern of a given SSR marker in a small set of progeny from each of the two mapping populations. Also, using the same screening material and criteria a set of 144 SSR markers previously developed using probe hybridization methods (Kresovich et al. 1995; Plieske and Struss 2001; Suwabe et al. 2002; Szewc-McFadden et al. 1996; Uzunova and Ecke 1999 Varghese et al. 2000; http://ukcrop.net/perl/ace/search/BrassicaDB) were screened for useful polymorphisms that could be used in mapping studies (SSR names available in Electronic Supplementary Material 2). The 144 SSR chosen represent a sample of the SSR available in the literature and there was no intentional selection and/or preselection from our part.

A set of 50 FITO SSR markers most of which amplified single fragment polymorphisms that were easy to score using agarose and/or acrylamide gel electrophoresis, was chosen to screen the diverse set of 19 *B. oleracea* morphotypes. A polymorphic information content value PIC was calculated for each FITO SSR markers according to the formula:

$$\mathrm{PIC}_i = 1 - \Sigma p_{ij}^2$$

where p_{ij} is the frequency of the *j*th SSR fragment for SSR *i* (Anderson et al. 1993). Mean PIC values across all markers were calculated for Brussels sprouts, broccoli, cauliflower and cabbage.

Results

Polymorphisms detected by SSR markers developed using sequence information

Five hundred and eighty-seven SSR containing sequences were detected by searching for SSR motifs from 3,500 genomic sequences, mainly GSS from *B. oleracea*. PCR primer pairs for the 587 SSR containing sequences were developed (ESM 1, 3) and tested for amplification and useful polymorphisms. Positive PCR amplifications were detected with 433 primer pairs (74%) for the *B. rapa* parental lines and with 524 (89%) primer pairs for the *B. oleracea* parental lines (Table 1). The majority of SSR markers tested amplified single fragments, (280 for *B. rapa* and 378 for *B. oleracea* parental lines; Table 1). The remaining SSR primer pairs amplified 2–4 fragments for each species. Based on screening the parental lines and small sets of segregating progenies, we detected a total 213 (32% of the total possible) and 196 (26%) loci that were polymorphic and could be mapped in the *B. rapa*, and *B. oleracea* populations, respectively. About one-half of the loci for each species were detected as single polymorphisms (Table 1). We did not find compound SSR loci in this study.

At least 46 of the 50 FITO SSRs primer pairs used (ESM 1, 3) to screen the 19 *B. oleracea* accessions gave positive PCR amplifications (Table 2). Polymorphic SSR markers were detected within morphotypes represented by multiple

Table 1 Summary of FITO SSR markers developed and screened for useful polymorphisms using two *Brassica oleracea* and two *Brassica rapa* parental lines of mapping populations

Material screened	No. of FITO primers tested	No. of "+" PCR amplification	No. of PCR fragments in each parent	Maximum no. of possible loci[c]	Actual no. of polymorphic loci detected[c]				Total mappable loci detected	%
					0	1	2	3		
	63	–	0	–	–	–	–	–	–	–
	378	378	1	378	282	96	–	–	96	25
Brassica oleracea[a]	99	99	2	198	47	48	4	–	56	28
	23	23	3	69	6	11	3	3	26	38
	24	24	4	96	7	11	1	5	28	29
Total	587	524 (89%)	–	741	342	166	8	8	206	28
	154	–	0	–	–	–	–	–	–	–
	280	280	1	280	188	92	–	–	92	33
Brassica rapa[b]	89	89	2	178	41	44	4	–	52	29
	47	47	3	141	16	23	3	5	44	31
	17	17	4	68	3	6	5	3	25	37
Total	587	433 (74%)	–	667	248	165	12	8	213	32

[a] The *B. oleracea* parental mapping lines used for the polymorphic screening were TO1000DH3 rapid cycling and "Early Big" broccoli

[b] The *B. rapa* parental mapping lines used for the polymorphic screening were IMB211 rapid cycling and R500 yellow sarson

[c] Discrepancies between the maximum number of possible loci and the actual number of polymorphic loci detected is due to co-segregation of parental fragments

Table 2 Number of PCR primer pairs producing amplification products, and number of polymorphic fragments and PIC values within morphotypes for the set of 50 FITO SSRs primer pairs used to screening 19 *B. oleracea* accessions

Morphotype		No. of accessions sampled	No. of primer pairs producing PCR products	No. of polymorphic fragments[a]	Mean PIC value[b]
Botanical name	Common name				
B. oleracea var. *medullosa* Thell.	Marrow Stem Kale	1	44	n.d.[c]	n.d.
B. oleracea var. *costata* DC.	Tronchuda	1	46	n.d.	n.d.
B. oleracea var. *alboglabra* L. H. Bailey.	Chinese Kale	1	41	n.d.	n.d.
B. oleracea ssp. *acephala* (DC) Schubler.	Kale	1	41	n.d.	n.d.
B. oleracea var. *gongylodes* L.	Kholrabi	1	36	n.d.	n.d.
B. oleracea var. *gemmifera* Zenker.	Brussels Sprouts	2	44	15	0.48
B. oleracea var. *italica* Plenck.	Broccoli	3	41	21	0.54
B. oleracea var. *botrytis* L.	Cauliflower	4	40	29	0.57
B. oleracea var. *capitata* L.	Cabbage	5	45	37	0.65

[a] 72 polymorphic fragments were detected across all 19 accessions

[b] 0.67 was the mean PIC value for all markers across all the 19 accessions

[c] n.d. = not determined

accessions and the mean PIC values and ranges across all the polymorphic SSR markers were 0.48 (0.25–0.75) for Brussels sprouts, 0.54 (0.1–0.8) for broccoli, 0.57 (0.25–0.93) for cauliflower and 0.65 (0.16–0.96) for cabbage (Table 2).

Characteristics of SSR markers

Five motif classes, mono-, di-, tri-, tetra- and pentanucleotide, were identified in this study (Table 3). Dinucleotide motifs were the most frequently found (226) followed by tri- (193), penta- (93), and tetranucleotide (74). Only one mononucleotide motif was detected. An average of one SSR was detected every six *B. oleracea* GSS surveyed. The average length of each *B. oleracea* GSS is about 667 bases (Ayele et al. 2005). Thus, the estimated frequency of SSR motifs is one every ~4 Kb. The PCR product size for di-, tri-, tetra-, and pentanucleotide were 266 (range 101–446), 272 (range 100–400), 273 (range 119–398), 266 (range 101–446), respectively. The repeat number and repeat length distributions varied among the different motif types detected. Most of the dinucleotide motifs were found in the range of 6–10 repeats with 11–20 nucleotides in length, although they were represented throughout the observed distribution. Tri- tetra- and pentanucleotides were mostly found in the range of 2–5 repeats with 11–20 nucleotides in length (Fig. 1).

There were no significant correlations between the number of repeats, repeat length and type of SSR motif detected (e.g., di- or trinucleotide) with informative polymorphism detected in the four parental lines tested (data not shown). The percentage of polymorphic loci was also similar across the distribution of repeat number and repeat length (Fig. 1). The mean percentages of loci that were polymorphic between the four parental lines for di-, tri-tetra; and pentanucleotide motifs were 41, 31.5, 34 and 39%, respectively.

Two-hundred and seventy-two out of the 587 SSR makers were derived from clones that contained primary homology scores to the *Arabidopsis* genome (as determined by blast results reported in the UKCrop *Brassica* database for each cloned surveyed). Most of these sequences corresponded to genes or hypothetical expressed sequences (Fig. 2), and the predominant motif types within these two categories were di- (38) and trinucleotides (44). Other genomic landmarks such as transposable elements, predicted pseudogenes, and non-coding regions were also found to contain FITO SSR markers (Fig. 2). The percentages of polymorphisms according to the type of sequences flanking SSR motifs ranged from 24–38%, and were similar for the two sets of mapping parents (Fig. 2). The predicted distributions of the SSR markers from primary homology to the *Arabidopsis* genome showed an even representation of SSR markers among the five *Arabidopsis* chromosomes (ESM 4).

Comparison of results from different SSR detection methods

Dinucleotide motifs were the most frequent SSR class detected using the methodology presented in this study. We surveyed and screened for mappable polymorphisms a set 144 SSR that were developed using traditional hybridization methods (Kresovich et al. 1995; Plieske and Struss 2001; Suwabe et al. 2002; Szewc-McFadden et al. 1996; Uzunova and Ecke 1999; Varghese et al. 2000; http://ukcrop.net/perl/ace/search/BrassicaDB) (ESM 5). Dinucleotide motifs were also found to be the largest class of

Table 3 Summary of different motif types per motif class examined for 587 primers pairs developed using *B. oleracea* sequence information

SSR motif types detected from sequence information							
Dinucleotide		Trinucleotide		Tetranucleotide		Pentanucleotide	
Motif	Primers detected	Motif	Primers detected	Motif	Primers detected	Motif	Primers detected
AC/TG	15	AAC/TTG	11	AAAC/TTTG	5	AAAAC/TTTTG	9
AG/TC	65	AAG/TTC	19	AAAG/TTTC	7	AAAAG/TTTTC	6
AT/TA	81	AAT/TTA	10	AAAT/TTTA	9	AAAAT/TTTTA	15
CA/GT	12	ACA/TGT	13	AACC/TTGG	1	AAACC/TTTGG	2
CT/GA	53	ACC/TGG	5	AAGG/TTCC	1	AAATC/TTTTC	2
–	Total 226	ACG/TGC	3	AAGT/TTCA	1	AAATT/TTTAA	3
–	–	AGA/TCT	24	AATA/TTAT	7	AACCG/TTGGC	1
–	–	AGC/TCG	5	AATC/TTAG	1	AAGAG/TTCTC	1
–		AGT/TCA	4	AATG/TTAC	1	AAGGA/TTCCT	1
	–	ATC/TAG	6	ACAA/TGTT	4	AATAA/TTATT	1
–		ATT/TAA	4	AGAA/TCTT	2	AATTT/TTAAA	1
		CAA/GTT	6	ATAA/TATT	6	ACAAA/TGTTT	1
	–	CAG/GTC	3	ATAG/TATC	4	ACCCG/TGGGC	1
–		CAT/GTA	4	ATCT/TAGA	1	ACCGA/TGGCT	1
	–	CCG/GGC	3	CAAA/GTTT	2	ACCTC/TGGAG	1
–		CCT/GGA	7	CATT/GTAA	1	AGCTT/TCGAA	1
	–	CGA/GCT	2	CCAA/GGTT	1	AGGCA/TCCGT	1
–		CGC/GCG	1	CCCT/GGGA	1	ATCTA/TAGAT	1
	–	CGG/GCC	1	CTGT/GACA	1	ATCTC/TAGAG	1
–	–	CTC/GAG	8	CGTA/GCAT	1	ATTTA/TAAAT	3
–	–	CTT/GAA	30	CTTC/GAAG	1	ATTTT/TAAAA	2
–		GAT/CTA	3	GAAA/CTTT	1	CAACT/GTTGA	3
	–	GGT/CCA	4	GAGT/CTCA	1	CAGAA/GTCTT	1
–	–	TAC/ATG	1	GATT/CTAA	1	CCCCT/GGGGA	1
–	–	TAT/ATA	7	GGAA/CCTT	2	CCGAA/GGCTT	1
–	–	TCC/AGG	4	TAAA/ATTT	1	CTCTT/GAGAA	1
–	–	TGA/ACT	5	TCAT/AGTA	1	CTTGC/GAACG	2
–	–	–	Total 193	TCTA/AGAT	1	CTTTT/GAAAA	1
–	–	–	–	TCTT/AGAA	1	GAGAG/CTCTC	1
–	–	–	–	TTCT/AAGA	4	GGAAG/CCTTC	1
–	–	–	–	TTGA/AACT	1	GGAGA/CCTCT	1
–	–	–	–	TTGT/AACA	2	GGGAA/CCCTT	2
–	–	–	–	–	Total 74	GGGCC/CCCGG	1
–	–	–	–	–	–	GTTTG/CAAAC	1
–	–	–	–	–	–	GTTTT/CAAAA	1
–	–	–	–	–	–	TATAA/ATATT	1
–	–	–	–	–	–	TCAAA/AGTTT	1
–	–	–	–	–	–	TCCCA/AGGGT	1
–	–	–	–	–	–	TCTCT/AGAGA	1
–	–	–	–	–	–	TCTTC/AGAAG	2
–	–	–	–	–	–	TGATG/ACTAC	1
–	–	–	–	–	–	TGATT/ACTAA	1
–	–	–	–	–	–	TGCAT/ACGTA	1
–	–	–	–	–	–	TGGAA/ACCTT	1
–	–	–	–	–	–	TGGTC/ACCAG	4
–	–	–	–	–	–	TTCGA/AAGCT	1

Table 3 continued

SSR motif types detected from sequence information							
Dinucleotide		Trinucleotide		Tetranucleotide		Pentanucleotide	
Motif	Primers detected	Motif	Primers detected	Motif	Primers detected	Motif	Primers detected
–	–	–	–	–	–	TTCGG/AAGCC	1
–	–	–	–	–	–	TTGAA/AACTT	1
–	–	–	–	–	–	TTGAT/AAGAT	1
–	–	–	–	–	–	TTGTC/AACAG	1
–	–	–	–	–	–	TTTCT/AAACA	1
–	–	–	–	–	–	–	Total 93

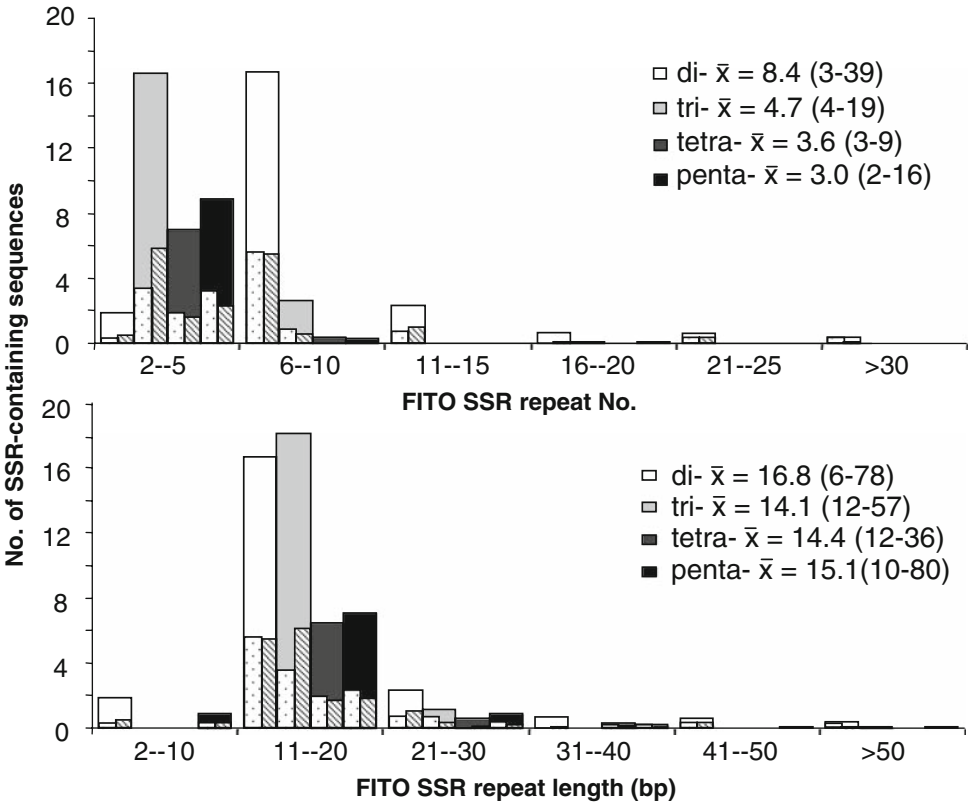

Fig. 1 Distributions of the range of four FITO SSR motifs (*boxes*) across SSR repeat number and repeat length (nucleotide). *Dotted* and *dashed boxes* within each of the four SSR classes (di-, tri-, tetra- and penta-), represent the number of polymorphic loci between two *B. oleracea* (TO1000DH3 rapid cycling and "Early Big" broccoli) and two *B. rapa* (IMB211 rapid cycling and R500 yellow sarson) parental mapping lines, respectively. The mean number and range of FITO SSR repeat and length for each motif type are provided in the legend

motif available in the set of 144 previously published SSR. Therefore, this motif class was used for comparing the sequence-based and clone hybridization-based SSR detection methodologies. The comparisons revealed differences in the proportions of motif types detected by each methodology (Fig. 3).

The polymorphisms detected for the two sets of mapping parents also differed between the methodologies (Fig. 3, dotted and dash boxes). The largest percentage of polymorphic loci detected in SSRs developed from traditional methods was found for the CT/GA (73%) motif type. Polymorphic loci percentages for the remaining motifs were similar (ranging within 10–15%). There were no polymorphisms detected for the AT/TA motif type for SSR markers developed using traditional discovery approaches. Percentages of polymorphisms detected using sequence-based information were similar (around 25–35%) for each of the different motif types examined, including the AT/TA motif type.

Discussion

A recent *B. rapa* SSR isolation study conducted by Suwabe et al. (2002) (hybridization-based methods) had similar PCR amplification efficiencies (~90%) to the one obtained

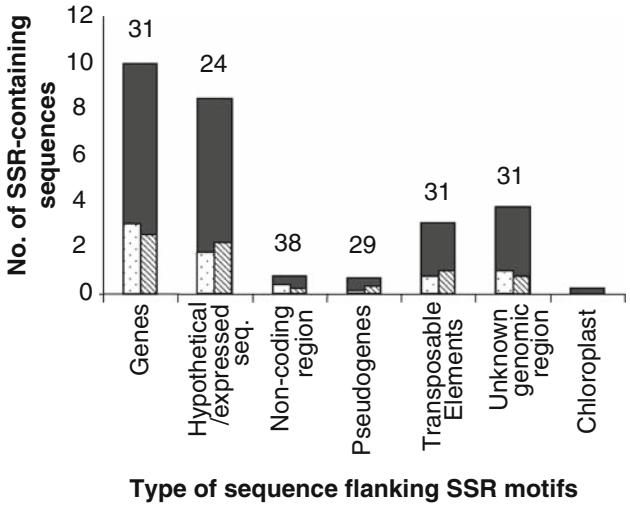

Fig. 2 Genomic distribution for 272 SSR containing sequences, for which there was primary homology to the *Arabidopsis* genome (*gray boxes*). *Dotted* and *dashed boxes* within each type of sequence flanking SSR motifs, represent the number of polymorphic loci between two *B. oleracea* (TO1000DH3 rapid cycling and "Early Big" broccoli) and two *B. rapa* (IMB211 rapid cycling and R500 yellow sarson) parental mapping lines, respectively. Numbers on tops of each bar illustrates the percentage of polymorphism according to the type of sequence flanking SSR motifs

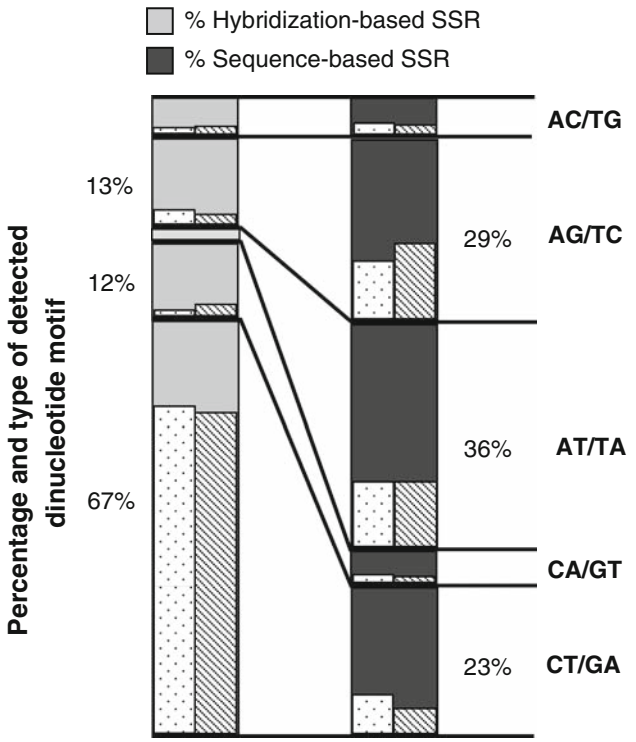

Fig. 3 Comparison between type of dinucleotide motifs detected and percentage of dinucleotides detected, using two different methods of SSR marker development (hybridization-based SSR detection *light gray* and sequence-based SSR detection *dark gray*). *Dotted* and *dashed boxes* within the two main boxes represent the number of polymorphic loci between two *B. oleracea* (TO1000DH3 rapid cycling and "Early Big" broccoli) and two *B. rapa* (IMB211 rapid cycling and R500 yellow sarson) parental mapping lines, respectively

in this study (85%). Lower efficiencies (15–20%) have been obtained in early *Brassica* SSR detection (using hybridization-based methods) studies (Kresovich et al. 1995 and Plieske and Struss 2001). The efficiency of detecting polymorphisms among our mapping parents was slightly lower for our sequence-based SSR markers (30%) than for the hybridization-based SSR markers obtained in the literature (~39%). However, SSR markers that are described in the literature (Kresovich et al. 1995; Plieske and Struss 2001; Szewc-McFadden et al. 1996; Suwabe et al. 2002; Uzunova and Ecke 1999; Varghese et al. 2000) had been previously selected for amplification and detection of polymorphism. Therefore, a much higher percentage was expected to show polymorphisms in the four parental lines as compared to the FITO unselected SSR markers. In this study we showed that SSR markers detected using sequence information may provide *Brassica* researchers with an efficient tool for the development of a large set of useful molecular markers.

The mean PIC values for the set of 50 FITO SSR markers that were used within Brussels sprout, broccoli, cauliflower and cabbage, showed that these markers are highly informative and can be used to explore the genetic diversity and heterozygosity among cultivated *B. oleracea* germplasm. This characteristic is of interest for *B. oleracea* public breeders due to the paucity of reported PCR-based markers for this species (Hale et al. 2006; Tonguc and Griffiths 2004). In addition, the PIC values obtained in our study are similar to values obtained in other *Brassica* SSR studies (Szewc-McFadden et al. 1996; Plieske and Struss 2001). Moreover, in agreement with Tonguc and Griffiths (2004), the cabbage cultivars were the most polymorphic (mean PIC value 0.65) followed by cauliflower, broccoli and Brussels sprouts. The high transferability of the FITO SSR markers between *B. rapa* and between and within *B. oleracea* reinforces the close phylogenetic relationship among these two species and their morphotypes (Warwick and Black 1993).

Our study reflects SSRs detected in a very small fraction of the available 0.44X (Ayele et al. 2005) sequencing pass of the genome. However, similar frequencies of SSR occurrence, to the one estimated in this study, have been reported for EST derived SSRs (1 every 5.46–6.2 Kb) in wheat and other cereals (Peng et al. 2005; Varshney et al. 2002). Wang et al. (1994) had estimated a lower frequency of occurrence (1 every 25.4 Kb) for *Brassica* SSRs, including mono- di-, tri- and tetranucleotide motifs. Also, lower SSR frequencies have been reported (1 SSR every 100–125 Kb), for *Brassica* GA/CT dinucleotide motif repeats (Kresovich et al. 1995; Lagercrantz et al. 1993). In our study, we estimated the frequency of GA/CT dinucleotides to be approximately 1 every 43.3 Kb. Lower SSR

frequencies reported in other studies could be due partly to the stringent limits set for SSR detection using probe hybridization methods. Other factors that may influence these low SSR frequencies are sampling variation, segments of the genome available for SSR detection (e.g., coding vs. non-coding) and library preparation. In order to achieve a more comprehensive estimation of SSR distribution in the core *Brassica* genome, more sequence information needs to become available. The sequence completion of the *B. rapa* cytogenetic chromosome 1 (linkage group R9) (Kim et al. 2006) will provide researchers with an exact determination of SSR occurrence on this chromosome.

Generally, SSR motifs detected using sequence information revealed a shorter number of repeats and length than those found in the literature (Lowe et al. 2004; Plieske and Struss 2001; Szewc-McFadden et al. 1996; Suwabe et al. 2004). The same observation was reported in studies aimed to isolate SSR motifs from EST databases of barley, cotton, maize and wheat (Park et al. 2005; Peng et al. 2005; Ramesh et al. 2002; Yu et al. 2004). A possible explanation for this phenomenon is that hybridization methods may fail to detect short repeats because the signal from this type of SSRs may be too weak. Another explanation may simply have to do with the region (from the *B. oleracea* genome) that was available for SSR detection and identification. The identification of shorter SSR motifs, however, did not influence our ability to find informative markers for the four parental lines. Indeed, the highest percentage of polymorphic loci was detected in repeats ranging from 2 to 10 with a length of 11–20 nucleotides. This was true for the four motif classes examined (Fig. 2 dotted and dashed boxes). This result seems to be in agreement with recent studies that have shown a lack of correlation on the size of repeats, both measured by length and number, and detection of polymorphic loci (Plieske and Struss 2001; Suwabe et al. 2004; Szewc-McFadden et al. 1996). Moreover, SSR polymorphisms are influenced by slippage mutation rates more than by the repeat length, with SSR evolutionary age being a key factor for SSR diversity (Lai and Sun 2003). We cannot, however, rule out the possibility that insertions and deletions (INDELs) at regions other than the SSR motifs may account for some of the polymorphism between the parental lines and germplasm surveyed in this study, specially for polymorphism observed between *Brassica* species or subspecies within the species. The most abundant dinucleotide motifs detected was the AT/TA type (36%). The most common trinucleotide motif detected was the CTT/GAA type (15%). The most common tetranucleotide motifs were AAAG/TTTC and ATAA/TATT (9% each), and the most common pentanucleotide motif was the AAAAAT/TTTTA type (16%). These motif types and proportions in the genome are in close agreement with the previous plant SSR studies (Fujimori et al. 2003; Morgante et al. 2002). AT and AG combinations of base pair motif types either in di- and trinucleotide configuration are the most abundant type in plants genomes (Cardle et al. 2000).

The set of 144 public SSR used here to compare the difference in detection methodology was composed to predominantly dinucleotide motifs. The detection methodology used in this study showed the CA/GT motif as the least represented one, indeed, this motif has been reported to be less frequent in plants than in mammals (Cardle et al. 2000; Fujimori et al. 2003; Morgante et al. 2002; Toth et al. 2000). On the contrary, SSR primers developed using hybridization methods showed a strong tendency to detect SSRs containing CT/GA motifs. Bias towards the detection of certain motifs over others, for example AT/TA, may be attributed to technical issues when hybridizing AT repeats (Tm and secondary structures, etc.).

For those SSR that had primary homology to *Arabidopsis* sequence, the use of sequence-based detection methods made it possible to predict the approximate location in the *B. oleracea* genome. This was possible through a comparative genomic approach with the putative position of markers in relation to syntenic *A. thaliana* genome segments previously described in Lukens et al. (2003) and Parkin et al. (2005). Also, most of the SSRs for which there was physical information were found to reside within genes. This result corroborates with Li et al. (2004) where the authors report on an increasing number of SSRs found and characterized within protein-coding genes and their untranslated regions both in mammals and plant genomes. The majority of this type of SSR exhibited trinucleotide motifs perhaps representing codon usage or avoidance of frame-shift mutations (Li et al. 2002, 2004).

Our study demonstrates that the use of genomic sequence for SSR detection produces a more balanced representation of SSR motifs in the *B. oleracea* genome than does the use of hybridization methods. The employment of sequence information alleviated the need to screen and select clones from genomic libraries in order to develop SSR markers. In addition, the use of genomic sequence allowed for the setting of specified parameters while designing the SSR primer pairs. The SSRs developed in this study have been used to generate genetic maps in two *Brassica* mapping populations (*B. oleracea* and *B rapa*). This will provide *Brassica* researchers with common sets of public mapping populations and PCR-based molecular markers.

References

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

Ayele M, Haas BJ, Kumar N, Wu H, Xiao Y, Van Aken S, Utterback TR, Wortman JR, White OR, Town CD (2005) Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in *Arabidopsis*. Genome Res 15:487–495

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S, Tomita M (2003) A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett 554:17–22

Hale AL, Farnham MW, Menz MA (2006) Use of PCR-based markers for differentiating elite broccoli inbreds. J Am Soc Hortic Sci 131:418–423

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeats markers and their use in determining relationships among *Lycopersicon eculentus* cultvars. Theor Appl Genet 106:363–373

Katari MS, Balija V, Wilson RK, Martienssen RA, McCombie WR (2005) Comparing low coverage random shotgun sequence data from *Brassica oleracea* and *Oryza sativa* genome sequence for their ability to add to the annotation of *Arabidopsis thaliana*. Genome Res 15:496–504

Kidwell KK, Osborn TC (1992) Simple plant DNA isolation procedure. In: Beckman J, Osborn TC (eds) Plant Genomes: methods for genetic and physical mapping. Kluwer Academic Publishers, The Netherlands, pp 1–13

Kim JS, Chung TY, King GJ, Jin M, Yang T-J, Jin Y-M, Kim H-I, Park B-S (2006) A sequence-tagged linkage map of *Brassica rapa*. Genetics 174:29–39

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple-sequence repeats (SSRs) 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 microsatellites motifs differs between plants and vertebrates. Nucleic Acid Res 21:1111–1115

Lai Y, Sun F (2003) The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol Biol Evol 20:2123–2131

Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465

Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet 108:1103–1112

Lukens L, Zou F, Parkin I, Lydiate D, Osborn T (2003) Comparison of the *Brassica oleracea* genetic map with the *Arabidopsis thaliana* physical map. Genetics 164:359–372

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Park Y-H, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:325–443

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781

Peng JH, Nora L, Lapitan V (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integ Genomics 5:80–96

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

Ramesh VK, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Robinson AJ, Love CG, Batley J, Barker G, Edwards D (2004) Simple Sequence marker loci discovery using SSR primer. Bioinformatics 20:1475–1476

Schloss S, Mitchell S, White G, Kukatla R, Bowers J, Paterson A, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 105:912–920

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

Suwabe K, Iketani H, Nunome T, Ohyama A, Hirai KM, Fukuoka H (2004) Characteristics of microsatellites in *Brassica rapa* genome and their potential utilization for comparative genomics in Cruciferae. Breeding Sci 54:85–90

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

Tonguc M, Griffiths PD (2004) Genetics relationships of *Brassica* vegetables determined using database derived simple sequence repeats. Euphityca 137:193–201

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981

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

Varghese JP, Rudolph B, Uzunova MI, Ecke W (2000) Use of 5′-anchored primers for the enhanced recovery of the specific microsatellite markers in *Brassica napus* L. Theor Appl Genet 101:115–119

Varshney RK, Thiel T, Stein N, Landrigde P, Garner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plants short tandem repeats. Theor Appl Genet 88:1–6

Warwick SI, Black LD (1993) Molecular relationships in subtribe Brassicinae (Cruciferar, tribe Brassiceae). Can J Bot 71:906–918

Yu J-K, La Rota M, Kantety RV, Sorrells ME (2004) EST Derived SSR Markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751