Theor Appl Genet (2009) 118:1121–1131
DOI 10.1007/s00122-009-0967-8

ORIGINAL PAPER

Development and genetic mapping of microsatellite markers from genome survey sequences in *Brassica napus*

Xiaomao Cheng · Jinsong Xu · Shu Xia · Jianxun Gu · Yuan Yang · Jie Fu · Xiaoju Qian · Shunchang Zhang · Jiangsheng Wu · Kede Liu

Received: 4 September 2008 / Accepted: 6 January 2009 / Published online: 4 February 2009
© Springer-Verlag 2009

Abstract Microsatellite or simple sequence repeat (SSR) markers are routinely used for tagging genes and assessing genetic diversity. In spite of their importance, there are limited numbers of SSR markers available for *Brassica* crops. A total of 627 new SSR markers (designated BnGMS) were developed based on publicly available genome survey sequences and used to survey polymorphisms among six *B. napus* cultivars that serve as parents for established populations. Among these SSR markers, 591 (94.3%) successfully amplified at least one fragment and 434 (73.4%) detected polymorphism among the six *B. napus* cultivars. No correlation was observed between SSR motifs, repeat number or repeat length with polymorphism levels. A linkage map was constructed using 163 newly developed BnGMS marker loci and anchored with 164 public SSRs in a doubled haploid population. These new markers are evenly distributed over all linkage groups (LGs). Given that the majority of these SSRs are derived from bacterial artificial chromosome (BAC) end sequences, they will be useful in the assignment of their cognate BACs to LGs and facilitate the integration of physical maps with genetic maps for genome sequencing in *B. napus*.

Communicated by H. C. Becker.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-0967-8) contains supplementary material, which is available to authorized users.

X. Cheng · J. Xu · S. Xia · J. Gu · Y. Yang · J. Fu · X. Qian · S. Zhang · J. Wu (✉) · K. Liu (✉)
National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China
e-mail: jiangshengwu@mail.hzau.edu.cn

K. Liu
e-mail: kdliu@mail.hzau.edu.cn

Introduction

The detection of DNA sequence variation is of great importance in genetic studies of the *Brassica* genomes. Restriction fragment length polymorphism (RFLP) markers have been extensively used to map agronomically important genes (Uzunova et al. 1995), comparative mapping of the *Brassica* genomes (Lagercrantz 1998) and also for comparative evolution in the related model crucifer, *Arabidopsis thaliana* (Lan et al. 2000; Parkin et al. 2005). However, RFLP requires a large amount of DNA and the procedure is inefficient. With the invention of polymerase chain reaction (PCR), a variety of PCR-based markers, such as random amplified polymorphic DNA, amplified fragment length polymorphisms, sequence-related amplified polymorphisms, simple sequence repeats (SSRs) and single nucleotide polymorphisms (Batley et al. 2007b) were successively developed. Among different types of molecular markers, SSRs are becoming the preferred markers of choice for tagging genes and assessing genetic diversity. This is largely because SSRs require only a small amount of DNA, are easily detectable by PCR, amenable to high-throughput analysis, codominantly inherited, multi-allelic, highly polymorphic, abundant and evenly distributed in genomes (Gupta and Varshney 2000). SSRs have been extensively used in tagging qualitative genes and dissecting the genetic basis of complex traits (Delourme et al. 2006; Padmaja et al. 2005; Zhao et al. 2006).

The abundance, characterization and the usefulness of microsatellite markers in *Brassica* species is well documented (Lowe et al. 2004; Plieske and Struss 2001; Suwabe et al. 2002). The number of publicly available *Brassica* microsatellite markers is increasing as several projects have been conducted to develop SSR markers in *Brassica* species including *B. rapa*, *B. oleracea*, *B. napus* and *B. nigra*

(Lowe et al. 2004; Plieske and Struss 2001; Suwabe et al. 2002), and several genetic linkage maps were constructed using these SSR markers for gene mapping (Padmaja et al. 2005; Saito et al. 2006) and QTL analysis (Chen et al. 2007b; Delourme et al. 2006; Lombard and Delourme 2001; Long et al. 2007; Qiu et al. 2006; Zhao et al. 2006). However, relatively few SSR markers are publicly available when compared to many other crops such as soybean and rice (McCouch et al. 2002; Shoemaker et al. 2008). The number of currently available mapped SSR markers is about 1,000 (Choi et al. 2007; Lagercrantz et al. 1993; Lowe et al. 2002, 2004; Piquemal et al. 2005; Suwabe et al. 2002) and insufficient for genetic analysis and map-based gene cloning in *Brassica*. Therefore, a great demand exists for many more SSR markers to construct high-density linkage maps for *Brassica* genome research and breeding.

The traditional procedure for developing SSR markers involves the generation of a small-insert genomic library, screening for SSR-containing clones by several rounds of hybridization with tandemly repeated oligonucleotides and sequencing the candidate clones. This process is time consuming, costly and labor-intensive. Development of SSR markers from publicly available DNA sequences has become a fast and cost-effective alternative for many crop species (Chen et al. 2007a; McCouch et al. 2002; Shoemaker et al. 2008; Shultz et al. 2007; Song et al. 2005). During the past few years, several genome projects have been conducted in *Brassica* species to decipher partial or whole genomes (Ayele et al. 2005; Hong et al. 2008) and to reveal gene expression profiles (Lim et al. 1996, 2000). A total length of 283 Mb whole genome shotgun sequences was obtained and deposited in the GenBank (Ayele et al. 2005). The *B. rapa* genome sequencing project (BrGSP), which seeks to decode the complete sequence of the *Brassica* A-genome using a bacterial artificial chromosome (BAC) by BAC sequencing approach, was launched based on the Chinese cabbage line Chiifu-401 (Hong et al. 2006). To date, a vast array of *B. rapa* genomic sequences, including BAC-end sequences (BESs) and seed BACs sequences have become available. Recently, great efforts have been made to develop GSS–SSRs and EST–SSRs for *B. rapa* (Choi et al. 2007; Ling et al. 2007), *B. oleracea* (Iniguez-Luy et al. 2008), *B. juncea* (Hopkins et al. 2007) and *B. napus* (Batley et al. 2007a). The utilization of publicly available *Brassica* sequence information for detecting SSR motifs provides a promising methodology for the development of a large number of useful molecular markers.

We describe the development of microsatellite markers in *B. napus* from publicly available genome survey sequences (GSSs) which mainly consist of BESs. The frequency and length variation of different types of di-, tri-, tetra- and pentanucleotide microsatellites in *B. napus* are analyzed. A total of 627 SSR markers were developed and evaluated for PCR amplification and polymorphisms using six cultivars. A subset of the polymorphic SSR markers was assigned to linkage groups (LGs) in the *B. napus* genome.

Materials and methods

Plant materials and DNA extraction

Six rapeseed cultivars or lines (S1, S2, M201, M202, No. 2127 and ZY821) were used for SSR marker polymorphism detection. These were chosen based on their use as parents of pre-existing populations that segregate for traits of interest. Using S1 with long siliques and large seeds, and S2 with short siliques and small seeds, an F_2 population of 186 individuals was produced to map genes controlling the silique length and seed weight. M201 and M202 have high and low seed oil content, respectively. With these two lines, a RIL population consisting of 157 lines was generated for mapping genes controlling oil content in seeds. No. 2127 is a purely yellow double-haploid (DH) line derived from a resynthesized *B. napus* line obtained from an interspecific cross of a light brown seeded *B. alboglabra* (a form of *B. oleracea*) and a yellow-seeded *B. rapa* var. *trilocularis* (yellow sarson) (Chen et al. 1988). ZY821 is a widely cultivated black-seeded elite cultivar. A DH population derived from the F_1 hybrid between No. 2127 and ZY821, which was previously used for mapping the genes controlling the seed coat color (Xiao et al. 2007), was used for the map construction. Total DNA was isolated from young leaves of 88 DH lines using the cetyltrimethylammonium bromide method (Li and Quiros 2001) and was used for the genetic mapping described herein.

Source of sequences

A total of 13,794 GSSs were downloaded from the NCBI. The majority of the GSSs (12,278) (DU110997-DU098720) are BESs from a *Bam*HI BAC library constructed with the inbred line “Tapidor”. The GSSs also include 957 (CZ886936-CZ887892) and 398 (CZ887893-CZ888290) genomic clones from the methylation filtered and unfiltered libraries, respectively. There were an additional 57 (CZ692860-CZ692804) and 104 (CZ906382-CZ906485) genomic sequences of RFLP probes from the pIJ2925 and the WG libraries, respectively.

Searching for microsatellites

A web-based tool, SSR Primer, that integrates the SSR repeat finder SPUTNIK (http://abajian.net/sputnik/) with the primer design program Primer3 (Rozen and Skaletsky 2000) within one pipeline, was used for mining SSR markers

from GSS sequences (GSS–SSR) as previously described (Robinson et al. 2004). Multiple FASTA formatted GSSs were submitted to the online SSR Primer tool (http://hornbill.cspp.latrobe.edu.au/ssrdiscovery.html) to screen each sequence for microsatellites. The criteria for SSR selection were as follows: dinucleotides, six repeats; tri-, tetra- and pentanucleotides four repeats. Mononucleotide repeats were not considered in this analysis due to the difficulty of distinguishing *bona fide* microsatellites from sequencing or assembly error.

Removal of redundant sequences and primer design

Microsatellites with repeat lengths of ≥16 bp for di-, ≥18 for tri- and ≥20 for tetra- and pentanucleotides were selected for SSR marker development. To eliminate redundancy and to avoid designing redundant sets of primers for the same locus, the selected SSR-containing sequences were assembled using the CAP3 software (Huang and Madan 1999). The resulting contigs and singletons were parsed to SSR Primer again for primer design with the following settings. The primer length was between 18 and 23 nucleotides, with an optimum size of 20 nucleotides. The melting temperatures ranged from 50 to 70°C, with an optimum temperature of 55°C. The optimum GC content was set to 50%, with a minimum of 30% and a maximum of 70%. The predicted PCR products ranged from 100 to 400 bp. The newly developed GSS–SSR markers were designated as “BnGMS”, representing ***B****rassica* ***n****apus* **g**enomic **m**icro**s**atellites. All primers were synthesized by GeneRay Biotech (Shanghai, China).

Polymorphism detection

All SSR markers developed in this study and 546 public SSR markers were subjected to polymorphism detection using six rapeseed cultivars. The SSR primers prefixed by “Ra”, “Ol”, “Na” and “Ni” were derived from genomic sequences of *B. rapa*, *B. oleracea*, *B. napus* and *B. nigra*, respectively, and obtained at http://ukcrop.net/ace/search/BrassicaDB (Lagercrantz et al. 1993; Lowe et al. 2002, 2004). Primer pairs prefixed “MR” were developed from SSR-containing genomic clones (Uzunova and Ecke 1999), “BRAS” and “CB” were from Piquemal et al. (2005), “MB” were developed by Lagercrantz et al. (1993). “BN” were developed from *B. napus* genomic sequences (Kresovich et al. 1995; Szewc-McFadden et al. 1996) and “BRMS” markers are derived from microsatellite-enriched genomic library and obtained from the National Institute of Vegetable and Tea Science, Japan (Suwabe et al. 2002). The markers prefixed by “sN”, “sR” and “sO” are developed by Agriculture and Agri-Food, Canada (http://brassica.agr.gc.ca/index_e.shtml), “nia” and “cnu” are from the Korea (Choi et al. 2007) and the FITO markers were obtained at http://www.osbornlab.agronomy.wisc.edu/research/maps/ssrs.html. PCR amplifications were performed in a volume of 10 µl containing 50 ng genomic DNA, 1× Taq buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM each primer and 0.25 U *Taq* DNA polymerase (Fermentas). The reaction mixture was initially denatured at 94°C for 5 min, followed by 40 cycles of amplification at 94°C for 30 s, 56°C for 60 s and 72°C for 45 s, and a final extension at 72°C for 5 min in a GeneAmp PCR system 9700 (Applied Biosystems). The PCR products were separated on 6% denaturing polyacrylamide gels. After electrophoresis, the gels were stained as previously described (Sanguinetti et al. 1994).

Linkage analysis and map construction

For map construction, the segregation was analyzed in the DH population for all SSR markers showing polymorphisms between the parental lines No. 2127 and ZY821. For markers that detected more than one polymorphic locus, the marker name is followed by a letter to distinguish the loci. For instance, the two loci detected by BnGMS147 were designated as BnGMS147a and BnGMS147b. Linkage analysis was performed using JoinMap3.0 (Van Ooijen and Voorrips 2001). The threshold for goodness-of-fit was set to ≤5.0, a recombination frequency of <0.4 and minimum logarithm of odds scores of 2.0. All genetic distances were expressed in centimorgan (cM) as derived by the Kosambi function (Kosambi 1944). Segregation of each marker in the DH population was analyzed by a chi-square test for “goodness-of-fit” to a 1:1 ratio. In order to assign the BnGMS markers to specific LGs and compare with published linkage maps, 164 public SSRs were used as anchor markers.

Results

Frequency and length variation of microsatellites in *B. napus*

To facilitate the development of new microsatellite markers for oilseed rape genome research and breeding, a total of 13,794 GSSs available in the GenBank database as of February 2007, representing a total length of 10.2 Mb of the *B. napus* genome, were searched for SSR-containing sequences (Rozen and Skaletsky 2000). A total of 2,113 GSSs containing 2,578 SSRs with repeat lengths of at least 12 nucleotides were identified, which represented a frequency of 15.7% of the *B. napus* GSSs. The frequency of microsatellites in these elements is one SSR every 4.0 kb. Of the total SSRs identified, the di- and trinucleotide repeat motifs are the most abundant repeat types and have

frequencies of 34 and 33%, respectively. The pentanucleotide repeats also occurred at a high frequency (19.2%) in the *B. napus* genome. The tetranucleotide repeat motifs had the lowest frequency (13.7%). Table 1 lists the number of the major SSR types identified from the *B. napus* GSSs. In dinucleotide repeats (DNRs), the most abundant repeat motif was $(AT)_n$ (20%), followed by $(AG)_n$ (12%) and $(AC)_n$ (2%). No $(GC)_n$ repeat SSRs were identified in the GSSs. All ten possible types of trinucleotide repeats (TNRs) occurred in the GSS–SSRs. Among the TNRs, $(AAG)_n$ motif was the most common (10.2%), followed by the $(AAT)_n$ (5.7%) and $(AAC)_n$ (5.1%) motifs. The GC-rich TNRs including $(ACG)_n$, $(AGC)_n$ and $(CCG)_n$ were the least abundant (Table 1). Among tetra- and pentanucleotide repeats in *B. napus*, $(AAAN)_n$ and $(AAAAN)_n$, the $(AAAT)_n$ and $(AAAAT)_n$, were especially more common than other combinations.

Development of SSR marker and detection of polymorphisms

In many organisms empirical data indicated that microsatellites with longer repeat lengths are highly polymorphic (Temnykh et al. 2001; Yi et al. 2006) and those having shorter repeat lengths tend to be less variable. From the 2,113 SSR-containing GSSs, all microsatellites having a repeat length ≥16 bp for di-, ≥18 for tri- and ≥20 for tetra- and pentanucleotides were selected for SSR marker development. The selected GSS–SSRs were assembled into contigs to remove redundant microsatellites, the results of which were 627 unique GSS–SSRs, designated as "BnGMS" markers hereafter. The majority of these BnGMS markers were from BESs. Only 48 markers were designed from the methylation filtered or unfiltered genomic library. None of these SSR markers were found in the sequences derived from RFLP probes. The BnGMS markers included 357 (56.9%) di-, 112 (17.8%) tri-, 28 (4.5%) tetra-, 48 (7.7%) pentanucleotide repeats and 82 composite repeat microsatellites (Table 2). The composite microsatellites either contain the same repeat motif that is interrupted by a short nonrepetitive sequence or consist of more than two repeat motifs.

All these 627 GSS–SSR markers were evaluated for successful PCR amplification and polymorphism by testing the genomic DNA of six oilseed rape cultivars. Among these markers, 591 (94.3%) successfully amplified at least one

Table 1 The distribution of the major SSR types identified from the GSS of *B. napus*

Motif	Number (%)	Range (bp)	Total length (bp)	Average length (bp)
Dinucleotide	871 (33.79)	12–178	16,795	19.28
AT	504 (19.55)	12–62	10,198	20.23
AG	305 (11.83)	12–178	5,619	18.42
AC	62 (2.40)	12–36	978	15.77
Trinucleotide	859 (33.32)	12–177	13,090	15.24
AAG	264 (10.24)	12–65	4,058	15.37
AAT	147 (5.70)	12–56	2,214	15.06
AAC	132 (5.12)	12–177	2,152	16.3
ATC	114 (4.42)	12–29	1,685	14.78
AGG	73 (2.83)	12–29	1,103	15.11
ACC	47 (1.82)	12–34	688	14.64
AGC	32 (1.24)	12–19	459	14.34
CCG	22 (0.85)	12–19	299	13.59
ACG	18 (0.70)	12–26	270	15
ACT	10 (0.39)	12–30	162	16.2
Tetranucleotide	352 (13.65)	13–174	5,488	15.59
AAAT	115 (4.46)	13–31	1,730	15.04
AAAG	42 (1.63)	13–30	634	15.1
AAAC	39 (1.51)	13–19	583	14.95
Others	156 (6.05)	13–174	2,541	16.29
Pentanucleotide	496 (19.24)	14–39	7,942	16.01
AAAAT	84 (3.26)	14–25	1,299	15.46
AAATT	75 (2.91)	14–26	1,119	14.92
Others	337 (13.07)	14–39	5,524	16.39
Total	2,578 (100)	12–177	43,315	16.8

Table 2 Characteristics of *B. napus* GSS–SSRs and efficiency of marker development

Motifs	Primers designed	Amplified primers[a] (%)	Polymorphic primers[b] (%)
Dinucleotide	357	333 (93.3)	267 (80.2)
AT	231	217 (93.9)	177 (81.6)
AG	114	107 (93.9)	81 (75.7)
AC	12	9 (75.0)	8 (88.9)
Trinucleotide	112	108 (96.4)	57 (52.8)
AAG	33	32 (97.0)	19 (59.4)
GC-rich TNRs	21	21 (100)	13 (61.9)
AAC	20	18 (90.0)	3 (16.7)
AAT	17	16 (94.1)	8 (50.0)
ATG	17	17 (100)	11 (64.7)
ACT	4	4 (100)	3 (75.0)
Tetranucleotide	28	28 (100)	23 (82.1)
Pentanucleotide	48	48 (100)	29 (60.4)
Composite	82	74 (90.2)	58 (78.4)
Total	627	591 (94.3)	434 (73.4)

[a] Percentage of successfully amplified GSS–SSRs per designed primer pair

[b] Percentage of polymorphic markers per amplified primer pair

fragment from the *B. napus* genome. A majority of the primer pairs amplified one or two main fragments, which might correspond to the amplification from either the A or C or both genomes (Plieske and Struss 2001). Among the successfully amplified markers, 434 primer pairs (73.4%) showed polymorphism among the six cultivars. There were 206, 205 and 238 polymorphic markers between parents S1 and S2, M201 and M202, and No. 2027 and ZY821, respectively. Eighty-four markers were polymorphic across the parents of all populations. The resynthesized *B. napus* line No. 2127 usually has band patterns distinct from the other five lines, indicating that No. 2127 is a desirable line to generate populations for the construction of high-density linkage maps and gene mapping. Information about the new markers is listed in Supplementary Table 1, which included the GenBank ID of the SSR-containing sequence, source library, SSR motif, primer sequences and corresponding T_m values, scorability of bands, and polymorphism survey results.

The efficiency of marker development was examined for each repeat motif. The success rate of PCR amplification and the level of polymorphism of the GSS–SSR markers for each SSR motif are listed in Table 2. The average success rate of PCR amplification was 94.3%. The tetra- and pentanucleotide motifs had the highest success rate (100%) of PCR amplification, followed by tri- (96.4%), di- (93.3%) and composite (90.2%) nucleotide repeats. All repeat motifs had success rates higher than 90%, except for the $(AC)_n$ repeat motif, which had a 75% success rate. BnGMS markers with tetra- (82.1%) and dinucleotides (80.2%) had the highest levels of polymorphism, followed by composite (78.4%), penta- (60.4%) and trinucleotide repeats (52.8%). The $(AC)_n$ motif had the highest level of polymorphism (88.9%), while the $(AAC)_n$ repeat motif had the lowest level of polymorphism (16.7%). The other motifs had polymorphism levels in excess of 50%. The relationships among repeat lengths, repeat numbers and polymorphism levels are shown in Fig. 1. There was no correlation between polymorphism level and repeat length of SSR markers in *B. napus* ($r = 0.21$ and $r = 0.41$, $P < 0.01$, respectively).

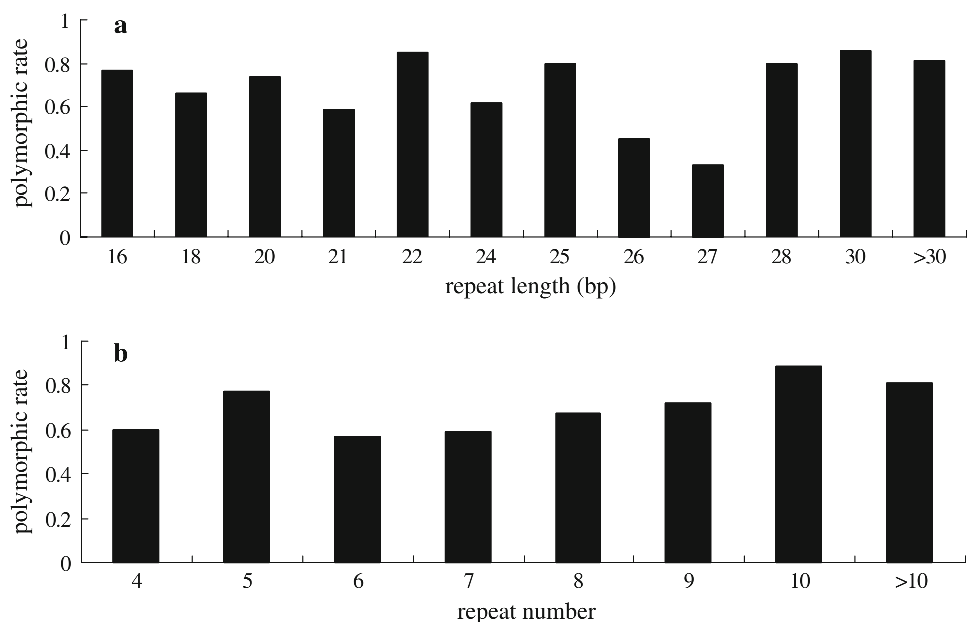


Fig. 1 Relationship between SSR length and polymorphism rate (**a**), and between repeat number and polymorphism rate (**b**). Polymorphism rate was calculated as polymorphic markers per amplified primer

Construction of the linkage map

One hundred and fifty-three GSS–SSRs that showed polymorphism between the parental lines No. 2027 and ZY821 were used to construct a SSR map. Among these, 136 new SSR markers, corresponding to 163 BnGMS loci, could be assigned to specific LGs and 17 failed to be integrated onto the map. From previously published linkage maps, 164 polymorphic SSRs were selected to serve as anchor markers on the basis of even chromosome distribution to ensure full coverage of the linkage map. The linkage map consisting of 19 LGs was constructed with 163 BnGMS loci and 196 anchor loci (Fig. 2). The linkage map had a total length of 1821.33 cM with an average interval of 5.07 cM between adjacent loci (Fig. 2). The 19 LGs were designated as A1–A10 and C1–C9 based on multiple anchor markers located on each chromosome. The marker order in our map was in perfect agreement with previously published linkage maps (Lowe et al. 2004; Chen et al. 2007b; Piquemal et al. 2005; Suwabe et al. 2008).

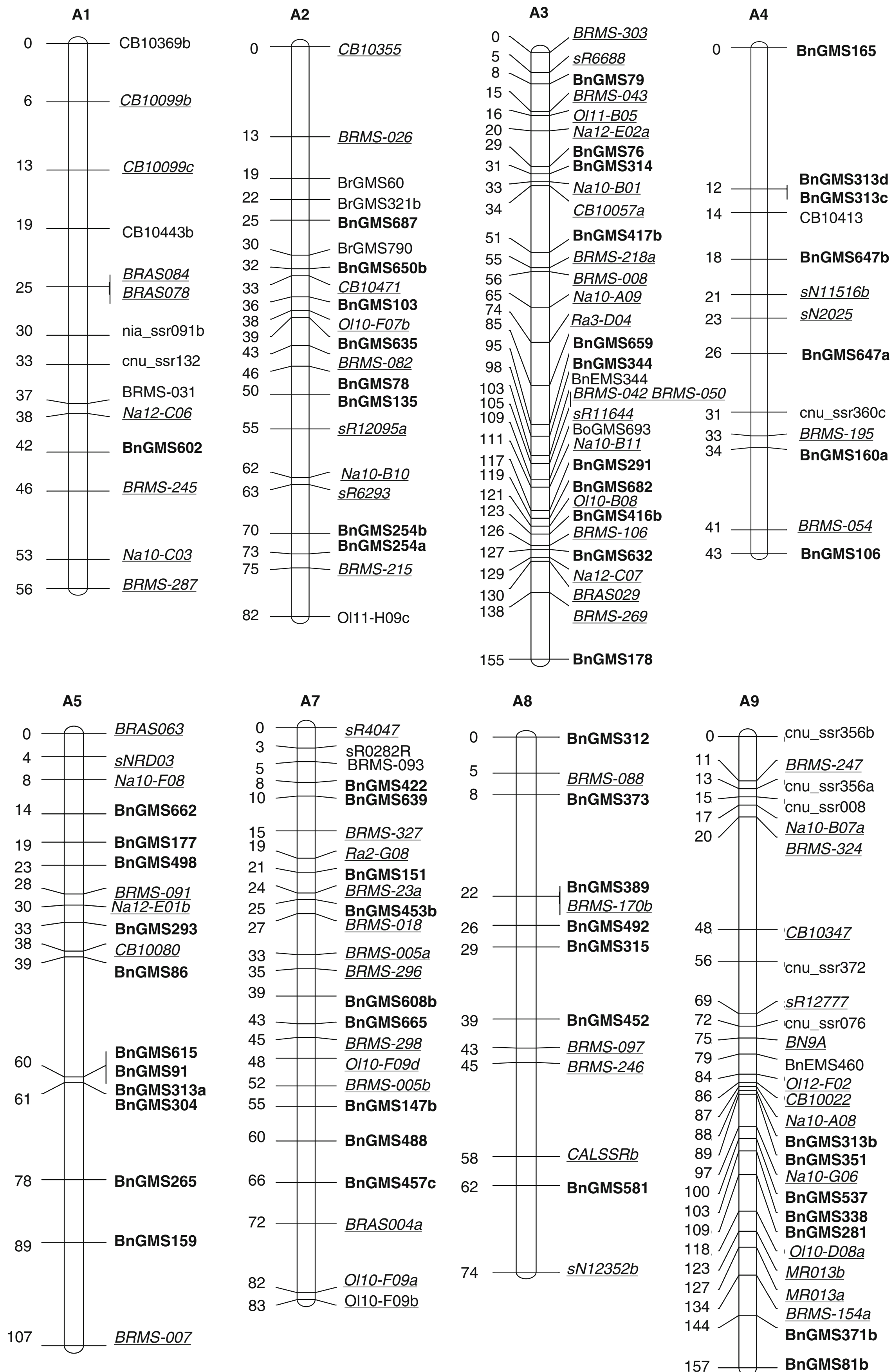
Of the mapped BnGMS loci, 75 and 88 loci were located on the A and C genomes, respectively, indicating that there was no genome bias for these GSS–SSR markers. The BnGMS loci were evenly distributed on most of the rapeseed LGs, especially for A3, A5, C4 and C9 (Fig. 2). LGs C3 and C9 had 17 and 18 GSS–SSR markers, respectively. A1 and C5 only contained one and two GSS–SSR markers, respectively. The other LGs had 4–14 GSS–SSR markers. The GSS–SSR markers greatly extended the length of C9 in the linkage map published by Piquemal et al. (2005). Most of the markers segregated with the expected 1:1 Mendelian ratio in the DH population. However, 124 (34.54%) markers deviated significantly ($P < 0.05$) from this ratio. Of these, 59 loci showed distorted segregation skewed towards the ZY821. Loci with skewed segregation tended to cluster on A2, A3, A4, C3, C4 and C9.

Fig. 2 An SSR-based *B. napus* linkage map. The map was constructed from 88 DH plants derived from a resynthesized *B. napus* line No. 2127 and ZY821 using 136 new GSS–SSR markers named "BnGMS" (*boldface*) with 164 public SSRs as anchors (*italics and underlined*) (Chen et al. 2007b; Piquemal et al. 2005; Suwabe et al. 2008). Positions of loci are given in cM. The 19 LGs were designated as A1–A10 and C1–C9 based on multiple anchor markers located on each chromosome

Discussion

In this study, we analyzed the frequency and length variation of microsatellites identified from a total length of 10.2 Mb of GSSs, which represents a random sampling of the 1132 Mb *B. napus* genome (Johnston et al. 2005). The frequency of SSR occurrence (one every 4.0 kb) in the *B. napus* GSSs is similar to that reported in the BESs of *B. rapa*, one of the diploid progenitors of *B. napus* (Hong et al. 2007), in the genome shotgun sequences of *B. oleracea*, the other diploid progenitor (Iniguez-Luy et al. 2008), as well as in the genome of rice (Temnykh et al. 2001). This frequency is much higher than the estimation based on SSR-containing clones obtained using traditional hybridization methods (Kresovich et al. 1995; Uzunova and Ecke 1999). Among the dinucleotide repeats, $(AT)_n$ is the most frequent repeat motif, followed by $(AG)_n$ and $(AC)_n$ repeats. This distribution is in good agreement with the patterns observed in *B. rapa* (Hong et al. 2007), *Arabidopsis* (Katti et al. 2001) and rice (Temnykh et al. 2001). But it is different from the observations in humans and *Drosophila*, in which $(AC)_n$ are the most frequent dinucleotide repeat sequence, followed by $(AT)_n$ and $(AG)_n$ (Katti et al. 2001). $(GC)_n$ repeats are extremely rare in eukaryotic genomes and this is also the case for *B. napus* (Katti et al. 2001). The distribution of trinucleotide repeats in *B. napus* GSS–SSRs is consistent with that in *B. rapa*, *A. thaliana* and *Oryza sativa* (Hong et al. 2007; Lawson and Zhang 2006), with $(AAG)_n$ being the most frequent repeat motif and GC-rich motifs being very rare (Table 2). In tetra- and pentanucleotide repeats, $(AAAN)_n$ and $(AAAAN)_n$, especially $(AAAT)_n$ and $(AAAAT)_n$, are more common than other combinations, which is consistent with the pattern observed in other plant genomes (Hong et al. 2007). These data suggest that the SSRs in the *B. napus* genome tend to skew toward AT rich motifs.

The relationships among repeat motifs, repeat lengths and polymorphic rates are very important for the selection of SSRs for marker development. Most of the current SSR markers in *Brassica* species were developed from SSR-containing clones (Lowe et al. 2004; Saal et al. 2001; Suwabe et al. 2002; Szewc-McFadden et al. 1996). These markers usually have repeat length ≥20 bp and show good PCR amplification and a high level of polymorphism (Lowe et al. 2004; Saal et al. 2001; Szewc-McFadden et al. 1996). In this study, two-thirds of the BnGMS markers were empirically designed from SSRs with repeat length ≥20 bp. In addition, one-third of the BnGMS markers were intentionally designed from SSRs having repeat length ≥16 and <20 bp to test their polymorphism levels. The results indicated that there were no significant correlations between the type of repeat motifs, the length and number of repeats and the polymorphism levels in *B. napus*. Similar results have also been observed in *Brassica* species (Iniguez-Luy et al. 2008; Plieske and Struss 2001; Suwabe et al. 2004; Szewc-McFadden et al. 1996), soybean (Shultz et al. 2007) and *Medicago truncatula* (Mun et al. 2006). Indeed, the SSRs with a repeat length of 11–20 nucleotides detected a even higher level



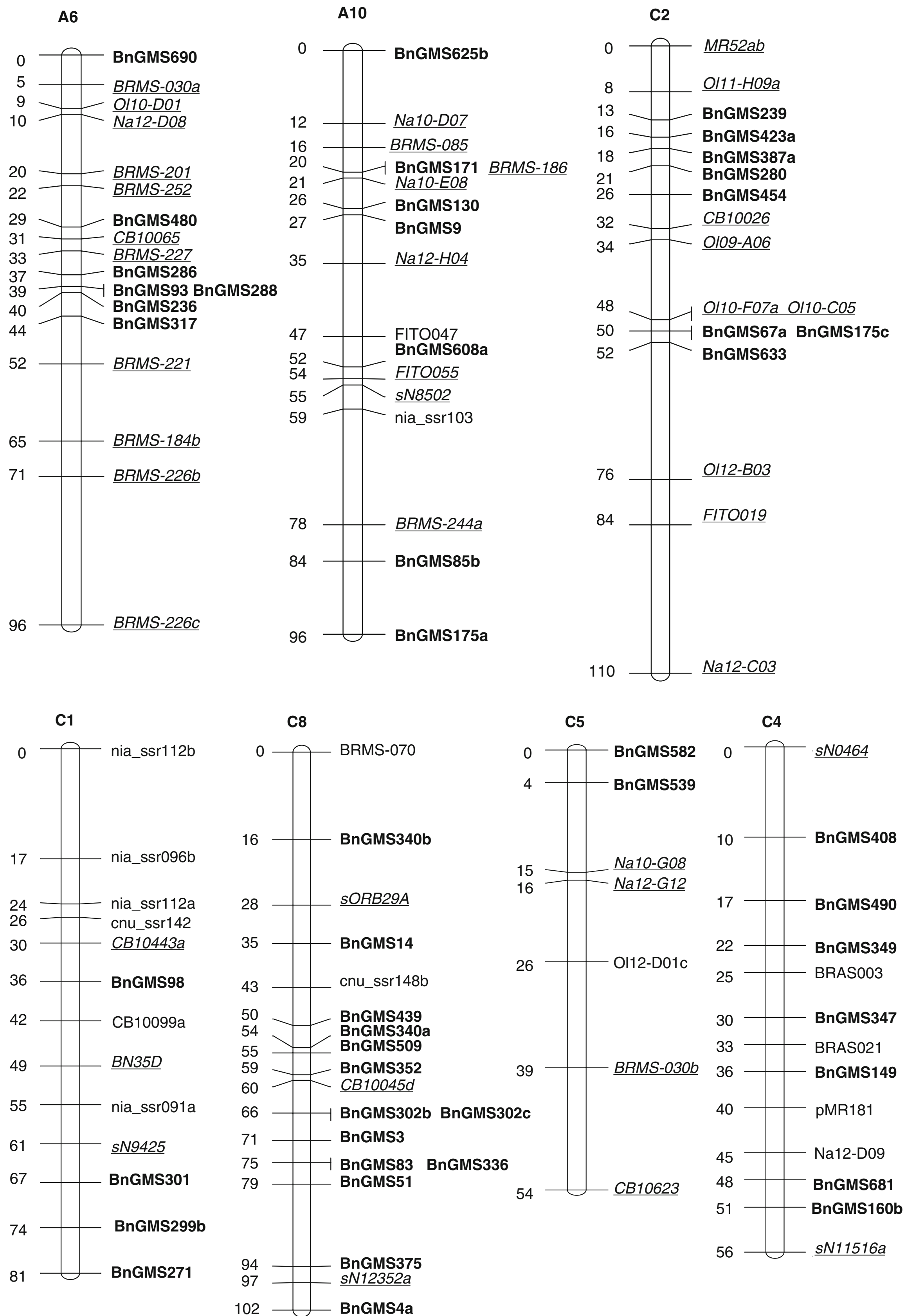


Fig. 2 continued

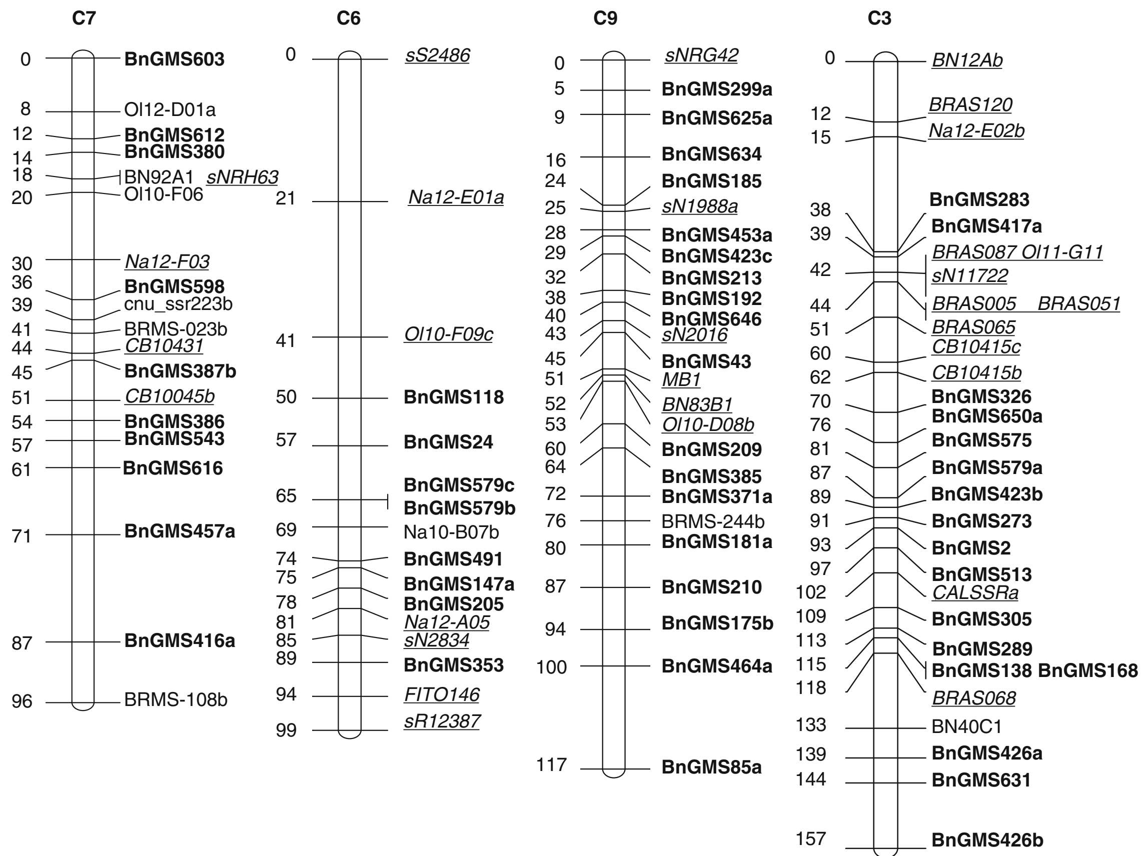


Fig. 2 continued

of polymorphism than SSRs with longer repeats in *Brassica* (Iniguez-Luy et al. 2008; Shultz et al. 2007). In contrast, the level of polymorphism of SSRs often increases with increasing SSR length and number of repeat units in maize, pepper and rice (Sharopova et al. 2002; Temnykh et al. 2001; Yi et al. 2006). However, the BnGMS markers do not include those having a repeat length of <16 and ≥12 bp, which are the most abundant in all organisms (Katti et al. 2001; La Rota et al. 2005; Temnykh et al. 2001). Thus, it is worthwhile to design more primers from those shorter SSRs to further test the relationship between the repeat length and the polymorphic rate in *Brassica*. If the relationship is true, the shorter SSRs that represent a large portion of the *B. napus* GSS–SSRs would be a rich potential source to develop a great number of polymorphic SSR markers for genetic analysis and molecular breeding.

Microsatellite markers developed from the A genome of *B. rapa* and the C genome of *B. oleracea*, the two diploid progenitors of *B. napus*, tend to only amplify a product in the original genome from which the markers were derived, rather than the other genome in *B. napus* (Lowe et al. 2004). These markers are usually mapped to the original genome, therefore are not suitable for the construction of a genetic linkage map covering all LGs of *B. napus*. For example, Suwabe et al. (2008) integrated the *Brassica* A genetic map with *B. napus* and *B. rapa* using 46 polymorphic BRMS markers derived from the A genome of *B. rapa*. Among these markers, 33 were mapped to the A1–A10 LGs, while only 13 mapped to the C1–C9 LGs (Suwabe et al. 2008). In contrast, the BnGMS markers are evenly distributed in both A and C genomes and do not skew to any genome, indicating SSR markers derived from BESs would be very useful for the construction of a saturated genetic linkage map (Chen et al. 2007a; Shoemaker et al. 2008). The mapping of the BES–SSR markers located their source BAC clones on *B. napus* chromosomes, which will facilitate the map-based cloning of tightly linked genes, comparative genome analysis and the integration of physical maps with genetic maps for future chromosome-based genome sequencing in *B. napus* (Choi et al. 2007). The genetic linkage map constructed in this study integrated the BnGMS markers with the SSR markers from all the existing SSR linkage maps, and will be very useful for their application in gene mapping and marker-assisted selection.

Acknowledgments This research was supported by the National Key Basic Research and Development Plan of China (2006CB101600) and the National Science Foundation of China (30623012). We thank M. Mazourek for critical review of this manuscript.

References

Ayele M, Haas BJ, Kumar N, Wu H, Xiao YL, 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

Batley J, Hopkins CJ, Cogan NOI, Hand M, Jewell E, Kaur J, Kaur S, Li X, Ling AE, Love C, Mountford H, Todorovic M, Vardy M, Walkiewicz M, Spangenberg GC, Edwards D (2007a) Identification and characterization of simple sequence repeat markers from *Brassica napus* expressed sequences. Mol Ecol Notes 7:886–889

Batley J, Jewell E, Edwards D (2007b) Automated discovery of single nucleotide polymorphism and simple sequence repeat molecular genetic markers. Methods Mol Biol 406:473–494

Chen BY, Heneen WK, Jönsson R (1988) Resynthesis of *Brassies napus* L. through interspecific hybridization between *B. alboglabra Bailey* and *B. campestris* L. with special emphasis on seed colour. Plant Breed 101:52–59

Chen C, Yu Q, Hou S, Li Y, Eustice M, Skelton RL, Veatch O, Herdes RE, Diebold L, Saw J, Feng Y, Qian W, Bynum L, Wang L, Moore PH, Paull RE, Alam M, Ming R (2007a) Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in *Brassicales*. Genetics 177:2481–2491

Chen W, Zhang Y, Liu X, Chen B, Tu J, Tingdong F (2007b) Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F_2 populations. Theor Appl Genet 115:849–858

Choi S, Teakle G, Plaha P, Kim J, Allender C, Beynon E, Piao Z, Soengas P, Han T, King G, Barker G, Hand P, Lydiate D, Batley J, Edwards D, Koo D, Bang J, Park B-S, Lim Y (2007) The reference genetic linkage map for the multinational *Brassica rapa* genome sequencing project. Theor Appl Genet 115:777–792

Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B, Specel S, Hanneton L, Dheu JE, Deschamps M, Margale E, Vincourt P, Renard M (2006) Genetic control of oil content in oilseed rape (*Brassica napus* L.). Theor Appl Genet 113:1331–1345

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Hong CP, Plaha P, Koo DH, Yang TJ, Choi SR, Lee YK, Uhm T, Bang JW, Edwards D, Bancrofts I, Park BS, Lee J, Lim YP (2006) A survey of the *Brassica rapa* genome by BAC-end sequence analysis and comparison with *Arabidopsis thaliana*. Mol Cells 22:300–307

Hong CP, Piao ZY, Kang TW, Batley J, Yang TJ, Hur YK, Bhak J, Park BS, Edwards D, Lim YP (2007) Genomic distribution of simple sequence repeats in *Brassica rapa*. Mol Cells 23:349–356

Hong CP, Kwon SJ, Yang TJ, Park BS, Lim YP (2008) Progress in understanding and sequencing the genome of *Brassica rapa*. Int J Plant Genomics. doi:10.1155/2008/582837

Hopkins CJ, Cogan NOI, Hand M, Jewell E, Kaur J, Li X, Lim GAC, Ling AE, Love C, Mountford H, Todorovic M, Vardy M, Spangenberg GC, Edwards D, Batley J (2007) Sixteen new simple sequence repeat markers from *Brassica juncea* expressed sequences and their cross-species amplification. Mol Ecol Notes 7:697–700

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Iniguez-Luy F, Voort A, Osborn T (2008) Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. Theor Appl Genet 117:977–985

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ (2005) Evolution of genome size in *Brassicaceae*. Ann Bot 95:229–235

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18:1161–1167

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterisation 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 (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228

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

Lan T-H, DelMonte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S, Paterson AH (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res 10:776–788

La Rota M, Kantety R, Yu J-K, Sorrells M (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23

Lawson M, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7:R14

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

Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwang I, Cho MJ (1996) Expressed sequence tags of Chinese cabbage flower bud cDNA. Plant Physiol 111:577–588

Lim JY, Shim CS, Chung EJ, Kim JS, Kim HU, Oh SJ, Choi WB, Ryou CS, Kim JB, Kwon MS, Chung TY, Song SI, Kim JK, Nahm BH, Hwang YS, Eun MY, Lee JS, Cheong JJ, Choi YD (2000) Analysis of expressed sequence tags from *Brassica rapa* L. ssp. *pekinensis*. Mol Cells 10:399–404

Ling AE, Kaur J, Burgess B, Hand M, Hopkins CJ, Li X, Love CG, Vardy M, Walkiewicz M, Spangenberg G, Edwards D, Batley J (2007) Characterization of simple sequence repeat markers derived in *silico* from *Brassica rapa* bacterial artificial chromosome sequences and their application in *Brassica napus*. Mol Ecol Notes 7:273–277

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

Long Y, Shi J, Qiu D, Li R, Zhang C, Wang J, Hou J, Zhao J, Shi L, Park B-S, Choi SR, Lim YP, Meng J (2007) Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis*. Genetics 177:2433–2444

Lowe AJ, Jones AE, Raybould AF, Trick M, Moule CL, 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

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

McCouch SR, Teytelman L, Xu YB, Lobos KB, Clare K, Walton M, Fu BY, Maghirang R, Li ZK, Xing YZ, Zhang QF, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Mun JH, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez AM, Kiss GB, Roe B, Young ND, Cook D (2006) Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. Genetics 172:2541–2555

Padmaja KL, Arumugam N, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK (2005) Mapping and tagging of seed coat colour and the identification of microsatellite markers for marker-assisted manipulation of the trait in *Brassica juncea*. Theor Appl Genet 111:8–14

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

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor Appl Genet 111:1514–1523

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

Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R, Zhuang X, Wang Y, Tan X, Dietrich E, Weihmann T, Everett C, Vanstraelen S, Beckett P, Fraser F, Trick M, Barnes S, Wilmer J, Schmidt R, Li J, Li D, Meng J, Bancroft I (2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. Theor Appl Genet 114:67–80

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

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

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

Saito M, Kubo N, Matsumoto S, Suwabe K, Tsukada M, Hirai M (2006) Fine mapping of the clubroot resistance gene, *Crr3*, in *Brassica rapa*. Theor Appl Genet 114:81–91

Sanguinetti CJ, Dias Neto E, Simpson AJ (1994) Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. Biotechniques 17:914–921

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

Shoemaker RC, Grant D, Olson T, Warren WC, Wing R, Yu Y, Kim H, Cregan P, Joseph B, Futrell-Griggs M, Will N, Davito J, Walker J, Wallis J, Kremitski C, Scheer D, Clifton SW, Graves T, Nguyen H, Wu X, Luo M, Dvorak J, Nelson R, Cannon S, Tomkins J, Schmutz J, Stacey G, Jackson S (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. Genome 51:294–302

Shultz J, Kazi S, Bashir R, Afzal J, Lightfoot D (2007) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor Appl Genet 114:1081–1090

Song Q, Shi J, Singh S, Fickus E, Costa J, Lewis J, Gill B, Ward R, Cregan P (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

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

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

Suwabe K, Morgan C, Bancroft I (2008) Integration of *Brassica* A genome genetic linkage map between *Brassica napus* and *B. rapa*. Genome 51:169–176

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

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

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

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

Van Ooijen JW, Voorrips RE (2001) JoinMap®3.0. software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Xiao SS, Xu JS, Li Y, Zhang L, Shi SJ, Shi SW, Wu JS, Liu KD (2007) Generation and mapping of SCAR and CAPS markers linked to the seed coat color gene in *Brassica napus* using a genome-walking technique. Genome 50:611–618

Yi G, Lee J, Lee S, Choi D, Kim B-D (2006) Exploitation of pepper EST–SSRs and an SSR-based linkage map. Theor Appl Genet 114:113–130

Zhao JY, Becker HC, Zhang DQ, Zhang YF, Ecke W (2006) Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor Appl Genet 113:33–38