Theor Appl Genet (2009) 119:1301–1311
DOI 10.1007/s00122-009-1135-x

ORIGINAL PAPER

Development and evaluation of single-nucleotide polymorphism markers in allotetraploid rapeseed (*Brassica napus* L.)

Peter Westermeier · Gerhard Wenzel · Volker Mohler

Received: 15 December 2008 / Accepted: 12 August 2009 / Published online: 16 September 2009

Abstract Single-nucleotide polymorphisms (SNPs) and insertion–deletions (INDELs) are currently the important classes of genetic markers for major crop species. In this study, methods for developing SNP markers in rapeseed (*Brassica napus* L.) and their in silico mapping and use for genotyping are demonstrated. For the development of SNP and INDEL markers, 181 fragments from 121 different gene sequences spanning 86 kb were examined. A combination of different selection methods (genome-specific amplification, hetero-duplex analysis and sequence analysis) allowed the detection of 18 singular fragments that showed a total of 87 SNPs and 6 INDELs between 6 different rapeseed varieties. The average frequency of sequence polymorphism was estimated to be one SNP every 247 bp and one INDEL every 3,583 bp. Most SNPs and INDELs were found in non-coding regions. Polymorphism information content values for SNP markers ranged between 0.02 and 0.50 in a set of 86 varieties. Using comparative genetics data for *B. napus* and *Arabidopsis thaliana*, an allocation of SNP markers to linkage groups in rapeseed was achieved: a unique location was determined for seven gene sequences; two and three possible locations were found for six and four sequences, respectively. The results demonstrate the usefulness of existing genomic resources for SNP discovery in rapeseed.

Communicated by A. Schulman.

This study is dedicated to the memory of Gerhard Schwarz.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1135-x) contains supplementary material, which is available to authorized users.

P. Westermeier (✉) · G. Wenzel · V. Mohler
Plant Breeding, Department of Plant Sciences, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Am Hochanger 4, 85350 Freising, Germany
e-mail: peter.westermeier@wzw.tum.de

Present Address:
G. Wenzel
Dean's Office, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Alte Akademie 8, 85350 Freising, Germany

Present Address:
V. Mohler
Institute for Crop Science and Plant Breeding, Bavarian State Research Center for Agriculture, Am Gereuth 8, 85354 Freising, Germany

Introduction

Oilseed rape (*Brassica napus* L.) is a recent allopolyploid species containing the two homologous but divergent genomes from turnip rape (*Brassica rapa* L.; A genome) and cabbage (*Brassica oleracea* L.; C genome) (U N 1935). This inter-specific hybridisation may have occurred less than 10,000 years ago (Rana et al. 2004). Rapeseed is one of the most important crops for oil production in the temperate climate zone. When breeding process started in the end of nineteenth century, the use of rapeseed oil was limited due to a high content of erucic acid (Charlton et al. 1975). In addition, the protein-rich residues of oil production were also not useable for feed purposes due to the presence of high contents of glucosinolates (Lewerenz et al. 1988). With the improvement of oilseed rape by reducing these undesirable components, the growing area in Germany increased up to 1.40 Mio ha in 2008 (Anonymous 2008). The breeding of modern oilseed rape varieties is increasingly supported by genomics tools through more effectively identifying and characterising genetic

variation from germplasm resources (Halldén et al. 1994; Diers et al. 1996; Becker et al. 1995; Seyis et al. 2003; Hasan et al. 2006) and surveying the transmission of target traits in breeding populations (Somers et al. 2001; Hu et al. 2006; Lei et al. 2007; Hasan et al. 2008).

Single-nucleotide polymorphisms (SNPs), the most basic genetic variation and most common class of DNA markers, and small insertions–deletions (INDELs) are numerously available only for a small fraction of cultivated plants such as rice (Nasu et al. 2002; Feltus et al. 2004; Monna et al. 2006) and maize (Bhattramakki et al. 2002; Batley et al. 2003; Bi et al. 2006). As for other crops, SNPs are not yet efficiently exploited for rapeseed. This situation has to be changed because SNPs are the basis for the development of most functional markers and the prerequisite to tap the full potential of association mapping. Since the latter application is carried out in diverse sample populations that generally show much lower level of linkage disequilibrium and thus much higher level of resolution than bi-parental mapping populations many polymorphisms are required to detect genotype–phenotype associations in the first place (Burke et al. 2007).

SNP discovery usually is achieved by in silico screening of comprehensive expressed sequence tag collections and re-sequencing of selected candidate genes from representative individuals of a gene pool (e.g. Somers et al. 2003; Cordeiro et al. 2006; Kota et al. 2008). Maize and rice are diploid crop species and, therefore, raise much less difficulties in developing SNP markers than polyploid species like bread wheat or rapeseed. An attempt to overcome this problem is to avoid the amplification of multiple fragments caused by orthologous and paralogous gene copies. Therefore, special strategies are needed for developing locus-specific SNP markers in polyploid species such as rapeseed.

The objectives of this study were (1) to develop a strategy for SNP discovery in complex genomes such as rapeseed, (2) to estimate SNP and INDEL frequencies in the rapeseed genome, (3) to assess the distribution of these markers in German rapeseed varieties, and (4) to map SNP/INDEL markers in the rapeseed genome by cross referencing to the *Arabidopsis thaliana* genome (in silico mapping).

Materials and methods

Plant material

Eighty conventional and six genetically modified German oilseed rape varieties including some old varieties that played an important role in German rapeseed breeding history were investigated in this study (Supplementary Table 1). Seeds were kindly provided by IPK Gatersleben (Germany), breeding companies mentioned in Supplementary Table 1 and the Bavarian State Research Center for Agriculture (Freising, Germany).

DNA extraction

DNA was isolated according to a modified CTAB protocol (Saghai-Maroof et al. 1984). 15–20 mg leaf tissue was homogenised in 1 ml CTAB extraction buffer [150 mM Tris (pH 7.5), 1.05 M NaCl, 15 mM EDTA (pH 8.0), 1.5% CTAB, 1.4% 2-mercaptoethanol] with a MM2 ball mill (Retsch, Haan, Germany). The samples were incubated at 60°C for 1 h and extracted two times with 1 vol chloroform/isoamyl alcohol (24:1). Genomic DNA was precipitated with 1 vol isopropyl alcohol and washed with 1 ml ethanol (75%). After drying at room temperature, the DNA was dissolved in 100 µl Tris buffer (10 mM, pH 7.5).

Polymerase chain reaction

Polymerase chain reaction (PCR) was carried out in a Perkin Elmer 9600 thermocycler (Applied Biosystems, Foster City, CA, USA) in a 20-µl reaction volume containing 10 ng genomic DNA, 0.5 U *Taq* DNA polymerase (Qiagen, Hilden, Germany), 200 µM of each dNTP, 0.2 µM of each oligonucleotide and 1.5 mM $MgCl_2$ in 1× PCR buffer (Qiagen). Cycling conditions were as following: 10 min initial denaturation at 95°C, 35 cycles of 30 s at 95°C, 1 min annealing step at primer-specific temperature (see Table 1) and 2 min at 72°C, and terminated by a final extension step at 72°C for 10 min. PCR amplicons were separated in 1.5% agarose gels and visualised on an UV transilluminator after staining with ethidium bromide. Oligonucleotide sequences are shown in Table 1.

SNP discovery

Oligonucleotides were developed based on genomic gene sequences retrieved from EMBL database. To reduce the probability for amplifying multiple fragments at least one of the two primers covered an intron–exon junction. Each primer pair was tested in a set of six *B. napus* genotypes (cvs. Express, Artus, Aviso, Smart, Talent and Falcon), one *B. oleracea* (cv. Marrow Stem) and *B. rapa* (cv. Buko) genotype each. Non-working primers and those producing multiple fragments in *B. napus* were excluded from further investigations. PCR products that showed polymorphisms between the two genomes of rapeseed were scanned for anonymous SNPs using denaturing high performance liquid chromatography (DHPLC) (Oefner and Underhill 1998). For that, PCR products from a mixed DNA sample

Table 1 Oligonucleotides developed for amplifying SNP/INDEL-containing fragments in *B. napus* (PCR-for/PCR-rev) and for SNP genotyping using primer extension method (PE-for/PE-rev)

EMBL accession	Primer	Sequence 5′ → 3′	Amplicon (bp)/SNP	Annealing (°C)
A62529	PCR-for	ACAACAGCGACCAACTCCAA	620	62.0
	PCR-rev	GATGCTGATTGGCTGTTCGT		
	PE-for	CCGTATCTCGTCCGTGGCTT	A/G	
	PE-rev	CGTCGTAGCTCAAAGCCGG	T/C	
ACGM060	PCR-for[a]	TACAAAGCTTCTCTATGTTGATGTTG	378	50.0
	PCR-rev[a]	TTTTAAAATTGTTAATACGTTTCTTAATG		
	PE-for	GGAAAGTATGGGAGGAAGAGGTC	G/A	
	PE-rev	CCATCCATCATTTCATGATCATTTA	C/T	
AF490462	PCR-for[b]	ACGTTCAGGTCTAGGCGATGA	1007	62.0
	PCR-rev[b]	GCCACCCAAACTGCACTGTTA		
	PE-for	GGGGCCGCTATTTTGCTCT	C/T	
	PE-rev	GTCTACGATCTCCAGGCTTGTTG	G/A	
BNA251988	PCR-for	AATTTATTATGGCTGTAACAGGTGGCT	698	63.0
	PCR-rev	GGCGGGTACTCGAGGAACTTAGA		
	PE-forA[c]	TAATTATGCAGGCGGCAAGG	G/A	
	PE-revA[c]	CATCTTCACGAGCTTGAACCCTAT	C/T	
	PE-forB[c]	GGAAAATGCGGTTACTCCGA	G/A	
	PE-revB[c]	CCAAAATCGACGGTTTACGAAA	C/T	
CD830352	PCR-for	CGGTTTCGAAATCGATCAAGG	378	60.0
	PCR-rev	CTTCTTAGGCAAAAGCACAGAGTTG		
	PE-for	CCGTCCTCGAGTACCTCGC	C/A	
	PE-rev	GAGAGAACGAAGACTTACTTCGGC	G/T	
D13987	PCR-for	GTTCTGTTTTTGTAGAGTTTTGGAA	411	58.0
	PCR-rev	ATAATAGCCACTCTTGTCTCTTGTC		
	PE-for	CGTTGATGAATCTCTGCAAAGAGAG	C/T	
	PE-rev	TGCGGAATGCAGCTTGAATCT	G/A	
D26381	PCR-for	GATAAAGGACTGTCTTGCTCAACTG	401	59.0
	PCR-rev	CCATCCATGAAGATAACTGAATCAA		
	PE-for	CGTTGATGAATCTCTGCAAAGAGAG	A/G[d]	
	PE-rev	TGCGGAATGCAGCTTGAATCT	C/A[d]	
L25404	PCR-for	TGGCTTGTGGAGGTTTAGTGATTT	310	63.0
	PCR-rev	GGGTAGATCTCCTCGTACTTCCTGT		
	PE-for	ACAGTATATCTCATAGACTGGTTCCTCAA	C/T	
	PE-rev	AAGTCTATTTCTTTCCACGTAGTTTCC	G/A	
L31891	PCR-for[a]	ATAAAGTCAAACAAGAATGCAGAAAA	355	57.0
	PCR-rev[a]	CTGCATGCTTACAAAAGTTCACTGT		
	PE-for	TCTTTGTATCCTTAGTTTGTATGCAAGTAA	A/C	
	PE-rev	GGCAATCTTCTTTCTTATGATAATATACTTAAG	T/G	
X55937	PCR-for	CGAGTTGGAGGTAATAAAAACAGAAGTT	324	61.0
	PCR-rev	GGTGGCTTGTAGGCGATGAAAC		
	PE-for	CTCAAGTGTTGAAGGAAGTCCAAGA	A/C	
	PE-rev	CGTTAGGGTACTCCGTTTTGCA	T/G	
X61610	PCR-for	CAGAGCTCTTGAGGTTAGTCTCTAACTC	534	60.0
	PCR-rev	AAGATGGTCAAGAAGATTCTCCAAA		
	PE-forA[c]	TATCCTAATGGGATTAGTTGAAGGC	T/C	
	PE-revA[c]	ATCAAGACCGTTGATGCGGA	A/G	
	PE-forB[c]	TTGAGGTTAGTCTCTAACTCTCACTCACT	C/T[d]	
	PE-revB[c]	TCACCTTGAGCTCAGCAAAAGTAA	G/C[d]	

Table 1 continued

EMBL accession	Primer	Sequence 5′ → 3′	Amplicon (bp)/SNP	Annealing (°C)
X62120	PCR-for1	CATACGTTGTTCAAGGTCGATATC	635	63.0
	PCR-rev1	CCAGACGGAACGTCTACAGAAC		
	PCR-for2	GCATTCTCCAGGGTGTAAGTACATT	309	63.0
	PCR-rev2	ATGTTGTATTTCGGAGCACCAT		
	PE-for1	GGTCAACCATGGCAGGGACA	A/G[e]	
	PE-rev1	CCTGTTGACCTTGTTGTCCCTG	C/T[e]	
	PE-for2	CCAGAATTTTAGTTAAAATATATTTTCTTAAAAC	T/A	
	PE-rev2	TCTTAGTATACATTAACTGTATTTTGATTTCGT	A/T	
X67142	PCR-for	TTCAGCCGTGGCATTAATCG	446	63.0
	PCR-rev	GGAGCTTGTTTAATAATCCCAGGA		
	PE-for	GTAACTAATAAGAATATGGCACAAGGTTT	T/C	
	PE-rev	ACATAAATGTATATATTTCAGATGTTAATTATATAATT	A/G	
X74782	PCR-for	ATAGCTTTACCGACATGTGATACTGCC	557	61.0
	PCR-rev	GATACTTTGCAAATGAAACAAGAGGATAGA		
	PE-for	CAGAAAATAATCTTCAAAATTTTAATTTACA	T/C	
	PE-rev	CAAATGGAAAAATGACGTCATCTATC	A/G	
X77576	PCR-for	TGCTGTTTTAGACAGGTACTTCACA	446	60.0
	PCR-rev	CTCTCAGCTTATCCTGCGTCA		
	PE-for	AAGAAGTTAGTTCGAGCCTGTGTTC	C/T	
	PE-rev	CTCACCACACCCACACCTGC	G/A	
X93461	PCR-for[a]	CTTGTTTCGTTTACTAATATAAGAGATTTGG	501	57.0
	PCR-rev[a]	AAGGAGTAAGCAACAAGTGGACAT		
	PE-for	CATTATGTTTCTGGGTTCTTGCTTG	A/C	
	PE-rev	TAAGATCAAGATCCATAATCGACAA	T/G	
Y12531	PCR-for	GATTCAAGAGAAAGAAAGGTACTAACTTCT	599	60.0
	PCR-rev	TTTCAAACAGAAAACAGGCAACTC		
	PE-for	ATCCATTTTATAAAAAATCAAATCAGAGT	A/C[d]	
	PE-rev	TAATTTTTTTCCTAAAATAACGTTTTTTT	A/G[d]	

[a] Primer development based on published sequence alignments of INRA project GOP T4 (Fourmann et al. 2002)

[b] Primer information published in Xiao and Lu (2005)

[c] Two different fragments from the same gene used

[d] INDEL used for scoring

[e] Different SNPs were used in forward and reverse reactions

(cvs. Express, Artus, Aviso, Smart, Talent and Falcon) were compared to corresponding PCR products from a single genotype (cv. Falcon). This approach considered the power of DHPLC method to detect allele frequencies between 5 and 10% in pooled DNA samples (Bäumler et al. 2003). PCR assays were denatured at 95°C for 3 min and re-annealed over 30 min by decreasing the temperature from 95 to 60°C in order to allow hetero-duplex formation. The samples were run on an automated WAVE® DHPLC instrument (Transgenomic, Omaha, NE, USA) equipped with a DNASep® HPLC column (diameter 4.6 mm, length 50 mm, filled with alkylated nonporous polystyrenedivinylbenzene copolymer microspheres). Melting temperature was calculated for each fragment by Navigator™/WAVE®-MD software, version 1.5.4 (Transgenomic). The mobile phase consisted of two eluents (pH 7.0): buffer A contained 0.1 M triethylammonium acetate with 2 vol.% acetonitrile; buffer B contained 0.1 M triethylammonium acetate with 25 vol.% acetonitrile. Fragments were eluted with a linear acetonitrile gradient calculated with Navigator™/WAVE®-MD software, version 1.5.4 (Transgenomic) depending on the sequence of the fragment to be analysed. Fragments showing either no polymorphism between different genotypes or sequence mixtures were discarded from further investigation. Polymorphisms were confirmed by sequence analysis from both ends on an ABI Prism 377 platform (Applied Biosystems) using the BigDye®-Terminator v1.1 chemistry

(Applied Biosystems) following manufacturers guide. Editing and aligning of DNA sequences was performed with Sequencing Analysis Software 3.2 and Sequence Navigator 1.0.1 (Applied Biosystems).

SNP genotyping

SNP genotyping in 86 rapeseed varieties was performed using a mini sequencing method (primer extension; Sokolov 1990). To remove residual dNTPs and primers PCR products were treated enzymatically with shrimp alkaline phosphatase (SAP) and exonuclease I (EXO I; both Fermentas GmbH, St. Leon-Rot, Germany). 3 µl PCR product were treated with 1 U SAP and 0.4 U EXO I by incubation at 37°C for 1 h, followed by an enzyme inactivation step at 75°C for 15 min. Minisequencing reactions were carried out in a final volume of 2.5 µl containing 1.25 µl SNaPshot™ Multiplex Ready Mix (Applied Biosystems), 0.5 pmol extension primer (designed for an annealing temperature of 60°C; Table 1) and 1 µl cleaned PCR product (0.01–0.1 pmol DNA). Extension was performed on a Perkin Elmer 9600 thermocycler for 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s and extension at 60°C for 30 s. Post-reaction extension products were treated with 0.5 U of SAP by incubation at 37°C for 1 h, followed by 15 min heat treatment at 75°C for SAP inactivation. Electrophoresis of extension products was carried out on an ABI PRISM® 377-96 DNA Sequencer (Applied Biosystems) on a 12 cm 5% Long Ranger® (BMA, Rockland, ME, USA), 6 M urea polyacrylamide gel. GS Run 12E-2400 or 12E-1200 Run Modules were used for performing electrophoresis. 3 µl cleaned extension products were mixed with 3 µl formamide-dextran blue solution and 1 µl of the mixture was loaded. Results were analysed using GeneScan® 3.1 software (Applied Biosystems).

Statistical analysis

Polymorphism information content (PIC) values of SNP markers were calculated according to Botstein et al. (1980) and Anderson et al. (1993):

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_i^2,$$

where k is the total number of alleles and p is the frequency of the ith allele at a given locus.

In silico mapping

Assignment of the current SNP markers to an RFLP map of *B. napus* (Parkin et al. 2005) started with a comparison of every SNP-containing fragment to the *A. thaliana* genome sequence using the BLASTN tool provided by The Arabidopsis Information Resource (www.arabidopsis.org). Corresponding gene IDs in *A. thaliana* were integrated into the data set of Parkin et al. (2005) that, based on the closest *A. thaliana* gene identified, shows the ordered blast results for 368 RFLP probes from rapeseed according to their respective position on the pseudo-chromosome sequence. This allowed targeting the homologous gene sequences, within each chromosome, to a certain block of RFLP markers. Each of these blocks is located in one or more linkage groups of the genetic map of rapeseed (Parkin et al. 2005). In all cases, the closest RFLP markers were used to identify the corresponding linkage blocks and, thus, all possible locations of the SNP-containing gene fragments. Information on genome specificity provided by fragment selection procedure and sequence analysis then allowed to determine the unambiguous or best possible location of SNP markers to linkage groups from either A or C genomes.

Results

SNP discovery

A set of 181 primer pairs was developed from 121 gene sequences present in the EMBL database and covering 86 kb of the rapeseed genome. Where possible, primers were chosen to target more than one fragment of each gene sequence. Of 160 primer pairs that successfully amplified DNA from rapeseed and its diploid ancestors, 134 generated a single PCR product as indicated by agarose gel electrophoresis. From this fraction, 70 markers revealed intergenomic and locus-specific polymorphisms: 45 showed genome-specific amplification, i.e. they were null in either of the 2 ancestor genomes of rapeseed (Fig. 1a), whereas 25 revealed length polymorphisms between the genomes of *B. oleracea* and *B. rapa* (Fig. 1b). For SNP scanning, a total of 67 locus-specific fragments together with 32 of apparently the same size between the ancestor A and C genomes were subjected to DHPLC analysis. The elution profiles of single-genotype and mixed-DNA sample PCR products for each primer pair were compared allowing copy number estimation and SNP discovery. Fifty-four markers failed to display a single peak from single-genotype PCR and, consequently, from the reaction using the pooled DNA sample (Fig. 2a), whereas 28 single-copy sequences did not show polymorphism for the PCR that used the mixture of diverse genotypes as DNA template (Fig. 2b). The remainder of 17 single-copy fragments corresponding to 16 genes showed intra-genomic polymorphisms, as exemplified in Fig. 2c. In addition, a singular fragment from gene accession D26381 that showed

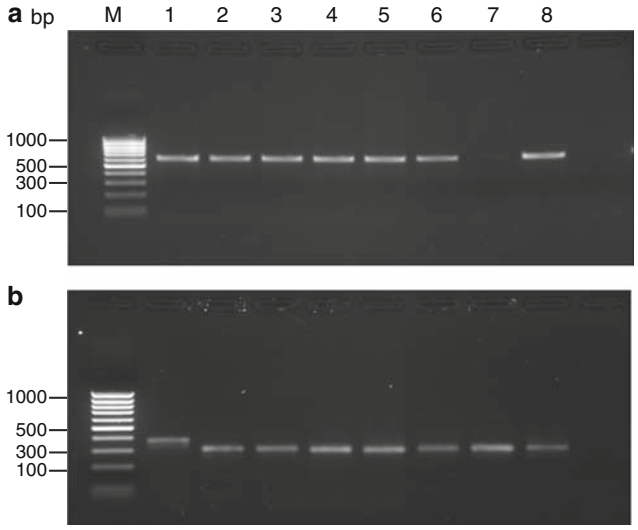

Fig. 1 Genome-specific amplification in *Brassica* species with primers **a** X62120-2 showing a null allele for *B. rapa* and **b** Y12531 showing a length polymorphism between the genomes of *B. oleracea* and *B. rapa*. *M* 100 bp ladder, *1 B. oleracea* cv. 'Marrow Stem', *2 B. napus* cv. 'Express', *3 B. napus* cv. 'Artus', *4 B. napus* cv. 'Aviso', *5 B. napus* cv. 'Smart', *6 B. napus* cv. 'Talent', *7 B. rapa* cv. 'Buko', *8 B. napus* cv. 'Falcon'

intra-genomic polymorphism already on agarose gel was included for further investigation. Altogether, using this strategy an SNP detection yield of 9.9% was achieved. Primers for amplifying SNP/INDEL-containing fragments were designated according to the EMBL accession number of the corresponding genes and are shown in Table 1.

SNP frequency

For SNP confirmation, the selected gene fragments spanning 8.9 kb were re-sequenced in the six genotypes that composed the mixed DNA sample for SNP discovery. Including the 28 single-copy sequences without polymorphisms a total of 21.4 kb of all sequences and individuals yielded 93 nucleotide polymorphisms consisting of 87 SNPs and 6 INDELs. The highest number of polymorphisms (28) was found for the 411-bp fragment from gene accession D13987, followed by X62120-1 exhibiting 12 SNPs. Detailed information about position and nature of SNPs for each fragment are given in Supplementary Table 2. Taking into account the 28 monomorphic locus-specific fragments (12.5 kb) from DHPLC analysis, a frequency of 1 SNP/247 bp and 1 INDEL/3,583 bp was estimated for rapeseed. The number of SNPs and INDELs varied between coding and non-coding regions: 50 SNPs and 4 INDELs were found in 7,559 bp of non-coding regions; 34 SNPs and 1 INDEL were detected in 7,376 bp of coding regions. One INDEL spanned an exon–intron junction and three SNPs could not be unambiguously attributed to either coding or non-coding regions. Hence, an average of one SNP every 151 bp and one INDEL every 1,890 bp in non-coding regions and one SNP every 216 bp and one INDEL every 7,376 bp in coding regions was detected.

Allele frequency

Eighteen SNP/INDEL-containing fragments were available for analysing the panel of 86 rapeseed cultivars. For 14 fragments, 1 SNP was enough for genotyping since all polymorphisms within a fragment were in linkage disequilibrium; this was also true for the two SNPs of the two different fragments from gene accession X62120 which provided the same genotyping information. For BNA251988, two independent SNPs were determined, whereas for X61610 a second band was detected in agarose gel that occured in 46 genotypes. Therefore, 19 SNPs/INDELs were available for PIC estimation in 86 varieties. SNP markers are typically bi-allelic, therefore, the PIC values for single SNPs cannot be greater than 0.50. PIC values for single SNP markers in rapeseed ranged between 0.02 and 0.50 with a mean of 0.18 (Table 2). The value of 0.02 for BNA251988-A was due to a point mutation occurring only once in the panel of rapeseed cultivars, whereas the alleles of the other SNP markers were well distributed (Table 2). Seventeen of 19 SNPs showed heterozygous genotypes occurring in frequencies between 2 and 93% of the varieties, two markers did not show any heterozygosity. Hybrid varieties carried a mean of 5.80 heterozygous SNPs, whereas line varieties showed a mean of 3.58 heterozygous SNPs. Mean comparison with Welch two-sample *t*-test revealed highly significant difference between hybrids and line varieties ($t = 3.7397$, df $= 26.151$, P-value $= 0.0009117$).

In silico mapping

For 15 *B. napus* genes, a single homologous counterpart each and thus a single genome location was found in *A. thaliana*; two genes with the accession numbers L25404 and X55937 were duplicated. Results of BLASTN analysis with corresponding gene IDs and gene functions in *A. thaliana* are given in Supplementary Table 3. Based on the study of Parkin et al. (2005), all possible loci for the 17 genes in the rapeseed genome were determined and information on genome specificity then allowed to confine their assignment to particular linkage groups: a unique location in the *B. napus* genome was obtained for seven gene sequences; two and three possible locations were determined for six and four sequences, respectively (Table 3). The possible locations of the SNP-containing gene fragments within and across the 19 linkage groups of rapeseed are shown in Fig. 3.

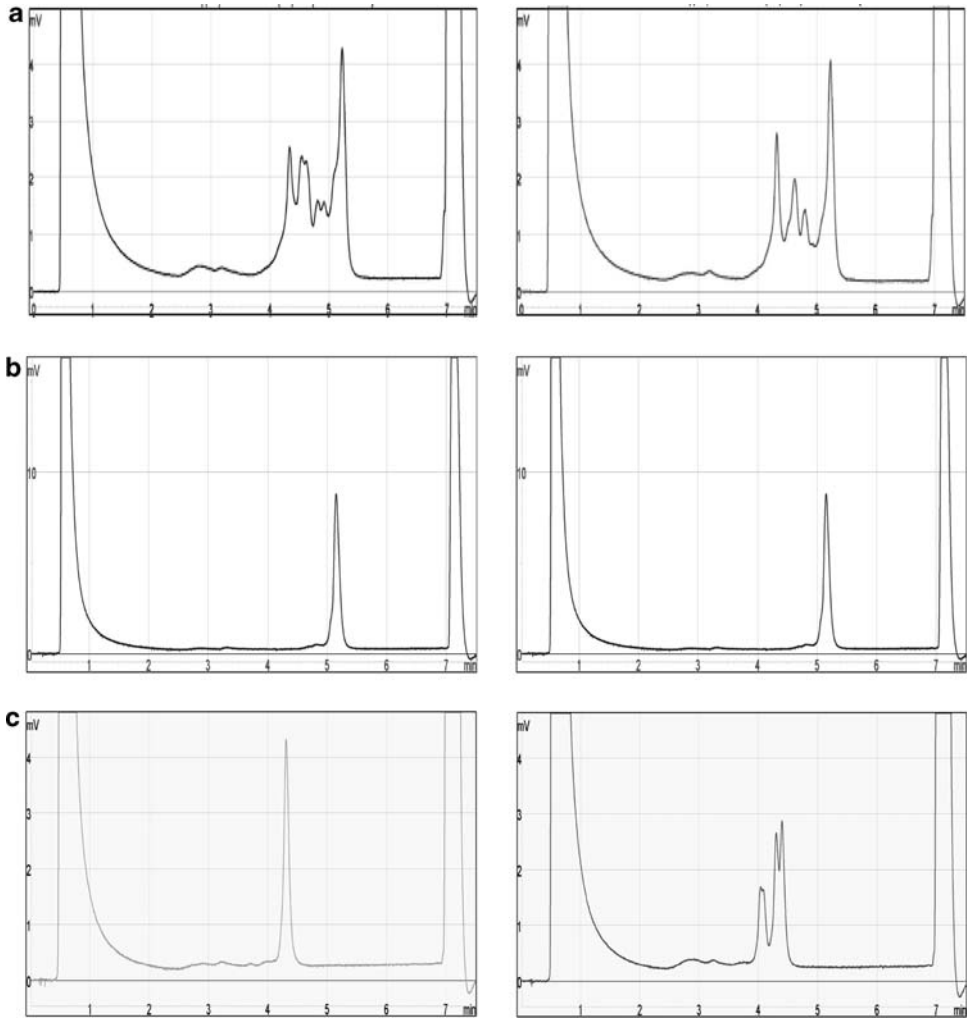

Fig. 2 Selection of single-locus PCR products using DHPLC method. Elution profiles of single-genotype and mixed-DNA sample PCR products are shown on the left and the right side, respectively: **a** low-copy PCR fragment, **b** single-copy PCR fragment showing no polymorphism between varieties and **c** single-copy PCR fragment showing polymorphism between varieties

Discussion

The complex genome structure of rapeseed poses an obstacle for discovering SNP markers in this crop species. Another barrier for detecting SNP markers in rapeseed is its relatively narrow genetic base compared to other crop species (Seyis et al. 2003). The two most important events (low erucic acid and low glucosinolate content) in breeding history of rapeseed were based on the utilisation of single plants that were exclusively used for crossing and selection.

To overcome these problems, a strategy involving two major steps for the selection of genome-specific fragments was applied: PCR primers were designed in intron–exon junctions thus making possible a higher efficiency for discriminating the A and C genomes in PCR. This step yielded 44% (70/160) genome-specific amplicons. DHPLC analysis then showed that 59% of genome-specific fragments proved to be single-copy, whereas only 19% of apparently genome non-specific PCR products did. Summarised, 83% of all SNP-containing fragments originated from locus-specific PCR products, and only three SNP markers were found in fragments of similar size between ancestor genomes of rapeseed. In addition, the concurrent SNP discovery reduced the amount of cost intensive and time consuming sequence analysis by around 90%.

In the survey of 18 candidate sequences across 6 rapeseed genotypes, we obtained 87 point mutations and 6 INDELs. The SNP frequency detected in our set of rapeseed accessions was 1 SNP/247 bp. This frequency is lower compared to those reported in maize (Tenaillon et al. 2001; 1 SNP/104 bp), rice (Monna et al. 2006; 1 SNP/113 bp), sunflower (Fusari et al. 2008; 1 SNP/69 bp) and barley (Kanazin et al. 2002; 1 SNP/189 bp). The INDEL frequency reported in this study was 1 INDEL/3,583 bp. Higher frequencies of INDELs were reported for maize (Bi et al. 2006, 1 INDEL/309 bp; Bhattramakki et al. 2002, 1 INDEL/124 bp) and melon (Morales et al. 2004; 1 INDEL/1,666 bp). As has also been reported by other studies, we found more sequence polymorphisms in non-coding regions than in coding regions. Ching et al. (2002) found 1 SNP/31 bp and 1 INDEL/85 bp in non-transcribed regions

Table 2 Genotype frequencies, allele frequencies and PIC values of SNP and INDEL markers in rapeseed

Marker		Genotype frequency			Allele frequency		
		Allele 1 (%)	Allele 2 (%)	Heterozygous (%)	Allele 1 (%)	Allele 2 (%)	PIC
A62529	SNP	57	28	15	64	36	0.46
ACGM060	SNP	76	14	10	81	19	0.31
AF490462	SNP	0	7	93	46	54	0.50
BNA251988-A	SNP	99	1	0	99	1	0.02
BNA251988-B	SNP	58	19	23	69	31	0.42
CD830352	SNP	47	1	52	73	27	0.39
D13987	SNP	52	35	13	59	41	0.48
D26381	INDEL	30	70	0	31	69	0.42
L25404	SNP	62	24	14	69	31	0.42
L31891	SNP	89	9	2	89	11	0.20
X55937	SNP	0	37	63	62	38	0.45
X61610-A	SNP	79	13	8	84	16	0.27
X61610-B	INDEL	47	0	53	73	27	0.39
X62120-1	SNP	64	10	26	76	24	0.36
X62120-2	SNP	64	10	26	76	24	0.36
X67142	SNP	73	19	8	78	22	0.35
X74782	SNP	56	24	20	66	34	0.45
X77576	SNP	67	19	14	74	26	0.38
X93461	SNP	52	26	22	62	38	0.47
Y12531	INDEL	84	6	10	89	11	0.20

Table 3 Genome specificity of SNP/INDEL-containing fragments and their possible chromosome locations

Marker	*B. napus*	*B. oleracea*	*B. rapa*	Possible chromosome locations
A62529	×	×	–	**N15**, N9, N7, **N17**, N6, N8, **N18**
ACGM060	×	–	×	**N4**, N14, **N5**
AF490462	×	×	–	N7, **N17**
BNA251988	×	+	×, +	**N1**, N11, N13, N17, **N8**
CD830352	×	×	×[a]	**N2**, N12, **N3**, N13, **N10**, N19
D13987	×	–	×	**N4**, N14
D26381	×	×	–	N3, **N13**
L25404	×	×	×[a]	**N2**, N12, N15, **N6**, N18
L31891	×	–	×	N14, **N9**, N19
X55937	×	–	×	**N2**, N12, **N4**, N14
X61610	×	×[a]	×	N2, **N12**, N3, **N13**, N10, **N19**
X62120-1	×	–	×	**N3**, N13, N17
X62120-2	×	+	×	**N3**, N13, N17
X67142	×	×	×[a]	**N3**, N13, **N10**, N19
X74782	×	–	×	**N4**, N14
X77576	×	–	×	**N5**, N15, **N6**
X93461	×	–	×	**N4**, N14
Y12531	×	×	–	N1, **N11**, N3, **N13**, **N17**

Unambiguous or best possible linkage group allocation of SNP markers is indicated in bold letters

× specificity of fragments, – null allele, + fragment length polymorphism between the genomes of *B. oleracea* and *B. rapa*

[a] Genome assignment based on similarity to sequences of either *B. oleracea* or *B. rapa*

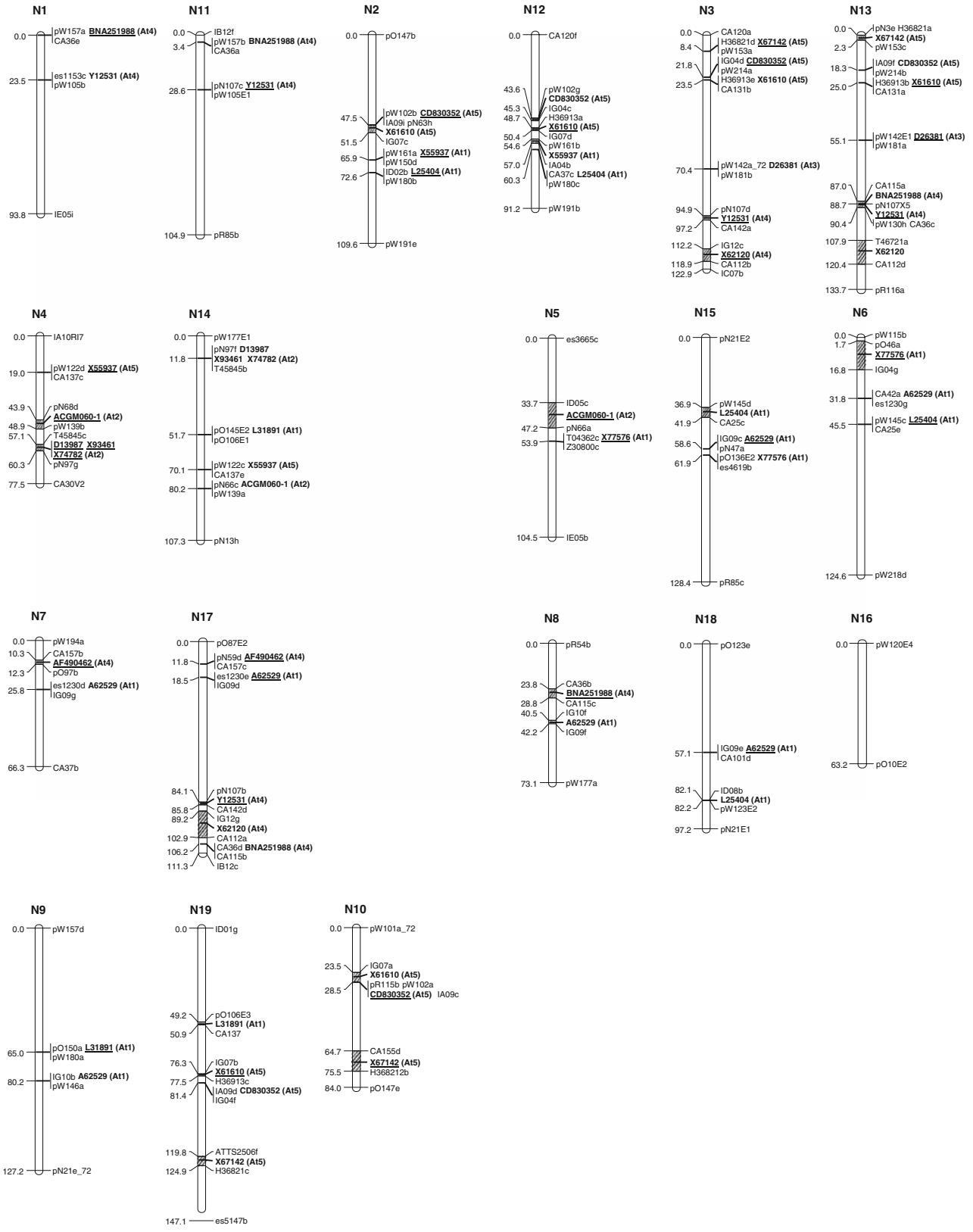

Fig. 3 In silico mapping of SNP and INDEL markers in rapeseed by cross referencing to the *A. thaliana* genome. All markers could be allocated to either the A or C genome. Unambiguous and best possible locations of SNP markers are *underlined*. Chromosome locations of gene sequences in *Arabidopsis* are given in *brackets*

and 1 SNP/124 bp and hardly any INDELS in transcribed regions of the maize genome. Fusari et al. (2008) found 1 SNP/90 bp in coding regions and 1 SNP/48 bp in non-coding regions of the sunflower genome. It is difficult to speculate about these discrepancies at the species level since they might represent the natural level of genetic polymorphism of the single species. However, gene sampling, fragment sizes and the number and genetic divergence of genotypes sequenced might have influenced the great differences observed. One possible reason for the low polymorphism rate in rapeseed may have arisen from breeding history. The two major events that led to today's success of oilseed rape were based on the selection of two single events (low erucic acid and low glucosinolate content) and further usage in breeding double-low (00) seed quality. This 'bottle-neck' phenomenon resulted in a considerably narrow genetic basis of current oilseed rape breeding material.

Except for G/A in BNA251988-A, all point mutations were found in a frequency that clearly exceeded 1% in our set of rapeseed varieties and met the definition of SNP markers (Kirk et al. 2002). PIC values of the SNP markers ranged between 0.20 and 0.50; for BNA251988-A, a PIC of 0.02 was obtained because it was found only in 1 of 86 varieties and, consequently, was classified as a rare allele. As expected, significantly more heterozygous SNP genotypes were found in hybrid varieties than in line varieties. Heterozygous marker genotypes in line varieties may account for residual heterozygosity. However, while only one-fourth of the varieties were hybrids, SNP analysis showed for four markers heterozygosity in 52–93% of the cultivars. This high amount cannot be explained by residual heterozygosity only, but is rather an indication for the amplification of more than one copy of these four gene fragments in a subset of genotypes. This observation further demonstrated the complexity of SNP analysis, in that case interpretation of genotyping data, due to gene copy variation among *B. napus* genotypes.

The study of Parkin et al. (2005) showed that physically linked markers in *A. thaliana* correspond to genetically linked loci in *B. napus*. On the basis of this information, in silico mapping revealed between two and seven possible loci for the current SNP markers in rapeseed. Sequences located on *Arabidopsis* chromosomes 1 and 5 occurred more frequently in the rapeseed genome than those located on chromosomes 2 and 4 (Fig. 3). The majority of sequences originating from orthologues on chromosomes 1 and 5 had four or more copies in rapeseed, whereas only two sequences from *Arabidopsis* chromosomes 2, 3 and 4 had more than three copies. This is in agreement with the study of Li et al. (2003) who investigated the colinearity between *Arabidopsis* and *B. oleracea*. The SNP discovery strategy allowed besides selection of single-copy PCR products the prediction of genome specificity. This information allowed for seven sequences the 'electronic' determination of a unique chromosome location; for the remaining sequences two or three possible chromosome positions could be suggested. However, in silico mapping of markers cannot be considered as definite, even if it provides useful information for further studies with mapping populations, due to rearrangements of single genes between *Arabidopsis* and *Brassica* genera (Qiu et al. 2009) and rearrangements between *B. rapa* and *B. oleracea* genomes after synthesis of *B. napus* (Udall et al. 2005).

Our approach is well suited for searching SNP markers in polyploid plant species like rapeseed. The selection of genome-specific amplicons combined with DHPLC analysis ensured a satisfying retrieval of single-copy fragments for inter-varietal polymorphism analysis.

Acknowledgments This study was funded by the Bavarian State Ministry of Environment, Public Health and Consumer Protection. The authors thank Friedrich Felsenstein from EpiGene GmbH for making the DHPLC instrument available for polymorphism screening.

References

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

Anonymous (2008) Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz. www.bmelv-statistik.de. Accessed 11 Sep 2008

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol 132:84–91

Bäumler S, Sierotzki H, Gisi U, Mohler V, Felsenstein FG, Schwarz G (2003) Evaluation of *Erysiphe graminis* f sp *tritici* field isolates for resistance to strobilurin fungicides with different SNP detection systems. Pest Manag Sci 59:310–314

Becker HC, Engqvist GM, Karlsson B (1995) Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. Theor Appl Genet 91:62–67

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC III, Tingey SV, Rafalski A (2002) Insertion-deletion polymorphisms in 3′ regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 48:539–547

Bi IV, McMullen MD, Sanchez-Villeda H, Schroeder S, Gardiner J, Polacco M, Soderlund C, Wing R, Fang Z, Coe EH (2006) Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map. Crop Sci 46:12–21

Botstein D, White RL, Skolnik M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Burke JM, Burger JC, Chapman MA (2007) Crop evolution: from genetics to genomics. Curr Opin Genet Dev 17:525–532

Charlton KM, Corner AH, Davey K, Kramer JKG, Mahadevan S, Sauer FD (1975) Cardiac lesions in rats fed rapeseed oils. Can J Comp Med 39:261–269

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Cordeiro GM, Eliott F, McIntyre CL, Casu RE, Henry RJ (2006) Characterisation of single nucleotide polymorphisms in sugarcane ESTs. Theor Appl Genet 113:331–343

Diers BW, McVetty PBE, Osborn TC (1996) Relationship between heterosis and genetic distance based on restriction fragment length polymorphism markers in oilseed rape (*Brassica napus* L.). Crop Sci 36:79–83

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14:1812–1819

Fourmann M, Barret P, Froger N, Baron C, Charlot F, Delourme R, Brunel D (2002) From *Arabidopsis thaliana* to *Brassica napus*: development of amplified consensus genetic markers (ACGM) for construction of a gene map. Theor Appl Genet 105:1196–1206

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol 8:7

Halldén C, Nilsson N-O, Rading IM, Säll T (1994) Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. Theor Appl Genet 88:123–128

Hasan M, Seyis F, Badani AG, Pons-Kühnemann J, Friedt W, Lühs W, Snowdon RJ (2006) Analysis of genetic diversity in the *Brassica napus* L. gene pool using SSR markers. Genet Resour Crop Evol 53:793–802

Hasan M, Friedt W, Pons-Kühnemann J, Freitag NM, Link K, Snowdon RJ (2008) Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. napus). Theor Appl Genet 116:1035–1049

Hu XY, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.). Theor Appl Genet 113:497–507

Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 48:529–537

Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F (2002) Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res 30:3295–3311

Kota R, Varshney RK, Prasad M, Zhang H, Stein N, Graner A (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics 8:223–233

Lei SL, Yao XQ, Bin Y, Chen W, Ma CZ, Tu JX, Fu TD (2007) Towards map-based cloning: fine mapping of a recessive genic male-sterile gene (*BnMs2*) in *Brassica napus* L. and syntenic region identification based on the *Arabidopsis thaliana* genome sequences. Theor Appl Genet 115:643–651

Lewerenz HJ, Plass R, Bleyl DWR, Macholz R (1988) Short-term toxicity study of allyl isothiocyanate in rats. Food 32:723–728

Li G, Gao M, Yang B, Quiros CF (2003) Gene for gene alignment between the *Brassica* and *Arabidopsis* genomes by direct transcriptome mapping. Theor Appl Genet 107:168–180

Monna L, Ohta R, Masuda H, Koike A, Minobe Y (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). DNA Res 13:43–51

Morales M, Roig E, Monforte AJ, Arús P, Garcia-Mas J (2004) Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). Genome 47:352–360

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 9:163–171

Oefner PJ, Underhill PA (1998) DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). In: Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE (eds) Current protocols in human genetics (Suppl 19). Wiley, New York, pp 7.10.11–17.10.12

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

Qiu D, Gao MQ, Li GY, Quiros C (2009) Comparative sequence analysis for *Brassica oleracea* with similar sequences in *B.*rapa and *Arabidopsis thaliana*. Plant Cell Rep 28:649–661

Rana D, van den Boogaart T, O'Neill CM, Hynes L, Bent E, Macpherson L, Park JY, Lim YP, Bancroft I (2004) Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. Plant J 40:725–733

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Seyis F, Snowdon RJ, Lühs W, Friedt W (2003) Molecular characterization of novel resynthesized rapeseed (*Brassica napus*) lines and analysis of their genetic diversity in comparison with spring rapeseed cultivars. Plant Breed 122:473–478

Sokolov BP (1990) Primer extension technique for the detection of single nucleotide in genomic DNA. Nucleic Acids Res 18:3671

Somers DJ, Rakow G, Prabhu VK, Friesen KRD (2001) Identification of a major gene and RAPD markers for yellow seed coat colour in *Brassica napus*. Genome 44:1077–1082

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebly JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). Proc Natl Acad Sci USA 98:9161–9166

Udall JA, Quijada PA, Osborn TC (2005) Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. Genetics 169:967–979

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

Xiao L, Lu CM (2005) Cloning of *fae1* gene partial sequence and SNP analysis in *Brassica* species. Scientia Agric Sinica 38:891–896