Theor Appl Genet (2006) 113:1221–1231
DOI 10.1007/s00122-006-0377-0

ORIGINAL PAPER

Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs

M. L. Flannery · F. J. G. Mitchell · S. Coyne · T. A. Kavanagh · J. I. Burke · N. Salamin · P. Dowding · T. R. Hodkinson

Received: 10 November 2005 / Accepted: 15 July 2006 / Published online: 15 August 2006

Abstract We report a new set of nine primer pairs specifically developed for amplification of *Brassica* plastid SSR markers. The wide utility of these markers is demonstrated for haplotype identification and detection of polymorphism in *B. napus, B. nigra, B. oleracea, B. rapa* and in related genera *Arabidopsis, Camelina, Raphanus* and *Sinapis*. Eleven gene regions (*ndhB-rps7* spacer, *rbcL-accD* spacer, *rpl16* intron, *rps16* intron, *atpB-rbcL* spacer, *trnE-trnT* spacer, *trnL* intron, *trnL-trnF* spacer, *trnM-atpE* spacer, *trnR-rpoC2* spacer, *ycf3-psaA* spacer) were sequenced from a range of *Brassica* and related genera for SSR detection and primer design. Other sequences were obtained from GenBank/EMBL. Eight out of nine selected SSR loci showed polymorphism when amplified using the new primers and a combined analysis detected variation within and between *Brassica* species, with the number of alleles detected per locus ranging from 5 (loci MF-6, MF-1) to 11 (locus MF-7). The combined SSR data were used in a neighbour-joining analysis (SMM, D_{DM} distances) to group the samples based on the presence and absence of alleles. The analysis was generally able to separate plastid types into taxon-specific groups. Multi-allelic haplotypes were plotted onto the neighbour joining tree. A total number of 28 haplotypes were detected and these differentiated 22 of the 41 accessions screened from all other accessions. None of these haplotypes was shared by more than one species and some were not characteristic of their predicted type. We interpret our results with respect to taxon differentiation, hybridisation and introgression patterns relating to the 'Triangle of U'.

Communicated by H. C. Becker.

M. L. Flannery · F. J. G. Mitchell · N. Salamin · P. Dowding · T. R. Hodkinson (✉)
Department of Botany, School of Natural Sciences, Trinity College, University of Dublin, Dublin, D2, Ireland
e-mail: trevor.hodkinson@tcd.ie

S. Coyne · T. A. Kavanagh
Smurfit Institute of Genetics, Trinity College, University of Dublin, Dublin, D2, Ireland

M. L. Flannery · J. I. Burke
Teagasc Crop Research Centre, Oak Park, Carlow, Ireland

N. Salamin
Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Introduction

Brassicaceae encompasses 380 genera and approximately 3,000 species (Mabberley 1993), many of which are of global economic importance including *Brassica napus* L. (oilseed rape), *Brassica rapa* L. (=*B. campestris* L., turnip rape), *Brassica oleracea* L. (Brussels sprouts, broccoli, cabbage, cauliflower, kale and kohlrabi) and *B. nigra* L. (black mustard). *Brassica nigra*, *B. oleracea* and *B. rapa* are known to be inter-fertile and have been hybridised in different combinations to produce three other species: *Brassica carinata* A. Braun, *Brassica juncea* (L.) Czern and *Brassica napus* (the Triangle of U; UN 1935; Erickson et al. 1983; Palmer et al. 1983; Song et al. 1988; Lowe et al. 2002). Chloroplast RFLP markers have provided further evidence for the multiple origins of some of these species (Palmer et al. 1983; Soltis and Soltis 1993).

Genomic analyses of Brassicaceae are advanced in comparison to most other plant families and have been aided by the publication of the complete 120 Mbp genome of *Arabidopsis* (The *Arabidopsis* Genome Initiative 2000). As with other model organisms, a range of molecular markers and methodologies have been applied to *Brassica* including RAPD, ISSR (Quiros et al. 1995; Bornet and Branchard 2004), AFLP (Negi et al. 2000; Hansen et al. 2003) and RFLP (Song et al. 1995; Cavell et al. 1998). Numerous nuclear simple sequence repeat (SSR; microsatellite) markers have also been developed and tested (Kresovich et al. 1995; Lowe et al. 2004 and references therein) but little attention has been given to the development of plastid (chloroplast) SSR markers for *Brassica* and its close relatives.

Chloroplast (and mitochondrial) genomes are uniparentally and maternally inherited in *Brassicas* and most angiosperms (Palmer et al. 1983; Soltis and Soltis 2000). They can therefore detect different patterns of variation to those revealed by nuclear markers. For example, they can document the maternal parent of hybrid plants (Hodkinson et al. 2002), define organelle haplotypes (Halldén et al. 1993) or detect introgression (Rieseberg and Soltis 1991; Cronn and Wendel 2003). However, chloroplast DNA generally exhibits lower mutation rates than nuclear DNA and therefore interspecific variation is low in comparison to the nuclear genome (Provan et al. 2001; Panda et al. 2003).

The chloroplast genome of *Arabidopsis* (154 kbp) encodes only 87 proteins (Sato et al. 1999) but many are of high agronomic performance value (Schulte et al. 1997; Martin et al. 2002). It is clear that markers that span the plastid genome of *Brassica* will have utility for plant breeders selecting for, or manipulating, plastid-encoded traits. For example *rbcL*, which encodes the large subunit of ribulose bisphosphate carboxylase, the principal carbon-fixing enzyme in plants, is located in the plastid genome. Another example, with perhaps more importance to Brassicaceae breeders is the plastid *aacD* gene a subunit that, together with three nuclear gene subunits, encodes acetyl-CoA carboxylase, a key plastid localised enzyme regulating the rate of de novo fatty acid biosynthesis and therefore oil yield (Schulte et al. 1997). Genes for chilling tolerance have also been shown to be maternally inherited and hence of plastid origin (Chung et al. 2003).

Simple sequence repeats consist of tandemly repeated multiple copies of mono-, di-, tri-, tetra- etc. nucleotide repeat motifs. Length polymorphism makes them highly informative markers, providing an efficient and accurate means of testing genetic variation (Powell et al. 1996). It has been estimated that on average, an SSR greater than 20 bp in length occurs every 33 kb in plant nuclear genomes compared with every 6 kb in mammals (Powell et al. 1996). SSR markers are capable of detecting genetic differences between closely related plants (Lowe et al. 2002, 2004) and they are now routinely used to investigate the genetic structure of natural populations (Balloux and Lugon-Moulin 2002). SSRs also have huge potential for characterisation, conservation and utilisation of crop diversity (Szewc-McFadden et al. 1996). Although an extensive list of Brassicaceae nuclear SSRs is publicly accessible (UKCropNet 2003), there is an absence of plastid SSR markers specifically designed for *Brassica* species. Plastid SSR markers are available for *Arabidopsis thaliana* (Provan 2000). There are also universal plastid primers available that amplify SSRs from a wide range of angiosperm taxa (Chung and Staub 2003).

This paper focuses on the development and use of a set of primers with high utility for detection of polymorphism in Brassicaceae at inter-generic, inter-specific and intra-specific taxonomic levels. The *Arabidopsis* primers developed by Provan (2000) were shown to amplify in *Brassica* (Provan 2000) but levels of polymorphism detected by these in *Brassica* were not reported. We chose to develop our own primers because of this and because we felt that primers designed specifically for *Brassica* had the potential to detect greater polymorphism and have a higher chance of amplification from a wide range of *Brassica* taxa and from a wide range of template DNA qualities.

We tested the ability of our new markers to define plastid haplotypes, discriminate taxa and to reconstruct evolutionary history. The markers were developed from intron and spacer regions of DNA from 11 plastid DNA loci (*ndhB-rps7* spacer, *rbcL-accD* spacer, *rpl16* intron, *rps16* intron, *atpB-rbcL* spacer, *trnE-trnT* spacer, *trnL* intron, *trnL-trnF* spacer, *trnM-atpE* spacer, *trnR-rpoC2* spacer, and *ycf3-psaA* spacer) that are known to show higher incidence of SSRs than coding regions (Provan et al. 2001). These markers are needed for a wide range of applications in plant biology such as plant breeding, evolutionary biology, population genetics, molecular ecology, and systematics.

Materials and methods

DNA sequencing and GenBank searching of intron and spacer DNA regions from a range of Brassicaceae

DNA from five intron and spacer regions of plastid DNA was sequenced from a range of Brassicaceae. These regions included the *atpB-rbcL* spacer, *rpl16*

intron, *rps16* intron, *trnL* intron, and *trnL-F* intergenic spacer (hereafter *trnL-F*). Plastid DNA regions from five species of Brassicaceae (Table 1) were amplified from total genomic DNA extractions (following a modified protocol of Doyle and Doyle 1987) and sequenced using primers c and f (Taberlet et al. 1991) for the *trnL-F* region, primers 16F and 2R (Oxelman et al. 1997) for the *rps16* gene region, primers F71 and R1661 (Jordan et al. 1996) for the *rpl16* gene region and primers 1R and 2R (Samuel et al. 1997) for the *atpB-rbcL* spacer region. Polymerase chain reaction (PCR) used an Applied Biosystems DNA Thermal Cycler (Geneamp® PCR System 9700). The thermal cycling for all PCRs comprised 32 cycles, each with 1 min denaturation at 97°C, 1 min annealing at 48°C, and an extension of 3 min at 72°C. A final extension of 7 min at 72°C was included. Amplified, double-stranded DNA fragments were purified using PCR mini-columns (GibcoBRL) and sequenced using *Taq* Dye-Deoxy Terminator Cycle Sequencing Kits (Applied Biosystems) on an ABI 310 automated DNA sequencer (Applied Biosystems). Sequences are deposited in GenBank, EMBL and DDBJ (Table 1).

Six further regions (*ndhB-rps7* spacer, *rbcL-accD* spacer, *trnE-trnT* spacer, *trnM-atpE* spacer, *trnR-rpoC2* spacer, *ycf3-psaA* spacer) were identified as sources of potentially variable SSRs by screening the GenBank/EMBL/DDBJ *Arabidopsis* sequence (NC_000932). Partial sequences of *Brassica napus* cv. Licosmos were obtained for these regions (Coyne 2002) and this allowed conserved SSR flanking regions to be identified for primer design.

Selection of SSR loci and primer design

All plastid DNA sequences were screened for loci containing SSRs for primer development using a modified version of the perl script misa.pl (Thiel 2003; available at http://www.2.unil.ch/phylo/). Sequences for the *atpB-rbcL* spacer, *ndhB-rps7* spacer, *rbcL-accD* spacer, *rpl16* intron, *rps16* intron, *trnE-trnT* spacer, *trnL* intron, *trnL-trnF* spacer, *trnM-atpE spacer, trnR-rpoC2* spacer, and *ycf3-psaA* spacer were each aligned in separate matrices with a GenBank sequence for *Arabidopsis* and relevant sections of other GenBank acquired sequences. This allowed conserved regions flanking the SSRs to be identified for primer design. SSRs that were interrupted, or absent in some of the sequenced taxa were excluded. Loci with long mononucleotide repeats were preferentially selected as they are more likely to detect polymorphism than shorter ones.

For primer design, conserved regions (containing no SSRs, AT or GC repeats) of 20–26 bp in length, flanking the SSR region and located 100–300 bp apart were selected. A total of ten sets of SSR primers were designed and tested using the same amplification conditions for the gene region from which they were derived (Table 2). Primer sets were developed to amplify SSRs within each of the sequenced regions. Chloroplast primers regularly cross-amplify in related species so we tested our primers on a range of *Brassicas* and related genera (*Arabidopsis, Camelina, Raphanus* and *Sinapis*).

Neighbour-joining and haplotype analysis

A neighbour-joining (NJ) analysis was carried out in PAUP*4 (Swofford 2000) using all plastid SSR markers (using the presence and absence of alleles) with genetic distances based on delta mu-squared distance (D_{DM}; Goldstein et al. 1995) which modifies the average squared distance by removing its dependence of population size. The D_{DM} distance was calculated in the computer package MICROSAT1.5 (Minch et al. 1996). D_{DM} distance was used because it is appropriate for the stepwise mutation of SSRs and because an assessment of our aligned DNA sequences for each gene revealed that much of the length variation, in most genes, was due to SSR length variation. However, some other non-SSR indels ranging in length from 1 to 34 bp were present in four of the gene region loci (MF-4, 1 indel of 7 bp; MF-5, 2 indels of 1 bp; MF-6, 1 indel of 6 bp; MF-10, 1 indel of 1 bp and 1 of 34 bp). Because the D_{DM} is not suitable for length variation caused by such indels we tested an alternative distance measure not based on the stepwise mutation model, that is the Nei–Li distance measure (Nei and Li 1979). However, the tree produced by NJ analysis based on the Nei–Li similarity measure did not differ significantly from the tree based on the D_{DM} measure and is not presented.

Multi-allelic haplotypes were compiled for each accession, and were analysed using parsimony analysis (PAUP*4; Swofford 2000). The parsimony analysis used heuristic search options including 1,000 replicates of random stepwise addition, keeping 100 trees per replicate, and using TBR branch swapping. The groupings of haplotypes closely reflected those found in the NJ analysis and are therefore not shown. Instead the haplotype code (Table 4) was mapped onto the NJ tree so that patterns of haplotype distribution could be assessed.

Results

Two hundred and fifteen plastid SSR alleles (with repeat unit greater than 7–10) were detected from a

Table 1 Accessions and sequences used for SSR marker development and genotyping

Taxon	ID code	Plant source	Gene region sequenced or (SSR markers[f])	GenBank number
Arabidopsis thaliana	sn[j]	Unspecified	Plastid genome	NC_000932[c]
Arabidopsis thaliana cv. Landsberg	154–157	Nott. SC[g]	SSR markers	NA
Arabidopsis thaliana cv. Landsberg erecta	158–160	Nott. SC	SSR markers	NA
Arabidopsis thaliana cv. Wassilewskija	162, 163	Nott. SC	SSR markers	NA
Arabidopsis thaliana cv. Wassilewskija	164, 165	Nott. SC	SSR markers	NA
Brassica cretica	sn	Isolate Gr14	*trnL-F*	Y15350[b]
		Isolate Gr01		Y15351[b]
Brassica juncea	sn	Unspecified	*trnL-F*	AF451575[d]
Brassica hilarionis	sn	Isolate Cy143	*trnL-F*	Y15353[b]
Brassica insularis	sn	Isolate Sal146	*trnL-F*	Y15355[b]
Brassica oleracea	sn	Isolate GB226	*trnL-F*	Y15359[b]
Brassica oleracea ssp. *capitata*	30	Comm.[h] leaf	*atpB-rbcL*	AY752702[a]
			rpl16	AY752709[a]
			rps16	AY752721[a]
			trnL-F	AY752712[a]
Brassica oleracea ssp. *capitata*	sn	Unspecified	*trnL-F*	AF451574[d]
Brassica oleracea ssp. *capitata*	30	Comm. leaf	SSR markers	NA
Brassica oleracea ssp. capitata cv. Durham early	86	Comm. seed	SSR markers	NA
Brassica oleracea ssp. *gemmifera*	33	Comm. leaf	*rpl16*	AY752710[a]
			trnL-F	AY752713[a]
			rps16	AY752719[a]
			atpB-rbcL	AY752703[a]
Brassica oleracea ssp. *gemmifera*	33	Comm. leaf	SSR markers	NA
Brassica oleracea ssp. *gemmifera*	56	Comm. leaf	SSR markers	NA
Brassica oleracea ssp. *gemmifera*	60	Comm. leaf	SSR markers	NA
Brassica oleracea ssp. *italica*	36	Comm. leaf	*rpl16*	AY752711[a]
			trnL-F	AY752714[a]
			atpB-rbcL	AY752704[a]
			rps16	AY752720[a]
Brassica oleracea ssp. *italica*	36	Comm. leaf	SSR markers	NA
Brassica montana	sn	Isolate ES301	*trnL-F*	Y15357[b]
		Isolate It101		Y15358[b]
Brassica napus	sn	Cultivar H165	*atpB-rbcL*	AF264734[e]
Brassica napus cv. Marinka	127	Comm. seed	*trnL-F*	AY752715[a]
			atpB-rbc	AY752707[a]
			Lrps16	AY752724[a]
			trnL-F	AY752716[a]
			atpB-rbcL	AY752708[a]
			rps16	AY752722[a]
Brassica napus ssp. *napus* cv. Marinka	128, 143 144, 146	Comm. Seed	SSR markers	NA
Brassica napus ssp. *napus* cv. Sheila	132, 137	Comm. seed	SSR markers	NA
Brassica napus ssp. *napus* cv. Marinka	134–140	Comm. seed	SSR markers	NA
Brassica napus ssp. *napus* cv. Jura	148	Comm. seed	SSR markers	NA
Brassica napus ssp. *napus* cv. Jura	149	Comm. seed	SSR markers	NA
Brassica napus ssp. *napobrassica* cv. Best of all	62	Comm. seed	SSR markers	NA
Brassica rapa ssp. rapifera cv. Snowball earl white stone	70	Comm. seed	SSR markers	NA
Brassica rapa ssp. rapifera cv. Purple top Milan	75	Comm. seed	SSR markers	NA
Brassica nigra	sn	USDA[i] PI254362	*trnL-F*	AF451578[d]
		USDA PI357368		AF451579[d]
Brassica nigra	90	Comm. seed	SSR markers	NA
Brassica rapa	46	Field collected	*atpB-rbcL*	AY752705[a]
			rps16	AY752723[a]
			trnL-F	AY752717[a]
Brassica rapa ssp. *chinensis*	sn	Cultivar Pong-hap	*trnL-F*	AF451571[d]
Brassica rapa ssp. *pekinensis*	sn	Unspecified	*trnL-F*	AF451572[d]

Table 1 continued

Taxon	ID code	Plant source	Gene region sequenced or (SSR markers[f])	GenBank number
Brassica rapa ssp. *rapa*	sn	Cultivar Goseki	*trnL-F*	AF451573[d]
Brassica rapa	46	Field collected	SSR markers	NA
Brassica rapa	53	Field collected	SSR markers	NA
Brassica rupestris	sn	Isolate It94	*trnL-F*	Y15360[b]
Brassica villosa	sn	Isolate It88	*trnL-F*	Y15361[b]
Camelina sativa	124	Field collected	*atpB-rbcL*	AY752706[a]
			rps16	AY752725[a]
Camelina sativa	124	Field collected	SSR markers	NA
Lepidium virginicum	sn	Unspecified	*trnL-F*	AF451582[d]
Raphanus sativus	10	Commercial seed	*atpB-rbcL*	AY752701[a]
			trnL-F	AY752718[a]
Raphanus sativus	sn	Cultivar Mei Hwa	*trnL-F*	AF451576[d]
		Cultivar Mei Nong		AF451577[d]
Raphanus sativus	3	Comm. seed	SSR markers	NA
Raphanus sativus cv. Scarlet Globe	66	Comm. seed	SSR markers	NA
Sinapis alba	sn	USDA PI296037	*trnL-F*	AF451580[d]
		USDA PI45890		AF451581[d]
Sinapis arvensis	2	Field collected	SSR markers	NA

[a] This paper

[b] Lanner (unpublished data)

[c] Sato et al. (1999)

[d] Yang et al. (2002)

[e] Zhang (unpublished data)

[f] SSR markers = accessions used for genotyping with the SSR markers developed in this paper

[g] Nottingham *Arabidopsis* Stock Centre (http://www.arabidopsis.info)

[h] Commercial

[i] USDA National Plant Germplasm System

[j] sn = without ID code

total screened DNA sequence of 163,720 bp. An average of 1.31 SSR loci, greater than 10 bp in length, was detected per 1,000 bp. Nearly all of the SSRs detected were mononucleotide (96.9%) but some di-, tri- and tetra-nucleotide repeats were recorded in *rps16*, *trnL-F* and *ndhB-rps7* regions. Adenine $(A)_n$ and thymine $(T)_n$ mononucleotide repeats were found to be more common (30.8 and 64.6%, respectively) than cytosine and guanine repeats, (0 and 1.5%, respectively). A or T repeats were also the longest and therefore potentially more polymorphic. They therefore represent the majority of loci used for primer design (Table 2).

Amplification was successful with all primers, except at the MF-10 locus. Locus MF-5 was monomorphic but all other loci were polymorphic (Table 2). There was an average of 6.5 alleles per locus (Table 3). The number of alleles detected per polymorphic locus ranged from 5 (loci MF-6, MF-1) to 11 (locus MF-7). The most variable locus (MF-7) is found within the *trnM-atpE* intergenic spacer. This is a mononucleotide repeat with $T_{7–16}$ in our reference sequences. Differences in marker length ranged from 9 bp (155–164) in MF-6 to 38 (273–311) in MF-3 (Tables 2, 3).

Individual loci vary in their ability to differentiate plastid genome types and taxa. For example, in our sample set, locus MF-7 was able to separate *A. thaliana, B. rapa* ssp. *rapifera*, *B. nigra* and *Camelina sativa* from all other *Brassica* species tested (Table 3). Primer MF-3 was found to differentiate *A. thaliana, C. sativa* and *R. sativus* and *Sinapis arvensis*, but could not unambiguously differentiate the *Brassica* taxa.

The combination of all plastid SSR markers tested in the NJ analysis with D_{DM} distances (Fig. 1), successfully differentiated *A. thaliana*, *B. oleracea, B. nigra, C. sativa, R. sativus* and *S. arvensis* individuals. The combined analysis failed to separate *B. rapa* and *B. napus* individuals into clearly identifiable groups (neither was monophyletic). *Brassica* (*B. napus, B. nigra, B. oleracea, B. rapa*) is monophyletic and its sister group is *Raphanus*.

A total number of 28 multi-allelic haplotypes were detected (Fig. 1 and Table 4) and these could differentiate 22 of the 41 accessions screened from all other accessions (i.e. 22 accessions had unique haplotypes). The haplotypes were also found to be species-specific across all taxa; no haplotype was shared by any taxa.

Table 2 Plastid microsatellite primer sets

SSR locus	Gene region	Repeat[a]	Primer sequences (5′-3′), forward (F) and reverse (R)	Fragment size range (bp)	No. of alleles[d]
MF-1	*trnL-F*	$(A)^{b}_{7-10}$	F TCAATTGCACATTCTAGAATTCTAAG R CAATTCAATATGGTTATATATTAGAG	157–185	5
MF-2	*rpl16*	$(T)_{8-13}$	F GGTTCCGTCGTTCCCATCGC R CATAATAATTAGATAAATCTGTTCC	157–185	7
MF-3	*trnE-trnT*	$(T)_{7-10}$	F AATGGTATGACTAGCTTATAAGG R CTTAACAATGAGATGAGGCAATC	273–311	8
MF-4	*psaA-ycf3*	$(C)_{3-8}(T)_{6-12}(T)_{7-9}$	F CGGATCTATTATGACATATCC R GAAATATGAATACACTAGATTAGG	127–155	8
MF-5	*trnT-rpoC2*	$(T)_{11-13}$	F CCTGGCGGTATCAAGATGCCACT R GCCATAATGGTACAGAACTAT	196	1
MF-6	*atpB-rbcL*	$(A)_{7-8}(T)_{5-6}$	F GAAGGAATAGTCGTTTTCAAG R CATAAATAGAGTTCCATTTCGG	155–164	5
MF-7	*TrnM-atpE*	$(T)_{7-16}$	F CGGCAGGAGTCATTGGTTCAAA R GATTTTGTAACTAGCTGACG	142–171	11
MF-8	*rbcL-accD*	$(T)_{8-10}$	F CTTATATTCATAAGCGAAGAAC R AATAACAATAGATGAATAGTCA	233–243	6
MF-9	*ndhB-rps7*	$(T)_{5-13}$	F GGGCCGTTATGCTCATTACG R TCCTATTCATGGGGATTCCG	283–311	8
MF-10	*rps16*	$(C)_{7-9}(A)_{7-8}(T)_{6-8}$	F GTGGTAGAAAGCAACGTGCGACTT R ACTTGAGTTACGAGAGTACGAATG	[c]	0

[a] Repeat number in reference sequences used to design primers

[b] $(A)_{7-10}$ is an adenine repeat between 7 and 10 bp in length that was detected in the reference sequences

[c] Not amplified

[d] See Table 3 for a detailed breakdown of alleles per taxon

Table 3 Allele size of the amplified SSR markers in Brassicaceae

	MF-1 *trnL-F*	MF-2 *rpl16*	MF-3 *trnE-trnT*	MF-4 *psaA-ycf3*	MF-6 *atpB-rbcL*	MF-7 *trnM-atpE*	MF-8 *rbcL-accD*	MF-9 *ndhB-rps7*
Arabidopsis thaliana	179[a]	179	296	127	157	171	233	308
	185	185					243	309
Brassica napus	157	157	283	142	157	145	233	283
	164	168	309	144	162	147	240	307
			310	145	164	151	242	308
						152	243	309
Brassica napus cv. *napobrassica*	164	168	283	142	157	147	241	308
B. nigra	164	175	283	155	157	157	242	311
B. oleracea ssp. *capitata*	164	168	284	138	157	142	239	309
				139		145	241	
							243	
B. oleracea ssp. *gemmifera*	164	168	283	139	157	145	241	309
			284			146	242	
B. oleracea ssp. *italica*	164	168	283	139	157	145	243	284
B. rapa	164	168	283	142	157	147	241	309
		171		144		152	243	
B. rapa ssp. *rapifera*	164	171	283	145	157	148	241	309
							242	
Camelina sativa	185	185	279	127	161	158	242	307
Raphanus sativus	164	169	280	139	161	155	241	300
							242	301
Sinapis arvensis	165	175	273	136	155	155	241	309

Locus MF-5 was not included in the table as it was monomorphic

[a] Numbers represent size of amplification products (bp)

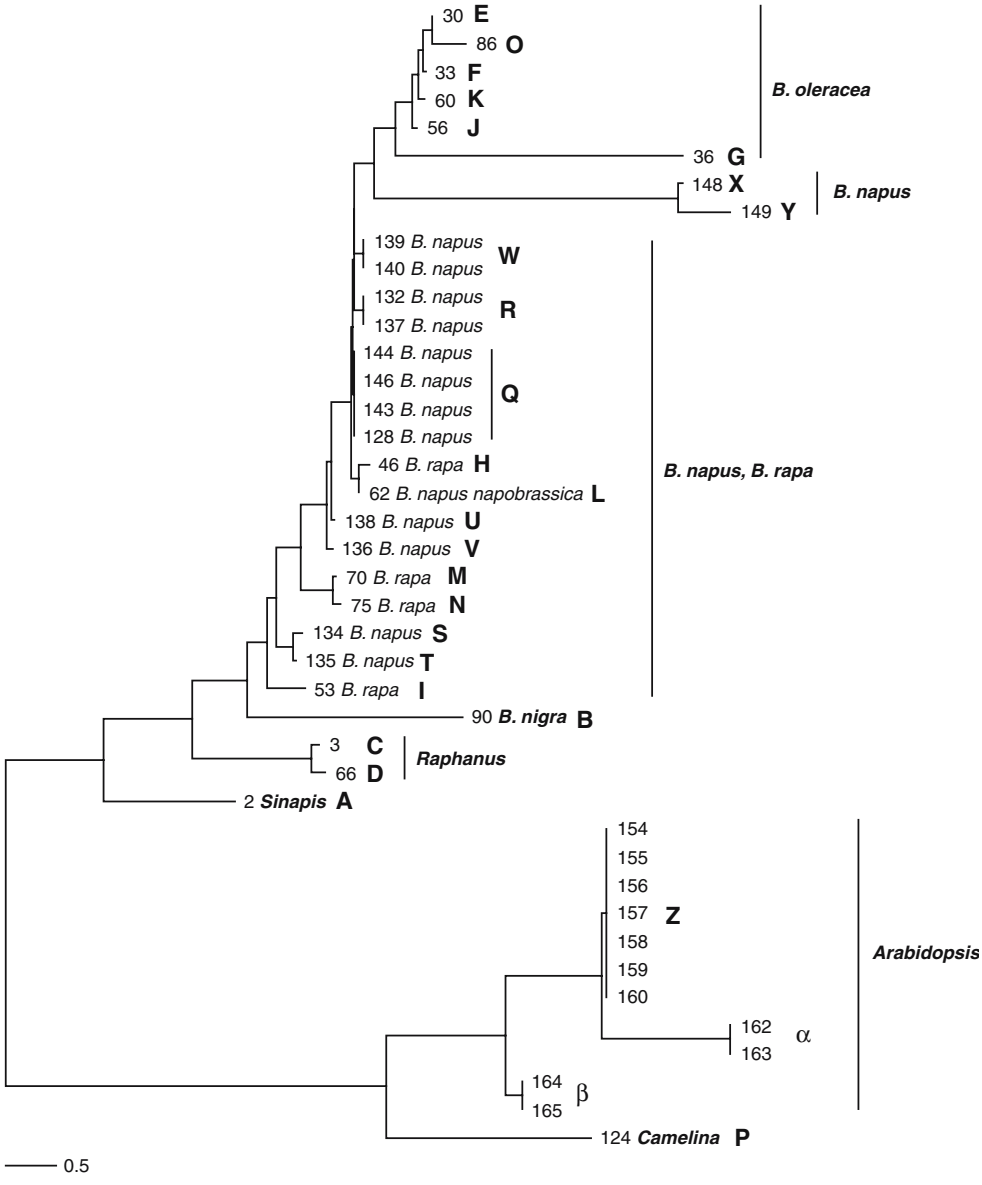

Fig. 1 Neighbour joining analysis using allele information from the eight plastid DNA loci. Haplotype codes from Table 4 have been plotted onto the tree (*bold letters*). Identification codes of plants correspond to those given in Table 1. Branch lengths are proportional to D_{DM} genetic distance. *Arabidopsis thaliana*, *Brassica oleracea, B. nigra, Camelina sativa, Raphanus sativus* and *Sinapis arvensis* individuals are grouped into their respective species groups. *Brassica rapa* and *B. napus* individuals are not differentiated

Discussion

Marker development and polymorphism detected in Brassicaceae

We report a new set of primers specifically designed for amplification of *Brassica* plastid SSR markers. These markers were shown to detect plastid DNA variation within and between *Brassica* species (*B. napus*, *B. nigra*, *B. oleracea*, *B. rapa*; Fig. 1 and Table 3). Twenty-eight haplopes were detected and found to be diagnostic for 22 of the 41 accessions screened. Therefore, more than half of the accessions had unique haplotypes. We also demonstrate the wider utility of these markers for plastid genome identification and detection of polymorphism in other Brassicaceae including *Arabidopsis, Camelina, Raphanus* and *Sinapis* (Fig. 1 and Table 3). Successful cross amplification was expected because studies using nuclear DNA have demonstrated cross-taxa (species and genera) SSR amplification (Westman and Kresovich 1998; Lowe et al. 2002). Furthermore, Provan (2000) showed that plastid SSRs could be developed from sequence information in Brassicaceae and used to detect variation among natural populations of *Arabidopsis* (detecting an average of four alleles per locus). Other studies using plastid DNA have focused on the differentiation of two or more *Brassica* species (Palmer et al. 1983; Scott and Wilkinson 1999). For example, plastid DNA has been shown to distinguish *B. napus* from *B. rapa* (Scott and Wilkinson 1999). In their study, universal chloroplast primers developed by Taberlet et al. (1991)

Table 4 Multi-allelic haplotypes detected from 41 accessions of Brassicaceae

Haplotype	Multi-allelic haplotype	Species	ID code
A	bdfafcab[a]	*Sinapis arvensis*	2
B	adgbiddh	*Brassica nigra*	90
C	abccfccd	*Raphanus sativus*	3
D	abbcfdcd	*Raphanus sativus*	66
E	aafbbced	*Brassica oleracea*	30
F	aafbbcdd	*Brassica oleracea*	33
G	aaabbedd	*Brassica oleracea*	36
H	aafbdcde	*Brassica rapa*	46
I	acfbgedf	*Brassica rapa*	53
J	aafbcded	*Brassica oleracea*	56
K	aafbcced	*Brassica oleracea*	60
L	aaebdcde	*Brassica napus*	62
M	acfbdddg	*Brassica rapa*	70
N	acfbdcdg	*Brassica rapa*	75
O	aafbabec	*Brassica oleracea*	86
P	dfdcfdba	*Camelina sativa*	124
Q	aaebcdde	*Brassica napus*	128, 143, 144, 146
R	aaebcede	*Brassica napus*	132, 137
S	aaebeedg	*Brassica napus*	134
T	aaebeddg	*Brassica napus*	135
U	aafbcddf	*Brassica napus*	136
V	aaebcddf	*Brassica napus*	138
W	aadbcdde	*Brassica napus*	139, 140
X	aaebadie	*Brassica napus*	148
Y	aaedadie	*Brassica napus*	149
Z	ceebiafa	*Arabidopsis thaliana*	154, 155, 156, 157, 158, 159, 160
α	dfebiafa	*Arabidopsis thaliana*	162, 163
β	ceebiefa	*Arabidopsis thaliana*	164, 165

[a] Alleles at each locus (MF-1, MF-2, MF-3, MF-4, MF-6, MF-7, MF-8, MF-9). Polymorphisms are coded as letters in the order of increasing molecular weight. For example haplotype A has allele b at SSR locus MF-1, allele d at SSR locus MF-2, allele f at SSR locus MF-3, etc

were used to amplify the *trnL* intron and *trnL-F* intergenic spacer DNA region and differences in length variation were used to differentiate the two *Brassica* species.

In this paper, ten SSR primer sets were developed, nine of which amplified SSRs successfully and eight of which were found to detect polymorphism when tested on a broad range of Brassicaceae material (Tables 3, 4). The locus that failed to amplify, MF-10, showed considerable length variation (1 indel of 1 bp and 1 indel of 34 bp) between the accessions sequenced and will undoubtedly be of use as a marker. New primers need to be designed and tested for this region. The average number of alleles detected at each of the other loci was 6.5 and the range was 1–11 (Table 2). Locus MF-7 (found within the *trnM-atpE* intergenic spacer) detected the most variation (11 alleles) and amplified in all accessions tested. This is a mononucleotide repeat with T_{7-16} in our reference sequences. It is worthy to note that this locus was the longest mononucleotide sequence found in the reference plants (*Brassica* and *Arabidopsis*). Longer SSRs are generally more likely to detect polymorphism than short ones (Provan et al. 2001). The markers detected variation (Table 3) within *B. napus* (21 alleles over all 8 loci), *B. rapa* (15 alleles) and *B. oleracea* (16 alleles), *Arabidopsis* (12 alleles), and *Raphanus* (10 alleles). Only a single individual of *C. sativa* and *S. arvensis* was included so intra-specific variation could not be assessed with these species. The markers also detected differences between *Brassica* species (Fig. 1 and Table 3) and related genera. Locus MF-5 (*trnT-rpoC2*) was a mononucleotide repeat of moderate length in our reference sequences (T_{13} in *Arabidopsis*) and T_{11} in *B. napus* cv. Licosmos (Coyne, 2002). However, no variation was detected in the *Brassica* or *Arabidopsis* accessions amplified. This discrepancy is likely to be due to sampling; if sufficient numbers of plants were screened we would expect variation to be detected. In spite of this, the locus clearly has lower potential than the others to provide polymorphic molecular markers in Brassicaceae.

The level of variation detected by the combination of primers is higher than that found in *Arabidopsis* by Provan (2000) who detected an average of four alleles per plastid SSR locus (compared to 6.5 alleles per locus here). The number of SSR loci per 1,000 bp of sequence was highest for the *rps16* intron and the *trnL-F* regions (5.67 and 3.16 SSRs per 1,000 bp, respectively). Length variation of the whole *trnL-F* intergenic spacer region has been shown previously by Scott and Wilkinson (1999) to discriminate *B. napus* from *B. rapa* samples. It is likely that the plastid SSR loci detected in this paper (MF-1) could have been responsible for the length variation they demonstrated in this gene region. The *rps16*, *trnL-F* regions have also proven useful for plastid SSR development in other organisms such as ash (Oleaceae; Harbourne et al. 2005) and orchids (Orchidaceae; Fay and Cowan 2001). Introns and spacer regions would be expected to have higher incidence of SSRs than coding DNA sequence (Provan et al. 2001).

Application of the markers for differentiation of Brassicaceae taxa and individuals

The loci were examined individually and in combination to test their ability to differentiate Brassicaceae genera, species and individuals. Individually the primers had some potential to differentiate taxa. For example, primer MF-3 (*trnE-trnT*) was found to differentiate

A. thaliana, *C. sativa*, *Raphanus sativus* and *S. arvensis* from each other and all other taxa. Primer set MF-7 (*trnM-atpE*) was found to differentiate *A. thaliana, B. rapa* ssp. *rapifera*, *B. nigra* and *C. sativa* from each other and all other taxa tested. However, it failed to differentiate between *B. oleracea* ssp., *B. napus* ssp. *napobrassica* and *B. napus* ssp. *napus.* The remaining six loci were not, on their own, able to unambiguously differentiate taxa. However, when the results from combined loci are considered by NJ (Fig. 1) and the multiallelic haplotypes (Table 4), the markers can differentiate plastid types and taxa more efficiently.

The NJ analysis (Fig. 1) using the combined marker dataset from eight loci (with multi-allelic haplotype information plotted onto the tree) was not intended to provide a detailed phylogenetic analysis of Brassicaceae species. However, they do indicate the overall nature and levels of genetic diversity detected by the combined data in a range of taxa. Because of the elevated mutation rate and nature of mutation at SSR loci (often a loss or gain of a repeat unit due to slipped strand mis-pairing; Provan et al. 2001), a higher degree of homoplasy (parallelism or reversal) can be expected at plastid SSR DNA loci than at non-SSR regions. Such homoplasy will obscure the phylogenetic utility of the markers particularly when wide taxonomic comparisons are made, such as between Brassicaceae genera. Furthermore, *Brassicas* are known to hybridise and undergo introgression (Palmer et al. 1983: Lowe et al. 2002; Soltis and Soltis 2000). Plastid genomes of *Brassicas* have also been manipulated by plant breeders via interspecific and intergeneric hybridisation (and somatic hybridisation processes) and can posses unexpected haplotype types (Halldén et al. 1993). Phylogenetic inferences based solely on plastid DNA may not, therefore, adequately represent the evolutionary history of the plants. For example, unexpected plastid DNA groupings can often be found because of plastid capture events (Hodkinson et al. 2002; Cronn and Wendel 2003). We conducted the NJ analysis here to show the general use of the new markers for plastid type identification and for a wide range of applications at infra-specific, infra-generic and intergeneric levels.

The combination of all plastid SSR markers tested in the NJ analysis (Fig. 1) was generally able to separate the individuals of Brassicaceae into taxon-specific groups. For example, *Arabidopsis*, *Camelina*, *Sinapis* and *Brassica* genera are all clearly differentiated. Within *Brassica*, *B. oleracea* is separated from *B. napus* and *B. rapa*. However, individuals of *B. napus* (including *B. napus* ssp. *napobrassica*) are not clearly separated from *B. rapa*. This pattern is also seen with the multi-allelic haplotype data (Fig. 1 and Table 4). *Brassica* is monophyletic in the NJ analysis and *Raphanus* is its sister group. The close relationship of *Raphanus* to *Brassica* has been reported (Warwick and Black 1991) especially to a *B. rapa/B.oleracea* lineage (Yang et al. 1998). Our results therefore support the conclusion that *Raphanus* is closely allied to *Brassica* and is possibly its sister genus. However, our results cannot support the hypothesis that it is more closely related to a *B. rapa*/*B. oleracea* lineage than a *B. nigra* lineage (as suggested by Yang et al. 1998).

The eight polymorphic SSR loci would be expected to offer potential to study maternal patterns of inheritance in Brassicaceae and hence document maternally inherited gene flow. *B. napus* (AC genome) is a polyploid resulting from the hybridisation of the diploids *B. oleracea* (C genome) and *B. rapa* (A genome) (UN 1935; Erickson et al. 1983; Palmer et al. 1983). This hybrid would be expected to share the plastid DNA markers of one but not both parents unless it had originated on multiple occasions. However, the multi-allelic haplotypes were taxon specific; no haplotype was shared by any taxa. In the NJ analysis *B. rapa* always groups with *B. napus* and we could infer that *B. rapa* is the more likely plastid genome donor of these allopolyploid hybrids.

Brassica oleracea haplotypes do not group with *B. napus* except accessions 148 and 149 in the NJ analysis but the branch lengths leading to these accessions are long, suggesting that they are considerably different from the *B. oleracea* types. Therefore, there is little evidence to support the hypothesis that *B. oleracea* is the plastid genome donor of any of these allopolyploid individuals. The failure of this marker set to resolve the maternal genome donor of the allopolyploid *B. napus* could also be due to homoplasy in the dataset (Doyle et al. 1998). The SSR loci evolve at a rapid rate and are therefore prone to parallel evolution (the same mutation can occur independently and provide no information regarding common ancestry). Furthermore, the hybrid origin of *Brassica* breeding material, and hence their plastid haploypes, is more complicated than simple triangle of U relationships. Plastids have been intentionally or unintentionally introgressed into a range of Brassicaceae. This was shown by Halldén et al. (1993) who used chloroplast RFLPs in alloplasmic breeding lines of *Brassica* to discover that most plastid genomes of *B. napus*, *B. rapa* and *B. oleracea* differed from their expected type. Our study therefore supports the findings of Halldén et al. (1993) and shows the high utility of SSRs for plastid type identification but warns against their utility for taxon identification in this group of plants.

In conclusion, a vast array of nuclear SSR marker systems has been developed for *Brassica* but relatively few have been developed for plastid DNA. The combination of the plastid SSR markers developed here with informative nuclear markers offers great potential for plant breeding programmes that require unambiguously identified plastid types (such as those involving nucleo-cytoplasmic interaction studies; Halldén et al. 1993) and detailed population genetic and diversity studies. We have also used a combination of plastid and nuclear SSR markers for documenting pollen and seed mediated gene-flow in oil-seed rape (*B. napus*; Flannery 2004) for risk assessment of genetically modified crops.

Acknowledgments MLF was supported by a Teagasc PhD Walsh Fellowship. TAK and SC were supported by EU contract LSHG-CT-2003–503238 (PLASTOMICS).

References

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol 11:155–165

Bornet B, Branchard M (2004) Use of ISSR fingerprints to detect microsatellites and genetic diversity in several related *Brassica* taxa and *Arabidopsis thaliana*. Hereditas 140:245–248

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

Chung S-M, Staub JE (2003) The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. Theor Appl Genet 107:757–767

Chung S-M, Staub JE, Fazio G (2003) Inheritance of chilling injury: a maternally inherited trait in cucumber. J Am Soc Hort Sci 128:526–530

Coyne S (2002) Transformation and marker gene removal strategies for the plastid genome of *Brassica napus*. PhD Thesis, Smurfit Institute of Genetics, University of Dublin, Trinity College, pp 172

Cronn R, Wendel JF (2003) Cryptic trysts, genomic mergers and plant speciation. New Phytol 161:133–142

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

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). Mol Biol Evol 15:215–218

Erickson LR, Strauss NA, Beversdorf WB (1983) Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphidiploids. Theor Appl Genet 65:201–206

Fay MF, Cowan RS (2001) Plastid microsatellites in *Cypripedium calceolus* (Orchidaceae): genetic fingerprints from herbarium specimens. Lindleyana 16:151–156

Flannery ML (2004) Risk assessment of genetically modified crops in Ireland: investigations of pollen dispersal from oilseed rape (*Brassica napus* L.) using molecular and pollen trapping techniques. PhD Thesis, Department of Botany, University of Dublin, Trinity College

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Halldén C, Gertsson B, Säll T, Lind-Halldén C (1993) Characterization of organellar DNA in alloplasmic lines of *Brassica napus* L. Plant Breed 111:185–191

Hansen LB, Siegismund HR, Jorgensen RB (2003) Progressive introgression between *Brassica napus* (oilseed rape) and *B. rapa*. Heredity 91:276–283

Harbourne ME, Douglas GC, Waldren S, Hodkinson TR (2005) Characterization and primer development for amplification of chloroplast microsatellite regions of *Fraxinus excelsior*. J Plant Res 118: 339–341

Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA (2002) The use of DNA sequencing (*ITS* and *trnL-F*), AFLP and fluorescent in situ hybridisation to study allopolyploid *Miscanthus* (Poaceae). Am J Bot 89:279–286

Jordan W, Courtney MW, Neigel JE (1996) Low levels of interspecific genetic variation at a rapidly evolving chloroplast DNA locus in north American duckweeds (Lemnaceae). Am J Bot 83:430–439

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

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

Mabberley DJ (1993) The plant-book: a portable dictionary of the vascular plants. Cambridge University Press, London

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of *Arabidopsis*, cyanobacteria and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA 99:12246–12251

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1996) Microsat (version 1.5b): a computer program for calculating various statistics on microsatellite allele data

Negi MS, Devic M, Delseny M, Lakshmikumaran M (2000) Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. Theor Appl Genet 101:146–152

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

Oxelman B, Liden M, Berglund D (1997) Chloroplast *rps*16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). Plant Syst Evol 206:257–271

Palmer JD, Shields CR Cohen DB, Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor Appl Genet 65:181–189

Panda S, Martín JP, Aguinagalde I (2003) Chloroplast and nuclear DNA studies in a few members of the *Brassica oleracea* L. group using PCR-RFLP and ISSR-PCR markers: a population genetic analysis. Theor Appl Genet 106:1122–1128

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Provan J (2000) Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana.* Mol Ecol 9:2183–2185

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147

Quiros CF, Truco MJ, Hu J (1995) Sequence comparison of two codominant RAPD markers in *Brassica nigra:* deletions, substitutions and microsatellites. Plant Cell Rep 15:268–270

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. Evol Trend Plant 5:65–84

Samuel R, Pinsker W, Kiehn M (1997) Phylogeny of some species of *Cyrtandra* (Gesneriaceae) inferred from the *atp*B-*rbc*L cpDNA intergene region. Bot Acta 110:503–510

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 6:283–290

Schulte W, Töpfer R, Stracke R, Schell J, Martini N (1997) Multifunctional acetyl-CoA carboxylase from *Brassica napus* is encoded by a multi-gene family: indication for plastidic localisation of at least one isoform. Proc Natl Acad Sci USA 94:3465–3470

Scott SE, Wilkinson MJ (1999) Low probability of chloroplast movement from oilseed rape (*Brassica napus*) into wild *Brassica rapa*. Nat Biotechnol 17:390–392

Soltis DE, Soltis PS (1993) Molecular data and the dynamic nature of polyploidy. Cr Rev Plant Sci 12:243–273

Soltis DE, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol Biol 42:45–75

Song KM, Osborn TC, Williams PH (1988) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 1 Genome evolution of diploid and amphidiploid species. Theor Appl Genet 75:784–794

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

Swofford DL (2000) PAUP4*: Phylogenetic analysis using parsimony (* and other methods), version 4.0b10. Sinauer Associates, Massachusetts

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

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Thiel T (2003) MISA—Microsatellite identification tool. http://www.pgrc.ipk-gatersleben.de/misa/

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

UKCropNet (2003) *Brassica* Database. http://www.ukcrop.net/perl/ace/search/BrassicaDB

Warwick SI, Black LD (1991) Molecular systematics of *Brassica* and allied genera (subtribe Brassicinae, Brassiceae) -chloroplast genome and cytodeme congruence. Theor Appl Genet 82:81–92

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

Yang Y-W, Tseng P-F, Tai P-Y, Chang C-J (1998) Phylogenetic position of *Raphanus* in relation to *Brassica* species based on 5S rRNA spacer sequence data. Bot Bull Acad Sin 39:153–160

Yang YW, Tai PY, Chen Y, Li WH (2002) A study of the phylogeny of *Brassica rapa*, *B. nigra*, *Raphanus sativus*, and their related genera using noncoding regions of chloroplast DNA. Mol Phylogenet Evol 23:268–275