Plant Omics Journal POJ
POJ 7(1):35-46 (2014) ISSN:1836-3644

Regularities in simple sequence repeat variations induced by a cross of resynthesized *Brassica napus* and natural *Brassica napus*

Caihua Gao[1†], Jiaming Yin[1†], Annaliese S. Mason[2], Zhanglin Tang[1], Xiaodong Ren[1], Chao Li[3], Zeshan An[1], Donghui Fu[4*], Jiana Li[1*]

[1]Engineering Research Center of South Upland Agriculture of Ministry of Education, College of Agronomy and Biotechnology, Southwest University, Chongqing, China
[2]Centre for Integrative Legume Research and School of Agriculture and Food Sciences, The University of Queensland, Brisbane 4072, Australia
[3]Institute of Oil Croups of Guizhou Province, Jinzhu Town, Xiaohe district, Guiyang 550006, China
[4]Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang 330045, China

***Corresponding authors: fudhui@163.com; ljn1950@swu.edu.cn**

[†]These authors contributed equally to this work.

Abstract

Interspecific hybridization can induce extensive variation in genome sequences, including simple sequence repeat (SSR) regions. To determine the characteristics of SSR variation induced by interspecific hybridization and the possible effect of SSR variation on gene function, we constructed a *Brassica napus* doubled-haploid (DH) population from a cross between natural *B. napus* and resynthesized *B. napus* (*B. oleracea* × *B. rapa*) and identified, located, sequenced and functionally annotated SSR variants. The results showed that novel SSR variants were generated in the F_1 generation and maintained in the introgressed DH population. Elimination of sequences carrying SSRs also occurred in the F_1 hybrids, with three times as many sequences lost in the introgressed DH population compared to in the F_1 hybrids, probably due to non-homologous recombination. The degree of SSR variation observed depended primarily on the number of SSR repeats and secondarily on the nucleotide composition of the SSR motifs. In the introgressed DH population, many genes containing SSRs exhibited frameshift mutations (62.5%) due to the expansion or contraction of the SSR motifs following deletion (25%) or insertion (12.5%) mutations. Most genes harboring SSR variants were associated with vital metabolic processes, such as protein or DNA metabolic processes. The SSR variation induced by interspecific hybridization reflects an intrinsic property of species adaptability post-hybridization through variation. This study is beneficial to understanding the origin of SSRs and the effects of SSR mutation on polyploid genomes.

Keywords: *Brassica napus*, Interspecific hybridization, Simple sequence repeats (SSRs), SSR variation, Functional annotation.

Abbreviation: SSR_Simple sequence repeat, DH_double haploid, E type_elimination type bands or fragments, N type_novel type bands or fragments, P type_parental type bands or fragments.

Introduction

Simple sequence repeats (SSRs), also called short tandem repeats, variable number tandem repeats, or microsatellites, comprise 1.0% of prokaryote genomes and more than 3.0% of eukaryote genomes (Morgante et al., 2002; Rahim, 2008; Subramanian et al., 2003). SSRs have been developed into stable and widely used PCR-based molecular markers (Dai et al., 2013; Fajolu et al., 2013; Huang et al., 2012; Kantartzi et al., 2009; Lawson and Zhang, 2006; Parida et al., 2009; Parida et al., 2010; Park et al., 2009; Sun et al., 2013; Wang et al., 2009). The distribution of SSRs in genomes is nonrandom (Huang et al., 2012; Labbe et al., 2011; Lawson and Zhang, 2006). In plant genomes, SSRs are situated preferentially in repetitive DNA regions (Morgante et al., 2002), such as centromeres (Cuadrado et al., 2008), and peri-centromeric heterochromatin SSRs are mostly associated with retrotransposons (Hong et al., 2007). In *Brassica rapa*, SSRs are also preferentially associated with gene-rich regions, but very few are located in coding regions (Hong et al., 2007). The number of SSRs in 5' untranslated regions (UTRs) is higher than that in 3' UTRs, exons, and introns in the genomes of *Arabidopsis thaliana*, *Oryza sativa* L. (Castillo et al., 2008; Hong et al., 2007; Parida et al., 2009), and *Triticeae* species (Cuadrado et al., 2008).

Allelic differences among SSRs predominantly involve variation in repeat numbers, but they can also involve variations in the length of the motif or in the flanking sequences of repeat units (Masters et al., 2010). The mutation rate per generation for SSR sequences has been estimated to be 7.7×10^{-4} per allele per generation for dinucleotide repeat motifs in maize (Vigouroux et al., 2002) and 2.03×10^{-3} per allele per generation for AT dinucleotide motifs in *A. thaliana* (Marriage et al., 2009). The SSR mutation rate is 100 to 10 000 times higher than the rate of base pair substitutions (Vigouroux et al., 2002) and 10 to 100 times higher than that of frameshift mutations within coding sequences (Eckert and Hile, 2009). Therefore, SSRs can be said to be a type of hypervariable sequence in genomes.

Organisms can balance the expansion of short SSRs and contraction of long SSRs to avoid infinite growth of these repeats

Table 1. Correlations between repeat number, size of motif, total SSR length and frequency of SSR variants in an introgressed DH population of *B. napus.*

	Repeat number	Size of motif
Size of motif	-0.5080 (P < 0.0001)	
Total SSR length		-0.0446 (P = 0.5345)
Frequency of SSRs		-0.4292 (P = 0.0127)

Note: This correlation analysis was based on a linear model. For each particular SSR variant, its corresponding repeat number, size of motif, total length and total number was collected to become a data array for the correlation analysis.

(Ellegren, 2000). There are several major factors that restrict or promote SSR expansion. Malfunction of DNA mismatch repair genes is associated with a loss of the ability of cells to detect and repair mutated SSRs (Golubov et al., 2010; Leung et al., 1998; Liu et al., 2000; Shin et al., 2002; Sol Mateo et al., 1998). Radioactive induction can lead to the gain or loss of a large number of repeat units and even the complete disappearance of DNA sequences (Kovalchuk et al., 2003). Interspecific hybridization is also a potential stressor that can induce SSR variation (Tang et al., 2008; Tang et al., 2009). After interspecific hybridization between wheat and rye, numerous SSR sequences were found to be eliminated and some novel SSR sequences appeared (Tang et al., 2008; Tang et al., 2009). In S5 lines of artificial *B. napus* derived from *B. rapa* and *B. oleracea*, most SSR markers were found to be altered relative to the parental lines (Gaeta and Chris Pires, 2010).

Most SSRs found in expressed sequence tags (ESTs) from pepper, peanut, wheat, barley, and yam crops are involved in gene regulation or expression (Castillo et al., 2008; Huang et al., 2012; Narina et al., 2011; Salmon et al., 2010; Wang et al., 2007). Whereas variation in SSRs can cause diseases in animals and humans (Eckert and Hile, 2009; Freimer and Slatkin, 1996; Liu et al., 2000; Shin et al., 2002; Shin et al., 2007), SSR variation provides the capacity to adapt to environmental change in bacterial species (Moxon and Wills, 1999) and many eukaryotic genomes (Kashi and King, 2006). However, the effect of SSR variation on gene function in plants has rarely been reported.

In order to determine the characteristics and regularity of SSR variation induced by interspecific hybridization, and the possible effect of SSR variation on gene function, we constructed a doubled-haploid (DH) population from a cross between natural *B. napus* (AACC, 2n=4X=38) and resynthesized *B. napus* (derived from interspecific hybridization between *B. rapa* (AA, 2n=2x=20) and *B. oleracea* (CC, 2n=2x=18)), hereafter referred to as the introgressed DH population. In this population, we detected a large number of SSR variants: appearance of novel SSR bands and elimination of parental SSR bands. We sequenced these identified SSR variants and analyzed the patterns of mutation, the frequency at which the SSRs mutated, the chromosomal localization of the SSRs, and the putative impact of these variants on the genome.

Results

Distribution of fragments harboring SSR variants in the introgressed DH population

"Parental" bands were present in the parent generation and in subsequent generations, whereas "abnormal" bands were either monomorphic between the parent lines but eliminated in subsequent generations ("eliminated" bands) or not present in the parent generation but present in subsequent generations ("novel" bands) (Fig 1). In the introgressed DH population, 15.4% (219) of the total bands (1420) amplified by 647 primer pairs were abnormal bands. All plants in the introgressed DH population produced abnormal bands, with a range of 1 - 55 novel bands and a range of 1 - 92 eliminated bands per plant (Fig 2). Detailed information for 430 bands produced by the 47 primer pairs generating at least one abnormal band in the introgressed DH population is presented in Supplementary Table 1. The frequency of eliminated bands was five times higher than that of novel bands. There were no significant differences (F = 0.85, P = 6386) between the frequency of abnormal bands amplified by 'H' type primers, 'Cen' type primers, 'A' type primers and 'Y' type primers (as described in the material and method section), demonstrating that the variation of the bands amplified by these primers did not vary by primer resource origin.

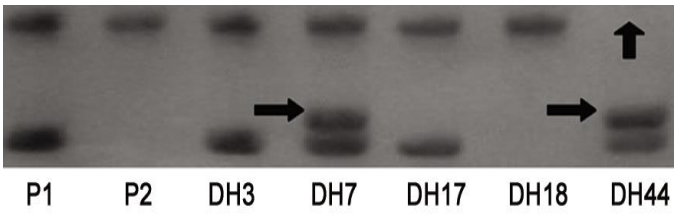

Fig 1. Novel and eliminated bands amplified by A125 primer pairs in an introgressed DH population of *B. napus*. Vertical arrows indicate eliminated bands (present in parents and some DH lines but absent in other DH lines). Horizontal arrows indicate novel bands (absent in the parents but present in some DH lines).

In the F_1 hybrids used to generate the introgressed DH population, 29 primer pairs (4.5% out of 647 primer pairs) presented with abnormal bands. These 29 primers amplified a total of 339 bands, including 270 parental bands (79.7%), 27 novel bands (8.0%) and 42 eliminated bands (12.8%). Thus, the proportion of novel bands generated in the F_1 hybrids was the same as that in the introgressed DH population (8.0% vs. 8.1%), while the ratio of eliminated bands in the introgressed DH population was three times higher than that in the F_1 hybrids (42.8% vs. 12.8%).

Chromosomal localization of fragments harboring SSR variants

Bands in the introgressed DH population amplified by 39 primer pairs (83.0% of the 47 primer pairs) could be mapped on to *B. rapa* (Fig 3a) or *B. oleracea* (Fig 3b) chromosomes. Analysis of the chromosomal localization of the variant sequences in the introgressed DH population revealed that the distribution of these fragments on chromosomes of *B. rapa* and *B. oleracea* was uneven. A total of 194 out of the 264 bands were mapped (73.5%): 142 bands onto chromosomes of *B. rapa* (53.8%) and 52 bands (19.7%) onto chromosomes of *B. oleracea*.

The most novel fragments were located on A1 (5.7% of all mapped bands), A3 (3.1%), and A9 (2.6%). The most eliminated fragments were mapped to A7 (12.4% of total mapped

Table 2. Comparative analysis of parental, novel, and eliminated fragments of mutated SSRs detected in an introgressed DH population of *B. napus.*

Motif type [a)]	Motif	No. E-motifs	No. N-motif	No. A-motif	No. P-motif	Total no. motifs	Percentage of total fragments
Mono-nucleotide	A	4	0	4	0	4	2.1
	T	4	6	10	1	11	5.7
Di-nucleotide	AC	1	0	1	0	1	0.5
	AG	11	2	13	8	21	10.9
	AT	18	2	20	2	22	11.5
	CT	7	0	7	0	7	3.6
	GA	8		8	1	9	4.7
	TA	12	2	14	4	18	9.4
	TC	5	0	5	0	5	2.6
	TG	3	0	3	0	3	1.6
Tri-nucleotide	ACA	1	0	1	0	1	0.5
	AGA	4	0	4	1	5	2.6
	ATG	3	0	3	0	3	1.6
	CAT	4	0	4	0	4	2.1
	CTC	3	0	3	0	3	1.6
	GAA	3	0	3	0	3	1.6
	GAT	2	4	6	6	12	6.3
	GGA	1	0	1	0	1	0.5
	GTG	3	0	3	0	3	1.6
	TAA	3	0	3	0	3	1.6
	TAT	3	0	3	0	3	1.6
	TGA	14	0	14	0	14	7.3
	TTA	1	0	1	0	1	0.5
Tetra-nucleotide	AAAT	1	4	5	3	8	4.2
	AAGA	1	4	5	3	8	4.2
	ATAA	1	0	1	0	1	0.5
	TGAT	1	4	5	6	10	5.2
Penta-nucleotide	CAAAT	1	0	1	0	1	0.5
	GTCTG	1	0	1	0	1	0.5
	TCCCT	1	0	1	0	1	0.5
	TGGTC	1	0	1	0	1	0.5
	TGGTT	3	0	3	0	3	1.6
	TTGAT	1	0	1	0	1	0.5
Total no. fragments	-	130	28	158	35	192	-
Percentage (%)	-	67.7	14.6	82.3	18.2	-	-

a)SSR mutation type. E-motif denotes the number of motifs in eliminated bands. N-motif refers to the number of motifs in novel fragments. A-motif refers to the number of motifs in abnormal fragments (the total number of E and N motifs). P-motif refers to the number of motifs in parental fragments.

bands), A9 (7.7%), and A1 (6.2%). Of the 52 mapped bands in the *B. oleracea* genome, none were novel bands, and the largest proportion of eliminated bands was located on linkage group C1 (6.7%), followed by C2 (5.7%), and C3 (3.6%). A single locus producing SSR variants in different plants of the introgressed DH population was categorized as a simple mutation locus. A total of 194 SSR variants were assigned to 71 simple mutation loci. One mutation cluster where all variants occurred together in a single chromosomal region was determined to be a single mutation event. In other words, the occurrence of one mutation event was predicted to lead to the generation of multiple linked SSR variants. Of the 71 simple mutation loci, 42 could be attributed to 27 mutation events (mutation clusters). In particular, one of the mutation events resulted in the emergence of 17 SSR loci variants on chromosome A1 of *B. rapa* (Fig 3a). In addition, six SSR variants could be attributed to a single mutation event on chromosome C2 of *B. oleracea* (Fig 3b).

Table 3. Nucleotide composition and 2-mer frequency in SSR flanking regions and SSR regions of natural and induced SSR fragments detected in an introgressed DH population of *B. napus.*

Nucleotide frequencies	Regions (Frequency ± SD)	
Nucleotide composition	Flanking	SSR
T(U)%	29.4 ± 5.3	33.0 ± 18.5
C	18.1 ± 4.0	2.9 ± 5.0
A	30.3 ± 4.9	41.4 ± 14.6
G	22.3 ± 5.1	22.7 ± 18.1
A&T	59.7	74.4
G&C	40.3	25.6
2-mer frequency		
si [a)]	7.8 ± 7.8	1.1 ± 1.7
sv [b)]	9.0 ± 9.7	1.8 ± 3.1
R	1.4 ± 1.5	0.7 ± 1.2
TT	46.2 ± 24.2	8.0 ± 7.0
AA	48.7 ± 27.4	10.5 ± 9.0
CC	26.8 ± 12.8	0.4 ± 0.8
GG	33.7 ± 16.4	6.3 ± 8.5
GC	0.8 ± 1.0	0.1 ± 0.3
TC	1.8 ± 2.1	0.2 ± 0.4
TA	1.8 ± 1.6	0.4 ± 0.7
TG	1.2± 1.5	0.3 ± 0.8
CT	1.8 ± 2.5	0.3 ± 0.7
CA	1.0 ± 1.2	0.2 ± 0.5
CG	0.8 ± 1.1	0.1 ± 0.2
AT	1.7 ± 2.1	0.4 ± 0.7
AC	1.1 ± 1.4	0.3 ± 0.7
AG	2.0 ± 2.2	0.5 ± 1.2
GT	1.0 ± 1.3	0.2 ± 0.5
GA	2.0 ± 2.3	0.3 ± 0.4
Total	171.9 ± 68.7	28.2 ± 19.9

The software MEGA 4 (Tamura et al. 2007) was used to detect the nucleotide composition of the fragments and transversion and transition ratios. [a)] si = Transitional Pairs. Transitions are interchanges within the purines (A and G) or pyrimidines (C and T). [b)] sv = Transversional Pairs. Transversions are interchanges between the purines (A or G) and pyrimidines (C or T). Transition mutations generated at higher frequency than transversions but are less likely to result in amino acid substitutions. Therefore, a higher sv indicates a higher possibility of an exchange in amino acid.

Dissection of SSR variants in the introgressed DH population

A total of 155 SSR variants that corresponded to 264 sequenced bands were detected: 116 perfect SSRs and 39 compound SSRs (≥ 2 motif compositions). One hundred and eleven SSRs with 41 different motifs were identified among the eliminated bands, whereas 20 SSR variants with eight different motifs were identified among the novel bands.

A total of 35 novel identified SSR fragments amplified by nine SSR primer pairs in the introgressed DH population were sequenced. We found that 23 (66%) of these novel fragments exhibited repeat number variation, 24% showed transition or transversion of the SSR motifs, and 10% displayed expansion or deletion of the flanking sequences of the repeat (see Supplementary Table 2 for details). For example, the SSR primer pair A125 gave rise to two parental bands (No. 1 and No. 3) and one novel band (No. 2) (Fig 4).

The highest frequency SSR variant motifs in the introgressed DH population were T (5.7%), AT (11.5%), AG (10.9%), TA

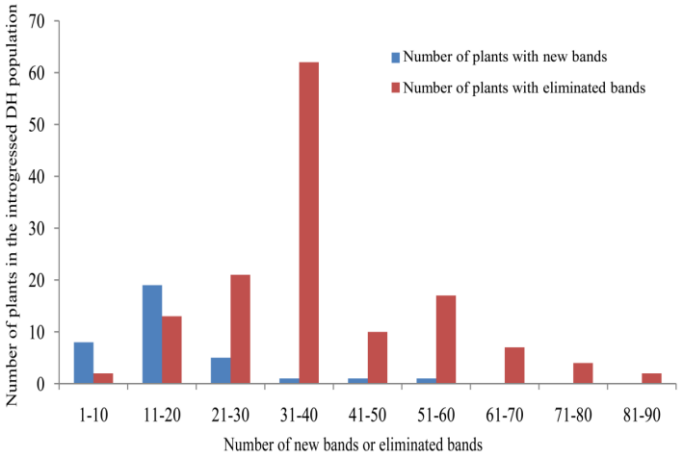

Fig 2. Frequency of plants with novel bands or eliminated bands present in an introgressed DH population of *B. napus*.

Table 4. DNA sequence diversity (± SD [a]) in both SSR regions and flanking regions of fragments subject to natural variation or induced variation in an introgressed DH population of *B. napus*.

Regions	Nucleotide Diversity	θ (per site) from η [b)]	Average number of nucleotide differences
Flanking; induced group	0.01 ± 0.01	0.01 ± 0.01	1.32 ± 1.13
Flanking; natural group	0.14 ± 0.16	0.16 ± 0.18	12.96 ± 13.35
SSR; induced group	0.01 ± 0.03	0.01 ± 0.03	0.4 ± 1.06
SSR; natural group	0.11 ± 0.11	0.11 ± 0.10	3.24 ± 5.38

[a)] SD= standard deviation. [b)] Eta (η) is the total number of mutations. Theta (θ) is Watterson's estimator of nucleotide diversity per site or per gene (based on Eta) and used as an index of evaluating mutability per site or per gene.

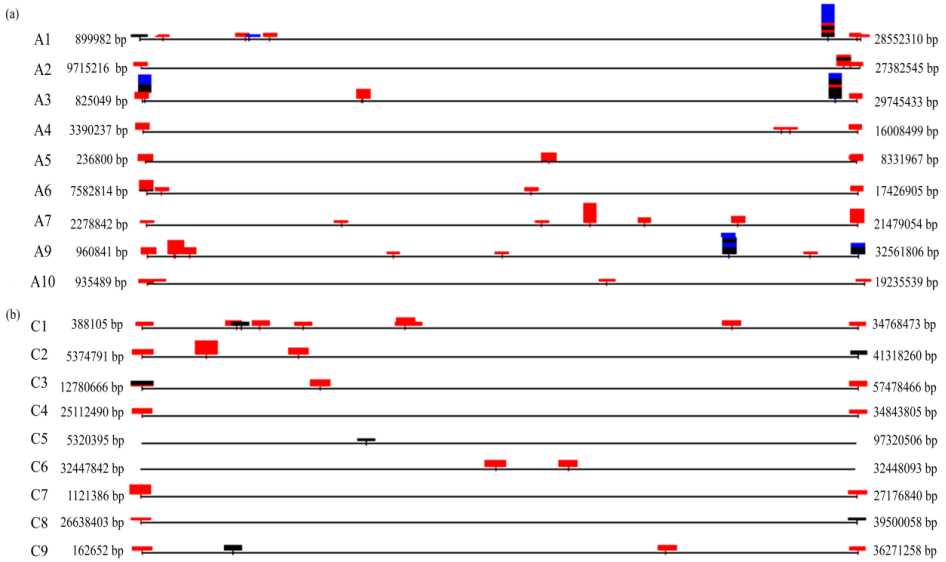

Fig 3. Localization of mutated fragments in an introgressed DH population mapped onto (a) *B. rapa* (2n=20) and (b) *B. oleracea* chromosomes (2n=18). The red, blue, and black symbols denote the eliminated fragments (E), novel fragments (N), and parental fragments (P), respectively. There are no parental or abnormal bands mapped onto A8 linkage group of *B. rapa*.

(9.4%), TGA (7.3%), GAT (6.3%), TGAT (5.2%), AAGA (4.2%) and AAAT (4.2%). The number of repeats of a given motif varied from 3 to 35, with a mode of 3 to 4. A significant negative correlation was observed between the number of repeats and the size of the motif (Table 1; r^2 = -0.5080, P < 0.0001). A negative correlation was also detected between size of the motif and the frequency of SSR variants (r^2 = -0.4292, P = 0.0127). Total SSR variant lengths ranged from 10 to 70 bp, with the highest frequency SSRs 12 bp long (44 variants). Long SSRs were rare (Fig 5), demonstrating that SSR variation was induced within a narrow range of SSR lengths.

SSRs with mononucleotide, dinucleotide, or trinucleotide motifs tended to vary in terms of the number of repeats; whereas for SSRs that contained tetranucleotide or pentanucleotide motifs, the flanking sequences and motif composition showed more variation. SSRs that contained dinucleotide motifs were most readily mutated, followed by SSRs with trinucleotide motifs and mononucleotide motifs (F = 5.91, P < 0.0001). A greater number of SSRs with dinucleotide or trinucleotide motifs were present in the parental species but disappeared in the DH lines than SSRs with other motifs. For SSRs with mononucleotide motifs, T was mutated more frequently than A, and no poly-C/G SSRs were mutated. Among dinucleotide-motif SSRs, SSRs that contained the AT motif were mutated the most frequently, followed by those that comprised TA and AG motifs. SSRs with A/T-rich tetranucleotide motifs generated novel SSRs readily, but G/C-rich pentanucleotide SSRs did not. These results indicated that SSRs with A/T-rich motifs mutated more readily than G/C-type SSRs (Table 2). Overall, the degree of SSR variation depended primarily on variation in repeat number and in the nucleotide composition of the SSR motifs, whereas the flanking regions of the SSRs were relatively conserved.

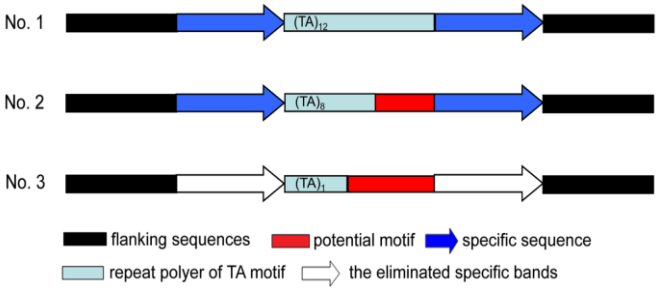

Fig 4. Patterns of SSR variants in an introgressed population and its parent *Brassica* species. The figure shows the two parental bands (No. 1 and No. 3) and a single novel band (No. 2) that were amplified by SSR primer pair A125 in line DH98. The results indicate that variation occurred not only in the number of repeat units but also within the flanking sequences.

Comparing natural SSR variants with induced SSR variants

To analyze the difference between natural SSR variants and induced SSR variants, we defined discrepancies between parental fragments that were amplified by a single primer pair

Table 5. Indel diversity (± SD [a]) in both SSR regions and flanking regions of fragments subject to natural variation or induced variation in an introgressed DH population of *B. napus*

Regions	Total Indel events	Average indel length	Indel Diversity	Theta per sequence[b]
Flanking; induced group	0.41 ± 0.59	44.56 ± 13.96	1.13 ± 0.35	1.13 ± 0.35
Flanking; natural group	2.69 ± 3.18	8.38 ± 8.75	2.05 ± 1.78	0.88 ± 1.63
SSR; induced group	0.86 ± 0.36	7.50 ± 6.02	1.00 ± 0.00	1.00 ± 0.00
SSR; natural group	1.50 ± 1.02	5.35 ± 3.78	1.39 ± 0.96	1.16 ± 0.77

[a] SD= standard deviation. [b] Eta (η) is the total number of mutations. Theta (θ) is Watterson's estimator of nucleotide diversity per site or per gene (based on Eta) and used as an index of evaluating mutability per site or per gene.

as natural SSR variants, and novel and eliminated fragments amplified by a single primer pair as induced SSR variants. We then calculated the composition and nucleotide pair frequencies of the SSRs and flanking regions for both the natural and induced fragment variants (Table 3). Composition analysis showed that the C/G content was higher in the flanking regions (40.3%) than in the SSR regions (25.6%) for both natural and induced fragments. The analysis of nucleotide 2-mer frequencies revealed a bias towards AA and TT in both the SSR motifs (10.5% AA and 8.0% TT) and the flanking regions (46.2% AA and 48.7% TT) for both natural and induced fragments (Table 3). Higher nucleotide diversity was detected in the flanking regions (mean value 1.7) than in the SSR regions (mean value 1.2). The calculated R (transition/transversion) value for transitional and transversional biases was lower in SSR regions (0.74 ± 1.16) than in flanking regions (1.38 ± 1.53), whereas the frequency of nucleotide mutation was higher in SSR regions than in flanking regions.

The nucleotide diversity of natural variants was higher than that of induced variants in both SSR motifs and flanking regions (Table 4). Indel (insertion and deletion) numbers and Indel diversity were also higher in both SSR regions and flanking regions of natural variants than in induced variants. However, the average Indel length within SSR regions was greater in induced variants than in natural variants (7.50 ± 0.62 vs. 5.35 ± 3.78). In flanking regions, the difference was even more striking: the average length of Indels in areas subject to induced variation was four times greater than in areas with natural SSR variants (44.56 ± 13.96 vs. 8.38 ± 8.75) (Table 5). The mean genetic distance of the natural variants group was greater than that of the induced variants group with respect to both SSR motifs and flanking regions (Table 6). The diversity of the fragments containing SSR motifs was higher than that of the flanking regions in both the natural variants group and the induced variants group. Thus, it can be seen that hybridization induced a more marked degree of variation than natural pressure, but the number and frequency of the incidences of variation accumulation were greater under natural pressure.

Functional annotation of mutated fragments in the introgressed DH population

All sequenced fragments were used to perform BlastN analysis against the *B. rapa* cDNA database ($E \leq 1.0 \times e^{-20}$) to judge whether these fragments were located in coding sequences. Abnormal SSRs were then compared with parental SSRs amplified by the same primer pair within coding sequences (47 primer pairs), and Blast2Go analysis of SSR variants used to determine putative effects on gene function. We found that while no SSR variants led to premature termination of the protein, 62.5% of SSR variants resulted in a frameshift mutation. In the remaining SSR variants, a deletion (25%) or insertion (12.5%) mutation resulted in a minor amino acid change

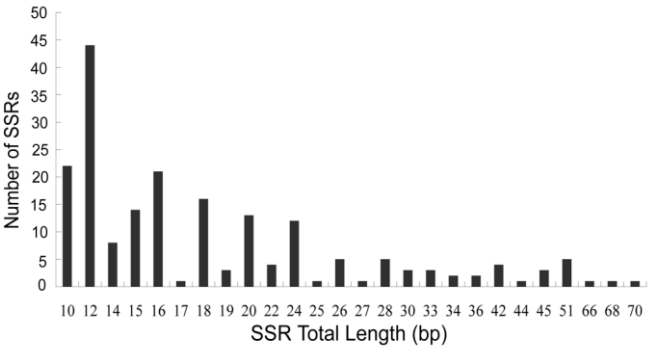

Fig 5. Number of SSRs of different total lengths (bp) detected in an introgressed population of *B. napus*.

without a change in protein composition. Blast2Go analysis told us that 75.0% of the SSR variants could result in changes in gene function, particularly the frameshift mutants. From this analysis, 198 SSR fragments (74.4%) were putatively involved in biological processes, 80 fragments (30.1%) had putative molecular functionality, and 150 fragments (56.4%) were putatively expressed in various cellular components.

The variant fragments harboring SSR variants within coding regions were involved in diverse biological processes (Fig 6a). Most fragments (44.2%) were related to metabolic processes, which included protein metabolic processes (7.3%) and DNA metabolic processes (6.3%) as well as other metabolic processes (33.3%). In addition, 8.3% of the fragments were involved in the regulation of metabolic processes, whereas 7.3% of the fragments participated in gene expression and 4.2% played a role in the regulation of gene expression. Other fragments were associated with other important biological processes, such as developmental processes (7.3%), reproduction (0.52%), and transcription (0.52%). For instance, the novel fragment Y66F145-7 was associated with the *AP2* domain transcription factor, which could affect the transcription of the AP2 gene and hence influence seed mass and seed yield (Jofuku et al. 2005). Some of the parental and abnormal fragments were associated with responses to stimulus (4.2%), such as the parental fragment cen39F196-3 and eliminated fragments cen39F197-5 and cen39F197-7 which may be related to *Fusarium oxysporum* resistance.

A high proportion of fragments harboring SSR variants within coding regions were involved in transferase (21.3%), indicating that important molecular functions, such as methyltransferase activity (the eliminated fragments A9F7-2, A9F8-2, A9F8-2 and A9F7-5 were related to histone methyltransferase) and the activation of retrotransposons (the eliminated fragments Y34F121-1 and Y34F121-2 were associated with the retroelement pol polyprotein), frequently occurred during the process of hybridization. Other common molecular activities of mutated fragments included catalytic

Table 6. Mean genetic distance (± SD) analysis of SSR-containing sequences in an introgressed DH population of *B. napus*.

Regions	Overall mean distance	Within groups		Between groups	Net between Groups
		Natural group	New group		
Flanking	2.60 ± 8.44	0.78 ± 1.59	0.06 ± 0.12	0.83 ± 1.73	0.41 ± 0.96
SSR	1.25 ± 4.86	6.30 ± 12.54	0.00 ± 0.00	4.65 ± 9.26	1.50 ± 2.99

Note: SD= standard deviation. The SSR motifs and flanking sequences of sequenced fragment were exacted. The genetic distances among SSR motifs and among the flanking sequences were calculated by MEGA 4 (Tamura et al. 2007) with the parameters set to the default values.

activity (15.0%), transporter activity (11.3%) (eliminated fragment Y34F123-8 was related to a potassium channel-like protein), kinase activity (8.8%), transcription factor activity (1.3%), and transcription regulator activity (1.3%) (Fig 6b). In addition, binding activities for macromolecules and micromolecules were poorly represented in general, but nucleic acid binding and ion binding activities were relatively well represented (6.3% and 5.0% respectively).

Analysis of the subcellular localization of the mutated fragments within coding regions showed that 11.0% of the mutated fragments were putatively expressed in the cytoplasm, 4.8% in the mitochondria, 3.2% in the cell wall, 3.2% in membranes, 2.0% in the nucleus, 0.4% in the vacuole, and the remainder in other cellular components (47.0%) or other organelles (29.0%) (Fig 6c). Overall, in the introgressed DH population, alterations (emergence of novel fragments or elimination of fragments) appeared to lead to functional alternation of many genes, which is likely a mechanism of species adaptability through generation of post-hybridization variation.

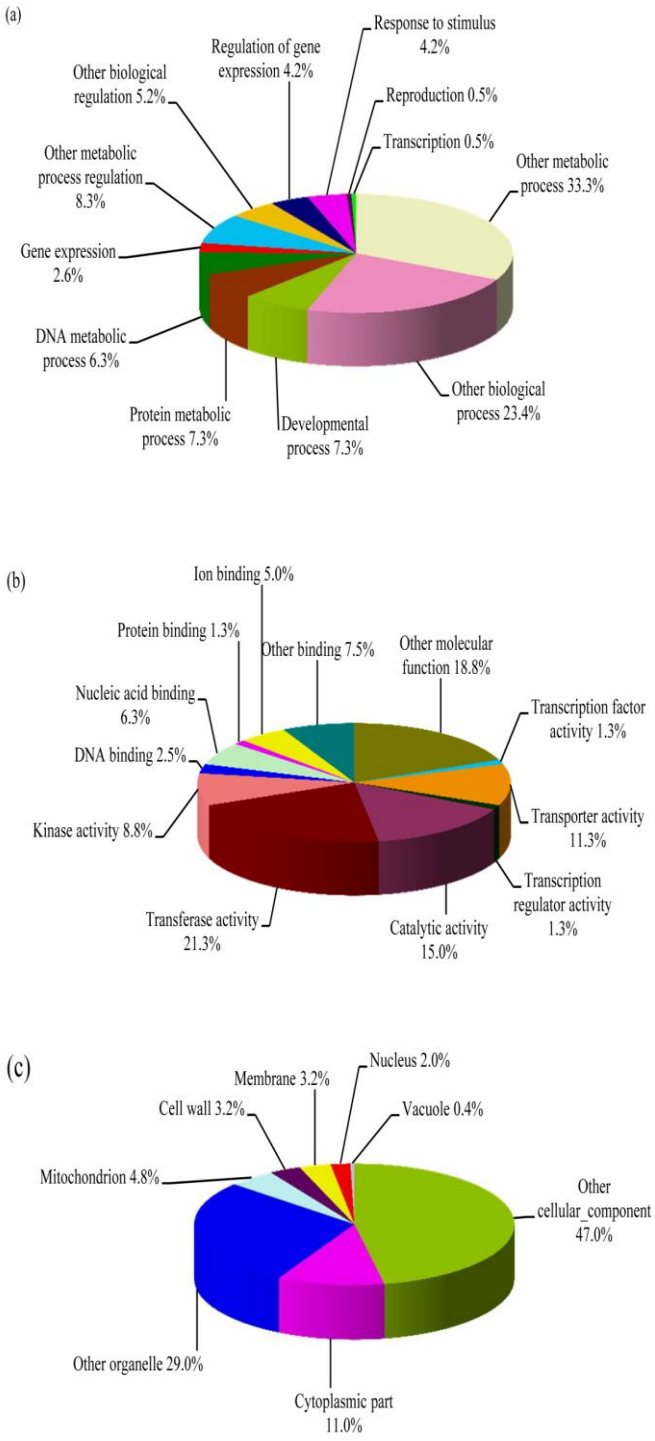

Fig 6. Putative functional annotation of mutated SSRs detected in an introgressed DH population of *B. napus* with respect to (a) biological process, (b) molecular function, and (c) cellular component localization by Blast 2GO.

Discussion

Interspecific hybridization is an important avenue for speciation and for genetic exchange between different species. Interspecific hybridization can induce major variation in genome sequences, including already mutation-prone SSR tracts (Tang et al., 2009). However, the characteristics of SSR variation induced by interspecific hybridization and the possible effects of SSR variation on gene function are not well understood. In this study, we found that novel SSR variants were consistently generated in both the F_1 generation and in the introgressed DH population, whereas the frequency of SSR-containing sequence elimination sharply increased from the F_1 hybrid generation to the introgressed DH population. Novel and eliminated fragments were unevenly distributed on chromosomes of *B. rapa* and *B. oleracea*, with the most novel fragments mapped to A1, A3 and A9, and the most eliminated fragments mapped to A7. SSR variation comprised mostly variation in repeat number, and varied as a result of nucleotide composition of the SSR motifs, with SSR flanking regions showing less variation. In the introgressed DH population, the genic sequences containing SSRs most commonly showed frameshift mutations, followed by deletion or insertion mutations. Most genes harboring SSR variants were related to key metabolic processes, such as protein or DNA metabolic processes. Hybridization induced a higher degree of SSR variation than natural pressure. These results help us to understand factors affecting SSR variation and the role of SSR mutation in interspecific hybridization.

Why interspecific hybridization can easily induce sequence variants

The elimination of many SSR fragments, as well as the appearance of novel SSR fragments, was detected in the F_1 hybrids

and in the DH population derived from hybridization between natural *B. napus* and resynthesized *B. napus*. These results were consistent with previous results of genomic change during interspecific hybridization detected using RFLP (Sharpe et al., 1995; Udall, 2003) and other markers (Szadkowski et al., 2010; Szadkowski et al., 2011). Hybridization between genetically distant species can induce substantial SSR variation (Tang et al., 2009). Similarly, rapid induced variation of tandem repeats, regulatory elements, and promoter regions has also been observed during interspecific hybridization between wheat and rye (Tang et al., 2008).

During the interspecific hybridization analyzed here, the elimination of sequences was observed more frequently than the appearance of novel sequences (42.8% vs. 8.1%), which was consistent with the results of studies in other species (Chen and Ni, 2006; Shaked et al., 2001). The rapid elimination of sequences is a general phenomenon in newly resynthesized allopolyploids of *Aegilops*, *B. napus* and *Triticum* (Gaeta and Chris Pires, 2010; Ma and Gustafson, 2008; Shaked et al., 2001; Sharpe et al., 1995; Szadkowski et al., 2010; Szadkowski et al., 2011; Udall, 2003).

So how does interspecific hybridization induce sequence variants? Interspecific hybridization has long been known to result in non-homologous recombination, which can easily result in chromosome rearrangements (inversions, translocations, duplications etc.) (Chen and Ni, 2006; Cheung et al., 2009; Gaeta and Chris Pires, 2010; Salmon et al., 2010; Szadkowski et al., 2010). Genetic control of non-homologous recombination is known to be poor in *Brassica* interspecific hybrids (Nicolas et al., 2009; Sharpe et al., 1995 Szadkowski et al., 2010; Udall, 2003). Hence, elimination of SSR alleles in our study is predicted to be due to chromosome fragment loss resulting from non-homologous recombination (putatively homoeologous recombination) (Gaeta and Chris Pires, 2010; Gaeta et al., 2007; Nicolas et al., 2007; Nicolas et al., 2009; Sharpe et al., 1995; Szadkowski et al., 2010; Szadkowski et al., 2011; Udall, 2003; Xiong and Pires, 2011). The mapping of SSR variants to the released *B. rapa* and *B. oleracea* genome sequence supports non-homologous recombination as the primary mechanism of sequence loss in our study.

Sequence elimination can also occur as a result of interspecific hybridization inducing transposon activations or bursts (Chen and Ni, 2006; Michalak, 2009; Parisod et al., 2009; Yaakov and Kashkush, 2011): the imprecise excision of transposable elements may cause excision of flanking sequences and subsequent loss of sequences (Chen and Ni, 2006; Huang et al., 2012; Wessler, 1988). Illegitimate recombination initiated by crossing over between two direct repeat sequences on the same monomer may also result in elimination of the sequences between the two repeats (circle excision and subsequent fragment loss) (Ozkan et al., 2001; Shaked et al., 2001).

Confounding factors that may contribute to novel and eliminated SSRs observed in this study include PCR artifacts and gene conversion. Chimeric PCR amplification products are generated from fusion of two partially homologous templates in one reaction, and this may occur as a result of PCR amplification of duplicated/homoeologous loci. In allotetraploid cotton (*Gossypium hirsutum*), 31.1% on average (0% - 89%) of products are affected by this mechanism (Cronn et al., 2002). Missing alleles may also result from competition for PCR amplification between different parental sequences present in a single amplification reaction (Cronn et al., 2002). Gene conversion between homoalleles may also induce sequence elimination of one allele. Gene conversion is a mode of genetic recombination in which one allele sequence replaces the homologous allele sequence, leading to loss of heterozygosity. This phenomenon was classically documented for homologous chromosomes but can also occur between homoeologous sequences in allopolyploids (Gaeta and Chris Pires, 2010; Salmon et al., 2010).

Finally, the frequent observation of eliminated bands in our study could be a result of one primer of the primer pair binding to a rearranged region, and the other primer binding to the original locus. If this happens, it is likely that no PCR products will be generated.

SSR variants can enable the development of new highly polymorphic SSR markers

In this study, the degree of SSR variation that was induced by genetically distant hybridization depended on the nucleotide composition of the SSR motif and the motif type. These conclusions are consistent with previous findings that SSR variation depends on the flanking sequence of the SSR, together with the number of repeat units, base substitutions, and Indels within repetitive and non-repetitive segments (Ellegren, 2000; Huang et al., 2012; Vigouroux et al., 2002).

With respect to induced SSR variants, SSRs that contained dinucleotide motifs mutated more readily than SSRs that contained trinucleotide motifs, whereas the opposite was true for natural variants in this study. We found that AT motifs were the most abundant, followed by TA and TGA motifs, among all the SSRs investigated in the DH population. In contrast, CCG repeats were found to be the most prevalent in the rice genome (Roorkiwal et al., 2009) whereas (GA)n and (TC)n were the most abundant in tea (Sharma et al., 2009) and (AT)n, (ATT)n, and (AAAT)n were the most abundant motifs in soybean (Song et al., 2010). In humans, the mutation rates in microsatellite loci that do not cause disease are lower for dinucleotide-motif SSRs, higher for trinucleotide-motif SSRs, and even higher for tetranucleotide-motif SSRs, whereas in SSRs that do cause disease, trinucleotide-motif SSRs have 3.9 - 6.9 times higher mutation rates than tetranucleotide-motif SSRs (Chakraborty et al., 1997). In any given species, some types of SSRs will be more prevalent than other types of SSRs. This suggests that certain types of SSRs may mutate more readily and spread across the genome of a particular species. Identification of these highly variable SSR types in different species might assist in the efficient design of highly polymorphic SSR markers. For instance, SSRs that contained AT, TA, or TGA motifs tended to mutate more readily than other SSRs in our study, and hence SSRs that contain AT and TA-rich motifs might generate more polymorphic markers in *Brassica*. With this new knowledge, it might be possible to avoid blind selection of SSRs and hence reduce associated screening costs and increase efficiency. Compared to the traditional method of SSR marker development (probe hybridization), development of SSR markers "*in silico*" from sequence data, especially from coding sequences, ESTs, unigenes or PUTs (Plant genomedatabase assembled unique transcripts) as in this study, is simple, less time-consuming and more informative (Iniguez-Luy et al., 2008).

Materials and Methods

Plant materials

Fig 7 displays the origin of an introgressed DH population. The resynthesized *B. napus* ($A^rA^rC^oC^o$, 2n=4x=38) was obtained from an interspecific cross between *B. rapa* (A^rA^r, 2n=2x=20) and *B. oleracea* var. *acephala* (C^oC^o, 2n=2x=18). This resynthesized *B. napus* was crossed with natural *B. napus* ($A^nA^nC^nC^n$, 2n=4x=38). Microspores from four hybrid F_1 plants

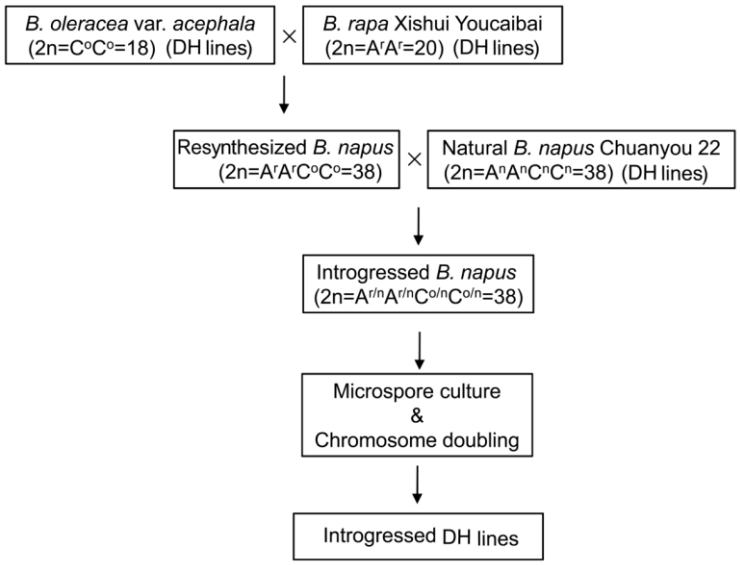

Fig 7. The origins of the introgressed *B. napus* DH population. An interspecific cross between *B. rapa* "Xishui Youcaibai" (A^rA^r, 2n=2x=20), a local seed-harvested double-haploid (DH) variety, and *B. oleracea* var. *acephala* (code M878) (C^oC^o, 2n=2x=18) which is a local yellow-seeded and horticulture-based DH variety, was obtained and subjected to chromosome doubling to produce the resynthesized *B. napus* ($A^rA^rC^oC^o$, 2n=4X=38) (code R029). This resynthesized *B. napus* was crossed with natural *B. napus* ($A^nA^nC^nC^n$, 2n=4x=38) (a DH line from a semi-winter cultivar Chuanyou 22, code N089). Microspores from four hybrid F_1 (codes H1, H6, H10 and H11) were then used for microspore culture with colchicine-induced chromosome doubling to generate an introgressed population with 112 lines.

(codes H1, H6, H10 and H11) were used for microspore culture with colchicine-induced chromosome doubling to generate the introgressed population. Resynthesized *B. napus* with a chromosomal number of 38 were selected after testing by cytological observation according to the method of Li et al. (1995) and Huang et al (2012). DNA from the introgressed population, the F_1 hybrids, and the parental lines, resynthesized *B. napus* (code R029) and natural *B. napus* (code N089), was analyzed to detect SSR variants. The introgressed DH population and their parental lines were grown in 2010 in the field at Xiema, located in Beibei, Chongqing, China.

Primer design

We screened 647 SSR primer pairs to detect polymorphic loci. Of these, 191 out of 314 SSR primer pairs came from the *B. rapa* genome (entitled 'H+code number') (http://www.brassica-rapa.org). In addition, 456 SSR loci were identified and appropriate primer pairs were designed by using the SSR Locator software (http://www.ufpel.tche.br/faem/fitotecnia/fitomelhoramento /faleconosco.html) (Da Maia et al., 2008). These primers included 38 pairs that were specific to centromeric regions (entitled 'Cen+code number'), 314 primer pairs that were designed from *B. rapa* genomic sequences available from the NCBI (The National Center for Biotechnology Information) (entitled 'A+code number'), and 104 primer pairs that amplified loci within seed-yield-specific genes (entitled 'Y+code number'). Primer sequences are provided in the Supplementary Table 3.

PCR amplification

All 647 SSR primer pairs were screened in resynthesized *B. napus* and the two parents to detect polymorphic loci. Of the 647 SSR primer pairs, 47 primer pairs amplified polymorphic alleles. Details of these primers are listed in the Supplementary Data (Supplemental Table 1).

The PCR reaction mix used to amplify the SSR regions contained 0.02 mmol L^{-1} dNTPs, 0.5 U of Taq DNA polymerase, 75 ng of template DNA, 5 mmol L^{-1} each SSR primer pair, and 1× DNA polymerase buffer (10 mM Tris pH 9.0, 50 mM KCl, and 1.5 mM $MgCl_2$). Touchdown PCR amplification was conducted with the following thermal profile: initial denaturation at 94°C for 5 minutes; 5 cycles of 30 s at 94°C, 45 s at 61°C with a decrease of 1°C per cycle, and 1 minute at 72°C; 26 cycles of 30 s at 94°C, 45 s at 57°C, and 1 minute at 72°C; and a final extension at 72°C for 10 minutes.

The PCR products were separated on 6% denaturing acrylamide gels and abnormal bands (bands that showed discrepancies between parents and progeny) were identified. Abnormal bands (the bands amplified in the DH population that were different from the bands amplified in the parents) were classified into eliminated bands, which were present in the parents but not in the DH lines, and novel bands, which were present in the DH lines but not in the parents. Bands present in both the parent cultivars and in the DH lines are hereafter referred to as "parental" bands and they were sequenced when abnormal bands occurred. All these bands were amplified repeatedly (eight times) to ensure that the abnormal bands really existed in the genome. The sequences of the 47 primer pairs that amplified abnormal bands are available in Supplementary Table 3, and the name and motif composition of each SSR locus is shown in Supplementary Table 2.

Sequencing and sequence analysis

Both the "abnormal" bands and the parental bands were cut out of the 6% denaturing acrylamide gel, 20 µL of deionized water was added and the gel slices incubated at 95°C for 10 minutes. The resulting supernatant was transferred to a PCR reaction mix for amplification with the original primers using the same PCR program described above. The PCR products were separated on 6% denaturing acrylamide gels and identified as target or non-target fragments. If the desired fragment was obtained, it was ligated into the vector pMD18-T (TaKaRa) and transformed into *Escherichia coli*. Positive clones were

selected for sequencing. The 264 sequences have been submitted to NCBI and the corresponding GenBank accession numbers ranged from JX131017 to JX131280. The software MEGA 4 (Tamura et al., 2007) was used to detect the nucleotide composition of the fragments, together with the genetic distance and transversion and transition ratios. DnaSP v5 (Librado PaR, 2009) was used to analyze the nucleotide diversity and Indel events in the sequenced fragments with the parameters set to the default values.

Detection of SSRs in the genome of introgressed B. napus

Simple sequence repeats (SSRs) were detected in the sequenced fragments by using the SSR Locator software (http://www.ufpel.tche.br/faem/fitotecnia/fitomelhoramento/faleconosco.html) (Da Maia et al., 2008) with the parameters set to the default values. All detected SSRs were classified into different types on the basis of the composition of their motif units and their flanking sequences.

For each particular SSR variant detected in the introgressed DH population, its corresponding repeat number, size of motif (motif type), total SSR length and frequency of SSR variants was determined. The data of all these SSR variants was collected to become a data array for a correlation analysis.

Chromosomal localization

The sequences of the abnormal and parental fragments obtained from the introgressed population were used as queries to perform a BlastN search against the *B. rapa* and *B. oleracea* DNA databases (http://brassicadb.org/brad/). Sequences with the highest scores and E values less than 10^{-10} were selected for *in silico* mapping onto *B. rapa* chromosomes (n=10) (Schranz et al., 2006) and *B. oleracea* chromosomes (n=9) (Panjabi et al., 2008; Xiong and Pires, 2011). After mapping, the chromosomal locations were drawn using the linDNA (linear map of DNA constructs) program of the Jemboss software, with the parameters set to default values (Rice et al., 2000). We determined the relationships of *Brassica* homoeologous chromosomes and blocks based on Parkin et al. (2003) and Panjabi et al. (2008). We combined the results of *in silico* mapping of sequences of bands amplified by a single primer pair with the results of *Brassica* homoeologous chromosomes and blocks to determine if bands were allelic (present on homologous chromosomes), homoeo-allelic (present on homoeologous chromosomes) or if they originated from non-hom(oe)ologous chromosomes.

Analysis of putative gene function

To determine whether these fragments were located in coding sequences, we used the sequenced fragments to perform BlastN analysis against the *B. rapa* and *B. oleracea* cDNA databases ($E \leq 1.0 \times e^{-20}$). Fragments located in coding regions were selected to conduct further analysis. We compared novel SSRs with amplified parental SSRs and translated to protein sequences using Jemboss (Rice et al., 2000). Multiple alignments of translated amino acids were carried out with ClustalW to analyze the structural variation of the SSR variants. In addition, functional changes were detected by Blastp analysis ($E \leq 1.0 \times e^{-10}$) in NCBI (http://blast.ncbi.nlm.nih.gov/). The putative functions of these abnormal sequences and corresponding parental sequences were analyzed by using Blast2GO (Conesa and Gotz, 2008) (www.hindawi.com/journals/ijpg/2008/619832.html) in order to investigate the potential role of the SSR variants. Putative functions were categorized by biological process involved, molecular functionality and cellular component localization.

Conclusions

In conclusion, interspecific hybridization induced extensive SSR variation. The SSR variants observed mainly comprised repeat number variants, followed by variants with transitions or transversions within the SSR motif, or variants with insertions or deletions within the SSR flanking sequences. Most mutated SSRs within coding regions were associated with vital metabolic processes, indicating that SSR mutations can affect gene function and the regulation of gene expression, which may be a means of increasing genome variation in organisms as an adaptive mechanism of species evolution. Analyses of SSR variants induced by interspecific hybridization could prove beneficial in understanding SSR origins and the effect of mutated SSRs on polyploid genomes, thereby enabling the efficient design of highly polymorphic SSR markers.

Acknowledgements

This work was supported financially by National Natural Science Foundation of China (code: 31260335), and Research Fund for the Doctoral Program of Higher Education of China (code: 20123603120002). ASM is supported by an Australian Research Council Discovery Early Career Researcher Award (DE120100668).

Supplementary data: This manuscript has Supplementary data which are available online.

References

Castillo A, Budak H, Varshney RK, Dorado G, Graner A, Hernandez P (2008) Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. BMC Plant Biol 8:97

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci 94 (3):1041-1046

Chen ZJ, Ni ZF (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28 (3):240-252

Cheung F, Trick M, Drou N, Lim YP, Park JY, Kwon SJ, Kim JA, Scott R, Pires JC, Paterson AH, Town C, Bancroft I (2009) Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. Plant Cell 21 (7):1912-1928

Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genom 2008:619832

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. Theor Appl Genet 104 (2-3):482-489.

Cuadrado A, Cardoso M, Jouve N (2008) Physical organisation of simple sequence repeats (SSRs) in *Triticeae*: structural, functional and evolutionary implications. Cytogenet Genome Res 120 (3-4):210-219

Da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FI, Costa de Oliveira A (2008) SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. Int J Plant Genom 2008:412-696

Dai LJ, Wang B, Zhao H, Peng JH (2013) Transferability of genomic simple sequence repeat and expressed sequence tag-simple sequence repeat markers from *sorghum* to

Miscanthus sinensis, a potential biomass crop. Crop Sci 53 (3):977-986
Eckert KA, Hile SE (2009) Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. Mol Carcinogen 48 (4):379-388
Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. Nat Genet 24 (4):400-402
Fajolu OL, Wadl PA, Vu AL, Gwinn KD, Scheffler BE, Trigiano RN, Ownley BH (2013) Development and characterization of simple sequence repeats for *Bipolaris sorokiniana* and cross transferability to related species. Mycologia 105 (5):1164-1173
Freimer NB, Slatkin M (1996) Microsatellites: evolution and mutational processes. Ciba Found Symp 197:51-67
Gaeta RT, Chris Pires J (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186 (1):18-28
Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19 (11):3403-3417
Golubov A, Yao Y, Maheshwari P, Bilichak A, Boyko A, Belzile F, Kovalchuk I (2010) Microsatellite instability in *Arabidopsis* increases with plant development. Plant Physiol 154 (3):1415-1427
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 Cell 23 (3):349-356
Huang P, Xu L, Liang W, Tam CI, Zhang Y, Qi F, Zhu Z, Lin S, Zhang B (2012) Genomic deletion induced by Tol2 transposon excision in zebrafish. Nucleic Acids Res. 41(2):e36
Iniguez-Luy FL, Voort AV, Osborn TC (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 (6):977-985
Jofuku KD, Omidyar PK, Gee Z, Okamuro JK (2005) Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proc Natl Acad Sci USA 102 (8):3117-3122.
Kantartzi SK, Ulloa M, Sacks E, Stewart JM (2009) Assessing genetic diversity in Gossypium arboreum L. cultivars using genomic and EST-derived microsatellites. Genetica 136 (1):141-147
Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends Genet 22 (5):253-259
Kovalchuk O, Kovalchuk I, Arkhipov A, Hohn B, Dubrova YE (2003) Extremely complex pattern of microsatellite mutation in the germline of wheat exposed to the post-Chernobyl radioactive contamination. Mutat Res 525 (1-2):93-101
Labbe JL, J., Murat C, Morin E, Le Tacon F, Martin F (2011) Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers. Curr Genet 57 (2):75-88
Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7 (2):R14
Leung SY, Chan TL, Chung LP, Chan AS, Fan YW, Hung KN, Kwong WK, Ho JW, Yuen ST (1998) Microsatellite instability and mutation of DNA mismatch repair genes in gliomas. Am J Pathol 153 (4):1181-1188
Li Z, Liu HL, Luo P (1995) Production and cytogenetics of intergeneric hybrids between *Brassica napus* and *Orychophragmus violaceus*. Theor Appl Genet 91:131-136
Librado PaR J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452
Liu T, Wahlberg S, Burek E, Lindblom P, Rubio C, Lindblom A (2000) Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer. Genes Chromosomes Canc 27 (1):17-25
Ma XF, Gustafson JP (2008) Allopolyploidization-accommodated genomic sequence changes in triticale. Ann Bot-London 101 (6):825-832
Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, Kelly JK (2009) Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (*Brassicaceae*). Heredity 103 (4):310-317
Masters BS, Johnson LS, Johnson BG, Brubaker JL, Sakaluk SK, Thompson CF (2010) Evidence for heterozygote instability in microsatellite loci in house wrens. Biol Lett 7 (1):127-130
Michalak P (2009) Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. Heredity 102 (1):45-50
Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30 (2):194-200
Moxon ER, Wills C (1999) DNA microsatellites: agents of evolution? Sci Am 280 (1):94-99
Narina SS, Buyyarapu R, Kottapalli KR, Sartie AM, Ali MI, Robert A, Hodeba MJD, Sayre BL, Scheffler BE (2011) Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (*Dioscorea alata* L.). BMC Genomics 12:100
Nicolas SD, Le Mignon G, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme R, Chalhoub B, Ryder CD, Chevre AM, Jenczewski E (2007) Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids. Genetics 175 (2):487-503
Nicolas SD, Leflon M, Monod H, Eber F, Coriton O, Huteau V, Chevre AM, Jenczewski E (2009) Genetic regulation of meiotic cross-overs between related genomes in *Brassica napus* haploids and hybrids. Plant Cell 21 (2):373-385
Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13 (8):1735-1747
Panjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes. BMC Genomics 9:113
Parida SK, Dalal V, Singh AK, Singh NK, Mohapatra T (2009) Genic non-coding microsatellites in the rice genome: characterization, marker design and use in assessing genetic and evolutionary relationships among domesticated groups. BMC Genomics 10:140
Parida SK, Yadava DK, Mohapatra T (2010) Microsatellites in *Brassica* unigenes: relative abundance, marker design, and use in comparative physical mapping and genome analysis. Genome 53 (1):55-67
Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in Spartina. New Phytol 184 (4):1003-1015
Park YJ, Lee JK, Kim NS (2009) Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. Molecules 14 (11):4546-4569
Parkin IAP, Sharpe AG, Lydiate DJ (2003) Patterns of genome duplication within the *Brassica napus* genome. Genome 46 (2):291-303

Rahim F (2008) In silico comparison of simple sequence repeats in high nucleotides-rich genomes of microorganism. Pakistan J Biol Sci 11 (20):2372-2381

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16 (6):276-277

Roorkiwal M, Grover A, Sharma PC (2009) Genome-wide analysis of conservation and divergence of microsatellites in rice. Mol Genet Genom 282 (2):205-215

Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol 186 (1):123-134

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the *Brassicaceae*: building blocks of crucifer genomes. Trends Plant Sci 11 (11):535-542

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13 (8):1749-1759

Sharma RK, Bhardwaj P, Negi R, Mohapatra T, Ahuja PS (2009) Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). BMC Plant Biol 9:53-76

Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38 (6):1112-1121

Shin J, Yuan Z, Fordyce K, Sreeramoju P, Kent TS, Kim J, Wang V, Schneyer D, Weber TK (2007) A del T poly T (8) mutation in the 3' untranslated region (UTR) of the CDK2-AP1 gene is functionally significant causing decreased mRNA stability resulting in decreased CDK2-AP1 expression in human microsatellite unstable (MSI) colorectal cancer (CRC). Surgery 142 (2):222-227

Shin KH, Park KH, Hong HJ, Kim JM, Oh JE, Choung PH, Min BM (2002) Prevalence of microsatellite instability, inactivation of mismatch repair genes, p53 mutation, and human *papillomavirus* infection in Korean oral cancer patients. Int J Oncol 21 (2):297-302

Sol Mateo M, Mollejo M, Villuendas R, Algara P, Sanchez-Beato M, Martinez-Delgado B, Martinez P, Piris MA (1998) Analysis of the frequency of microsatellite instability and p53 gene mutation in splenic marginal zone and MALT lymphomas. Mol Pharmacol 51 (5):262-267

Song QJ, Jia GF, Zhu YL, Grant D, Nelson RT, Hwang EY, Hyten DL, Cregan PB (2010) Abundance of SSR Motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci 50 (5):1950-1960

Subramanian S, Madgula VM, George R, Kumar S, Pandit MW, Singh L (2003) SSRD: simple sequence repeats database of the human genome. Comp Funct Genom 4 (3):342-345

Sun LD, Yang WR, Zhang QX, Cheng TR, Pan HT, Xu ZD, Zhang J, Chen CG (2013) Genome-wide characterization and linkage mapping of simple sequence repeats in Mei (*Prunus mume* Sieb. et Zucc.). PLoS One 8 (3)

Szadkowski E, Eber F, Huteau V, Lode M, Coriton O, Jenczewski E, Chevre AM (2011) Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. New Phytol 191 (3):884-894

Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chevre AM (2010) The first meiosis of resynthesized Brassica napus, a genome blender. New Phytol 186 (1):102-112

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599

Tang Z, Fu S, Ren Z, Zou Y (2009) Rapid evolution of simple sequence repeat induced by allopolyploidization. J Mol Evol 69 (3):217-228

Tang ZX, Fu SL, Ren ZL, Zhou JP, Yan BJ, Zhang HQ (2008) Variations of tandem repeat, regulatory element, and promoter regions revealed by wheat-rye amphiploids. Genome 51 (6):399-408

Udall J (2003) A genetic study of oilseed *Brassica napus* L. Mapping chromosome rearrangements and quantitative trait loci. PhD dissertation University of Wisconsin

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JS, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19 (8):1251-1260

Wang A, Yu Z, Ding Y (2009) Genetic diversity analysis of wild close relatives of barley from Tibet and the Middle East by ISSR and SSR markers. C R Biol 332 (4):393-403

Wang HY, Wei YM, Yan ZH, Zheng YL (2007) EST-SSR DNA polymorphism in durum wheat (*Triticum durum* L.) collections. J Appl Genet 48 (1):35-42

Wessler SR (1988) Phenotypic diversity mediated by the maize transposable elements *Ac* and *Spm*. Science 242 (4877):399-405

Xiong ZY, Pires JC (2011) Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. Genetics 187 (1):37-49

Yaakov B, Kashkush K (2011) Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. Genome 54 (1):42-49