Quantity, Distribution, and Evolution of Major Repeats in *Brassica napus*

6

Nomar Espinosa Waminal, Sampath Perumal, Shengyi Liu, Boulos Chalhoub, Hyun Hee Kim and Tae-Jin Yang

Abstract

Repetitive elements (REs) play major roles in genome organization, size, and evolution, but are often underrepresented in genome assemblies. The recent genome assembly of the allotetraploid *Brassica napus* genome revealed that 48% of the genome comprised REs, including transposons and tandem repeats. In the present work, we show the overall quantity and comparative analyses of major repeat families in both the assembled and unassembled portions of the reference *B. napus* genome. We surveyed the abundance, distribution, diversity, and dynamics of ten major REs in the *B. napus* genome, which represented less than 1% of the total 1130 Mb *B. napus* genome in the current assembly. However, in silico mapping of raw whole-genome sequence reads from nine *B. napus* accessions revealed about 11% of the genome as represented by these ten repeat families. Comparative analyses of these major repeats showed their evolutionary dynamics in the *B. rapa* (A_r), *B. oleracea* (C_o), and *B. napus* (A_nC_n) genomes as well as a considerable inter- and intraspecies repeat diversity among different *B. napus* accessions. Cytogenetic mapping of these major repeats showed their genomic abundance and distribution, with some families having a conserved subgenomic distribution pattern in the *B. napus* genome. Finally, the impact of genetic changes to REs and their corresponding epigenetic readjustments during *B. napus* evolution are also discussed in this chapter.

N. E. Waminal · S. Perumal · T.-J. Yang (✉)
Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea
e-mail: tjyang@snu.ac.kr

N. E. Waminal · H. H. Kim
Department of Life Science, Chromosome Research Institute, Sahmyook University, Seoul 01795, Republic of Korea

S. Liu
Key Laboratory of Biology and Genetic Improvement of Oil Crops, the Ministry of Agriculture, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430062, China

B. Chalhoub
Institute of Crop Science, Zhejiang University, Hangzhou 310058, China

B. Chalhoub
Institut National de Recherche Agronomique (INRA)/Université d'Evry Val d'Essone, Unité de Recherche en Génomique Végétale, UMR1165, Organization and Evolution of Plant Genomes, 2 rue Gaston Crémieux, 91057 Evry, France

6.1 Introduction

We have come to a pinnacle in the history of genomics when enormous volumes of nucleotide sequence information can be gathered cost

S. Liu et al. (eds.), *The Brassica napus Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-43694-4_6

effectively in a short time (Soltis et al. 2013; Schatz et al. 2012; Zhou et al. 2010), allowing the genome assembly of many important crop species at an unprecedented pace and accuracy. This permits a better understanding of the overall genome landscape and provides sufficient evidence leading to the negation of the pejorative 'junk' status of repetitive DNA fractions in a genome (Eddy 2012). In contrast, these repeats are fundamental components for the holistic function of a cell (Fedoroff 2012; Plohl et al. 2008; Freeling et al. 2015).

Despite advancements in next-generation sequencing (NGS) technology and assembly algorithms, efficient genome assemblies of repetitive sequences are still hindered especially when using short NGS sequences (Alkan et al. 2011; Claros et al. 2012; Hamilton and Buell 2012). This is even more difficult in plants, which often harbor abundant repetitive DNA and experienced various extents of polyploidization, (for reviews, see Schatz et al. 2012; Claros et al. 2012; Mihai et al. 2002; Renny-Byfield and Wendel 2014). Due to abundant REs such as tandem repeats (TRs) and transposable elements (TE), NGS-based genome assembly causes shrinkage of the actual repeat copies in a genome, even down to a few copies in an assembly (Macas et al. 2007; Alkan et al. 2011; Claros et al. 2012; Mihai et al. 2002; Schatz et al. 2012; Tang et al. 2015). Moreover, polyploidy or whole-genome duplication (WGD) and small-scale duplications further exacerbate this challenge by creating duplicate copies of genes or larger genomic regions. This redundancy drives mis-assemblies that could occur along these large-scale duplications (Paterson and Wendel 2015; Claros et al. 2012).

For example, in the two diploid progenitors of *Brassica napus*, *B. rapa* and *B. oleracea*, about 40 and 38% of respective genomes have not been included in pseudo-molecules, mainly due to REs (Waminal et al. 2015, 2016). Evidently, REs and polyploidy greatly influence the quality of genome assemblies and ultimately the acquisition of high-resolution pseudo-molecules. Although longer reads are offered by single-molecule, or third-generation sequencing technologies, they are still insufficient in resolving mega-base length tandem repeat regions (Schatz et al. 2012; Schadt et al. 2010). Consequently, densely heterochromatic regions such as the centromere and pericentromere have very little representation, or none at all, in some reference genome assemblies, even for model plants such as rice and *Arabidopsis* (Gao et al. 2015).

Although REs remain largely unassembled and unexplored in many sequenced plant genomes (Michael and Jackson 2013; Liu et al. 2014; Wang et al. 2011b), it holds a plethora of information about chromosome and genome dynamics, gene regulation, genome evolution, and epigenetic control (Biemont 2010; Biémont and Vieira 2006; Nowak 1994; Chadwick 2009; Melters et al. 2013; Mehrotra and Goyal 2014). Hence, they deserve a fair genome-wide analysis. Understanding their genomic distribution would provide a more enhanced comprehension of the holistic genome landscape and origin. Moreover, they complement studies in structural and functional genomics (Biemont 2010; Wang et al. 2011a; Choi et al. 2014).

B. napus ($2n = 4x = 38$, $A_nA_nC_nC_n$ genome) is an allopolyploid oilseed crop that formed within the past 7,500 years through hybridization between its progenitor genomes, *B. rapa* ($2n = 2x = 20$, A_rA_r genome) and *B. oleracea* ($2n = 2x = 18$, C_oC_o genome). The recent release of the allopolyploid *B. napus* genome revealed an aggregated 72 × genome multiplication since the origin of angiosperms (Chalhoub et al. 2014). This advance has provided a suitable foundation for deeper understanding of the dynamics of its REs through comparative studies with its progenitor diploid species, *B. rapa* and *B. oleracea*. In the present study, we surveyed the genomic abundance, chromosomal distribution, diversity, and dynamics of the major *Brassica* repeats in nine *B. napus* accessions (Table 6.1). We further discussed the role of epigenetic readjustments and its interplay with genetic changes in response to allopolyploidization.

Table 6.1 Summary of *Brassica napus* accessions used for the survey of major repeats[a]

ID	Accession/cultivar	Origin/type	Amounts (Mbp)	Genome coverage (x)
Bn-1	Zhongsuang11	Winter rapeseed	2,126.6	1.9
Bn-2	M083	Semi-winter rapeseed	1,273.8	1.1
Bn-3	Aburamasari	Asian (Japan) oilseed rape	13,900.7	12.3
Bn-4	Aviso	European oilseed rape	8,679.2	7.7
Bn-5	Darmor-bzh	European winter oilseed rape	6,029.7	5.3
Bn-6	Siberian kale	Kale	13,323.1	11.8
Bn-7	*B. napus* 'H165'	Resynthesized	15,928.4	14.1
Bn-8	Rutabaga	Swede sensation	14,923.5	13.2
Bn-9	Yudal	Asian (Korea) oilseed rape	14,105.3	12.5

[a]All the sequences above were provided by Shengyi Liu and Boulos Chalhoub

6.2 Repeats in the A_r, C_o, and A_nC_n Reference Genomes

Assembled sequences represented 58, 82, and 75% of total genome sizes of 485 Mb, 630 Mb, and 1130 Mb for *B. rapa*, *B. oleracea*, and *B. napus*, respectively (Fig. 6.1) (Wang et al. 2011a; Chalhoub et al. 2014; Liu et al. 2014). Of these sequences, repetitive DNAs represented 23, 41, and 35% for *B. rapa*, *B. oleracea*, and *B. napus*, respectively (Table 6.2). Relative to their respective estimated genome sizes, these values were reduced to 13, 32, and 23% (Table 6.2 and Fig. 6.1). Non-REs representing euchromatic regions covered 45, 50, and 52% of total estimated genome sizes of *B. rapa*, *B. oleracea*, and *B. napus*, respectively (Fig. 6.1).

Of the repeats in the reference assemblies, TEs represented the bulk fraction with 97, 96, and 97% followed by TR with 2, 3, and 2%. The rest were unclassified sequences (Fig. 6.1). While retrotransposons (Class I TE) were more abundant than DNA transposons (Class II TE) in *B. oleracea* and *B. napus*, the reverse was observed in *B. rapa* (Table 6.2). Among Class I TEs, LTR retrotransposons represented the majority with *Ty1/Copia* being more abundant than *Ty3/Gypsy* in both diploids (4.13 vs. 3.42% in *B. rapa* and 10.85 vs. 8.86% in *B. oleracea*), while about the same amount was present in *B. napus* (8.05 vs. 8.18%). Both diploid progenitors had different major Class II TEs. *Helitrons* were more abundant than *CACTA* elements in *B. rapa* (3.74 vs. 1.94%), but the reverse was observed in *B. oleracea* (3.96 vs. 5.55%). Accordingly, both elements had a similar representation in *B. napus* (3.69 vs. 3.83%).

Obviously, a considerable proportion of REs have not yet been included in the assembled sequences. In the following sections, we analyzed the genome proportion of several reported RE families in the current assembly and in the total whole-genome sequences (WGS), and we checked for types of repeats captured in both assembled and unassembled fractions. The RE families that were used in this analysis represented less than 1% of the current assembly; therefore, the values we obtained in this work mostly reflect the portions in the unassembled fraction. While TEs were more abundant in the assembled fractions, TRs were most prevalent in the unassembled, representing 41, 47, and 35% in *B. rapa*, *B. oleracea*, and *B. napus*, respectively (Fig. 6.1).

6.3 The Major Repeats of the *Brassica* Genomes

The difficulty in accounting for RE sequences in genome assembly scaffolds often leaves them in the 'hidden fraction' of a genome. REs can be subcategorized into dispersed or tandem repeats based on genomic distribution (Cizkova et al. 2013; Heslop-Harrison 2000; Plohl et al. 2012).

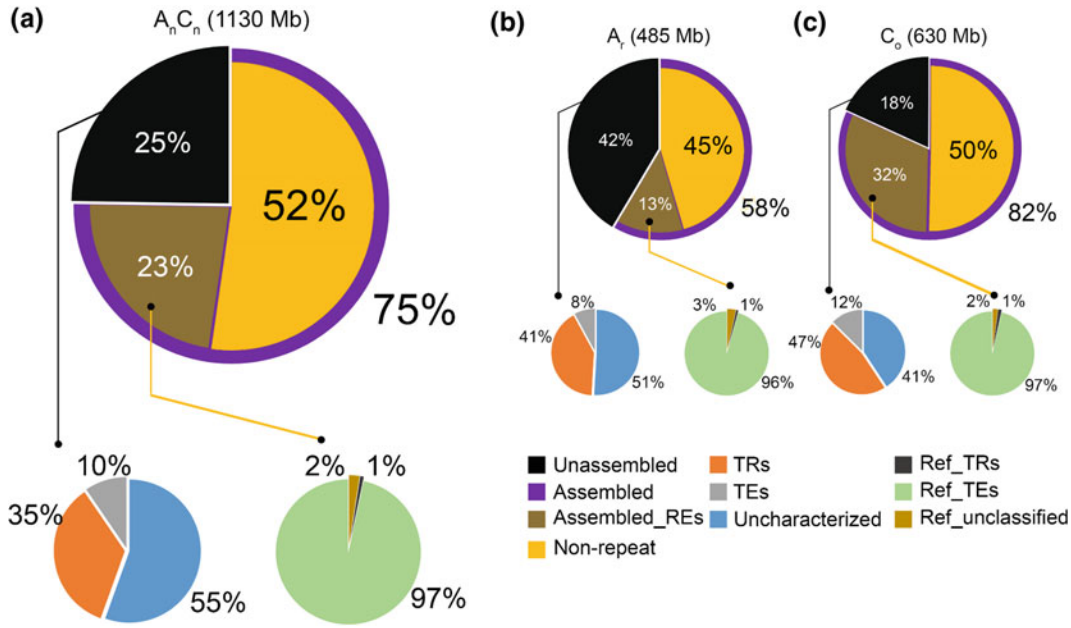

Fig. 6.1 Genomic proportions of both the assembled and unassembled sequences in the *Brassica napus* and its diploid progenitors. **a** Large inner pie chart represents the estimated total non-repeat genic fraction (*yellow slice*) and repeat fraction (*brown slice*) of *B. napus* genome. The outer doughnut chart represents the percentage of assembled (*purple slice*) and unassembled (*black slice*) fractions relative to the estimated genome size of 1130 Mb as calculated by Chalhoub et al. (2014). Smaller pie charts at the bottom left and right summarize the REs in the unassembled and assembled genome fractions, respectively. **b**, **c** Same diagrams for *B. rapa* and *B. oleracea*, respectively

Dispersed repeats include TEs, which are distributed in the entire genome, subgenome, or specific chromosomal regions (Choi et al. 2014; Lim et al. 2007). TRs are organized in a head-to-tail arrangement in distinct chromosomal regions (Coluccia et al. 2011; Sharma et al. 2013).

Most sequenced plant genomes contain a large proportion of Class I TE, mostly of the LTR superfamily (Michael and Jackson 2013). While up to 50% of TRs (Plohl et al. 2008) has been reported; the highest representation are mostly those with centromeric origins (Melters et al. 2013). In this chapter, we focused our survey on the major repeats representing Class I and Class II TEs, structural satellite repeats, and housekeeping ribosomal RNA genes in the *B. napus* genome. The analysis of major repeats in its progenitor genomes provides an enhanced understanding in choosing major repeat elements for the survey of the *B. napus* genome (Waminal et al. 2015, 2016). The major repeats include centromeric satellite repeats (CentBn1 and CentBn2), rDNA tandem repeats (5S and 45S), subtelomeric repeats (BnSTRa and BnSTRb), centromeric retrotransposon in *Brassica* (CRB; Lim et al. 2007), pericentromeric retrotransposon specific to *B. rapa* (pCRBr; Lim et al. 2007), and dispersed LTR and TIR elements specific to the *B. oleracea* genome, BoCopia and BoCACTA, respectively.

Centromeric repeats of *B. napus* are categorized into two groups, CentBn1 and CentBn2, and are homologous to their diploid progenitors (CentBr1/CentBr2 and CentBo1/CentBo2 from *B. rapa* and *B. oleracea*, respectively) (Perumal et al. 2017). Collectively, we refer to them as centromeric repeats of *Brassica* (CentB). The 5S and 45S rDNA sequences are conserved among the A_r, C_o, and A_nC_n. The subtelomeric satellite repeats, BnSTRa and BnSTRb, have orthologs in both A_r (Waminal et al. 2015) and C_o (Waminal et al. 2016) genomes. Collectively, we refer to them as *Brassica* subtelomeric repeats, BSTRa

Table 6.2 Total RE fractions in reference genomes of *Brassica rapa*, *B. oleracea,* and *B. napus*

	Brassica rapa 'Chiifu'[b]				*Brassica oleracea* 'capitata'[c]					*Brassica napus* '*Darmor-bzh*'[d]		
Transposon categories[a]	Copy numbers	Coverage (bp)	% in Ref.	GP (%)	Copy numbers	Coverage (bp)	% in Ref.	GP (%)	Copy numbers	Coverage (bp)	% in Ref.	GP (%)
Class I												
LTR/*Copia*	30,349	11,292,047	4.13	2.33	68,842	51,827,340	10.85	8.23	97,152	59,466,274	8.05	5.26
LTR/*Gypsy*	19,229	9,327,740	3.42	1.92	48,169	42,322,494	8.86	6.72	68,354	60,377,324	8.18	5.34
LTR/Unknown	11,358	3,768,473	1.38	0.78	21,405	11,626,899	2.43	1.85	31,172	16,050,602	2.17	1.42
SINE	4,248	549,493	0.20	0.11	8,359	1,122,599	0.23	0.18	12,053	1,638,766	0.22	0.15
LINE total	7,090	3,260,523	1.19	0.67	10,345	6,792,364	1.42	1.08	16,003	10,516,160	1.42	0.93
Subtotal	72,274	28,198,276	10.33	5.81	157,120	113,691,696	23.79	18.05	224,734	148,049,126	20.05	13.10
Class II												
hAT	14,722	2,915,918	1.07	0.60	22,113	5,022,703	1.05	0.80	36,356	7,950,833	1.08	0.70
CACTA	17,742	5,289,213	1.94	1.09	40,258	26,517,864	5.55	4.21	58,323	27,277,074	3.69	2.41
PIF/Harbinger	2,057	581,852	0.21	0.12	3,994	1,375,937	0.29	0.22	5,853	1,790,014	0.24	0.16
Tc1/Mariner	13,307	3,022,686	1.11	0.62	18,119	5,622,275	1.18	0.89	34,488	8,342,337	1.13	0.74
MITE/Tourist	16,867	2,720,339	1.00	0.56	29,593	6,967,515	1.46	1.11	51,770	9,356,495	1.27	0.83
DNA-Unknown	29,919	7,453,903	2.73	1.54	43,933	13,054,872	2.73	2.07	75,309	19,547,930	2.65	1.73
Helitron	46,182	10,206,949	3.74	2.10	65,310	18,932,278	3.96	3.01	107,645	28,297,585	3.83	2.50
Subtotal	140,796	32,190,860	11.79	6.64	223,320	77,493,444	16.22	12.30	369,744	102,562,268	13.89	9.08
Tandem repeats	1,879	565,650.00	0.20	0.12	262,840	53,622	1.61	0.53	331,850	97,713.30	0.01	0.01
Unclassified		2,183,677	0.80	0.45		4,105,469	0.86	0.65		6,383,948	0.86	0.56
Subtotal	1,879	2,749,327	1.00	0.57	262,840	4,159,091	0.86	1.18	331,850	6,481,661	0.88	0.57
Total	214,949	63,138,463	23.11	13.02	643,280	195,344,231	40.87	31.53	926,328	257,093,055.30	34.82	22.75

[a]Adapted from Chalhoub et al. (2014)
[b]Estimated genome size: 485 Mb, assembly size: 283.8 Mb, without N gaps: 273.1 Mb
[c]Estimated genome size: 630 Mb, assembly size: 515.34 Mb, without N gaps: 477.85 Mb
[d]Estimated genome size: 1130 Mb, assembly size: 850.3 Mb, without N gaps: 738.3 Mb

and BSTRb (Perumal et al. 2017). Centromeric retrotransposon of *Brassica* (CRB), a *Ty1/copia* LTR retrotransposon, which is mostly associated with heterochromatic regions and is intermingled with CentB (Lim et al. 2007) in the Oleracea lineage (A_r and C_o) chromosomes, remained conserved in the A_nC_n genome. Sequences of genome-specific transposons such as the pericentromeric retrotransposon of *B. rapa* (pCRBr), *B. oleracea Ty1/Copia* retrotransposon (BoCopia), and *B. oleracea* CACTA transposon (BoCACTA) remained conserved in A_nC_n genome (Lim et al. 2007; Perumal et al. 2017; Waminal et al. 2016).

6.4 Genomic Abundance and Distribution of Major Repeats in *B. napus*

Plant genomes sequenced to date have considerable amount of unassembled fractions (Michael and Jackson 2013). Assembly statistics often only provide a general view of what was efficiently anchored in the genome assembly; hence, it does not provide the proportional genomic abundance of these elements based on the actual genome content (Waminal et al. 2015; Schatz et al. 2012). Reasonably, estimating their abundance can be achieved by read mapping of WGS reads on the repeat unit sequence and by molecular cytogenetic mapping through fluorescence in situ hybridization (FISH). These approaches will elucidate their proportional abundance, genomic distribution, and impact for evolution (Waminal et al. 2015, 2016, 2018; Schatz et al. 2012; Lee et al. 2017).

6.4.1 In Silico WGS Read Mapping and Cytogenetic Mapping

Based on current available data, only about 75% of the 1130 Mbp *B. napus* genome was assembled into scaffolds (Fig. 6.1). Of these scaffolds, only 57% (645 Mbp) was anchored unambiguously into pseudo-chromosomes (Chalhoub et al. 2014). Among the ten repeat families included in this survey (Table 6.3), BSTRb and BSTRa had the highest and second-highest genome proportion (GP) in the pseudo-chromosomes, respectively, followed only, but with considerably lower copy numbers, by CentB1 and CentB2. No rDNA or TE sequences were represented in the anchored assembly, except for the truncated BoCopia element. From the unanchored sequences in the current assembly, several copies of 5S rDNA, and a single copy each of *CACTA* and pCRBr elements were added to the total captured REs (Table 6.3). Nevertheless, still no 45S rDNA and CRB were included in the total assembly. Their long sequence, highly repetitive nature, and pericentromeric location could help explain their exclusion in the current assembly (Pop and Salzberg 2008; Wang et al. 2011a). Altogether, these major repeats covered only about 2.3 Mbp or less than 1% in the current assembly (Table 6.3). As expected, this assembly did not provide robust information about the proportional abundance of major REs in the *B. napus* genome.

Upon read mapping of WGS to representative sequences of these elements, the captured elements increased dramatically to about 124 Mb, which is equivalent to about 11% of the genome (Table 6.3). All elements were well represented, even the 45S rDNA, which was not represented in the assembly, was about double the copies of both BSTRs together. In fact, it had the second-highest genome proportion (2.7%), second only to CentB1 (3.6%). In terms of copy number, the shortest elements, CentB1 and CentB2, had the most numbers (228,030 and 51,093, respectively), and the dispersed BoCopia had the least (284 copies). Overall, based on total accumulated length of all ten repeat elements, the in silico WGS read mapping captured 55 times more than what was present in the current assembly. This corroborates the observed accumulation of major repeats in the unassembled portions.

In the current *B. napus* genome assembly, about 35% comprises TEs alone—excluding TRs (Chalhoub et al. 2014). This represented 97% of the total REs included in the assembly (Fig. 6.1). With the RE families in this work, we captured 11% of the total genome which represented 45%

Table 6.3 Summary of major repeat content in the *Brassica napus* 'Darmor-*bzh*' reference genome assembly and in WGS sequence of nine accessions

Repeat element	Unit length (bp)	Amounts in *B. napus* reference genome							Estimated amounts in 1 × WGS (1130 Mbp)[a]			GP, FISH (%)	GA (%)[c]	Element source
		Pseudo-molecule (645 Mb)		Unanchored scaffolds (205 Mb)		Total (850 Mb)		GP[b] (%)	Copy	Length (Kb)	GP[b] (%)			
		Copy	Kb	Copy	Kb	Copy	Kb							
CentB1	177	307	51.3	29	4.7	336	55.9	0.005	228,030.9 (±60748.1)	40,361.5 (±10752.4)	3.572	8.0	0.14	Liu et al. (2014)
CentB2	177	313	51.1	205	33.7	518	84.8	0.008	51,092.9 (±16975.8)	9043.4 (±3004.7)	0.800	2.0	0.94	Liu et al. (2014)
5S nrDNA	501	–	–	45	22.1	45	22.1	0.002	5146.9 (±1712.7)	2578.6 (±858.1)	0.228	0.9	0.86	Waminal et al (2015)
45S nrDNA	7456	–	–	–	0.0	–	–	0.000	4088.8 (±2353.2)	30,485.1 (±17546.5)	2.698	5.3	0.00	Waminal et al (2015)
BSTRa	350	1217	408.0	300	101.0	1517	509.0	0.045	20,348.8 (±10066.5)	7122 (±3523.3)	0.630	2.7	7.15	This study
BSTRb	351	3716	1251.5	916	317.6	4632	1569.1	0.139	23,141.8 (±12753.7)	8122.7 (±4476.6)	0.719	4.1	19.32	This study
CRB	5908	–	–	–	0.0	–	–	0.000	1168.4 (±303.1)	6901.5 (±1790.1)	0.611	2.8	0.00	Liu et al. (2014)
pCRBr	8395	–	–	1	8.1	1	8.1	0.001	960.4 (±406.9)	8216.6 (±3468.8)	0.727	1.9	0.10	Lim et al. (2007)
BoCopia	6711	1	5.6	–	0.0	1	5.6	0.000	284.4 (±52.6)	1909.8 (±353.3)	0.169	1.0	0.29	This study
BoCACTA	7675	–	–	1	7.6	1	7.6	0.001	1265.6 (±171.6)	9713.1 (±1318.5)	0.860	2.5	0.08	Alix et al. (2008)
Total		5554	1767.5	1496	496.7	7050	2254.1	0.199	335,528.9 (±89796.8)	124,454.5 (±17634.4)	11.014	31.2	1.81	

[a]Repeats were estimated based on the *in silico* read mapping analysis on 9 *B. napus* accessions and the average values were represented to 1 × WGS coverage (850 Mb)
[b]Genome proportion (GP): (total amounts in Kb/1130,000 Kb) × 100
[c]Genome assembly (GA) repeats: percentage of repeat amounts in reference genome compared with the estimated amounts in 1 × WGS
Kb Amounts in kilo base pair, *SD* standard deviation

of the total unassembled fractions. The remaining 55% of this fraction needs to be analyzed further. Considering that only four TEs and six TRs were used in this survey could partially explain why only 11% of the genome was captured. Apparently, it may be necessary to also check other elements, especially other non-LTR retrotransposons, *Ty1/Copia*, other *CACTA* elements, miniature TEs, and *Helitron*s, which are well represented in the current assembly (Chalhoub et al. 2014). Other members may not have been captured in the assembly and not included in our current analysis. However, an underestimation is possible considering several limitations and biases in identifying and estimating REs through computational analyses (Macas et al. 2007; Schatz et al. 2012; Treangen and Salzberg 2011).

To address this concern in quantifying genomic abundance while simultaneously checking genomic distribution of major REs, we carried out molecular cytogenetic analysis using FISH with the ten REs as probes. By signal-to-whole-chromosome area ratio, FISH facilitated estimation of this hidden fraction to about 31% (Table 6.3). CentB1 occupied the largest genomic portion followed by 45S rDNA, a pattern in congruence with that observed through WGS mapping. There is a general proportional increase of repeats estimated through molecular cytogenetics compared with those from in silico analysis. However, it is important to note that it is possible that these results could be an overestimate considering the wider area covered by fluorescence than the actual physical size. Thus, there is room for the development of more accurate RE quantification approaches. However, presently both WGS read mapping and FISH present a plausible approach toward this objective.

6.4.2 Cytogenetic Mapping of Major Repeats Discriminates Subgenomes and Individual Chromosomes

Identification of individual chromosomes is necessary for integrating genetic linkage groups and physical maps (Jiang and Gill 2006), and in understanding the dynamics of genomes in comparative cytogenomic studies (Iourov et al. 2008), particularly in the context of crop improvement. However, this has often been difficult especially among crops with small chromosome sizes, monomorphic chromosome arm ratios, and similar chromosome lengths (Waminal et al. 2012; Pich et al. 1995). Even with the availability of several routinely used cytogenetic markers such as 5S and 45S rDNA, which are localized to only a few chromosomes, other chromosomes are often difficult to identify. This is further aggravated by polyploidy, which increases chromosome number (Vrána et al. 2015). For instance, identifying subgenomes in *B. napus* has proven difficult due to the high homology of the A_n and C_n subgenomes (Snowdon et al. 1997). Although genomic in situ hybridization (GISH) studies have discriminated these two subgenomes, clear distinct discriminating signals are often difficult to obtain (Snowdon et al. 1997; Howell et al. 2008). Meanwhile, genomic distribution of major repetitive DNA has shown potential for identifying individual chromosomes and in resolving subgenomes without GISH (Macas et al. 2007; Hribova et al. 2010; Alix et al. 2008; Choi et al. 2014).

The C_n subgenome-specific hybridization of BoCACTA and BoCopia elements enabled easy and accurate discrimination between A_n and C_n subgenomes without many background signals using a general FISH procedure without the need for block DNA such as needed in GISH (Fig. 6.2; Alix et al. 2008). This was particularly useful in discriminating the overlapping chromosome lengths of the shorter chromosomes of C_n from longer chromosomes of A_n. It is important to note that although C_n chromosomes are generally longer than A_n, the shorter chromosomes of C_n such as C_n09 could be difficult to distinguish from those in A_n such as A_n07.

Another important method to accurately identify chromosomes is multicolor-FISH (Koo et al. 2004; Wang et al. 2012). This allows mapping of several probes (five probes in this case) in one FISH experiment (e.g., Kato et al. 2004), and if chromosomes are in good condition,

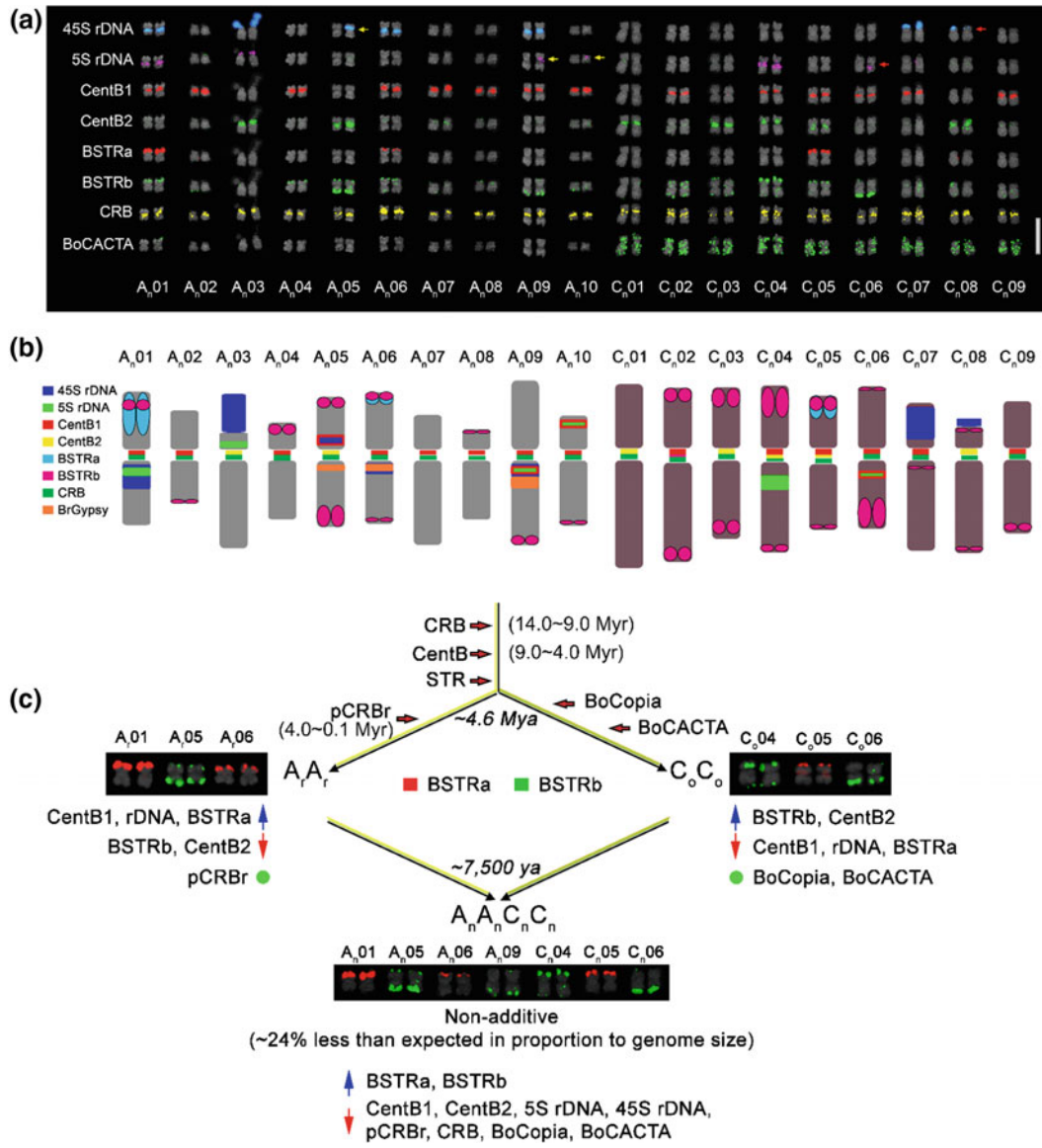

Fig. 6.2 Genomic distribution and evolution of major repeats in *Brassica napus*. **a** Karyogram of *B. napus* based on the distribution of major DNA repeats. Yellow and red arrows indicate major chromosomal rearrangements within the A_nA_n and C_nC_n subgenomes, respectively. Note that the C_nC_n chromosomes, although fewer in number, are generally larger than those of the A_nA_n, reflecting the genomic difference between the two diploid species. CRB is seen in all chromosomes while BoCACTA elements are specific to C_nC_n subgenome. Bar = 10 μm. **b** Karyotype idiogram of *B. napus* rDNA with red border represents hemizygous loci, most likely from homeologous unequal crossover. Darker chromosomes of the C_n subgenome indicate preferential hybridization of BoCACTA and BoCopia transposable elements. **c** Evolutionary dynamics of *Brassica* major repeats. Blue and red arrows indicate GP increase and decrease, respectively. Green circles indicate subgenome specificity of repeats

slides can be reprobed (Jiang and Gill 2006) up to four or five times; thus, increasing the number of probes ($5 \times 4 \sim 5 = 20 \sim 25$) to be analyzed in a shorter period of time while allowing more accurate characterization of individual chromosomes from a single chromosome spread. In this

approach, illegitimate recombinations involving REs could also be easily detected by comparing signal patterns from different probes. For example, an apparent loss of a 45S rDNA locus in one A_n05 homolog resulted to a hemizygous 45S rDNA A_n05 locus (Fig. 6.2). Additionally, another 45S rDNA locus at C_n08 had an unbalanced copy number between the two homologs as manifested by a considerably reduced signal in one homolog. Similar patterns were observed in some 5S rDNA loci (A_n09 and A_n10). Compared with diploid *B. rapa*, the A_n06 STRa locus was more reduced (Waminal et al. 2015). These physically observed changes in locus size could be explained by homologous recombination-mediated unequal crossovers, and tandem repeats are hot spots of these events (Kolomietz et al. 2002; Plohl et al. 2012).

Hence, cytogenetic mapping of these ten REs enabled an estimation of their genome abundance, easy discrimination of subgenomes, and identification of individual chromosomes and some associated illegitimate recombinations. A summary of individual chromosome features observed through FISH analysis is listed in Table 6.4.

Table 6.4 Features of individual *Brassica napus* chromosomes based on cytogenetic mapping of major repeats

A_n01	5S and 45S rDNA loci colocalized on the pericentromeric area of the long arm, CentBo1 on centromere, major BoSTRa/b locus at the telomeric area of the short arm, a weak proximal BoSTRc signal on the short arm, CRB
A_n02	CentBo1 on centromere and a weak BoSTRc on the subtelomere of the long arm, CRB
A_n03	NOR at the short arm, 5S rDNA at proximal region of short arm, CentBo2 on centromere, no STR signals, CRB
A_n04	CentBo1 on centromere, BoSTRc on subtelomeric region of both arms with stronger signal on short arm, CRB
A_n05	CentBo2 on centromere, major BoSTRc at telomeric region of long arm, and another weaker BoSTRc on the subtelemeric region of the short arm, hemizygous 45S rDNA translocation on pericentric region of short arm, pCRBr, CRB
A_n06	45S rDNA locus at the pericentromeric area of long arm, CentBo1 on centromere, the major BoSTRa/b locus at telomeric region of short arm is reduced compared to its ortholog in *B. rapa* (A_r), weak colocalized BoSTRc locus on short arm, pCRBr, CRB
A_n07	CentBo1, CRB
A_n08	CentBo1, weak BrSTRb at telomeric region of short arm, CRB
A_n09	Increased 45S rDNA signal at the intercalary region of the long arm when compared with its ortholog in *B. rapa*, CentBo1 on centromere, BoSTc at telomeric region of long arm, pCRBr, CRB
A_n10	Weak 5S rDNA locus at intercalary region of short arm, CentBo1, weak BoSTRc on long arm, CRB
C_n01	Increased CentBo2 signal compared to its paralog in *B. oleracea*, no other readily observable repeat signals, CRB
C_n02	CentBo1 signal, weak centromeric BoSTRc and telomeric BoSTRc on both arms, CRB
C_n03	CentBo2, telomeric BoSTRc at both arms, CRB
C_n04	With 5S rDNA at the pericentromeric area of long arm, CentBo1 and CentBo2, telomeric BoSTRc at both arms with major signal on short arm, CRB
C_n05	With CentBo1 and CentBo2, centromeric and telomeric BoSTRa/b on short arm being the only major BoSTRa/b signal, weak telomeric BoSTRc, CRB
C_n06	CentBo1, major BoSTRc signal on long arm, CRB
C_n07	45S rDNA on short arm, CentBo1, CRB
C_n08	Weak 45S rDNA on short arm, CentBo2, weak BoSTRc on both arms, CRB
C_n09	CentBo1 and weak CentBo2, weak telomeric BoSTRc on long arm, CRB

6.5 Comparative Repeatomics Reveals the Dynamics of Major Repeats in *Brassica* Species

The ancestral karyotype of the family Brassicaceae consisted of eight chromosomes which underwent several rounds of genome duplication and subsequent lineage-specific rearrangements, particularly involving REs, resulting to nine chromosomes of *B. oleracea* being larger than those of the ten chromosomes of *B. rapa* (Liu et al. 2014; Lysak et al. 2006). Comparative analysis of major repeats between *B. oleracea* and *B. rapa* genomes provides two scenarios of RE dynamics. The first suggests a continuous amplification of TEs and TRs in the *B. oleracea* genome over time after its divergence with *B. rapa* about 4.6 million years ago, increasing the genome size of *B. oleracea* (630 Mbp) to more than that of *B. rapa* (540 Mbp) (Liu et al. 2014). The second posits a rapid loss of transposable elements, e.g., BoCACTA, in *B. rapa* during divergence from *B. oleracea*, which was possibly driven by a slower reestablishment of epigenetic control that could have prevented homology-dependent, illegitimate recombination-induced repeat loss in *B. rapa* (Fedoroff 2012; Kelly et al. 2015).

During allopolyploidization, the merging of two genomes often results in genomic shock (Fedoroff and Bennetzen 2013; Fedoroff 2012; Renny-Byfield et al. 2013). This consequently initiates genome reprogramming by altering epigenetic makeup. Although the exact mechanisms and timeframe by which these events happen is not yet fully understood (Fedoroff 2013a), we know that this often leads to genome downsizing through elimination of DNA segments (Renny-Byfield et al. 2013; Renny-Byfield and Wendel 2014), often repetitive DNA fragments, a process aimed at reestablishing stable meiotic pairing and fertility in incipient allopolyploids (Fedoroff 2012; Renny-Byfield and Wendel 2014). In the absence of genome downsizing and element amplification in allopolyploids, an additive number of elements relative to the diploid progenitors can be expected. However, genome downsizing after allopolyploidization seems to be a rule rather than an exemption, although increased genome sizes have been reported (Renny-Byfield et al. 2013). DNA loss could even be biased toward a specific subgenome such as those observed in *Nicotiana tabacum* (Renny-Byfield et al. 2012), and allopolyploid cotton (Paterson et al. 2012). Consequently, the resulting allopolyploid often has a unique genomic make up relative to the diploid progenitors.

Accordingly, eight of the ten *B. napus* repeat elements in this survey showed a non-additive reduction of size; in fact, about a 24% GP reduction than what was expected relative to the genome sizes of the diploid progenitors (Figs. 6.2 and 6.3). The assembled centromeric repeats were the most reduced, followed by 45S rDNA, pCRBr, 5S rDNA, BoCACTA, BoCopia, and CRB (Fig. 6.3). However, BSTRs show non-additive amplification in the *B. napus* genome compared with its diploid progenitors, with BSTRb having more copies than BSTRa (Figs. 6.2 and 6.3). Satellite DNA regions are amplified/contracted in a very short evolutionary time as a result of unequal crossover between homologous sequences (for review on satellite DNA evolution, Plohl et al. 2012). Moreover, 45S rDNA loci are often targets of rapid locus elimination and reorganization among polyploids (Pellicer et al. 2010b, c). An increase of BSTRs in *B. napus* may have added benefits and consequently could have undergone positive selection, whereas other extra elements of other repeat families may not be necessary at all (Plohl et al. 2012).

Aside from interspecific variations with its diploid progenitors, copy number and GR size variation among the nine *B. napus* accessions were also observed (Fig. 6.4). Seven of the nine accessions showed relatively similar amounts of REs. However, two accessions, Bn-1 and Bn-2, generally had much lower TR copies, although they have much more 45S rDNA, compared with the other seven accessions (Fig. 6.4a, c). Additionally, centromeric and pericentromeric retrotransposons were more abundant in these two accessions (Fig. 6.4b, d). A similar intra-species repeat number variation was reported among

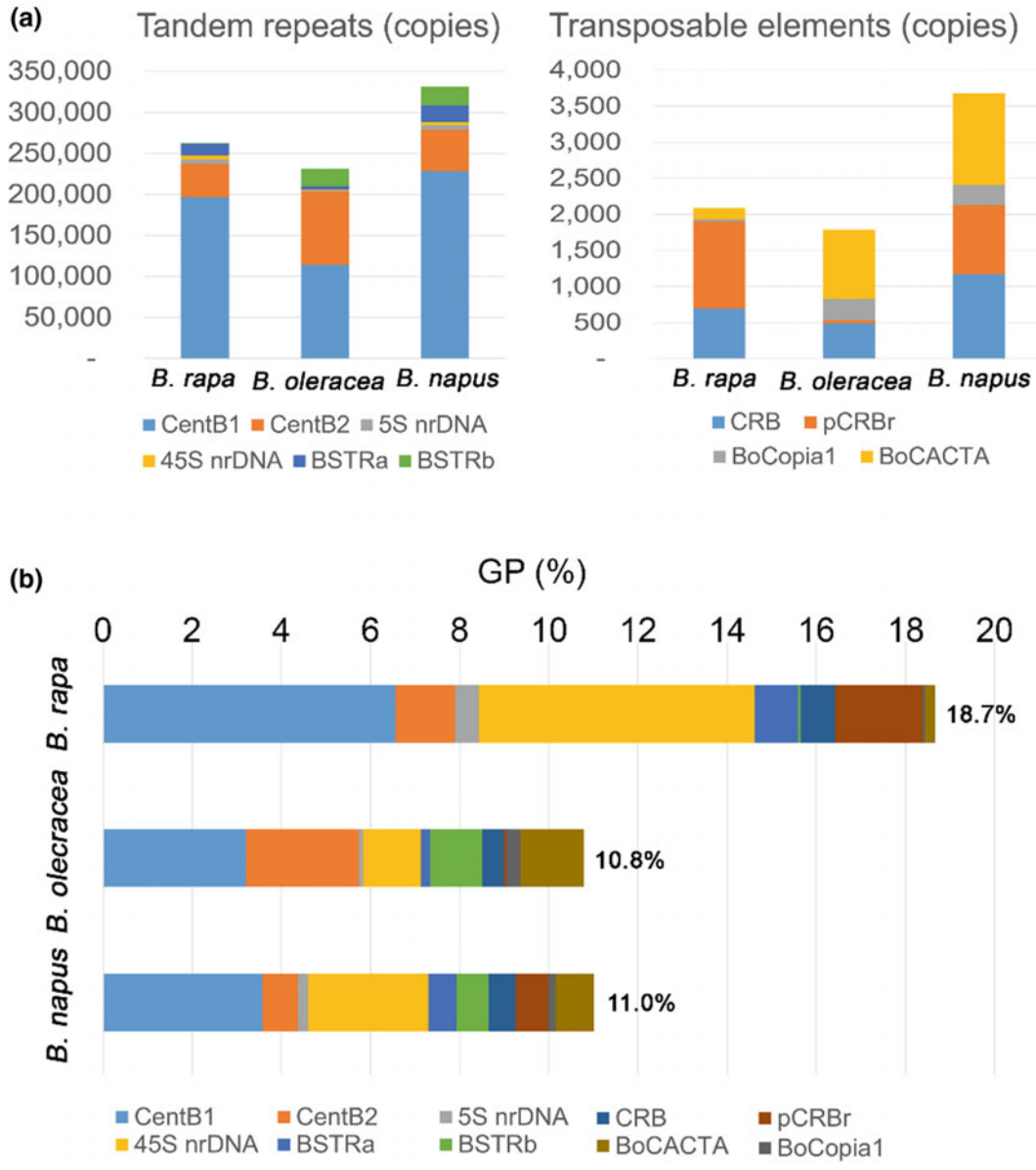

Fig. 6.3 Comparative analysis of major repeats in *B. rapa*, *B. oleracea*, and *B. napus*. **a** Copy number of each repeat family in each genome analyzed. Values for TR and TE are shown in separate charts. **b** Genome proportions of major repeats for each genome analyzed

several *B. oleracea* morphotypes (Perumal et al. 2017). In this previous study, some morphotypes, such as cauliflower and broccoli, had more CentBo1 than CentBo2. Some morphotypes, or accessions, apparently have unique RE compositions. We are aware of the limitations of in silico analysis in quantifying these repeats, which could have contributed to the observed value differences, especially considering the fewer WGS reads used in Bn-1 and Bn-2 (Table 6.1). However, the stark higher abundance of 45S rDNA and total TE in these two *B. napus* accessions indicate RE abundance independent from the amount of random WGS reads used.

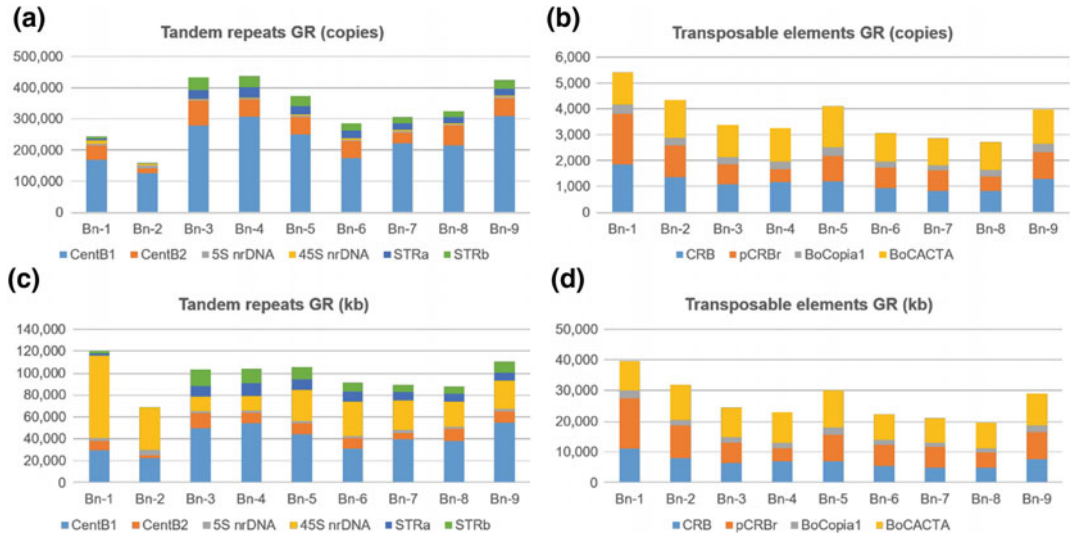

Fig. 6.4 Summary of major repeat composition in the nine *B. napus* accessions based on reference mapping. **a**, **b** Copy number-based genomic representation of tandem repeats and transposable elements among nine *B. napus* accessions. **c**, **d** Corresponding length in kilobase from **a**, **b**

Consequently, the impact of variation in RE abundance warrants further analyses, especially when taking into account a previous report that demonstrated a link between TE abundance variation and environmental adaptation (Kalendar et al. 2000).

Meanwhile, genome specificity of some TEs, as observed in the diploid progenitors, has been retained within the *B. napus* genome. BoCopia and BoCACTA retained their C_n specificity, and pCRBr its A_n specificity (Figs. 6.2 and 6.3). How certain elements are retained in different subgenomes, in the context of allopolyploidization, can be explained by epigenetic control mechanisms (Plohl et al. 2012; Fedoroff 2012), which will be discussed further in the following section.

6.6 Evolutionary Implications for Fluctuation of RE Fraction

Genome size variation within and among species, regardless of organism complexity or the number of protein coding genes, is commonly known as the C-value paradox (Pagel and Johnstone 1992; Freeling et al. 2015; Rosbash et al. 1974; Eddy 2012), and is attributed, as mentioned above, mainly to the size fluctuation of REs copy number. For instance, differential accumulation of several TE families, including tandem repeats, has defined the genomes of several *Fritillaria* species in the absence of WGD (Kelly et al. 2015). This phenomenon also caused intraspecies genome size variations in *Helianthus annuus* and other plants (Price and Johnston 1996; Wendel and Wessler 2000). Moreover, growing evidence supports the importance of this genomic fraction in proper genome function and evolution (Wei et al. 2013; Nowak 1994; Shapiro and von Sternberg 2005; Pardue and DeBaryshe 2003; Hall et al. 2005; Biémont and Vieira 2006; Freeling et al. 2015; Kalendar et al. 2000), particularly regarding their significant roles in chromosome segregation, gene expression, and heterochromatin maintenance (Pardue and DeBaryshe 2003; Biemont 2010; Sarilar et al. 2011; Sampath et al. 2013; Wolfgruber et al. 2009; Goodier et al. 2012; Peng and Karpen 2008). An important study by Kalendar et al. (2000) revealed the link between RE content fluctuation among individuals within a species and environmental adaptation. These examples demonstrate the adaptive and evolutionary importance of REs.

Studying the fluctuation of RE fraction is therefore an invaluable approach in understanding phylogenetic relationships, since genomic changes are quantifiable and can reveal variations among accessions and species. For example, the RE composition of *B. rapa* and *B. oleracea* are unique for each species (Waminal et al. 2015), so are the RE compositions among accessions in *B. oleracea* (Waminal et al. 2016) and *B. napus*. However, in *Brassica*, no studies have yet shown the direct phenotypic impact of RE fraction size variation to crop biology.

In *Brassica*, heterochromatins are mostly localized in centromeric and pericentromeric regions (Lim et al. 2007), where most REs are localized (Fig. 1A, B). CRB is a common centromeric component of the *B. rapa* (A_r), *B. nigra* (B_n), and *B. oleracea* (C_o) genomes. However, the absence of CentB hybridization in *B. nigra* supports the earlier divergence of the B_n genome from the A_r and C_o genomes (Lim et al. 2007; Koo et al. 2011; Arias et al. 2014). However, FISH analysis of *Brassica* STR showed genome-specific evolution of these subtelomeric repeats (Waminal et al. 2016) since their divergence. BSTRa seemed to be 'preferentially' selected in the *B. rapa* genome compared with BSTRb, while the opposite was observed in *B. oleracea*. This eventually led to a greater abundance of BSTRb than BSTRa in *B. napus* after the genome merger. Mechanisms that control their retention or elimination are being studied in more detail (Fedoroff 2012; Fablet and Vieira 2011).

Understanding how REs are controlled is necessary to exploit their underlying potential for crop improvement. Studies on sophisticated plant epigenetic control mechanisms, (Haag and Pikaard 2011; Slotkin and Martienssen 2007; Fedoroff 2012; Bennetzen and Wang 2014) for example, have elucidated this objective. DNA and histone modifications, which have a central feedback control mechanism involving siRNAs, are at the core of genome dynamics regulation to ensure genome homeostasis (see Haag and Pikaard 2011; Peng and Karpen 2008; Fedoroff and Bennetzen 2013; Fedoroff 2012). Events such as abiotic stress responses (Petit et al. 2010), polyploidization, or small-scale duplications (Renny-Byfield et al. 2013; De Smet et al. 2013) that disrupt this homeostasis can initiate TE and TR removal or accumulation. The trade-off between removal and accumulation of repeat elements depends on the temporal reestablishment of the epigenetic mechanisms that buffer the adverse effects of TEs and TRs, such as aneuploidy—or worse, sterility (Kelly et al. 2015; Fedoroff 2012). After genomic shock, rapid reestablishment of epigenetic control enables regulation of REs, locking them to recombinationally inert heterochromatin, resulting in larger genomes than when epigenetic mechanisms were reestablished more slowly. The latter provides more opportunities for homologous and illegitimate recombinations that removes DNA fragments and causes genome downsizing to occur (Kelly et al. 2015; Fedoroff 2012).

The same mechanisms (i.e., unequal crossovers of homologous sequences and repeat transposition) that are responsible for DNA segment deletion are also models that explain the homogenization and spread of repeats between sister chromatids, homologous chromosomes, and non-homologous chromosomes (Walsh 1987; Cohen et al. 2003; Hall et al. 2005; Charlesworth et al. 1994; Dover 1982). Unequal crossovers usually result in higher-order repeat units consisting of more than one type of element and variation in lengths of arrays (Hall et al. 2005; Talbert and Henikoff 2010; Plohl et al. 2012). Other mechanisms such as gene conversion, repeat transposition, and rolling circle replication may amplify satellite arrays and cause their spread into non-homologous chromosomes (Hall et al. 2005; Dover 1986; Plohl et al. 2012). Epigenetic control is an active cellular mechanism that controls when recombination and transposition should occur. Nonetheless, clear reasons regarding how and why they happen in response to abiotic stresses are unknown (Fedoroff 2013b).

6.7 Summary and Perspectives

As demonstrated in previous studies in *Pisum sativum* (Macas et al. 2007), *Musa acuminata* (Hribova et al. 2010), and some *Brassica* species (Waminal et al. 2015; Perumal et al. 2017; Waminal et al. 2016), a survey of plant genomes using NGS data and reference-guided mapping (Kim et al. 2015) together with FISH analysis is an excellent approach for quantifying and physically mapping repetitive genomic elements that are mostly omitted during assembly. This approach captured about 11% of the *B. napus* genome and enabled comparative 'repeatomics' analysis with its diploid progenitors. The fluctuating pattern of total RE fraction between *B. napus* and its diploid progenitors, as well as among different *B. napus* accessions further demonstrates that RE dynamics is responsible for the huge genome size variations among accessions of the same species (Wendel and Wessler 2000) or species in the same genus (Kelly et al. 2015; Renny-Byfield et al. 2013). We know that epigenetic control is at the center of this fluctuation; nevertheless, even with the current advances in genomics and epigenetics, accurate reasons for how and why these REs respond to abiotic stresses remain unknown. However, with further research, a robust explanation of the mechanisms that underlie the interconnectedness of the environment, genome, and organisms will be determined.

The empirically demonstrated correlation of TE size variation and environmental adaptation within species (Kalendar et al. 2000) is interesting, but whether repeatomics could have a predictive value in relation to agronomically favorable traits is questionable, but perhaps worth pursuing, particularly in the context of crop improvement, such as the oil content in *B. napus* (Delourme et al. 2006). Additionally, the power of WGS and FISH estimation approaches may further be corroborated by optical mapping (Tang et al. 2015; Lam et al. 2012) to provide accurate, single-molecule resolutions of mega-base tandem repeats, which represented a large portion of the unassembled fractions of the three species in this work. This would be particularly useful in analyzing RE fractions to support sequencing projects of species with large genomes, such as *Allium* species (Jakse et al. 2008), *Fritillaria* species (Kelly et al. 2015), and *Paris japonica* (Pellicer et al. 2010a).

References

Alix K, Joets J, Ryder CD, Moore J, Barker GC, Bailey JP, King GJ, Heslop-Harrison JS (2008) The CACTA transposon Bot1 played a major role in Brassica genome divergence and gene proliferation. Plant J 56(6):1030–1044. https://doi.org/10.1111/j.1365-313X.2008.03660.x

Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nat Methods 8(1):61–65. https://doi.org/10.1038/nmeth.1527

Arias T, Beilstein MA, Tang M, McKain MR, Pires JC (2014) Diversification times among Brassica (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. Am J Bot 101(1):86–91. https://doi.org/10.3732/ajb.1300312

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530. https://doi.org/10.1146/annurev-arplant-050213-035811

Biemont C (2010) A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186(4):1085–1093. https://doi.org/10.1534/genetics.110.124180

Biémont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. Nature 443(7111):521–524

Chadwick BP (2009) Macrosatellite epigenetics: the two faces of DXZ4 and D4Z4. Chromosoma 118(6):675–681. https://doi.org/10.1007/s00412-009-0233-5

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome.

Science 345(6199):950–953. https://doi.org/10.1126/science.1253435
Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371(6494):215–220. https://doi.org/10.1038/371215a0
Choi HI, Waminal NE, Park HM, Kim NH, Choi BS, Park M, Choi D, Lim YP, Kwon SJ, Park BS, Kim HH, Yang TJ (2014) Major repeat components covering one-third of the ginseng (*Panax ginseng* C. A. Meyer) genome and evidence for allotetraploidy. Plant J (For Cell and Molecular Biology) 77(6):906–916. https://doi.org/10.1111/tpj.12441
Cizkova J, Hribova E, Humplikova L, Christelova P, Suchankova P, Dolezel J (2013) Molecular analysis and genomic organization of major DNA satellites in banana (Musa spp.). PLoS ONE 8(1):e54808. https://doi.org/10.1371/journal.pone.0054808
Claros MG, Bautista R, Guerrero-Fernandez D, Benzerki H, Seoane P, Fernandez-Pozo N (2012) Why assembling plant genome sequences is so challenging. Biology 1 (2):439–459. https://doi.org/10.3390/biology1020439
Cohen S, Yacobi K, Segal D (2003) Extrachromosomal circular DNA of tandemly repeated genomic sequences in Drosophila. Genome Res 13(6A):1133–1145. https://doi.org/10.1101/gr.907603
Coluccia E, Pichiri G, Nieddu M, Coni P, Manconi S, Deiana AM, Salvadori S, Mezzanotte R (2011) Identification of two new repetitive elements and chromosomal mapping of repetitive DNA sequences in the fish Gymnothorax unicolor (Anguilliformes: Muraenidae). Eur J Histochem (EJH) 55(2):e12. https://doi.org/10.4081/ejh.2011.e12
De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci 110(8):2898–2903
Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B, Specel S, Hanneton L, Dheu J, Deschamps M (2006) Genetic control of oil content in oilseed rape (*Brassica napus* L.). Theoret Appl. Genetics 113(7):1331–1345
Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299(5879):111–117
Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165
Eddy SR (2012) The C-value paradox, junk DNA and ENCODE. Curr Biol 22(21):R898–R899
Fablet M, Vieira C (2011) Evolvability, epigenetics and transposable elements. Biomol Concepts 2(5):333–341. https://doi.org/10.1515/BMC.2011.035
Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and genome evolution. Science 338(6108):758–767
Fedoroff NV (2013a) McClintock and epigenetics. In: Plant transposons and genome dynamics in evolution. Wiley, pp 61–70. https://doi.org/10.1002/9781118500156.ch4
Fedoroff NV (2013b) Molecular genetics and epigenetics of CACTA elements. Methods Mol Biol 1057:177–192. https://doi.org/10.1007/978-1-62703-568-2_13
Fedoroff NV, Bennetzen JL (2013) Transposons, genomic shock, and genome evolution. In: Plant transposons and genome dynamics in evolution. Wiley, pp 181–201. https://doi.org/10.1002/9781118500156.ch10
Freeling M, Xu J, Woodhouse M, Lisch D (2015) A solution to the C-value paradox and the function of junk DNA: the genome balance hypothesis. Mol Plant 8(6):899–910. https://doi.org/10.1016/j.molp.2015.02.009
Gao D, Jiang N, Wing RA, Jiang J, Jackson SA (2015) Transposons play an important role in the evolution and diversification of centromeres among closely related species. Front Plant Sci 6
Goodier JL, Cheung LE, Kazazian HH Jr (2012) MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. PLoS Genet 8(10):e1002941. https://doi.org/10.1371/journal.pgen.1002941
Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol 12 (8):483–492. https://doi.org/10.1038/nrm3152
Hall SE, Luo S, Hall AE, Preuss D (2005) Differential rates of local and global homogenization in centromere satellites from arabidopsis relatives. Genetics 170(4):1913–1927. https://doi.org/10.1534/genetics.104.038208
Hamilton JP, Buell CR (2012) Advances in plant genome sequencing. Plant J (For Cell and Molecular Biology) 70(1):177–190. https://doi.org/10.1111/j.1365-313X.2012.04894.x
Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell 12(5):617–636
Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ (2008) A and C genome distinction and chromosome identification in *Brassica napus* by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics 180(4):1849–1857. https://doi.org/10.1534/genetics.108.095893
Hribova E, Neumann P, Matsumoto T, Roux N, Macas J, Dolezel J (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biol 10:204. https://doi.org/10.1186/1471-2229-10-204
Iourov IY, Vorsanova SG, Yurov YB (2008) Molecular cytogenetics and cytogenomics of brain diseases. Curr Genomics 9(7):452–465. https://doi.org/10.2174/138920208786241216
Jakse J, Meyer JD, Suzuki G, McCallum J, Cheung F, Town CD, Havey MJ (2008) Pilot sequencing of onion genomic DNA reveals fragments of transposable elements, low gene densities, and significant gene enrichment after methyl filtration. Mol Genetics Genomics (MGG) 280(4):287–292. https://doi.org/10.1007/s00438-008-0364-z

Jiang JM, Gill BS (2006) Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome/Nat Res Counc Can (Genome/Conseil national de recherches Canada) 49 (9):1057–1068. https://doi.org/10.1139/G06-076

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97(12):6603–6607. https://doi.org/10.1073/pnas.110587497

Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci 101(37):13554–13559. https://doi.org/10.1073/pnas.0403659101

Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novak P, Neumann P, Lysak MA, Day PD, Berger M, Fay MF, Nichols RA, Leitch AR, Leitch IJ (2015) Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol 208(2):596–607. https://doi.org/10.1111/nph.13471

Kim K, Lee S-C, Lee J, Yu Y, Yang K, Choi B-S, Koh H-J, Waminal NE, Choi H-I, Kim N-H, Jang W, Park H-S, Lee J, Lee HO, Joh HJ, Lee HJ, Park JY, Perumal S, Jayakodi M, Lee YS, Kim B, Copetti D, Kim S, Kim S, Lim K-B, Kim Y-D, Lee J, Cho K-S, Park B-S, Wing RA, Yang T-J (2015) Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. Sci Rep 5:15655. https://doi.org/10.1038/srep15655, http://www.nature.com/articles/srep15655#supplementary-information

Kolomietz E, Meyn MS, Pandita A, Squire JA (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes Chromosom Cancer 35(2):97–112

Koo DH, Plaha P, Lim YP, Hur Y, Bang JW (2004) A high-resolution karyotype of *Brassica rapa* ssp. pekinensis revealed by pachytene analysis and multicolor fluorescence in situ hybridization. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 109(7):1346–1352. https://doi.org/10.1007/s00122-004-1771-0

Koo DH, Hong CP, Batley J, Chung YS, Edwards D, Bang JW, Hur Y, Lim YP (2011) Rapid divergence of repetitive DNAs in Brassica relatives. Genomics 97 (3):173–185. https://doi.org/10.1016/j.ygeno.2010.12.002

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30(8):771–776

Lee J, Waminal NE, Choi HI, Perumal S, Lee SC, Nguyen VB, Jang W, Kim NH, Gao LZ, Yang TJ (2017) Rapid amplification of four retrotransposon families promoted speciation and genome size expansion in the genus *Panax*. Sci Rep 7(1):9045. https://doi.org/10.1038/s41598-017-08194-5

Lim KB, Yang TJ, Hwang YJ, Kim JS, Park JY, Kwon SJ, Kim J, Choi BS, Lim MH, Jin M, Kim HI, de Jong H, Bancroft I, Lim Y, Park BS (2007) Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related Brassica species. Plant J (For Cell and Molecular Biology) 49 (2):173–183. https://doi.org/10.1111/j.1365-313X.2006.02952.x

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang T-J, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Perumal S, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park B-S, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee T-H, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5. https://doi.org/10.1038/ncomms4930

Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci USA 103(13):5224–5229. https://doi.org/10.1073/pnas.0510791103

Macas J, Neumann P, Navratilova A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genom 8:427. https://doi.org/10.1186/1471-2164-8-427

Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genomics Proteomics Bioinf 12(4):164–171. https://doi.org/10.1016/j.gpb.2014.07.003

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, Garcia JF, Derisi JL, Smith T, Tobias C, Ross-Ibarra J, Korf I, Chan SW (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14 (1):R10. https://doi.org/10.1186/gb-2013-14-1-r10

Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome-Us 6(2). https://doi.org/10.3835/plantgenome2013.03.0001in

Mihai P, Salzberg SL, Shumway M (2002) Genome sequence assembly: algorithms and issues. Computer 35(7):47–54. https://doi.org/10.1109/MC.2002.1016901

Nowak R (1994) Mining treasures from 'junk DNA'. Science 263(5147):608–610

Pagel M, Johnstone RA (1992) Variation across species in the size of the nuclear genome supports the junk-DNA

explanation for the C-value paradox. Proc R Soc Lond B 249(1325):119–124

Pardue ML, DeBaryshe PG (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. Annu Rev Genet 37:485–511. https://doi.org/10.1146/annurev.genet.38.072902.093115

Paterson AH, Wendel JF (2015) Unraveling the fabric of polyploidy. Nat Biotech 33(5):491–493. https://doi.org/10.1038/nbt.3217

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, Yoo MJ, Byers R, Chen W, Doron-Faigenboim A, Duke MV, Gong L, Grimwood J, Grover C, Grupp K, Hu G, Lee TH, Li J, Lin L, Liu T, Marler BS, Page JT, Roberts AW, Romanel E, Sanders WS, Szadkowski E, Tan X, Tang H, Xu C, Wang J, Wang Z, Zhang D, Zhang L, Ashrafi H, Bedon F, Bowers JE, Brubaker CL, Chee PW, Das S, Gingle AR, Haigler CH, Harker D, Hoffmann LV, Hovav R, Jones DC, Lemke C, Mansoor S, ur Rahman M, Rainville LN, Rambani A, Reddy UK, Rong JK, Saranga Y, Scheffler BE, Scheffler JA, Stelly DM, Triplett BA, Van Deynze A, Vaslin MF, Waghmare VN, Walford SA, Wright RJ, Zaki EA, Zhang T, Dennis ES, Mayer KF, Peterson DG, Rokhsar DS, Wang X, Schmutz J (2012) Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492(7429):423–427. https://doi.org/10.1038/nature11798

Pellicer J, Fay MF, Leitch IJ (2010a) The largest eukaryotic genome of them all? Bot J Linn Soc 164(1):10–15. https://doi.org/10.1111/j.1095-8339.2010.01072.x

Pellicer J, Garcia S, Canela MA, Garnatje T, Korobkov AA, Twibell JD, Valles J (2010b) Genome size dynamics in Artemisia L. (Asteraceae): following the track of polyploidy. Plant Biol (Stuttg) 12(5):820–830. https://doi.org/10.1111/j.1438-8677.2009.00268.x

Pellicer J, Garnatje T, Hidalgo O, Tagashira N, Valles J, Kondo K (2010c) Do polyploids require proportionally less rDNA loci than their corresponding diploids? Examples from *Artemisia* subgenera *Absinthium* and *Artemisia* (Asteraceae, Anthemideae). Plant Biosyst 144(4):841–848. https://doi.org/10.1080/11263504.2010.522783

Peng JC, Karpen GH (2008) Epigenetic regulation of heterochromatic DNA stability. Curr Opin Genet Dev 18(2):204–211. https://doi.org/10.1016/j.gde.2008.01.021

Perumal S, Waminal NE, Lee J, Lee J, Choi BS, Kim HH, Grandbastien MA, Yang TJ (2017) Elucidating the major hidden genomic components of the A, C, and AC genomes and their influence on *Brassica* evolution. Sci Rep 7(1):17986. https://doi.org/10.1038/s41598-017-18048-9

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186(1):135–147. https://doi.org/10.1111/j.1469-8137.2009.03140.x

Pich U, Meister A, Macas J, Doležel J, Lucretti S, Schubert I (1995) Primed in situ labelling facilitates flow sorting of similar sized chromosomes. Plant J 7(6):1039–1044

Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409(1–2):72–82. https://doi.org/10.1016/j.gene.2007.11.013

Plohl M, Mestrovic N, Mravinac B (2012) Satellite DNA evolution. Genome Dyn 7:126–152. https://doi.org/10.1159/000337122

Pop M, Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. Trends Genet 24(3):142–149. https://doi.org/10.1016/j.tig.2007.12.006

Price HJ, Johnston JS (1996) Influence of light on DNA content of *Helianthus annuus* Linnaeus. Proc Natl Acad Sci USA 93(20):11264–11267

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Bot 101(10):1711–1725

Renny-Byfield S, Kovarik A, Chester M, Nichols RA, Macas J, Novak P, Leitch AR (2012) Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. PloS one 7(5). https://doi.org/10.1371/journal.pone.0036963 (ARTN e36963)

Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien MA, Leitch AR (2013) Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. Plant J 74(5):829–839. https://doi.org/10.1111/tpj.12168

Rosbash M, Ford PJ, Bishop JO (1974) Analysis of the C-value paradox by molecular hybridization. Proc Natl Acad Sci USA 71(9):3746–3750

Sampath P, Lee SC, Lee J, Izzah NK, Choi BS, Jin M, Park BS, Yang TJ (2013) Characterization of a new high copy Stowaway family MITE, BRAMI-1 in Brassica genome. BMC Plant Biol 13:56. https://doi.org/10.1186/1471-2229-13-56

Sarilar V, Marmagne A, Brabant P, Joets J, Alix K (2011) BraSto, a Stowaway MITE from Brassica: recently active copies preferentially accumulate in the gene space. Plant Mol Biol 77(1–2):59–75. https://doi.org/10.1007/s11103-011-9794-9

Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19(R2):R227–R240. https://doi.org/10.1093/hmg/ddq416

Schatz MC, Witkowski J, McCombie WR (2012) Current challenges in de novo plant genome sequencing and assembly. Genome Biol 13(4):243. https://doi.org/10.1186/gb4015

Shapiro JA, von Sternberg R (2005) Why repetitive DNA is essential to genome function. Biol Rev Camb Philos Soc 80(2):227–250

Sharma A, Wolfgruber TK, Presting GG (2013) Tandem repeats derived from centromeric retrotransposons. BMC Genom 14(1):142. https://doi.org/10.1186/1471-2164-14-142

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8(4):272–285. https://doi.org/10.1038/nrg2072

Snowdon RJ, Köhler W, Friedt W, Köhler A (1997) Genomic in situ hybridization in Brassica amphidiploids and interspecific hybrids. Theoret Appl Genetics 95(8):1320–1324. https://doi.org/10.1007/s001220050699

Soltis DE, Gitzendanner MA, Stull G, Chester M, Chanderbali A, Chamala S, Jordon-Thaden I, Soltis PS, Schnable PS, Barbazuk WB (2013) The potential of genomics in plant systematics. Taxon 62(5):886–898

Talbert PB, Henikoff S (2010) Centromeres convert but don't cross. PLoS Biol 8(3):e1000326. https://doi.org/10.1371/journal.pbio.1000326

Tang H, Lyons E, Town CD (2015) Optical mapping in plant comparative genomics. Genome/Nat Res Counc Can (Genome/Conseil national de recherches Canada) 1:8

Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13(1):36–46. https://doi.org/10.1038/nrg3117

Vrána J, Kubaláková M, Číhalíková J, Valárik M, Doležel J (2015) Preparation of sub-genomic fractions enriched for particular chromosomes in polyploid wheat. Biol Plant 59(3):445–455. https://doi.org/10.1007/s10535-015-0522-1

Walsh JB (1987) Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 115(3):553–567

Waminal N, Park HM, Ryu KB, Kim JH, Yang TJ, Kim HH (2012) Karyotype analysis of Panax ginseng C.A. Meyer, 1843 (Araliaceae) based on rDNA loci and DAPI band distribution. Comp Cytogenet 6 (4):425–441. https://doi.org/10.3897/compcytogen.v6i4.3740

Waminal NE, Perumal S, Lim K-B, Park B-S, Kim HH, Yang T-J (2015) Genomic survey of the hidden components of the *B. rapa* Genome. In: *The Brassica rapa* genome. Springer, pp 83–96

Waminal NE, Perumal S, Lee J, Kim HH, Yang T-J (2016) Repeat evolution in *Brassica rapa* (AA), *B. oleracea* (CC), and *B. napus* (AACC) genomes. Plant Breeding Biotechnol 4(2):107–122

Waminal NE, Pellerin RJ, Jang W, Kim HH, Yang T-J (2018) Characterization of chromosome-specific microsatellite repeats and telomere repeats based on low-coverage whole-genome sequence reads. Plant Breeding Biotechnol 6(1):1–8

Wang X, Torres MJ, Pierce G, Lemke C, Nelson LK, Yuksel B, Bowers JE, Marler B, Xiao Y, Lin L, Epps E, Sarazen H, Rogers C, Karunakaran S, Ingles J, Giattina E, Mun JH, Seol YJ, Park BS, Amasino RM, Quiros CF, Osborn TC, Pires JC, Town C, Paterson AH (2011a) A physical map of *Brassica oleracea* shows complexity of chromosomal changes following recursive paleopolyploidizations. BMC Genom 12:470. https://doi.org/10.1186/1471-2164-12-470

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z, *Brassica rapa* Genome Sequencing Project C (2011b) The genome of the mesopolyploid crop species *Brassica rapa*. Nature genetics 43 (10):1035–1039. https://doi.org/10.1038/ng.919

Wang DO, Matsuno H, Ikeda S, Nakamura A, Yanagisawa H, Hayashi Y, Okamoto A (2012) A quick and simple FISH protocol with hybridization-sensitive fluorescent linear oligodeoxynucleotide probes. RNA 18(1):166–175. https://doi.org/10.1261/rna.028431.111

Wei L, Xiao M, An Z, Ma B, Mason AS, Qian W, Li J, Fu D (2013) New insights into nested long terminal repeat retrotransposons in Brassica species. Mol Plant 6(2):470–482. https://doi.org/10.1093/mp/sss081

Wendel JF, Wessler SR (2000) Retrotransposon-mediated genome evolution on a local ecological scale. Proc Natl Acad Sci 97(12):6250–6252. https://doi.org/10.1073/pnas.97.12.6250

Wolfgruber TK, Sharma A, Schneider KL, Albert PS, Koo D-H, Shi J, Gao Z, Han F, Lee H, Xu R, Allison J, Birchler JA, Jiang J, Dawe RK, Presting GG (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genet 5(11):e1000743. https://doi.org/10.1371/journal.pgen.1000743

Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J (2010) The next-generation sequencing technology and application. Protein Cell 1(6):520–536. https://doi.org/10.1007/s13238-010-0065-3