Genome Size Variation and Comparative Genomics Reveal Intraspecific Diversity in *Brassica rapa*

Julien Boutte[1]*, Loeiz Maillet[1], Thomas Chaussepied[1], Sébastien Letort[2],
Jean-Marc Aury[3], Caroline Belser[3], Franz Boideau[1], Anael Brunet[1], Olivier Coriton[1],
Gwenaëlle Deniot[1], Cyril Falentin[1], Virginie Huteau[1], Maryse Lodé-Taburel[1],
Jérôme Morice[1], Gwenn Trotoux[1], Anne-Marie Chèvre[1], Mathieu Rousseau-Gueutin[1]
and Julie Ferreira de Carvalho[1]*

[1] IGEPP, INRAE, Institut Agro, Univ Rennes, Le Rheu, France, [2] IRISA/INRIA, Campus de Beaulieu, Rennes, France,
[3] Génomique Métabolique, Genoscope, Institut de biologie François-Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay,
Evry, France

Traditionally, reference genomes in crop species rely on the assembly of one accession, thus occulting most of intraspecific diversity. However, rearrangements, gene duplications, and transposable element content may have a large impact on the genomic structure, which could generate new phenotypic traits. Comparing two *Brassica rapa* genomes recently sequenced and assembled using long-read technology and optical mapping, we investigated structural variants and repetitive content between the two accessions and genome size variation among a core collection. We explored the structural consequences of the presence of large repeated sequences in *B. rapa* 'Z1' genome vs. the *B. rapa* 'Chiifu' genome, using comparative genomics and cytogenetic approaches. First, we showed that large genomic variants on chromosomes A05, A06, A09, and A10 are due to large insertions and inversions when comparing *B. rapa* 'Z1' and *B. rapa* 'Chiifu' at the origin of important length differences in some chromosomes. For instance, lengths of 'Z1' and 'Chiifu' A06 chromosomes were estimated *in silico* to be 55 and 29 Mb, respectively. To validate these observations, we compared using fluorescent *in situ* hybridization (FISH) the two A06 chromosomes present in an F1 hybrid produced by crossing these two varieties. We confirmed a length difference of 17.6% between the A06 chromosomes of 'Z1' compared to 'Chiifu.' Alternatively, using a copy number variation approach, we were able to quantify the presence of a higher number of rDNA and gypsy elements in 'Z1' genome compared to 'Chiifu' on different chromosomes including A06. Using flow cytometry, the total genome size of 12 *Brassica* accessions corresponding to a *B. rapa* available core collection was estimated and revealed a genome size variation of up to 16% between these accessions as well as some shared inversions. This study revealed the contribution of long-read sequencing of new accessions belonging to different cultigroups of *B. rapa* and highlighted the potential impact of differential insertion of repeat elements and inversions of large genomic regions in genome size intraspecific variability.

Keywords: ***Brassica*, genome evolution, transposable elements, LTR Gypsy, intraspecific diversity, ribosomal DNA**

INTRODUCTION

Genetic and phenotypic diversity are the drivers of plant evolution and adaptation. While ecology and population genetics provide important knowledge on the molecular processes underlying the astonishing plant biodiversity, comparative genomic analyses are essential to understand the large-scale genomic variations that can be observed within species and how these structural variations may impact plant phenotypes. Resequencing of accessions and construction of pangenomes of plant species such as wheat (Montenegro et al., 2017), maize (Gage et al., 2019), rice (Schatz et al., 2014; Wang et al., 2018; Zhao et al., 2018), or soybean (Lam et al., 2010; Li et al., 2014) have revealed the involvement of structural variations in various agronomically relevant traits. In maize, structural variants are implicated in plant architecture, flowering, and disease resistance (Walker et al., 1995; Chia et al., 2012; Lu et al., 2015). Similarly, using pangenomes, numerous structural variations have been revealed in *Brassica napus* varieties correlating with genes and phenotypic traits such as silique length and seed weight (Gazave et al., 2016), disease resistance (Dolatabadian et al., 2020), as well as flowering time and vernalization (Song et al., 2020). Structural variants identified as underlying flowering time were associated with the insertion of transposable elements (TEs) in three *Flowering Locus C* (or *FLC*) genes (Song et al., 2020). However, structural diversity in the progenitors of this allopolyploid species has, so far, been particularly overlooked.

Only recently, the development of *Brassica* pangenomes has started to reveal the extensive number of structural variants within one of the two diploid progenitors *Brassica oleracea* (Golicz et al., 2016) and assessed in regard to the domestication process of the subsequent allotetraploids *Brassica juncea* (Paritosh et al., 2019) and *B. napus* (Song et al., 2020). These studies highlight the importance of structural variants and gene copy number variation (CNVs) in adaptive phenotypes. While core genes represent 80% of the assembled genomes, each new accession brings around 20% of novel genes. In most of these analyses, however, the occurrence and importance of the repeat compartment are undervalued.

TEs form an important internal source of genetic diversity as a result of their ability to create mutations, alter gene expression, and promote chromosomal mispairing (Kidwell and Lisch, 2000). These ancient, ubiquitous, and dynamic components of eukaryotic genomes comprise up to 58% of the allopolyploid genome of *B. napus* (Song et al., 2020). TEs comprise two classes that have contrasted mode of replication, each divided in several families: retrotransposons (Class I) transpose via reverse transcription of their messenger RNA and DNA transposons (Class II), encompassing all other types of TEs (Wicker et al., 2007). Overall, TEs have major impact on genome organization and function and have been associated with up-regulated genes in response to various abiotic stresses, environmental gradients, reproductive mode (asexuality, selfing vs. outcrossing), interspecific hybrid formation, and polyploidy (Wright and Schoen, 1999; Kalendar et al., 2000; Ungerer et al., 2006; Makarevitch et al., 2015; Vicient and Casacuberta, 2017). Several studies from different biological systems indicate that the

patterns (position and number of copies) of a single TE family can vary within species (Quadrana et al., 2016). However, the study of TEs inserted in complex genomes of various populations is challenging and rely on the amplification of specific TE sequences (for instance using Sequence-Specific Amplification Polymorphism transposon display, Parisod et al., 2009) or low-coverage sequencing strategies (Ferreira de Carvalho et al., 2016). The utilization of long-read sequencing and optical mapping in several varieties of the same species is allowing fine comparative genomic analysis of the differential proliferation of TEs and other repeats. This overlooked component of plant genomes is getting accessible and provides important insights on the structural genomic diversity standing within species.

In the present study, we conducted a fine comparative genomic analysis of two varieties of *Brassica rapa* 'Chiifu' and 'Z1.' *B. rapa* (AA, 2n = 20) is one of the diploid progenitors of the important allotetraploid oilseed crops, *B. juncea* (AABB, 2n = 36), and *B. napus* (AACC, 2n = 38) (U, 1935). Therefore, *B. rapa* varieties can be used to broaden the genetic diversity of these domesticated species. *Brassica* diploid species arose from a whole genome triplication that occurred about 15–20 million years ago (Murat et al., 2015). Diploid *Brassica* are thus considered as paleohexaploids. Thereafter, their duplicated genomes rapidly underwent extensive gene deletion through a fractionation process (Murat et al., 2015). The majority of the genes are tending toward a return to single-copy status (Wang et al., 2011), and overall gene content in *Brassica* diploid genomes has so far been reduced by about half. Yet, different cultigroups within *Brassica* species have selected different alleles from the same duplicated genes explaining part of the astonishing and extensive number of morphotypes existing within each species (Cheng et al., 2014, 2016b; Golicz et al., 2016). *B. rapa* speciation started after the divergence with *B. oleracea* estimated 2.18 Mya (Li et al., 2017), but its diversification occurred mainly during the domestication process of *B. rapa* crops, 3,500 years BC in the Mediterranean basin with a secondary center of diversification in Asia (Guo et al., 2014). This resulted in genetically (Aissiou et al., 2018) and phenotypically diverse morphotypes (leafy, root or fodder, and oilseed types). We focus on two subspecies, ssp. *pekinensis* (Chinese cabbage with 'Chiifu') and ssp. *trilocularis* (yellow sarson with 'Z1'). Chinese cabbage has originated in central China by natural hybridization between Pak-choi (ssp. chinensis) and turnip rape (ssp. *oleifera*) 1,200–2,100 years ago (Song et al., 1990; Qi et al., 2017). Yellow sarson originated from India, but its precise origin is still unknown. Overall, the underlying structural genomic diversity has never been explored within this important crop.

Using established protocols (flow cytometry and cytogenetics), along with state-of-the-art whole genome sequencing and optical mapping for assembly of complex *Brassica* genomes, we compared structural rearrangements between two varieties of *B. rapa* that have contrasted morphotypes, evolutionary history and reproductive mode. Specifically, we identified differential insertion of ribosomal DNA and LTR retrotransposons and reconstructed the evolution of these repeats during *B. rapa* domestication. Finally, we unraveled wide variation in DNA

content among different accessions belonging to *B. rapa* core collection.

MATERIALS AND METHODS

Biological Material

Two *B. rapa* genotypes, *B. rapa* var. *pekinensis* cabbage 'Chiifu' (pure line), and *B. rapa* var. *trilocularis* 'Z1' (doubled haploid line), as well as their F1 hybrid ('Z1' × 'Chiifu'), were grown in semi-controlled conditions in a greenhouse. Seven to 10 individuals of each *B. rapa* accession representing biological replicates were harvested for leaves and roots. Leaf tissue was used for molecular work and flow cytometry, whereas root meristem samples were used to perform fluorescent *in situ* hybridization (FISH) experiments. In addition, 10 accessions representing the different groups of *B. rapa* core collection (Zhao et al., 2010; Cheng et al., 2016b) were chosen and grown in a similar environment for assessing genome size variation. Fresh leaves from three to eight individuals, used as biological replicates, were harvested per accession: from group 1, one accession of *ssp. pekinensis*; from group 2, one accession of *ssp. narinosa*; from group 3, one accession of *ssp. perviridis* and one accession of *ssp. nipposinica*; from group 4, two different accessions of *ssp. oleifera*; from group 5, one accession of *ssp. trilocularis*; from group 6, three different accessions of *ssp. rapa* (for thorough descriptions see Cheng et al., 2016c). Leaf tissue is then directly processed for flow cytometry analyses (**Supplementary Table 1**).

Annotations of Transposable Elements

The latest genome assemblies available for *B. rapa* 'Z1' (Belser et al., 2018) and 'Chiifu' v.3.0 (Zhang et al., 2018) were retrieved from the Genoscope website[1] and the *Brassica* database[2], respectively. These genomes were subjected to the same TE annotation package REPET V2.5 (Quesneville et al., 2005; Flutre et al., 2011; Hoede et al., 2014). Briefly, the REPET pipeline was used for the detection, classification (TEdenovo) and annotation of TEs (TEannot). The TEdenovo pipeline identifies TE copies present on a reference genome using a curated database of TEs: Repbase (V 20.05 for both nucleotides and amino acid sequences). Then, TEdenovo creates *de novo* consensus sequences when at least five copies of the same TE family are retrieved within the genome. Next, TEannot pipeline uses these consensus sequences as well as Repbase to annotate the provided reference genomes. Default parameters were set for this last step of annotation. Finally, two files containing the structural annotation of 'Z1' and 'Chiifu' were output in GFF3 format (Generic Feature Format version 3). The resulting GFF3 files were hand curated using custom python scripts (Python version 2.7.12).

Comparative Structural Analyses

To evaluate large-scale genome structural rearrangements, the assembled genomes of *B. rapa* 'Z1' and 'Chiifu', as well as

[1]http://www.genoscope.cns.fr/plants
[2]http://brassicadb.org/

the A subgenome of *B. napus* 'Darmor Bzh' V4.1, were retrieved from the *Brassica* database[2] and compared. The package MUMmer V3.23 (Kurtz et al., 2004) was used on each of the 10 chromosomes using the following settings: -id 95, -l 4,000. Inversions or deletions observed in 'Z1' when compared to 'Chiifu' were further verified using BioNano raw data from Belser et al. (2018). To better characterize these genomic regions, gene annotations were retrieved from GFF files. Gene Ontology (GO) enrichment was performed on GO terms of the identified genes present in specifically inserted regions of 'Z1' using Agrigo V2 (Tian et al., 2017). Finally, we assessed the duplication status of these genes using proto-Calepineae karyotype (PCK; Murat et al., 2015) and custom-made python scripts. In addition, among all inversions, three showing defined breakpoints in both 'Z1' and 'Chiifu' were chosen and validated using custom-made polymerase chain reaction (PCR) primers designed in flanking regions of inversions and specific to either 'Z1' or 'Chiifu' accessions (**Supplementary Table 2**).

Validation of Rearrangements Using Cytogenetics

To validate the presence of repeated sequences on the chromosome A06 of *B. rapa* 'Z1', FISH was carried out in 'Z1', 'Chiifu' and the F1 hybrid 'Z1' × 'Chiifu' according to protocols detailed in Książczyk et al. (2011). The two BAC clones KBrB022P06 and KBrH003P24 hybridizing on each of the A06 chromosome arms in *B. rapa* (Xiong and Pires, 2011) were labeled by random priming with biotin-14-dUTP (Invitrogen, Life Technologies). The ribosomal probe used in this study was pTa-71 (Gerlach and Bedbrook, 1979), which contained a 9-kb *Eco*RI fragment of rDNA repeat unit (18S-5.8S-26S genes and spacers) isolated from *Triticum aestivum*. pTa-71 was labeled with Alexa-488 dUTP by random priming. Biotinylated probes were immunodetected by Texas red avidin DCS (Vector Laboratories), and the signal was amplified with biotinylated anti-avidin D (Vector Laboratories). The chromosomes were mounted and counterstained in Vectashield (Vector Laboratories) containing 2.5 µg/mL 4′,6-diamidino-2-phenylindole (DAPI). Fluorescence images were captured using an ORCA-Flash4 (Hamamatsu, Japan) on an Axioplan 2 microscope (Zeiss, Oberkochen, Germany) and analyzed using Zen 2 PRO software (Carl Zeiss, Germany). Finally, the lengths of both A06 chromosomes in the hybrid 'Z1' × 'Chiifu' were estimated on 14 cells at the mitotic metaphase stage and measured using the visualization software Zen 2 PRO (Carl Zeiss, Germany).

Copy Number Estimation of Highly Repeated Sequences: Ribosomal DNA and Gypsy LTR Retrotransposon

CNVs of 45S rDNA (otherwise called 25S or 35S rDNA) and gypsy LTR Retrotransposons were obtained using Shotgun data for *B. rapa* var. *pekinensis* 'Chiifu' (from[2]) and *B. rapa*

var. *trilocularis* 'Z1' (from[3]), applying a method based on a calibration using the CNV (based on read depth approach) of single copy genes (scripts available at github/juboutte). First, 17,954 putative single copy genes in 'Chiifu' and 'Z1' were identified using custom-made python scripts based on PCK genes from *B. rapa* (Murat et al., 2015). Second, 38 rDNA sequences, corresponding to *Brassica* rDNA full-length sequences downloaded from the NCBI (last accessed August 5, 2019) and 96 gypsy sequences downloaded from the gypsy database (last accessed August 5, 2019)[4], were retrieved as reference sequences. Then, rDNA and gypsy sequences of 'Chiifu' and 'Z1' were identified by performing BLASTn and tBLASTn algorithms (*e*-value threshold of 10^{-5}; Altschul et al., 1997) between the rDNA and gypsy reference sequences and the two assembled genomes 'Chiifu' v.3.0 (Zhang et al., 2018) and 'Z1' (Belser et al., 2018). Finally, consensus sequences (two rDNA and 211 gypsy) were obtained overall for 'Z1' and 'Chiifu' using python custom script (sequences presenting more than 90% identity were combined together) and aligned with Mafft v.7.245; option –auto (Katoh and Standley, 2013). Visual checking of the alignments was done using Geneious v.10.0.9 (Kearse et al., 2012). 'Chiifu' and 'Z1' genomic reads were then mapped on the two 45S rDNA consensus, the 211 gypsy RT consensus and the 17,954 putative single copy genes using Bowtie 2, v.2.2.7 (Langmead and Salzberg, 2012) using the parameters G, 52, 6 and G, 58, 3 for 'Chiifu' and 'Z1,' respectively.

Phylogenetic Analyses

To study the evolutionary relationships of 45S rDNA and gypsy sequences of *B. rapa* 'Chiifu' and 'Z1,' phylogenetic analyses were performed. First, 45S rDNA and gypsy sequences (first translated to protein sequences and merged to the 96 gypsy database sequences) of 'Chiifu' and 'Z1' were aligned separately using Mafft alignment v.7.245; option –auto (Katoh and Standley, 2013). Second, beginning and end of the matrices with less than 25% of the sequences were deleted, as well as indels present in more than 25% of the sequences. Phylogenetic analyses were performed using Geneious tree builder with the Jukes–Cantor model and the neighbor-joining method (45S rDNA clade support was determined by 1,000 bootstrap replicates). To avoid Long Branch Attraction error, sequences presenting long branches were deleted, and new phylogenetic analyses were performed using the parameters described above.

Genome Size Estimation by Flow Cytometry

In order to explore the intraspecific variability of genome size among the different varieties of *B. rapa*, we assessed by flow cytometry, the DNA content, and estimated genome size for the focal accessions 'Chiifu' and 'Z1,' as well as 10 other accessions representing *B. rapa* intraspecific diversity. Briefly, approximately 4 mg of fresh leaves of *B. rapa* and *Pisum sativum* (var. Victor; used as an internal reference standard) were harvested and

transferred to a Petri dish. This material was chopped using a sharp razor blade in 500 µL of staining buffer (from Cystain PI OxProtect) and incubated at room temperature for 30–90 s. The solution was then filtered through a 50 µm nylon mesh, and 1.5 mL of solution (0.0166 mg of RNase A and 10 µL of propidium iodide) was added per sample. Incubation at room temperature was made for 30–60 min, protected from light. Estimation of genome size for each accession was obtained using a CyFlow space cytometer (Partec Inc.). This instrument was equipped with a 488 nm blue laser 50 mW and a band-pass filter LP590 used as an emission filter. Also, prior to running the samples, gain and linearity of the instrument were adjusted by using DNA control PI from Sysmex. Finally, G1 peaks in *B. rapa* accessions and *P. sativum* were collected for each sample to calculate nuclear DNA content (1C in pg) and haploid genome size (in Mbp).

Statistical Analyses

Proportions of repeat elements in both *B. rapa* 'Z1' and *B. rapa* 'Chiifu' and proportions of sequences within each LTR gypsy clade were investigated using χ^2-test. Genome size variations were investigated using analysis of variance (ANOVA) and Tukey pairwise comparisons [honestly significant difference (HSD)] tests. Normality of data was first tested by Shapiro test. Statistical tests were performed using R software v.3.5.1 (R Core Team, 2018).

RESULTS

Structural Analysis of Two Highly Contiguous *B. rapa* Genomes

Structural comparison of the *B. rapa* 'Z1' genome against both *B. rapa* 'Chiifu' genome and *B. napus* 'Darmor Bzh' A subgenome revealed six large sequences specific to *B. rapa* 'Z1' on A01, A05 (two regions), A06, A08, and A09 (**Table 1**, see also **Figure 1**). These regions were qualified as "specific" because no orthologous match could be found in the assembled *Brassica* genomes tested in this study. These six regions with a mean length of 12,179,136 bp (median = 9,811,381 bp) were composed of a mean number of TEs, rDNA, and genes of 4,244.5 (median = 2,878), 25.5 (median = 19), and 160.5 (median = 135.5), respectively (**Table 1**). A total of 963 genes were found in the regions specific to 'Z1.' On these 963 genes, 439 genes were only detected in the 'Z1' genome without any homologous similarity in 'Chiifu' genome. GO enrichment was performed for the genes localized in the 'Z1'-specific regions; however, no specific pathway was revealed. Similarly, we were not able to detect enrichment of duplicated genes that may be localized in these regions. Moreover, these six 'Z1' regions contained an important percentage of "N" sequences (mean = 30.59%, median = 26.38%). The observed "N" percentage in these regions was associated with the scaffolding of long reads due to optical mapping. Thus, we might lack genetic information, but thanks to optical mapping, we are confident on the size of these structural variants.

Additionally, 81 inverted regions were identified on all chromosomes between 'Z1' (cumulative length = 23,066,463 bp)

TABLE 1 | Position and length of the *B. rapa* 'Z1' genome-specific regions (in comparison to *B. rapa* 'Chiifu').

Chromosome	Region	Start	End	Length (in bp)	Percent of "N"' sequence	Percent of GC	No. of genes	Cumulate length of gene (in bp)	No. of TEs	Cumulate length of TEs (in bp)	No. of rDNA	Cumulate length of rDNA (in bp)
A01	Centromeric	16,062,102	23,330,349	7,268,248	25.51	30.39	230	360,211	3,207	4,608,900	47	157,414
A05	Telomeric	1	6,383,565	6,383,565	24.99	35.01	118	709,602	2,549	4,337,199	7	22,711
A05	Centromeric	19,127,969	33,629,471	14,501,503	21.90	36.61	153	947,497	6,431	11,118,253	18	57,395
A06	Pericentromeric	12,173,306	42,412,226	30,238,921	29.87	32.49	419	2,408,139	12,363	19,669,728	61	193,702
A08	Centromeric	6,540,349	8,868,417	2,328,069	27.25	29.90	8	1,953	580	1,554,720	0	0
A09	Centromeric	21,830,260	34,184,774	12,354,515	54.00	22.20	35	268,743	3,370	5,288,282	20	64,209

For each region, the number and cumulative length of genes, transposable elements, and ribosomal DNA were identified.

and 'Chiifu' (cumulative length = 176,703,831 bp) reciprocally (**Supplementary Table 3**). Inversions in 'Z1' and 'Chiifu' had a mean length of 281,771 bp (median = 72,650 bp) and 2,181,528 bp (median = 107,269 bp), respectively. A total of 22,274 (mean = 274.99, median = 14) and 21,780 (mean = 268.89, median = 13) genes were identified within 'Z1' and 'Chiifu' inversions, respectively. A total of 9,240 (mean = 114.07, median = 35) and 112,578 (mean = 1,389.85, median = 63) TEs were identified within 'Z1' and 'Chiifu' inversions, respectively. The cumulative length of TEs within the inversions corresponded to 49.86 and 44.15% of the total length of the inversions in 'Z1' and 'Chiifu,' respectively. By contrast, a very low number of rDNA sequences were identified in these inversions (with overall three and five rDNA sequences in 'Z1' and 'Chiifu,' respectively; **Supplementary Table 3**). Among these inversions, we were able to validate three inversions between 'Z1' and 'Chiifu,' of which two were present on chromosome A05 (located from 6.48–6.50 and 6.65–7.03 Mb) and one on chromosome A10 (located from 18.24–19.57 Mb) using PCR (**Supplementary Figure 1**).

Repeat Element Content in *B. rapa* 'Z1' and 'Chiifu'

The proportions of repeat elements (including TEs and ribosomal DNA) were estimated in *B. rapa* 'Z1' and *B. rapa* 'Chiifu.' TEs were annotated along the 10 chromosomes of both genomes with higher proportions in the centromeric regions. Considering a genome size of 529 Mb for *B. rapa* 'Z1' and 'Chiifu' after Johnston et al. (2005), TEs may represent 31.73 and 22.16% of 'Z1' and 'Chiifu' genomes, respectively. However, considering the genome sizes determined in this study by flow cytometry (577 and 545 Mb on average for 'Z1' and 'Chiifu,' respectively, see below), TEs may represent 29.09 and 21.50% of 'Z1' and 'Chiifu' genomes respectively. LINEs and LTR (including Copia and gypsy) TEs represent 3.02 and 8.58% of the 'Z1' genome, respectively. These proportions were also determined for the 'Chiifu' genome with 2.92 and 4.32% for LINEs and LTR TEs, respectively.
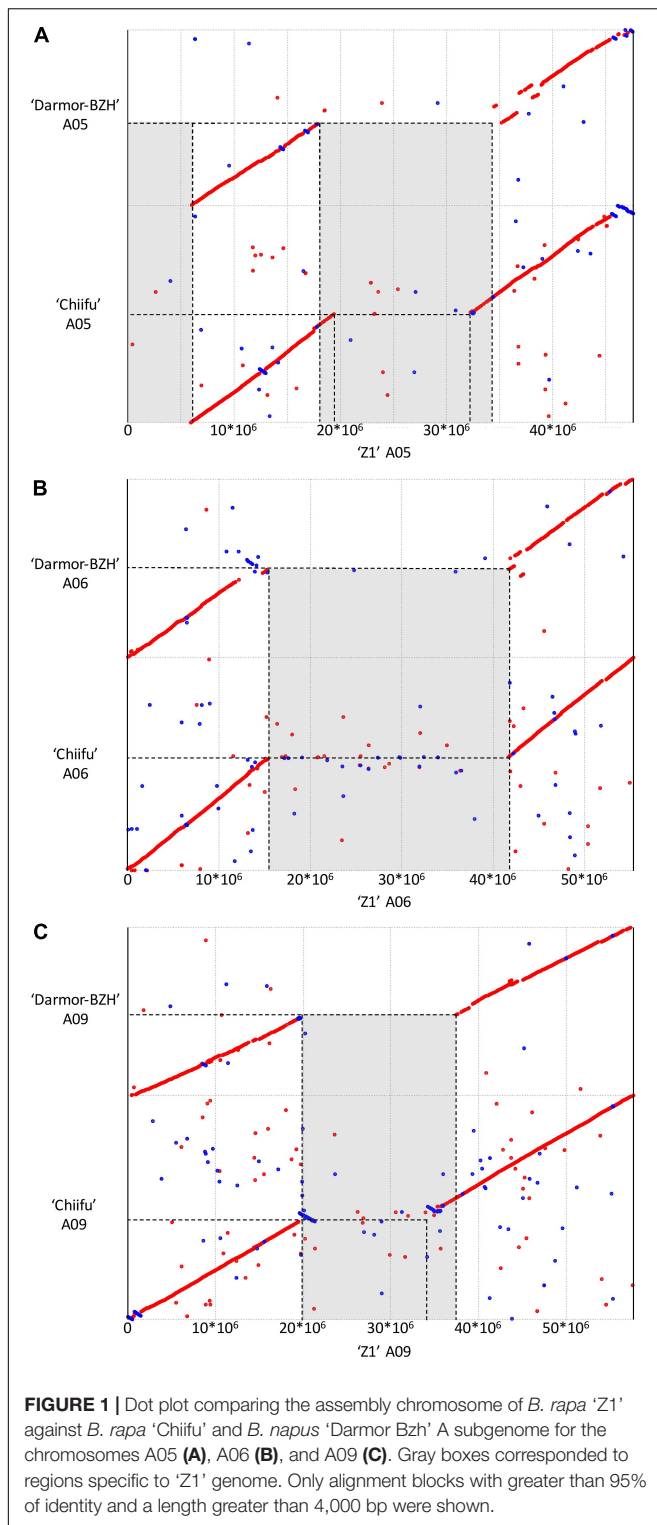
In *B. rapa* 'Z1,' several chromosomes (especially the A05 and A06 chromosomes) had a higher proportion of TEs compared to 'Chiifu' assembled chromosomes. As expected, regions rich in TEs had a lower proportion of genes. Ribosomal sequences were observed in the 10 chromosomes of both genomes with a higher proportion on the A01, A05, A06, and A09 compared

to other chromosomes of 'Z1' (**Figure 2**). As the short arm of the chromosome A03 (bearing the nucleolar organizer region, including a large domain of 45S rDNA) was not assembled, no rDNA was observed (Mun et al., 2010). These results were confirmed by CNV analysis, which allowed estimating the number of 45S rDNA copies to 1,416 and 1,750 in 'Chiifu' and 'Z1,' respectively. Interestingly, the regions presenting a high number of rDNA and TEs corresponded to the six regions specific to the 'Z1' genome (**Table 1** and **Figure 1**). Proportions of each TE order in the two genomes were also investigated using χ^2 test. No significant difference was observed between 'Chiifu' chromosomes; however, statistical differences were observed between 'Z1' chromosomes (χ^2-test; α = 0.05). Indeed, chromosomes A05 and A06 had a higher percentage of LTR TEs in comparison to the other chromosomes (**Figure 3** and **Supplementary Table 4**). CNV analysis confirmed that the proportion of LTR gypsy was higher in the 'Z1' genome (1,626 copies) compared to the 'Chiifu' genome (1,347 copies).

As the large specific sequences of 'Z1' were made predominantly of rDNA and LTR gypsy, phylogenies of both repeated elements were performed to explore their evolutionary dynamics. Ribosomal DNA phylogeny of the 'Chiifu' and 'Z1' genomes allowed us to identify three clades, one common to the two *B. rapa* samples (clade 3) and two containing only 'Z1' sequences (clades 1 and 2). These two clades contained 67 and 96 'Z1' rDNA sequences, respectively, coming from chromosomes A01, A05, A06, and A09 (**Figure 4A**). LTR gypsy phylogeny allowed us to identify six major clades (**Figure 4B**). Proportions of sequences within each LTR gypsy clade were not significantly different between 'Chiifu' and 'Z1' (χ^2-test; a = 0.05).

Validation of the A06 Chromosome Length Variation in B. *rapa* 'Z1' vs. 'Chiifu' Using Cytogenetic Approaches

Fluorescence *in situ* hybridization was performed on 'Z1,' 'Chiifu,' and F1 'Z1' × 'Chiifu' individuals to compare their A06 chromosomes, presenting high length variation. The 45S probe (green fluorescence) on *B. rapa* marked five different chromosome pairs (**Figure 5**). The strong FISH signal located on the A03 chromosomes likely reflected a large number of genes in the array of the 45S rDNA sequence. A second rDNA-rich locus was located on A01 chromosome, close to the centromere.

FIGURE 1 | Dot plot comparing the assembly chromosome of *B. rapa* 'Z1' against *B. rapa* 'Chiifu' and *B. napus* 'Darmor Bzh' A subgenome for the chromosomes A05 **(A)**, A06 **(B)**, and A09 **(C)**. Gray boxes corresponded to regions specific to 'Z1' genome. Only alignment blocks with greater than 95% of identity and a length greater than 4,000 bp were shown.

BAC-FISH analysis with two specific BACs of A06 chromosome arms (red fluorescence) revealed that the 45S rDNA probe hybridized to a major site on *B. rapa* 'Chiifu' (**Figures 5A–C**), a minor site on *B. rapa* 'Z1' (**Figures 5D–F**) and mixed-intensity sites, one major and one minor, on hybrid 'Z1' × 'Chiifu'

(**Figures 5G–I**). Moreover, in *B. rapa*, the individual mitotic metaphase chromosome A06 ranged from 2.1 to 5.6 μm in length (observed from 14 metaphase sets). The chromosome A06 from 'Chiifu' was on average 17.6% (±6.96%) smaller compared to the chromosome A06 of 'Z1' (**Supplementary Table 4**).

Genome Size Variation and Presence of Inversions Among B. *rapa* Accessions

DNA content was obtained for 'Z1,' 'Chiifu,' and 10 accessions representing *B. rapa* core collection. After checking normality of data (Shapiro test, $P = 0.487$), the significative impact of accessions was tested (ANOVA; $df = 11$, $P = 4.19e\text{-}13$), followed by Tukey pairwise comparisons and group assignment based on $P < 0.05$. Of particular interest, the two focal varieties 'Chiifu' and 'Z1' have significantly different 1C DNA content with 0.557 (median = 0.565) and 0.590 (median = 0.588), respectively (Tukey HSD; $P = 0.019$). All accessions can be divided in seven subspecies and six groups structured within the core collection (**Figure 6**). The variation in genome size does not seem to be associated with any of those phylogenetic structures.

Finally, we explored the presence of these inversions in the same core collection of *B. rapa* accessions (**Supplementary Figure 1**). Although it appeared that some nucleotide divergence might have prevented amplification, some accessions seem to share preponderantly either one or the other genetic structure (inversion on A05 zone II: populations 1, 2, 4, 5, and 10 might seem alike 'Z1,' whereas populations 3, 7, 8, and 9 seem alike 'Chiifu'; inversion on A10: populations 3, 4, 5, 6, 7, 9, and 10 seem more alike 'Z1').

DISCUSSION

The large phenotypic diversity within *B. rapa* species has always triggered intensive investigations from the community. Ancient whole genome duplication followed by subgenome parallel selection has been shown to play a role in morphotype diversification in *B. rapa* (Cheng et al., 2016b). In addition, contrasted levels of cytosine methylations correlating with various TE densities have been proposed to shape subgenome dominance and fractionation levels within the genome of *B. rapa* (Cheng et al., 2016a). However, the extent of TE diversity within the species as well as the structural genomic variants that might additionally explain within-species diversity has merely been touched upon.

Detection of Structural Variations Between Two Accessions of B. *rapa* and Impact on Genome Size
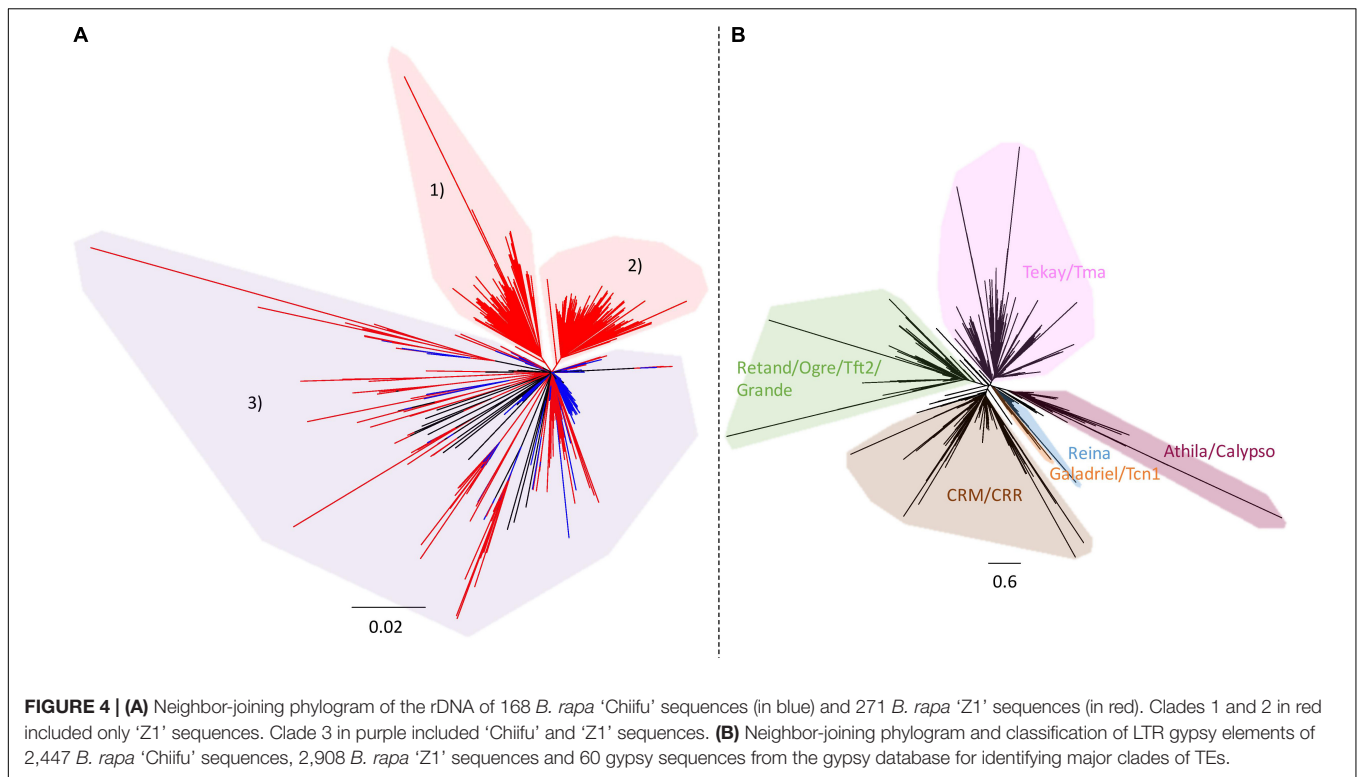
A recent study on whole-genome resequencing of eight *B. napus* accessions revealed millions of genomic variants, resulting in 3.2–6.8% (for a total of 77.2–149.6 Mb) of the genome presenting such variations (Song et al., 2020). Similarly, chromosome-level assemblies of seven *Arabidopsis thaliana* genomes unravel overall 13–17 Mb of rearranged genomic variants and up
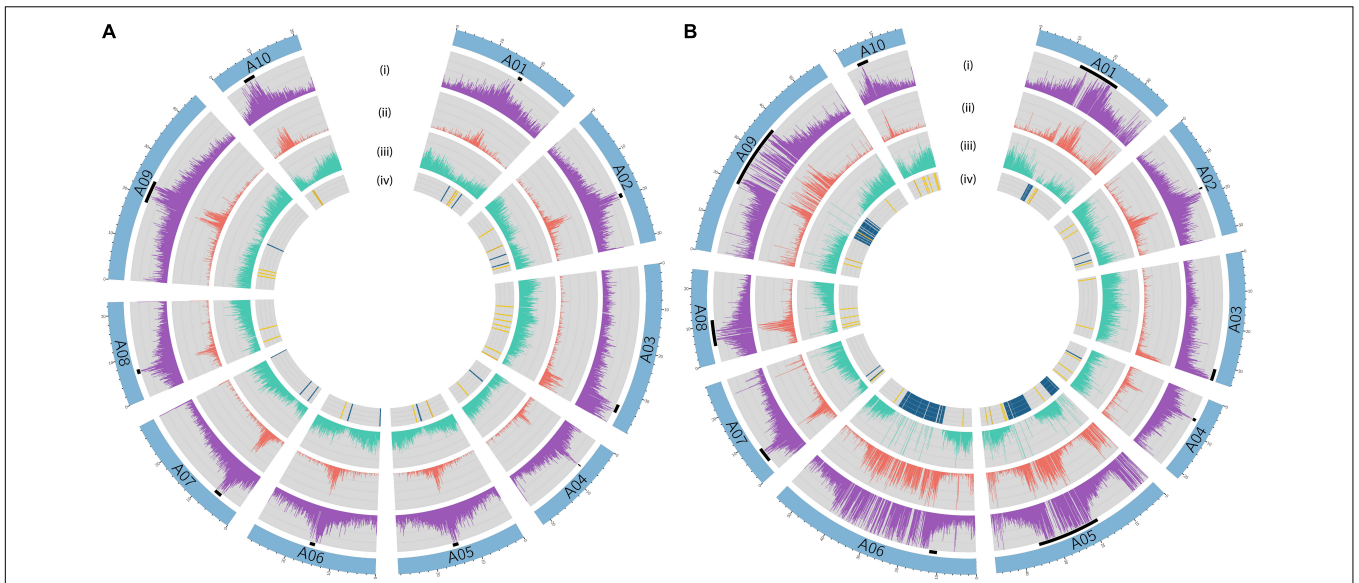
FIGURE 2 | Circular diagrams of the chromosomes of **(A)** *B. rapa* 'Chiifu' and **(B)** *B. rapa* 'Z1.' Black boxes represent chromosomes centromeric regions. Rings represent: (i) density of transposable elements, (ii) density of LTR, (iii) density of genes, and (iv) presence–absence of ribosomal DNA (45S in blue and 5S in yellow).
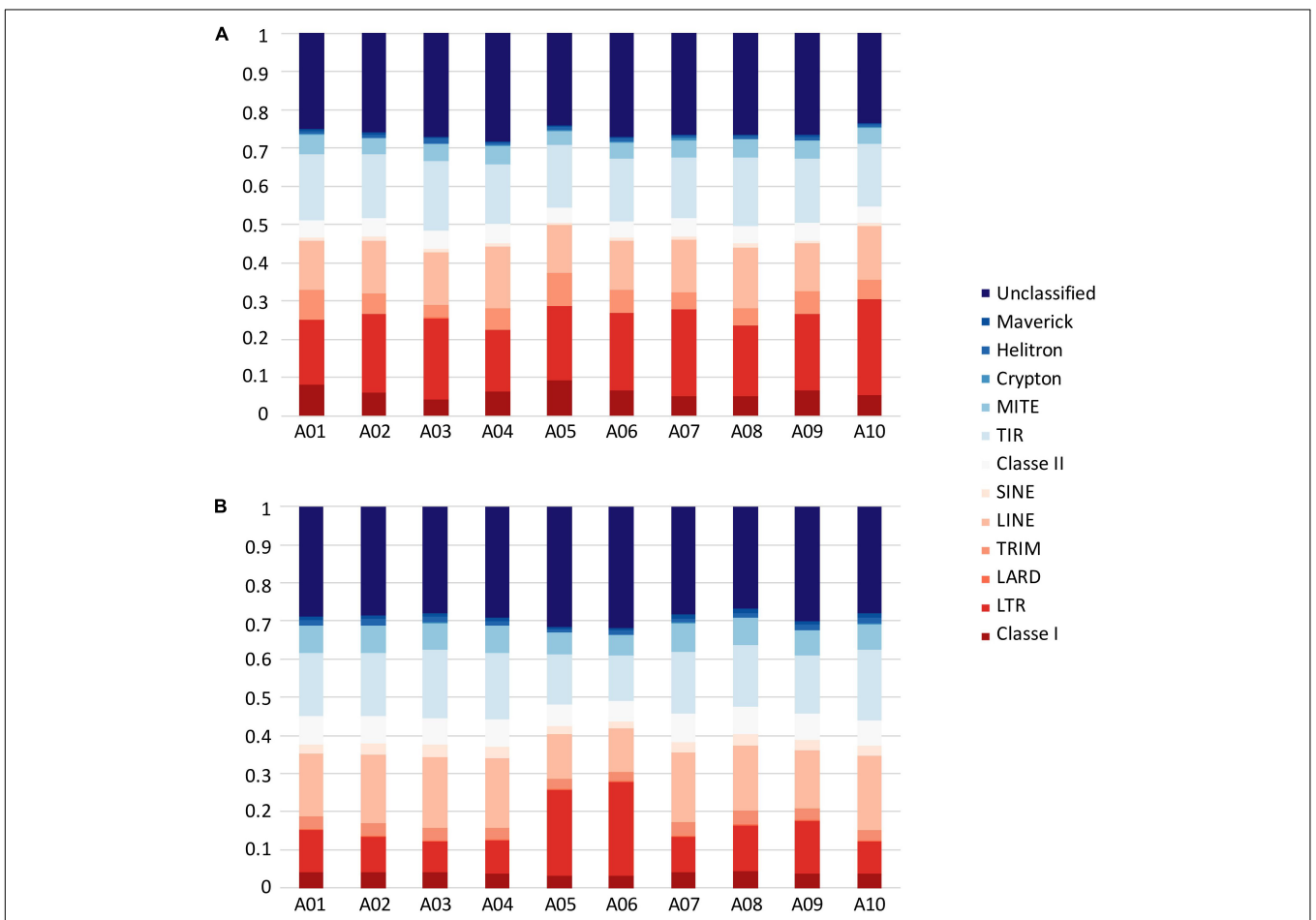
FIGURE 3 | Transposable element proportions in the 10 chromosomes of **(A)** *B. rapa* 'Chiifu' and **(B)** *B. rapa* 'Z1.' Class I and Class II corresponded to unclassified TEs within their respective Class.

FIGURE 4 | (A) Neighbor-joining phylogram of the rDNA of 168 *B. rapa* 'Chiifu' sequences (in blue) and 271 *B. rapa* 'Z1' sequences (in red). Clades 1 and 2 in red included only 'Z1' sequences. Clade 3 in purple included 'Chiifu' and 'Z1' sequences. **(B)** Neighbor-joining phylogram and classification of LTR gypsy elements of 2,447 *B. rapa* 'Chiifu' sequences, 2,908 *B. rapa* 'Z1' sequences and 60 gypsy sequences from the gypsy database for identifying major clades of TEs.

to 6.5 Mb (or 5.5% of the reference genome) of genome-specific sequences based on all possible pairwise comparisons (Jiao and Schneeberger, 2020). In the present study, we identified 23 Mb of inversions (representing 4.3% of the genome) and 73 Mb of specific insertions (representing 13.8% of the genome) in 'Z1' compared to 'Chiifu.' These results show the large unexplored genomic diversity present in only two accessions of *B. rapa*. Although we cannot directly address their role in phenotypic diversification within the species, these insertions include 963 coding regions for which almost half of them cannot find an orthologous copy in the 'Chiifu' genome. These genes might be part of the dispensable gene pool as explained in a study gathering 10 accessions of *B. oleracea*, where 81.3% of genes represent the core, whereas the remaining genes (18.7%) represent dispensable genes (Golicz et al., 2016).

Besides the functional consequences of such variants, the whole genome structure could be particularly impacted. The variation in genome sizes between our focal accessions has been previously addressed with estimations varying between 529 and 443 Mb for 'Z1' and 'Chiifu,' respectively (Belser et al., 2018; Zhang et al., 2018). However, the disparate methods used, either by flow cytometry or k-mer analysis, can reveal wide differences (see Zhang et al., 2018). Here, using internal control and repeated evaluations of DNA content by flow cytometry, we estimated an increase of ca. 6% of genome size between 'Z1' and 'Chiifu.' In addition, we were able to validate the difference of the A06 chromosome lengths between the focal accessions using a ribosomal DNA probe in a BAC-FISH experiment showing firsthand the possible additional impact of these elements on chromosome and genome size.

Extending our analysis to the *B. rapa* core collection, we explored the variation in genome sizes among different accessions of *B. rapa* representatives of the phenotypic diversity in the species. Interestingly, up to 16% variation in genome size was observed among the 12 accessions without visible structuration according to their phylogenetic relationships. These differences are probably due to intraspecific variation in repeat content and insertion/deletion of sequences as their expansion correlates with the diversification of morphotypes (ca. 1 Mya, Zhao et al., 2013).

Very few studies tried so far to identify intraspecific genome size variation in Brassicaceae (with same ploidy level and same number of chromosomes). In Johnston et al. (2005), *B. rapa* species (including unknown accessions) exhibited wider standard error than any other Brassicaceae species. By comparison in *A. thaliana*, the analyses of DNA content in 10 ecotypes showed little variation with less than 1% among the different accessions (Johnston et al., 2005), which seemed to be corroborated with recent assemblies of various *A. thaliana* accessions with low genome size variance (Jiao and Schneeberger, 2020).

Contrasted Repeat Landscapes Between the Focal Accessions

Usually, genome size variation is often associated with TE content and ribosomal repeats as these elements could be highly dynamic. The large specific insertions identified in 'Z1' allowed improving our understanding of centromeric and pericentromeric regions that are usually difficult to sequence and assemble in complex genomes such as plant genomes, because of their high density in repeats. Interestingly, apart from some dispensable genes,
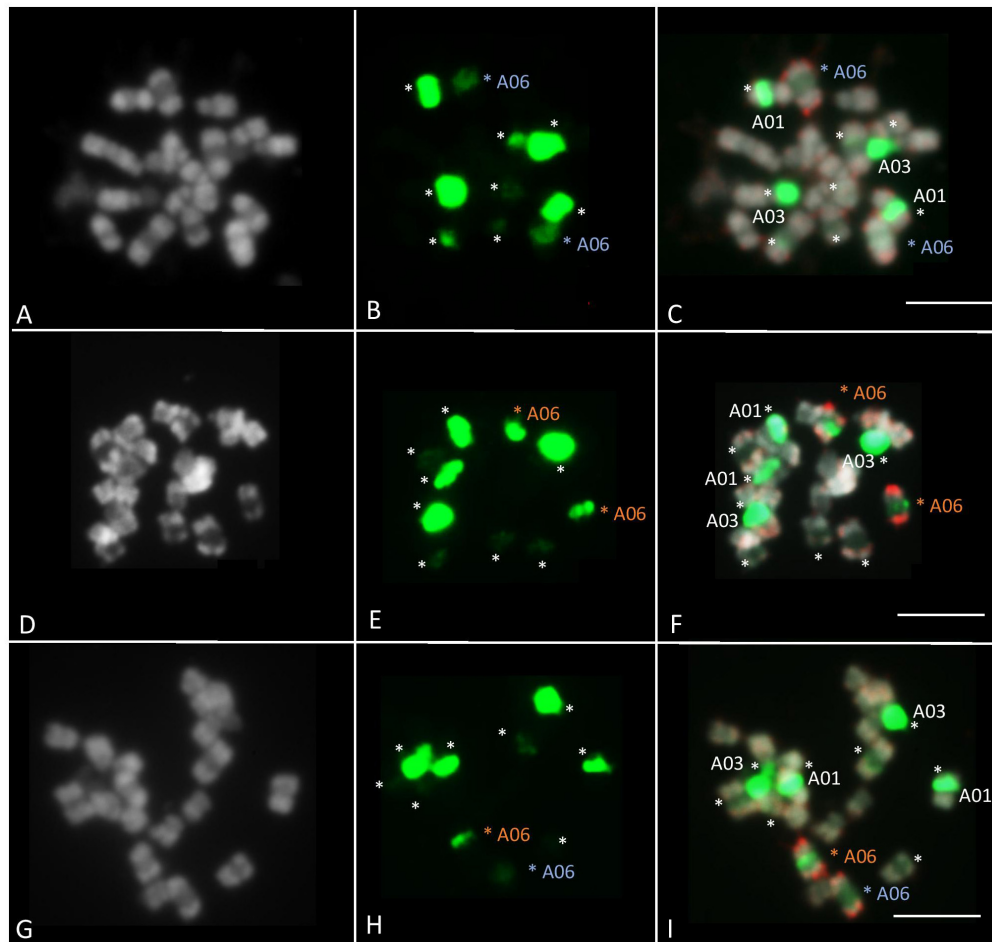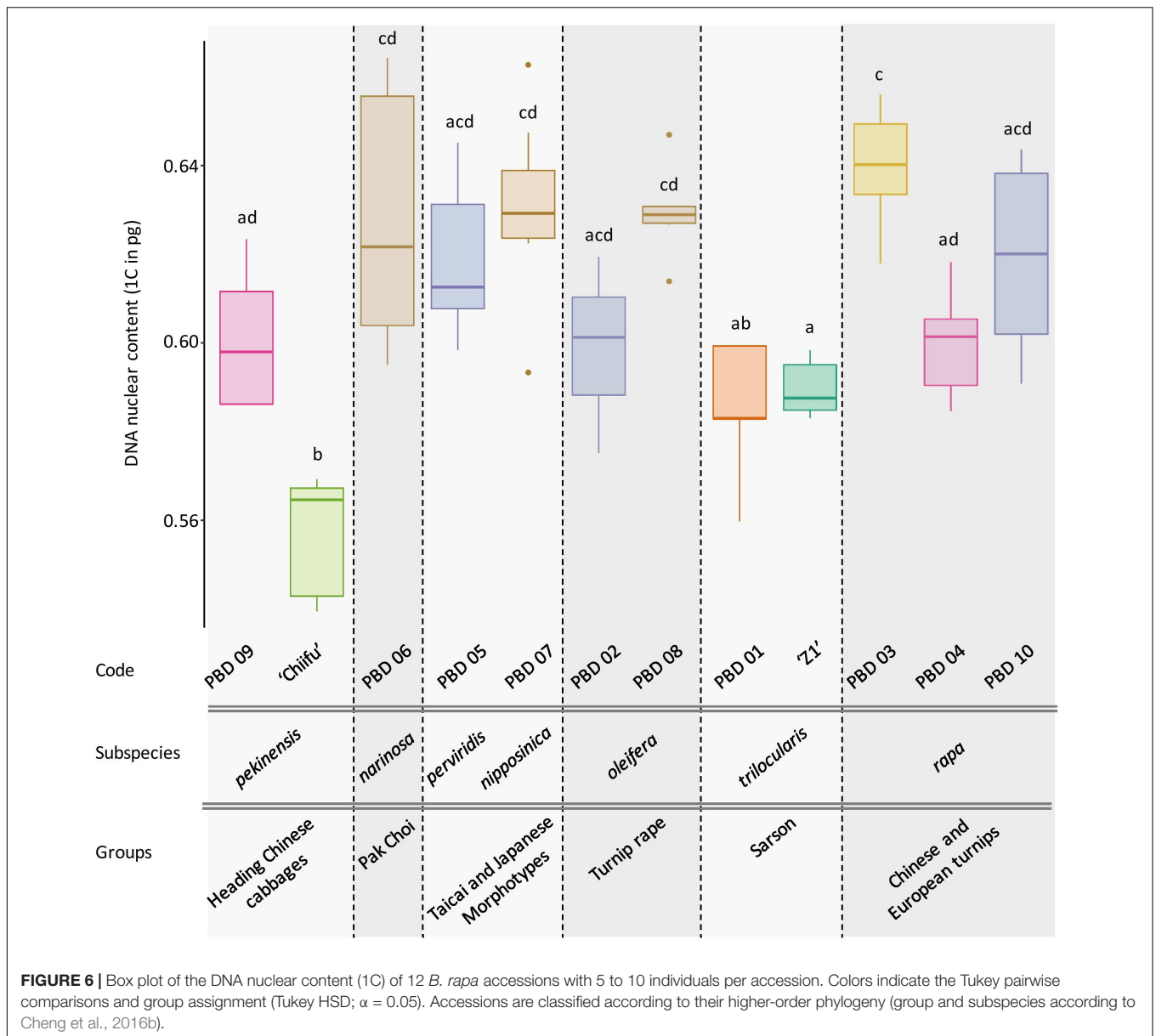
FIGURE 5 | Fluorescent *in situ* hybridization was carried out using 45S rDNA (in green) and KBrB022P06 + KBrH003P24 BAC probes (A06) (in red). FISH analyses of somatic metaphase chromosomes were carried out in *B. rapa* 'Z1' **(A–C)**, *B. rapa* 'Chiifu' **(D–F)**, and hybrid 'Z1' × 'Chiifu' **(G–I)**. The 45S rDNA probe on *B. rapa* marked five different chromosome pairs (stars). A06 chromosomes carrying 45S rDNA were indicated by a blue star in 'Z1' and an orange star in 'Chiifu.' Chromosomes were counterstained with DAPI (gray). Bars represent 5 μm.

specific 'Z1' insertions exhibited mostly both rDNA copies and LTR elements.

In *Brassica* species, the number of 45S rDNA loci is variable. For example, *B. oleracea*, *B. nigra*, and *B. rapa* have 4, 6, and 10 rDNA loci, respectively (Hasterok et al., 2001). However, the number of 45S (and 5S) rDNA loci varies among varieties within each species. Hasterok et al. (2006) identified 10 45S rDNA loci in all *B. rapa*, except in *B. rapa* subsp. *oleifera* (DC.) 'Schneeball' (six loci). Here, we investigated the 45S rDNA in two other *B. rapa* varieties, 'Z1' and 'Chiifu.' The number of 45S rDNA loci identified in these two varieties (ten loci, **Figure 5**) is identical to the number of 45S rDNA loci identified in the majority of other *B. rapa* varieties (Hasterok et al., 2006). However, based on the signal intensity, as shown by Xiong and Pires (2011) for the chromosome A05 of *B. rapa* 'Chiifu' vs. A05 of 'IMB218,' the loci on the chromosome A06 in *B. rapa* 'Z1' had low signal intensity but showed a decondensation, while the loci on the chromosome A06 in 'Chiifu' are fully condensed. As the A03 locus has been shown to be the only one

carrying transcriptionally active ribosomal genes in the nucleolar organizer region (Hasterok and Maluszynska, 2000), it could be of interest to test if the 45S rDNA genes localized on the A06 in 'Z1' are transcribed and active.

Although the number of loci between 'Z1' and 'Chiifu' is identical, the number of 45S rDNA copies is different between these two varieties. The number of copies estimated in 'Z1' (1,750 copies) is highly similar to the number of rDNA copies previously estimated in this variety (1,709 copies, Sochorová et al., 2017) and is more abundant than the number of copies identified in 'Chiifu' (1,416 copies). This difference could be the result of rearrangements with loss of rDNA copies in *B. rapa* 'Chiifu,' as demonstrated in polyploid species such as *B. napus* or *Nicotiana tabacum* (Lim et al., 2000; Książczyk et al., 2011; Sochorová et al., 2017). However, rDNA phylogeny of 'Chiifu' and 'Z1' (**Figure 4A**) has shown that among the three clades identified, two included only 'Z1' sequences (163 sequences). Thus, the most parsimonious scenario would indicate that these two clades most probably result from duplication events in 'Z1' after the

FIGURE 6 | Box plot of the DNA nuclear content (1C) of 12 *B. rapa* accessions with 5 to 10 individuals per accession. Colors indicate the Tukey pairwise comparisons and group assignment (Tukey HSD; α = 0.05). Accessions are classified according to their higher-order phylogeny (group and subspecies according to Cheng et al., 2016b).

divergence between the two focal accessions and not from loss of copies in 'Chiifu.'

Regarding TE content, assembly of the 'Z1' genome has allowed a first comparative exploration of the repetitive compartment between 'Z1' and 'Chiifu' (Belser et al., 2018); however, differences were mainly attributed to methodological dissimilarities between genome assemblies. Here, using a consistent analysis pipeline for TE detection and annotation in these two accessions, we can reveal the impact of TEs on *B. rapa* intraspecific diversity, taking into account the structural variations.

Using genome size estimation for *B. rapa* (i.e., 529 Mb according to Johnston et al., 2005), TEs represent 31.73 and 22.16% of 'Z1' and 'Chiifu' genomes, respectively. However, considering the genome sizes obtained in this study (577 and

545 Mb on average for 'Z1' and 'Chiifu,' respectively), TEs may represent 29.09 and 21.50% of the genome in 'Z1' and 'Chiifu,' respectively. Overall, our methodology allowed us to retrieve similar numbers of TEs compared to what is expected with 29 and 24% of TEs in accessions 'Z1' and 'Chiifu,' respectively (Belser et al., 2018). Interestingly, a tangible difference in TE content is consistently observed between these two accessions, and we proposed that this pattern is not widely distributed along the genome but focused on specific chromosomes and on identified structural variants.

The detailed analysis of the structural variants in 'Z1' showed that while inversions do not significantly sustain a higher proportion of TEs compared to the rest of the genome, specific insertions in 'Z1' carry a large proportion of LTR gypsy and LINEs; this is especially clear on chromosome A05 and A06 in

'Z1,' which harbor the largest genomic variants. TE proportions could be underestimated in these regions; as a large part of the inserted sequences were unsequenced data, we are nonetheless confident on the length of these sequences, thanks to reliable optical mapping scaffolds.

TE intraspecific diversity has been previously shown to be driven by an expansion of recent LTR retrotransposons in *B. rapa* with rapid nucleotide substitution (Zhao et al., 2013). However, we do not observe such patterns using a phylogenetic approach. Instead, all LTR gypsy clades seem to be represented in the inserted regions. This tendency seems to be particularly exacerbated in *B. rapa* and by extension in the A subgenome of *B. napus* compared to *B. oleracea* and the C subgenome of *B. napus* as demonstrated in Song et al. (2020). In addition, the inserted LTR retrotransposons do occur in clusters on chromosomes A05, A06, and A09, and their spreading predates the divergence between *B. rapa* and *B. oleracea* (Song et al., 2020).

CONCLUSION AND PERSPECTIVES

Our results unveil the necessity for long-read sequencing strategies and developing various whole-genome references for agronomically important species (Danilevicz et al., 2020). This recent endeavor in constructing pangenomes is revealing the importance of structural variants and the diversity in repeat content within species. This is especially relevant in species showing broad phenotypic diversity such as *Brassica* sp. Besides the role of these variants in *Brassica* diversification, we can only wonder at their role in other processes such as recombination. In that case, structural rearrangements (especially inversions) act as modifiers of recombination that create different patterns of recombination rate across the genome (Ortiz-Barrientos et al., 2016), leading to segregation distortion and partial lethality (Lynch and Force, 2000; Fransz et al., 2016). Similarly, TEs have been shown to cause post-zygotic barriers and reproductive isolation in a diversity of taxa (reviewed in Serrato-Capuchina and Matute, 2018). First hints suggest that both structural variants and repeat content seem highly dynamic among the different accessions of *B. rapa* composing the core collection. Thus, exploring this diversity will deeply improve our understanding of the species adaptive processes and could be used in the creation of new interspecific and intraspecific *Brassica* hybrid varieties.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.genoscope.cns.fr/plants and http://brassicadb.org.

AUTHOR CONTRIBUTIONS

JF and JB conceived and designed the experiments with inputs from LM. TC, SL, and LM implemented the REPET pipeline and performed the work on TE annotation. JB, LM, CF, and JM performed the bioinformatics analyses. J-MA and CB analyzed the generated optical maps and verified regions of interests. AB, VH, and OC performed the cytological analyses. GT performed the flow-cytometry experiments. GD and ML performed the DNA extractions and PCR to verify inversions. A-MC and MR-G helped supervise the project and participated in critical thinking of the results with LM, FB, CF, OC, JB, and JF. JF and JB wrote the manuscript. All authors approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.577536/full#supplementary-material

Supplementary Figure 1 | PCR amplifications and agarose gels validating inversions in three genomic regions on **(A)** chromosome A05 first inversion, **(B)** chromosome A05 second inversion and **(C)** chromosome A10. Both sides of each inversion have been investigated (L for left side and R for right side) in the focal accessions and *B. rapa* core collections (named here as PBD1 to PBD10), ordered on the gel as follows: 'Chiifu,' 'Z1,' PBD001 to PBD010.

Supplementary Table 1 | Accession identification and origin of the *B. rapa* accessions.

Supplementary Table 2 | Primer identifiers, sequences and target amplification.

Supplementary Table 3 | Position and length of the *B. rapa* 'Chiifu' vs. 'Z1' inversions. For each region, the number and cumulative length of genes, Transposable Elements and ribosomal DNA were identified.

Supplementary Table 4 | Transposable Element proportions in the 10 chromosomes of *B. rapa* 'Chiifu' and *B. rapa* 'Z1.' Class I and Class II correspond to TEs for which order's annotation remains unknown.

Supplementary Table 5 | A06 individual chromosome lengths (in μm) at mitotic metaphase in F1 hybrid ('Z1' × 'Chiifu').

REFERENCES

Aissiou, F., Laperche, A., Falentin, C., Lodé, M., Deniot, G., Boutet, G., et al. (2018). A novel *Brassica rapa* L. genetic diversity found in Algeria. *Euphytica* 214:241. doi: 10.1007/s10681-018-2318-9

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Cheng, F., Sun, C., Wu, J., Schnable, J., Woodhouse, M. R., Liang, J., et al. (2016a). Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytol.* 211, 288–299. doi: 10.1111/nph.13884

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., et al. (2016b). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat. Genet.* 48:1218. doi: 10.1038/ng.3634

Cheng, F., Wu, J., Cai, C., Fu, L., Liang, J., Borm, T., et al. (2016c). Genome resequencing and comparative variome analysis in a *Brassica rapa* and *Brassica oleracea* collection. *Sci. Data* 3:160119. doi: 10.1038/sdata.2016.119

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.* 1:14024. doi: 10.1038/hortres.2014.24

Chia, J.-M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44, 803–807. doi: 10.1038/ng.2313

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E., and Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* 54, 18–25. doi: 10.1016/j.pbi.2019.12.005

Dolatabadian, A., Bayer, P. E., Tirnaz, S., Hurgobin, B., Edwards, D., and Batley, J. (2020). Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* 18, 969–982. doi: 10.1111/pbi.13262

Ferreira de Carvalho, J., de Jager, V., van Gurp, T. P., Wagemaker, N. C. A. M., and Verhoeven, K. J. F. (2016). Recent and dynamic transposable elements contribute to genomic divergence under asexuality. *BMC Genomics* 17:884. doi: 10.1186/s12864-016-3234-9

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS One* 6:e16526. doi: 10.1371/journal.pone.0016526

Fransz, P., Linc, G., Lee, C.-R., Aflitos, S. A., Lasky, J. R., Toomajian, C., et al. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *Plant J.* 88, 159–178. doi: 10.1111/tpj.13262

Gage, J. L., Vaillancourt, B., Hamilton, J. P., Manrique-Carpintero, N. C., Gustafson, T. J., Barry, K., et al. (2019). Multiple maize reference genomes impact the identification of variants by genome-wide association study in a diverse inbred panel. *Plant Genome* 12, 1–12. doi: 10.3835/plantgenome2018.09.0069

Gazave, E., Tassone, E. E., Ilut, D. C., Wingerson, M., Datema, E., Witsenboer, H. M. A., et al. (2016). Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. *Front. Plant Sci.* 7:525. doi: 10.3389/fpls.2016.00525

Gerlach, W. L., and Bedbrook, J. R. (1979). Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res.* 7, 1869–1885. doi: 10.1093/nar/7.7.1869

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., et al. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7:13390. doi: 10.1038/ncomms13390

Guo, Y., Chen, S., Li, Z., and Cowling, W. A. (2014). Center of origin and centers of diversity in an ancient crop, *Brassica rapa* (Turnip Rape). *J. Hered.* 105, 555–565. doi: 10.1093/jhered/esu029

Hasterok, R., Jenkins, G., Langdon, T., Jones, R. N., and Maluszynska, J. (2001). Ribosomal DNA is an effective marker of Brassica chromosomes. *Theor. Appl. Genet.* 103, 486–490. doi: 10.1007/s001220100653

Hasterok, R., and Maluszynska, J. (2000). Nucleolar dominance does not occur in root tip cells of allotetraploid Brassica species. *Genome* 43, 574–579. doi: 10.1139/g00-005

Hasterok, R., Wolny, E., Hosiawa, M., Kowalczyk, M., Kulak-Ksiazczyk, S., Ksiazczyk, T., et al. (2006). Comparative analysis of rDNA distribution in chromosomes of various species of brassicaceae. *Ann. Bot.* 97, 205–216. doi: 10.1093/aob/mcj031

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., et al. (2014). PASTEC: an automatic transposable element classification tool. *PLoS One* 9:e91929. doi: 10.1371/journal.pone.0091929

Jiao, W.-B., and Schneeberger, K. (2020). Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat. Commun.* 11:989. doi: 10.1038/s41467-020-14779-y

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., et al. (2005). Evolution of genome size in brassicaceae. *Ann. Bot.* 95, 229–235. doi: 10.1093/aob/mci016

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

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kidwell, M. G., and Lisch, D. R. (2000). Transposable elements and host genome evolution. *Trends Ecol. Evol.* 15, 95–99. doi: 10.1016/s0169-5347(99)01817-0

Książczyk, T., Kovarik, A., Eber, F., Huteau, V., Khaitova, L., Tesarikova, Z., et al. (2011). Immediate unidirectional epigenetic reprogramming of NORs occurs independently of rDNA rearrangements in synthetic and natural forms of a polyploid species *Brassica napus*. *Chromosoma* 120, 557–571. doi: 10.1007/s00412-011-0331-z

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.-L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, P., Zhang, S., Li, F., Zhang, S., Zhang, H., Wang, X., et al. (2017). A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important *Brassica* species comprising the triangle of U. *Front. Plant Sci.* 8:111. doi: 10.3389/fpls.2017.00111

Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979

Lim, K. Y., Matyášek, R., Lichtenstein, C. P., and Leitch, A. R. (2000). Molecular cytogenetic analyses and phylogenetic studies in the Nicotiana section Tomentosae. *Chromosoma* 109, 245–258. doi: 10.1007/s004120000074

Lu, F., Romay, M. C., Glaubitz, J. C., Bradbury, P. J., Elshire, R. J., Wang, T., et al. (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat. Commun.* 6:6914. doi: 10.1038/ncomms7914

Lynch, M., and Force, A. G. (2000). The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* 156, 590–605. doi: 10.1086/316992

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., et al. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* 11:e1004915. doi: 10.1371/journal.pgen.1004915

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C.-K. K., et al. (2017). The pangenome of hexaploid bread wheat. *Plant J.* 90, 1007–1013. doi: 10.1111/tpj.13515

Mun, J.-H., Kwon, S.-J., Seol, Y.-J., Kim, J. A., Jin, M., Kim, J. S., et al. (2010). Sequence and structure of *Brassica rapa* chromosome A3. *Genome Biol.* 11:R94. doi: 10.1186/gb-2010-11-9-r94

Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., et al. (2015). Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biol.* 16:262. doi: 10.1186/s13059-015-0814-y

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

Ortiz-Barrientos, D., Engelstädter, J., and Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends Ecol. Evol.* 31, 226–236. doi: 10.1016/j.tree.2015.12.016

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M., and Ainouche, M. (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in Spartina. *New Phytol.* 184, 1003–1015. doi: 10.1111/j.1469-8137.2009.03029.x

Paritosh, K., Yadava, S. K., Singh, P., Bhayana, L., Mukhopadhyay, A., Gupta, V., et al. (2019). A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. *bioRxiv*[Preprint] doi: 10.1101/681080

Qi, X., An, H., Ragsdale, A. P., Hall, T. E., Gutenkunst, R. N., Pires, J. C., et al. (2017). Genomic inferences of domestication events are corroborated by written records in Brassica rapa. *Mol. Ecol.* 26, 3373–3388. doi: 10.1111/mec.14131

Quadrana, L., Bortolini Silveira, A., Mayhew, G. F., LeBlanc, C., Martienssen, R. A., Jeddeloh, J. A., et al. (2016). The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* 5:e15716. doi: 10.7554/eLife.15716

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., et al. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* 1:e22. doi: 10.1371/journal.pcbi.0010022

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Schatz, M. C., Maron, L. G., Stein, J. C., Wences, A. H., Gurtowski, J., Biggers, E., et al. (2014). Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biol.* 15:506. doi: 10.1186/s13059-014-0506-z

Serrato-Capuchina, A., and Matute, D. R. (2018). The role of transposable elements in speciation. *Genes (Basel)* 9:6914. doi: 10.3390/genes9050254

Sochorová, J., Coriton, O., Kuderová, A., Lunerová, J., Chèvre, A.-M., and Kovařík, A. (2017). Gene conversion events and variable degree of homogenization of rDNA loci in cultivars of Brassica napus. *Ann. Bot.* 119, 13–26. doi: 10.1093/aob/mcw187

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Song, K., Osborn, T. C., and Williams, P. H. (1990). Brassica taxonomy based on nuclear restriction fragment length polymorphisms. *Theor. Appl. Genet.* 79, 497–506. doi: 10.1007/bf00226159

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Ungerer, M. C., Strakosh, S. C., and Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16, R872–R873. doi: 10.1016/j.cub.2006.09.020

Vicient, C. M., and Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* 120, 195–207. doi: 10.1093/aob/mcx078

Walker, E. L., Robbins, T. P., Bureau, T. E., Kermicle, J., and Dellaporta, S. L. (1995). Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. *EMBO J.* 14, 2350–2363. doi: 10.1002/j.1460-2075.1995.tb07230.x

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165

Wright, S. I., and Schoen, D. J. (1999). Transposon dynamics and the breeding system. *Genetica* 107, 139–148. doi: 10.1007/978-94-011-4156-7_16

Xiong, Z., and Pires, J. C. (2011). Karyotype and identification of all homoeologous chromosomes of allopolyploid Brassica napus and its diploid progenitors. *Genetics* 187, 37–49. doi: 10.1534/genetics.110.122473

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., et al. (2018). Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic. Res.* 5, 1–11. doi: 10.1038/s41438-018-0071-9

Zhao, J., Artemyeva, A., Del Carpio, D. P., Basnet, R. K., Zhang, N., Gao, J., et al. (2010). Design of a *Brassica rapa* core collection for association mapping studies. *Genome* 53, 884–898. doi: 10.1139/G10-082

Zhao, M., Du, J., Lin, F., Tong, C., Yu, J., Huang, S., et al. (2013). Shifts in the evolutionary rate and intensity of purifying selection between two *Brassica* genomes revealed by analyses of orthologous transposons and relics of a whole genome triplication. *Plant J.* 76, 211–222. doi: 10.1111/tpj.12291

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z