A high-continuity genome assembly of Chinese flowering cabbage (Brassica rapa var. parachinensis) provides new insights into Brassica genome structure evolution

Guangguang Li[1], Juntao Wang[2], Yi Liao[1], Ding Jiang [1], Yansong Zheng[1], Xiuchun Dai[1], Hailong Ren[1], Hua Zhang[1], and Changming Chen[1]

[1]Affiliation not available
[2]South China Agricultural University

September 11, 2020

Abstract

Chinese flowering cabbage (Brassica rapa var. parachinensis) is a popular and widely cultivated leaf vegetable crop in Asia. Here, we performed a high quality de novo assembly of the 384 Mb genome of 10 chromosomes of a typical cultivar of Chinese flowering cabbage with an integrated approach using PacBio, Illumina, and Hi-C technology. We modeled 47,598 protein-coding genes in this analysis and annotated 52% (205.9/384) of its genome as repetitive sequences including 17% in DNA elements and 22% in long terminal retrotransposons (LTRs). Phylogenetic analysis reveals the genome of the Chinese flowering cabbage has a closer evolutionary relationship with the AA diploid progenitor of the allotetraploid species, Brassica juncea. Comparative genomic analysis of Brassica species with different subgenome types (A, B and C) reveals that the pericentromeric regions on chromosome 5 and 6 of the AA genome have been significantly expanded compared to the orthologous genomic regions in the BB and CC genomes, largely drive by LTR-retrotransposon amplification. This lineage-specific expansion may play a role in the species divergence in the Brassica genus. Furthermore, we found that a large amount of structural variations (SVs) identified within B. rapa lines that could impact coding genes, suggesting the functional significance of SVs on Brassica genome evolution. Overall, our high-quality genome assembly of the Chinese flowering cabbage provides a valuable genetic resource for deciphering the genome evolution of Brassica species and it can potentially serve as the reference genome guiding the molecular breeding practice of B. rapa crops.

A high-continuity genome assembly of Chinese flowering cabbage (*Brassica rapa* var. *parachinensis*) provides new insights into *Brassica* genome structure evolution

Running title: A chromosome-level genome assembly of Chinese flowering cabbage

Guangguang Li[1#], Juntao Wang[2#], Yi Liao[3#], Ding Jiang[1], Yansong Zheng[1], Xiuchun Dai[1], Hailong Ren[1], Hua Zhang[1*], and Changming Chen[2*]

[1] Guangzhou Institute of Agriculture Science, Guangzhou, 510308, China

[2] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (South China), Ministry of Agriculture and Rural Affairs, College of Horticulture, South China Agricultural University, Guangzhou, Guangdong, 510642, P.R. China

[3] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, 92697, USA.

[#] Equal contributors

Correspondence should be addressed to Changming Chen, cmchen@scau.edu.cn and Hua Zhang, Huangz123@163.com

Abstract: Chinese flowering cabbage (*Brassica rapa* var.*parachinensis*) is a popular and widely cultivated leaf vegetable crop in Asia. Here, we performed a high quality de novo assembly of the 384 Mb genome of 10 chromosomes of a typical cultivar of Chinese flowering cabbage with an integrated approach using PacBio, Illumina, and Hi-C technology. We modeled 47,598 protein-coding genes in this analysis and annotated 52% (205.9/384) of its genome as repetitive sequences including 17% in DNA elements and 22% in long terminal retrotransposons (LTRs). Phylogenetic analysis reveals the genome of the Chinese flowering cabbage has a closer evolutionary relationship with the AA diploid progenitor of the allotetraploid species, *Brassica juncea* . Comparative genomic analysis of *Brassica* species with different subgenome types (A, B and C) reveals that the pericentromeric regions on chromosome 5 and 6 of the AA genome have been significantly expanded compared to the orthologous genomic regions in the BB and CC genomes, largely drive by LTR-retrotransposon amplification. This lineage-specific expansion may play a role in the species divergence in the *Brassica* genus. Furthermore, we found that a large amount of structural variations (SVs) identified within *B. rapa* lines that could impact coding genes, suggesting the functional significance of SVs on *Brassica* genome evolution. Overall, our high-quality genome assembly of the Chinese flowering cabbage provides a valuable genetic resource for deciphering the genome evolution of *Brassica* species and it can potentially serve as the reference genome guiding the molecular breeding practice of *B. rapa* crops.

Keywords: Chinese flowering cabbage; *Brassica rapa* var.*parachinensis* ; genome structure evolution; assembly; PacBio; Hi-C

Introduction

Brassica , which belongs to the *Brassicaceae* family, is among the most economically important genus, since it contains a wide range of staple vegetables and oilseed crops. Over the course of its evolution, *Brassica* experienced an additional genome-wide triplication (WGT) event after it splitted with *Arabidopsis* from a common ancestor(Cheng et al., 2016; Lysak, Koch, Pecinka, & Schubert, 2005). Thus, species in the*Brassica* genus not only display great morphological and phytochemical diversity but also karyotype diversity(Cheng et al., 2016; Wang et al., 2019). Among the most agriculturally important*Brassica* species, there are three diploid genome types including*Brassica rapa* (AA), *Brassica nigra* (BB) and *Brassica oleracea* (CC), and three allopolyploid species which were generated by the pair combinations of the former three diploid species, including*Brassica napus* (AACC), *Brassica juncea* (AABB) and*Brassica carinata* (BBCC). These six species and their evolutionary origination and relationship with each other are well defined in a 'triangle of U' model(Wang et al., 2019;Yang et al., 2016).

Due to the rapid recent advances in sequencing technology, especially the next-generation sequencing (NGS), a large number of *Brassica*species have been sequenced, but most are only on a primitive level of quality. These sequenced genomes, for example those sequenced with illumina/Roche 454 technology, including *B. rapa* var.*pekinensis* Chiifu (Wang et al., 2011),*B. oleracea* 02-12(Liu et al., 2014), *B. oleracea*TO1000DH(Parkin et al., 2014), *B. nigra*YZ12151(Yang et al., 2016),*B. napus* (Bayer et al., 2017; Chalhoub et al., 2014; Sun et al., 2017), and *B. juncea* (Wang et al., 2019; Yang et al., 2016) had a relatively low continuity which may impede the genomic analysis especially at the complex genomic parts such as pericentromeric and centromeric regions. Only until recently, the application of long-read sequencing technologies, including Oxford Nanopore Technology (ONT) and Pacific Biosciences (PACBIO), to genome assembling has greatly improved continuity of the assembled contigs.There are at least four*Brassica* genomes that were reported to be sequenced with long read technology with a resulting contig N50 up to megabase size, including *B. oleracea* cultivars HDEM, *Brassica rapa* Z1 (yellow sarson)(Belser et al., 2018), *B. oleracea* var. botrytis(Sun et al., 2019) and *B. napus* (Song et al., 2020). These studies demonstrated great success in the assembly of high continuity genome assemblies (i.e. N50>5Mb)(Belser et al., 2018) with long read technology in *Brassica* genomes. Since the great morphological and phytochemical diversity in the*Brassica* species, genome information from a wide range of representative *Brassica* species will be helpful and needed to deeply decipher the genomic variants that may contribute to the great diversity that not only phenotype

but also karyotype various cultivars of the species.

The Chinese flowering cabbage (*Brassica rapa* var.*parachinensis*), locally known as Caixin, Tsai Tai, Choy Sum, bok choy, or Tsai Hsin(Tan, Fan, Kuang, Lu, & Reiter, 2019; Xiao et al., 2019), is an important leafy and bolting stem vegetable widely grown in Asia, particularly in China, Japan, and Korea(Kamran et al., 2020). This vegetable has high nutritional value and is rich in vitamins, minerals, secondary metabolites and dietary fiber, which confer human health-promoting effects(Xiao et al., 2019). Unlike other *B. rapa* vegetables, Chinese flowering cabbage can bolt and flower easily without strict vernalization under low temperature. Therefore, it is very important to conduct this genome sequencing and assembly to further uncover the genomic information and molecular mechanisms involved in the formation of special morphological and phytochemical characteristics of this cultivar.

In this study, we report a high continuity (N50 = 7.2 Mb) and chromosome level genome assembly for Chinese flowering cabbage *(Brassica rapa*). It was assembled with an integrated approach using Illumina sequencing, PacBio and high-throughput chromosome conformation capture (Hi-C) technology. The assembly resolved a large part of the pericentromeric regions of this species. In addition, genome comparison and evolutionary analysis of this genome and other representative*Brassica* species were conducted. The results provide novel insights into the *Brassica* genome structure evolution.

Materials and methods

Sample collection

Young leaves were collected from a single plant of B. rapa var. parachinensis cv. Youlv 701(Fig. 1), which is a highly inbred line issued by the Guangzhou Institute of Agriculture Science, in Guangzhou, Guangdong, China. The collected young leaves were soon frozen in liquid nitrogen and stored at -80*C for DNA and RNA extraction.

DNA extraction and sequencing

For Illumina sequencing, the phenol/chloroform extraction protocol was used to extract DNA from 2g of young leaves. An Illumina sequencing library for an insertion length of 250 bp was prepared using the TruSeq Nano DNA LT Library Preparation Kit (Illumina Inc., USA). DNA purity and size range were evaluated with Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). An Illumina sequencing library (PE) with an insertion length of 300-350 bp was constructed and sequenced using the Illumina HiSeq 2000 platform.

The DNA extracted from the young leaves was also used for the PacBio sequencing library construction. According to the manufacturer's protocol (Pacific Biosciences, USA), 10 μg of Chinese flowering cabbage genomic DNA were used for 30-kb template library preparation using the BluePippin Size Selection system (Sage Science, USA). The library was sequenced on the PacBio SEQUEL II platform.

The PacBio platform was used to generate long genomic reads for the construction of a reference genome for the Chinese flowering cabbage. After removing adaptor sequences, more than 113Gb of subreads were obtained with 219 times sequence coverage. The sequencing data were used for the following genome assembly operations.

Genome size estimation based on NGS sequencing data

The HTQC package(Xi Yang et al., 2013) was used to filter low-quality bases and reads. Briefly, three steps were performed to clean the NGS data. First, the adapter sequences were removed from the reads; second, the reads with more than 10% N bases were eliminated; and third, reads with more than 50% low-quality bases (<=5) were discarded. Lastly, we obtained 42.3 Gb (˜86X) of cleaned data for the Kmer-based analysis. We also randomly picked 10,000 read pairs and blasted them against the NCBI non redundant nucleotide (nt) database to check for obvious sample contamination.

De novo assembly of the Chinese flowering cabbage genome

The MECAT2 package(C.-L. Xiao et al., 2017) was used for the Chinese flowering cabbage genome assembly. Long reads had a length cutoff of 10 kb. We applied two rounds of polishing using NGS short reads with a Pilon (Walker et al., 2014). TRF (tandem repeats finder)(Benson, 1999) was used to identify the series repeats, and the series with the ratio of more than 60% of the series repeats are removed. The completeness of the assembled genome was evaluated using BUSCO v3.0 analysis(Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015).

Hi-C library preparation and data analysis

In the present study, 8 g of young leaf tissue collected from the same*B. rapa* var. *parachinensis* plant was used for Hi-C library construction. The Hi-C experiment consisted of the following steps: crosslinking, lysis, chromatin digestion, biotin marking, proximity ligations, cross linking reversal, and DNA purification(Xuefen Yang et al., 2019). The purified and enriched DNA was used for the sequencing library construction; the DNA was sequenced using the Illumina HiSeq 2000 platform (Illumina, USA). The overlapping group was hitched to the scaffold level using Juicer(Durand et al., 2016) and 3D-DNA(Dudchenko et al., 2017). MCScanX(Y. Wang et al., 2012) was used to make a collinear comparison between scaffolds and the existing *B. rapa*genome(Cai et al., 2017). The sequence was given a new name after being manually assembled to the chromosome.

We used bwa mem(Vasimuddin, Misra, Li, & Aluru, 2019) to map two paired reads to the chromosome level genome sequence alone with these parameters "-A1 -B4 -E50 -L0". Then HiCExplorer kit(Wolff et al., 2018) was used to build a Hi-C contact map. Parameters for the step hicCorrectMatrix were set to "–filterThreshold -3.5 5" and the rests were kept at default settings.

Single molecule RNA sequencing (Iso-seq) experiment and data analysis

For gene annotation of the genome, transcriptome sequencing was performed with mixed tissues of a young seedling (14 day after imbibition). RNA was extracted with the TRIzol Reagent (Invitrogen, USA). The RNA quality was checked by a spectrophotometer (LabTech, USA) and a 2100 Bioanalyzer (Agilent Technologies, USA). The verified RNA was used for transcriptome sequencing library construction. Briefly, the mRNA was reversely transcribed using a Clontech SMARTer cDNA synthesis kit. A BluePippin Size Selection System (Pacific Biosciences of California, Menlo Park, CA, USA) was used to perform the size selection for the two libraries, sized 0–3 kb and 2–6 kb, respectively, after cDNA amplification and purification. The SMRTbell libraries were constructed according to the manufacturer's protocol, and sequenced on the PacBio SEQUEL II platform (Pacific Biosciences of California, Menlo Park, CA, USA). Last, we used SMRTLink 7.0 (https://www.pacb.com/support/software-downloads/) to produce all the mRNA sequences for genome annotation.

Repetitive element annotation and construction of Circos picture

The extended de-novo TE Annotator (EDTA)(Ou et al., 2019) was used to annotate the DNATE and LTR type sequences of the genome. TRF (tandem repeats finder) (Benson, 1999) was used to identify the centromere sequence with 20000 points as the threshold. Finally, the repeat sequences were annotated with MAKER(Cantarel et al., 2008). MCScanX(Y. Wang et al., 2012) was used to find the collinearity from the comparison results and generate link files. Four tracks were constructed from the outer to the inner of Circos(Krzywinski et al., 2009), showing gene density, LTR density, DNATE density and TE density respectively, and the collinearity within the genome was shown in the inner circle.

Protein coding gene prediction

The Isoseq3 pipeline (https://github.com/pacificbiosciences/isoseq) was used to process the full-length transcriptome data of Chinese flowering cabbage to obtain the transcriptome sequence. At the same time, in order to obtain a more complete gene annotation, we integrated the annotation content of *B. juncea(J. Yang et al., 2016)* , *B. napus(Chalhoub et al., 2014)* , *B. oleracea(Liu et al., 2014)* , *B. rapa(Zhang et al., 2018)* and *B. nigra(W. Wang et al., 2019)* as the reference gene sequence using CD-HIT-EST (https://github.com/weizhongli/cdhit) to remove the sequence redundancy. The results of repeats sequence

found by EDTA(Ou et al., 2019) and TRF(Benson, 1999) were used as reference repeats to enter into MAKER(Cantarel et al., 2008) for 5 rounds of gene and repeat sequence annotation.

Phylogenetic analysis

The phylogenetic relationships between Chinese flowering cabbage and other *Brassica* plants were analyzed using the orthologs from single-copy genes. The Orthofinder package was used to find orthogroups and single-copy genes. All of the single-copy genes in one species were concatenated into a super alignment, then run through multiple sequence alignment using the mafft program(Katoh & Toh, 2010). Easyspecietree (https://github.com/Davey1220/EasySpeciesTree) was used to generate the phylogenetic relationship between the species using the maximum likelihood method.

Structural variants analysis

Structural variations were detected using an assembly-based pipeline based on LASTZ/CHAIN/NET/NETSYNTENY tools(Harris, 2007; Kent, Baertsch, Hinrichs, Miller, & Haussler, 2003; Liao, Zhang, Chakraborty, & Emerson, 2020; Schwartz et al., 2003) which is publicly available at https://github.com/yiliao1022/LASTZ_SV_pipeline. Insertion times of LTR-retrotransposons were estimated by the divergence time (T) between the two LTRs of each intact element with the formula: T= K/2r, where Ks refers to the sequence difference between the 5'-LTR and 3'-LTR of an intact LTR element and r refers to the average mutation rate. Here we used the neutral substitution rate of 1.5X 10-8 per synonymous site per generation(Koch, Haubold, & Mitchell-Olds, 2000).

Figure 1. Overview of the assembly pipeline for Brassica rapa var. parachinensis genome. The steps include assembly of PacBio reads followed by scaffolding using Hi-C, and extensive QC using high coverage of Illumina short reads followed by de novo repeat annotation and gene annotation using ISO-seq sequencing.

Results

A highly continuous genome assembly of Chinese flowering cabbage (*B. rapa* var.*parachinensis*)

A highly inbred line of Chinese flowering cabbage (*B. rapa* var.*parachinensis* , Fig.1) was used for the genome sequencing and assembly with deep coverage long reads and Hi-C data. The assembly pipeline for *Brassica rapa* var. *parachinensis* genome was shown in Fig.1. DNA samples from a single plant were prepared for PacBio, Illumina and Hi-C sequencing to avoid potential genome variability between different plants. Overall, we obtained a total of 113Gb PacBio and 47.5Gb Illumina raw reads (Table S1), corresponding to 219 and 86 depth of the estimated genome size (515 Mb), respectively. A preliminary survey of the genome size, heterozygosity, GC and transposon elements (TEs) content of this inbred line was carried out with 32GB clean illumina reads (Table 1; ˜83 coverage) using Kmer-based method (Liu et al. 2013). The genome size was estimated to be about 515Mb with an overall GC content of 38.9% and transposon elements (TE) content of 64.1% (Table S1). Remarkably, the heterozygosity is very low with only 0.16% that would facilitate assembly.

We applied an integrated strategy to assemble the genome. Firstly, the MECAT2 package(C.-L. Xiao et al., 2017) was used for the Chinese flowering cabbage genome assembly. Secondly, long reads with a length cutoff of 10 kb were polished using NGS short reads with a Pilon(Walker et al., 2014). Finally, we obtained the final contig assembly of 384Mb with a contig N50 length of 7.2Mb. The genome contained 450 contigs, and the longest contig was 19.9Mb (Table 1). The GC content for the genomic contigs were 37.6% (Table 1). The results of coverage statistics by SAM tools suggested that the assembly of this genome is credible (Table S2). Furthermore, we found that 97.8% and 0.8% of the completed and partial genes of the total of 1,440 BUSCO genes were detected in the genome, respectively, which validated the completeness of the genome (Table S3).

Furthermore, high-throughput chromatin conformation capture (Hi-C) data was used to scaffold the contigs into chromosome-level assembly. We obtained a total of 66 Gb cleaned Hi-C paired-end (PE) reads which is about 128 depth of the genome. Of which, 98.27% (434M/442M) were mappable to the current assembly and ˜33.18% (147M/442M) were mapped to different contigs. Using contact frequency calculated from the PE reads, 180 contigs were further scaffolded into 10 pseudo-chromosomes (Fig. 1A). These 180 contigs represent 87.93% (338 Mb/384Mb) of the total assembled sequence and 40% (180/450) of the total contigs. The final assembly contains 69 scaffolds with a scaffold N50 of 32Mb and the longest scaffold is 47.5Mb in length (Table 1). The Circos map of the genome shows that each position is collinear with the other two, indicating that the annotation is complete (Fig.1B). A large number of corrected repeat regions on A05 and A06 chromosomes were identified (Fig.1C), which indicated that there might be a large region of DNA transposons and LTR transposons at this region.

We also performed *de novo* gene prediction with guidance by homologs from related species, transcriptome from short read data and full-length transcripts from ISO-seq sequencing from the present study using the MAKER pipeline(Cantarel et al., 2008). We annotated 47,598 protein-coding genes in the Chinese flowering cabbage genome with an average gene length of 2060 bp (Table 1). The average number of exons per gene is 6.13, with a mean length of 199 bp (Table 1). Approximately 53.2% of the genome is annotated as repetitive sequences, which is consistent with the estimation of Kmer-based method. LTR retrotransposons (22.26 %) and DNA transposons (17.62 %) are the most abundant families (Table S4).

In conclusion, we provide, to our knowledge, so far the most contiguous and the first chromosome-level genome assembly of this species.

Table 1. Statistics and annotated analysis of the Chinese flowering cabbage genome assembly

	Number	Size	Sequence coverage(X)	Percentage(%)
Estimate of genome size		515 Mb		
PacBio reads	4,448,280	113,068 Mb	219.31	
illumina reads	322,016,292	42,330 Mb	82.10	
HiC reads	441,545,786	66,231 Mb	128.46	
Total reads		221,630 Mb	429.89	
Contigs	450	384 Mb		74.50
N50 of contigs		7.2 Mb		
Longest contig		19.9 M		
scaffold	69	384 Mb		74.62
N50 of scaffold		32.2 Mb		
Longest scaffold		47.5 Mb		
GC content		144.4 Mb		37.61
Total repetitive sequences		170.3 Mb		44.26
Total protein-coding genes	47598	47.3 Mb		12.31
Average length per gene		2,060 bp		
Average exons per gene	6.13	199 bp		

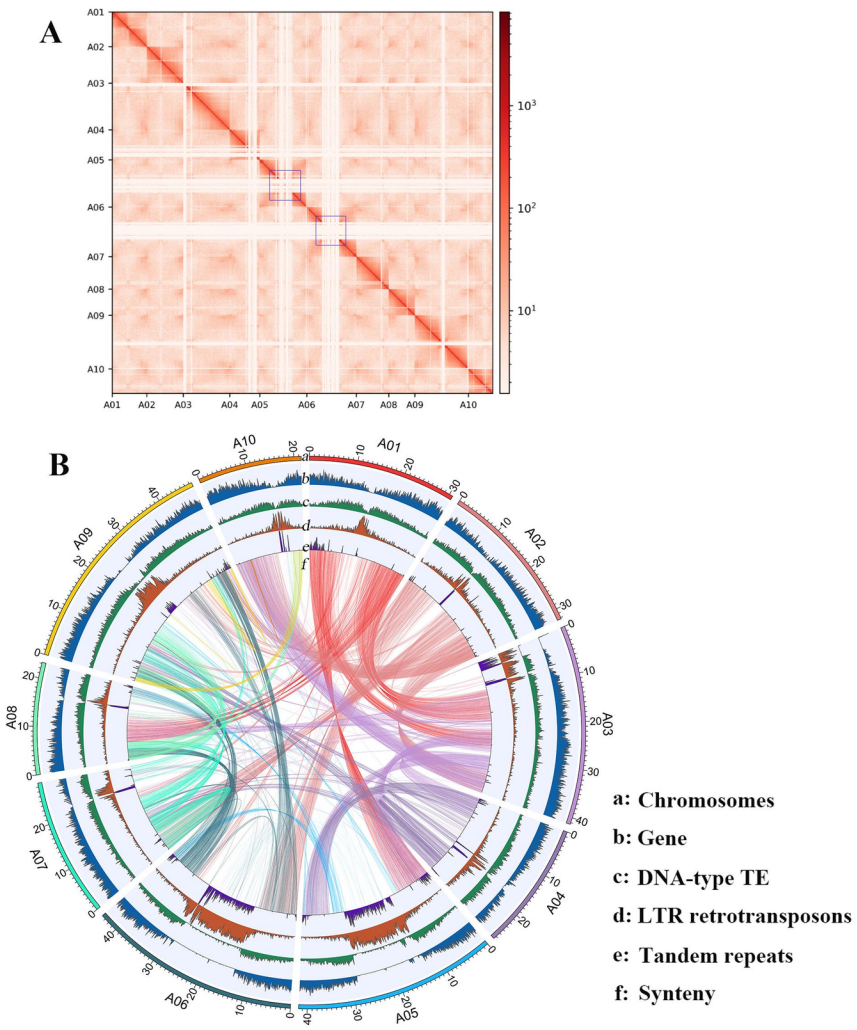

Figure 2. A highly continuous genome assembly of Chinese flowering cabbage (*B. rapa* var. *parachinensis*). (A) Hi-C contact map of the Chinese flowering cabbage assembled chromosomes; Density of Hi-C contacts are highest at the diagonals, suggesting consistency between assembly and the Hi-C map; Blue squares indicate highly repetitive pericentromeric regions on A05 and A06 chromosomes. (B) Circos diagram of sequence features on the chromosomes of *B. rapa* var. *parachinensis* ; A01, 02, 03, 04, 05, 06, 07, 08, 09 and 10 indicate the ten assembled chromosomes of *B. rapa* var.*parachinensis* .

Gene duplication analysis across 20 eudicot genomes reveals the current *B. rapa* var. *parachinensis* genome is among the most high-quality assemblies of *Brassica* genomes

To assess the completeness of genome assembly and gene models, we used Orthofinder(Emms & Kelly, 2015) to construct the ortholog group across 20 eudicot species and separate them into three categories: ortholog group with a single copy gene, two genes and multiple (more than two) genes. The frequency of each group among the 20 eudicot species revealed that the *Brassica* species (i.e. *B. napus* , *B. rapa* ,*B. juncea* and *B. nigra*) harbor more duplicated orthologs than *Arabidopsis* species (Fig. 3A,B), which is consistent with the fact that *Brassica* species experienced an extra whole genome triplication (WGT) event compared with the model plant *Arabidopsis thaliana* (Liu et al., 2014). Additionally, more duplicated orthologs are identified in the current *B. rapa* var. *parachinensis* genome assembly than in the two other assemblies of this species with

a relative lower N50 (Fig.3A), suggesting that we obtained a higher quality of genome assembly and gene annotation than previous studies(Belser et al., 2018; Zhang et al., 2018). BUSCO analysis suggested that all the 12*Brassica* species have a high quality of genome assembly and the current *B. rapa* var. *parachinensis* has the highest BUSCO value (Fig. 3B).

Next, we compared the overlap of gene models among *B. rapa* var.*parachinensis* and two other *B. rapa*genomes(Belser et al., 2018; Zhang et al., 2018). A total of 19,042 genes are shared by all three genomes. The Chinese flowering cabbage genome (Fig.3C) has more specific gene models, which may be caused by the difference of assembly quality among these three genomes or specific gene amplification history.

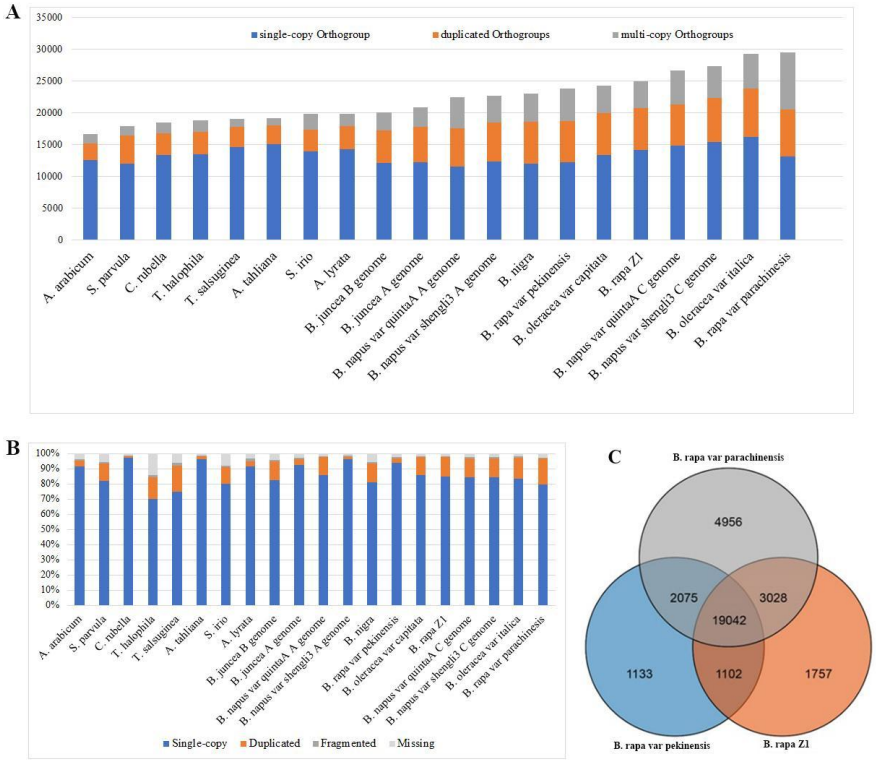

Figure 3. Distribution of genes in *B. rapa* var.*parachinensis* and other representative plant species. (A) Distribution of ortholog groups: single copy (blue), two copies (orange), and multiple copies (grey) across 20 eudicot genomes; (B) BUSCO analysis of genome assembly of the 20 eudicot genomes; (C) Venn diagram showing the overlap of gene families among Chinese flowering cabbage and two other assemblies of *B.rapa* species.

Phylogenetic analysis of a collection of *Brassica* genomes reveals Chinese flowering cabbage has a closer evolutionary relationship with the diploid progenitor of the allotetraploid species, *B. juncea*

The Brassicaceae family serves as a useful model for studying polyploidy and chromosome evolution. The evolutionary relationship of six ecologically important *Brassica* species including three diploid species (*B. rapa* , *B. oleracea* , and *B. nigra*) and three allotetraploid species (*B. napus* , *B. juncea* , and*B. carinata*) was well described in a classical U triangle model(Cheng et al., 2016) . To elucidate the evolutionary distance of the current Chinese flowering cabbage genome to other *Brassica* genomes, we constructed a phylogenetic tree (Fig. 4) for 12 collected *Brassica* genomes and eight related Brassicaceae species using the coding sequences of 434 single-copy genes that are present in all of the species. The result shows that the three *Brassica*

genome types are clearly separated from each other among the investigated species. The current Chinese flowering cabbage has a AA genome type which is closer to the AA genome of the allotetraploid species, *B. juncea* , than the AA genome of another *B. rapa* line, *B. rapa* var *pekinensis* in the phylogenetic tree, suggesting Chinese flowering cabbage is evolutionarily closer to the diploid progenitor of the allotetraploid species, *Brassica juncea* . Also, in the CC genome clade, *B. oleracea* var *capitata* was clustered firstly with two *B. napus* CC genomes and then with *B. oleracea* var *italica* , implying *B. oleracea* var *capitata* has a CC genome that is closer to the donor of CC genome of the *B. napus* . Similarly,*B. rapa* Z1 was clustered firstly with *B. napus* AA genome and then other AA genomes, pointing to it as being evolutionarily closer to the AA genome progenitor of *B. napus* .

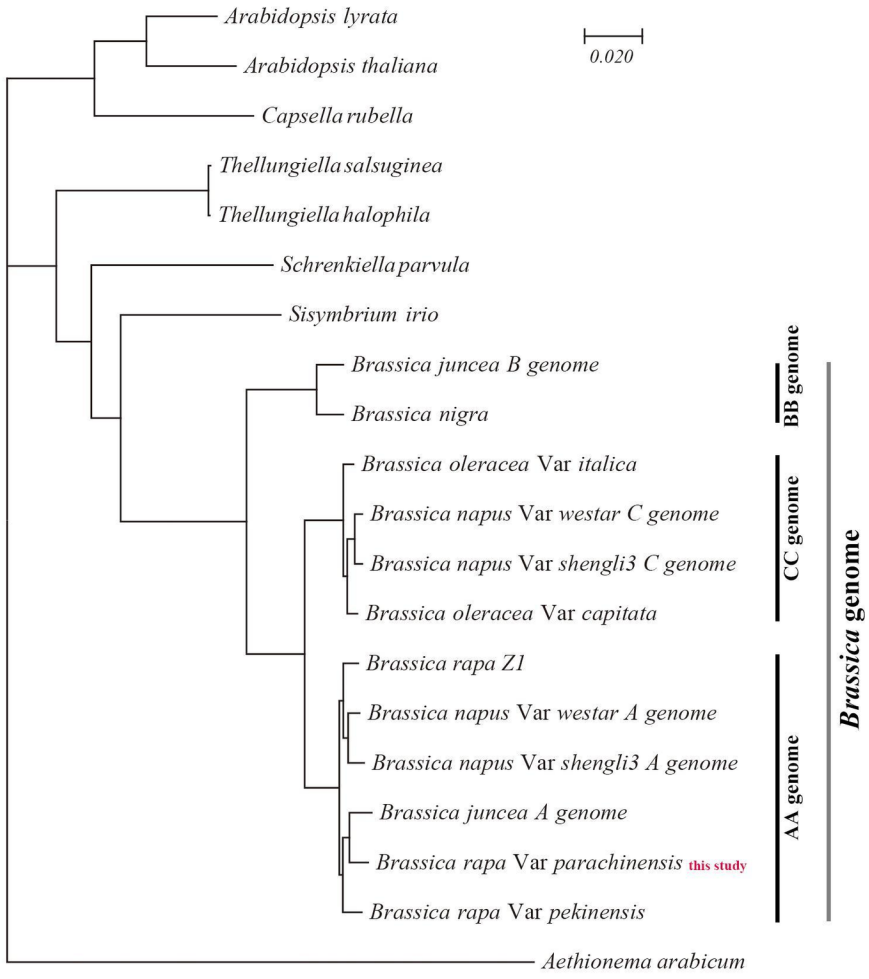

Figure 4 The phylogenetic relationship of *B. rapa* var.*parachinensis* with other Brassicaceae plants.

Extensive chromosomal arrangements between *Brassica*species

Genome-wide synteny analysis was conducted using syntenic orthologous genes both within and between species for *Brassica rapa* . Firstly, the genome of Chinese flowering cabbage was compared to two published genome assemblies of different strains of this species, *B. rapa*Z1(Belser et al., 2018) and*B. rapa* var. *pekinensis* (Zhang et al., 2018). The SyMAP map reveals that these three *Brassica rapa* assemblies retain well conserved overall genome architecture except a translocation event between chromosome 1 and chromosome 3 that

differentiates our assembly to the other two assemblies (Fig. 5A). Next, we performed the comparison between *B. rapa* var. *parachinensis* and two highly continuous assemblies of the *B. oleracea* genome (Belser et al., 2018; Liu et al., 2014). Besides the different chromosome numbers (i.e.*B. rapa* var. *parachinensis* ; AA genome, n=10 and *B. oleracea* ; CC genome, n=9), we observed extensive chromosomal rearrangements between these two species (Figure 5B). Only 2 chromosomes (Chr1 and Chr2) showed minimal changes since their divergence from a common ancestor. The extensive chromosomal rearrangements that occurred during the course of *Brassica* genome evolution is different from the observation in *Oryza* , one of the well-studied genus models in monocot, in which the karyotype of most diploid species is well-conserved, even over 15 million years evolutionary history(Stein et al., 2018).

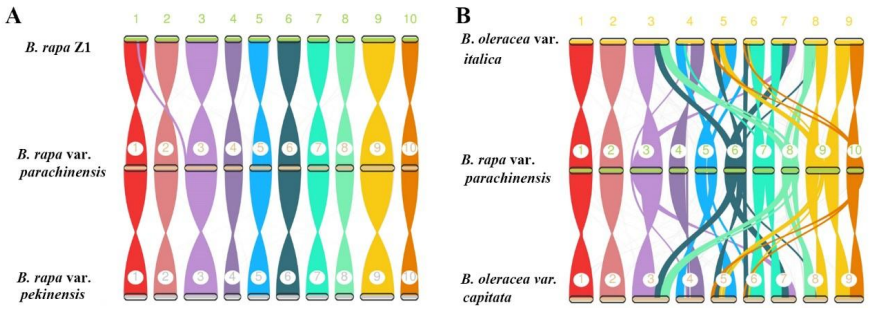

Figure 5. Genome synteny based on orthologous genes within and between species for *B. rapa* var. *parachinensis*. (A) Genome synteny between *B. rapa* var. *parachinensis* and two other *B. rapa* genome assemblies (*B. rapa*Z1(Belser et al., 2018) and*B. rapa* var. *pekinensis* (Zhang et al., 2018)); (B) Genome synteny between *B. rapa* var. *parachinensis* and two highly continuous assemblies of the *B. oleracea* genome (*B. oleracea* var. *capitata* (Belser et al., 2018; Liu et al., 2014) and *B. oleracea* var. *italica* (Belser et al., 2018; Liu et al., 2014)). Homologous chromosomes are labelled with the same number.

Genome structure evolution in *Brassica*: insight from pericentromeric regions

The pericentromeric regions of plant genomes are among the most rapidly evolving genomic parts, which are found to be largely driven by some major mechanisms such as LTR-retrotransposons proliferation, gene conversions, and segmental duplications(Liao et al., 2018). Comparison of the pericentromeric regions among three assemblies of the *B. rapa* with different assembly quality (Supplementary Fig. 1E) revealed that the current assembly resolved a larger part of pericentromeric repetitive regions than other two assemblies (Supplementary Fig. 1A,B,C,D). A large part of the pericentromeric regions was missed in the other two assemblies, especially the*B. rapa* var. *pekinensis* assembly. This result shows that high contiguous genome assemblies are required for comparative genomic analysis of highly repetitive regions.

Thus, for interspecies comparison, we selected highly contiguous assemblies for two closely related *Brassica* species, *B. nigra* and *B. oleracea* , which represent two other *Brassica*genome types (BB and CC), and compared the genome structure and sequence features at the pericentromeric regions of all chromosomes among these three *Brassica* species or genome types. We found that the pericentromeric regions of chromosome 5 and 6 in *B. rapa*experienced a lineage-specific LTR-retrotransposon amplification history. For example, comparison of chromosome 5 between *B. rapa*and *B. nigra* (Fig. 6A) showed that *B. rapa* has a clear enrichment of LTR retrotransposon compared to the orthologous pericentromeric regions of *B. nigra* although the syntenic relationship of the whole chromosome is well retained between these two species. This difference is more likely to be caused by lineage specific LTR retrotransposon amplification history since their divergence. While comparison between *B. rapa* and *B. oleracea* (Fig. 6B) showed that the synteny of chromosome 5 breaks at the centromere region (see also Fig. 5B) and the break event is more likely to occur in the*B. oleracea* lineage since the *B. rapa* share the synteny block with *B. nigra* (Fig. 6A), while the *B. oleracea* does not (Fig, 6C). Thus, chromosome rearrangements may be an alternative cause for the different

genome structure features observed in the pericentromeric regions. Similarly, the comparison of chromosome 6 revealed an analogous pattern (Fig. 6D,E,F).

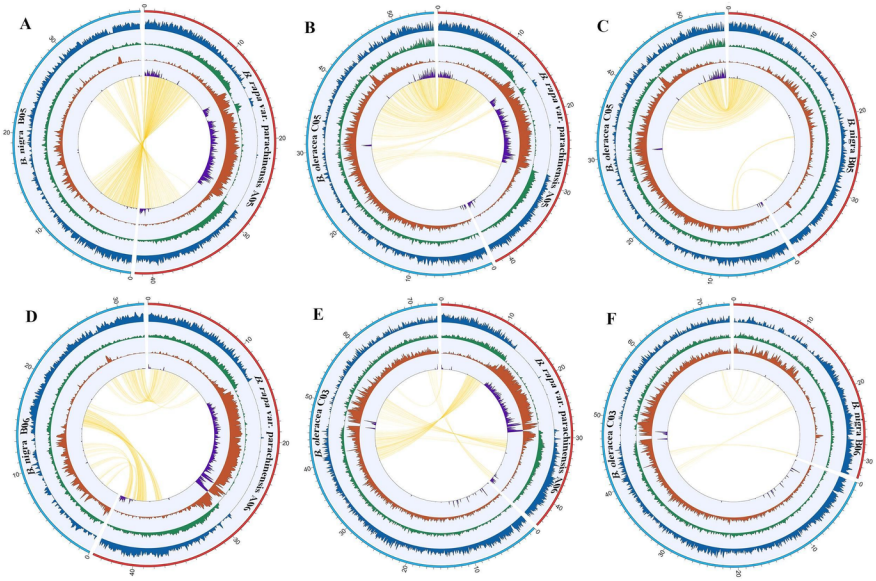

Figure 6. Comparative analysis of sequence features and systeny at the pericentromeric regions on chromosome 5 and 6 among three*Brassica* genome types: Chinese flowering cabbage (AA genome),*B. nigra* (BB genome) and *B. oleracea* (CC genome) . (A) Synteny map of Chr05 between *B. nigra* (BB genome) and *B. rapa* var. *parachinensis* (AA genome); (B) Synteny map of Chr05 between *B. oleracea* (CC genome) and *B. rapa var. parachinensis* (AA genome); (C) Synteny map of Chr05 between *B. oleracea* (CC genome) and *B. nigra* (BB genome). (D) Synteny map of Chr06 between *B. nigra* (BB genome) and *B. rapa* var.*parachinensis* (AA genome); (E) Synteny map of Chr03 of *B. oleracea* (CC genome) and Chr06 of *B. rapa* var.*parachinensis* (AA genome); (F) Synteny map of Chr03 of *B. oleracea* (CC genome) and Chr06 of *B. nigra* (BB genome). Tracks in the circos plot from outer to inner represent: a: Chromosomes; b: Gene; c: DNA-type TE; d: LTR retrotransposons; e: Tandem repeats; f: Synteny.

Structural variants in *Brassica* genomes

Structural variation (SV) is generally defined as genomic alterations that are 50bp or larger in size, typically including insertions (INSs), deletions (DELs), duplications (DUPs), inversions (INVs) and translocations (TRAs). SVs greatly impact the genes encoded in the genome and are responsible for diverse agronomically important phenotypes/traits. Compared to single nucleotide polymorphism (SNP) and short insertions and deletions (InDels), SVs are less commonly explored due to the difficulty in fully identifying them with short reads.*De novo* genome assemblies, especially with high contiguity, can facilitate in-depth genome-wide identification of all forms of structural variations. To the best of our knowledge, no work so far has been conducted to identify SVs based on high-contiguous genome assemblies in *Brassica* genomes. To close this knowledge gap and have a first glimpse of SVs differing within *Brassica rapa*genomes, we identified SVs using the genomes of *B. rapa*Z1(Belser et al., 2018) and*B. rapa* var. *parachinensis* (this study), each with genome assembly contig N50, 5.51 Mb and 7.26 Mb, respectively. As shown in Fig. 5A, these two genomes are different only in a single translocation and do not exist in large chromosomal rearrangements. Using the whole genome alignment approach, we identified a total of 27,190 insertions, 26,002 deletions, 1,374 duplications in *parachinensis* assembly, 1,368 duplications in Z1 *assembly,* and 46 medium-sized inversions with sizes ranging from 5.2Kb to 1,431.6 Kb, and 8,565 complex SVs with imprecise breakpoints between Z1 and *parachinensis* (Fig. 7A). Of the insertion events, 845 and 847 are found to be newly occurred LTR insertions specifically in *parachinensis* and Z1 assembly, respectively, which are consistent with their

relatively recent estimated insertion times (Fig. 7B). A large proportion of insertions and deletions detected was found to overlap with the gene regions based on the gene annotation. In Fig. 7C, two cases of local tandem duplication are shown to overlap with gene fragments or full genes. Additionally, comparative genomic analysis can also provide insights into the mutational mechanisms of structural variations. Of the 46 inversions identified, we found that repeat sequences, especially inverted repeat sequence features prevail at the flanking regions, highlighting the causal role of sequence features on small-size inversion formation (Fig. 7D). Taken together, our analysis of genomic structural variations based on these highly contiguous genome assemblies provide the first glimpse of SVs in the *Brassic* a genomes and their functional significance on gene structure and thus the potential effect on phenotype.

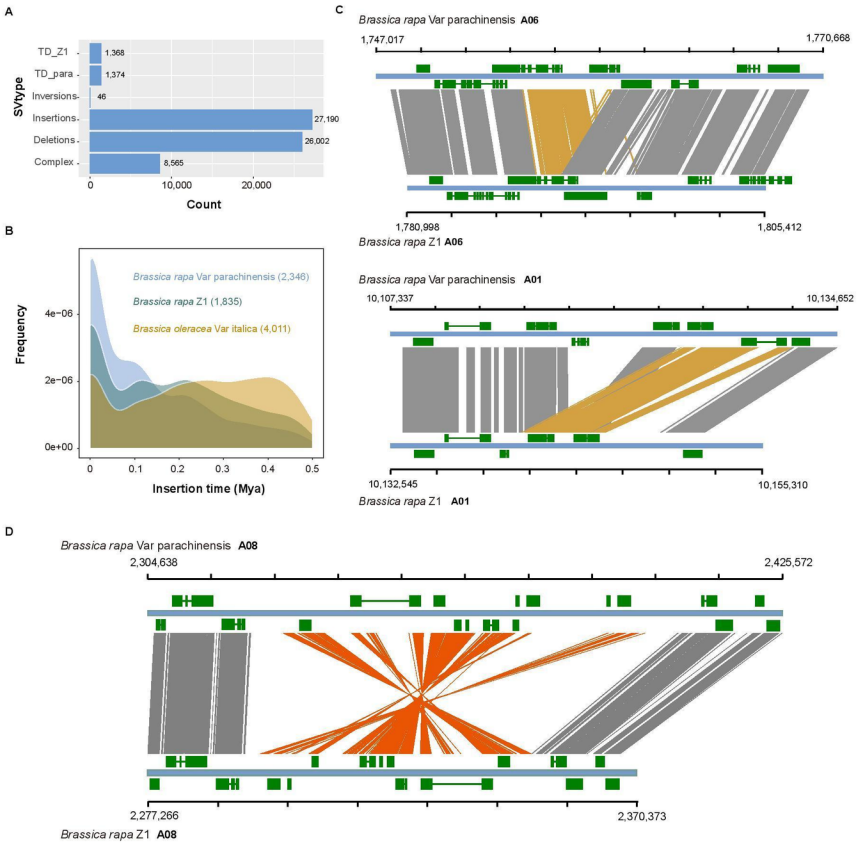

Figure 7. Structural variations between two *B. rapa*lines. (A) Total number of structural variations identified using highly continuous assemblies between *Brassica rapa* Z1 and*Brassica rapa* var. *parachinensis* . TD_Z1, tandem duplications in Z1 assembly relative to the *parachinensis*assembly and TD_pare vice versa. Complex SVs indicate their breakpoints are imprecise. (B) Distribution of insertions times of LTR-retrotransposons in three highly continuous *Brassica* genome assemblies. (C) Examples of tandem duplication impacting genes. (D) Example of medium size genomic inversions between *Brassica rapa*Z1 and *Brassica rapa* var. *parachinensis,* which prevails in*Brassica* genome evolution.

Discussion

Chinese flowering cabbage (*B. rapa* var. *parachinensis*) is an important leafy and bolting stem vegetable with high nutritional value which has been widely grown in Asia(Tan et al., 2019). Among the abundant ecological types of *Brassica rapa* that are planted as vegetables in China, Chinese flowering cabbage is the one that is well-adapted to the high temperature and high humidity climate in the south of China. It can

be planted all year round for tender flower products without the need for a strict vernalization process. In this study, we report the first chromosome-level genome assembly of this important ecological *B. rapa* strain, Chinese flowering cabbage, which provides a valuable genomic data resource for evolutionary studies for *B. rapa* and related *Brassica* species. This present study is the first to report on the genome size, heterozygosity, and repeat content of the Chinese flowering cabbage genome.

Highly continuous genome assembly is critical for genome-wide marker development and gene model prediction. Enormous studies have demonstrated that recent long-read sequencing technologies can greatly improve the continuity of genome assembly(Song et al., 2020; Wang et al., 2019; Belser et al., 2018; Zhang et al., 2018). In this study, we used PacBio long reads to assemble the *B. rapa* var. *parachinensis* genome. Because of the low heterozygous ratio (0.16%) of the plants used in this genome sequencing, we obtained the contig N50 length of 7.26 Mb, which is longer than the two *B. rapa* genomes sequenced recently by PacBio and Nanopore technology(Belser et al., 2018; Zhang et al., 2018), and much longer than the genomes of*B. rapa* and *B. oleracea* sequenced using Illumina technology(Liu et al., 2014; Wang et al., 2019). We applied the Hi-C technique to scaffold more than 545 Mb contigs onto 10 chromosomes. The scaffold N50 length of the final assembly reached 32.3 Mb, with the maximum size of 47.4Mb, which was similar to the *B. rapa* Z1 genome sequenced with Nanopore technology(Belser et al., 2018) (Table S5). The completeness of the genome (97.8%) was validated using the BUSCO analysis in the present study, and surpassed most of the genome of related *Brassica* species sequenced thus far, including *B. oleracea*HDEM(Belser et al., 2018),*B. oleracea* var.*botrytis* (Sun et al., 2019) and *B. rapa*Z1(Belser et al., 2018) (Table S5).

In the present study, the assembly of the Chinese flowering cabbage genome resolved most of the pericentromeric regions of the *B. rapa* . Among them, the pericentromeric regions of chromosome 5 (A05) and 6(A06) were found to be significantly expanded in comparison to other pericentromeric regions and very few genes were annotated in this region (Fig. 2B; Fig. 6). This observation can further be verified by the Hi-C contact map in which the pericentromeric regions of chromosome 5 and 6 have a clear sparse Hi-C contact signal that is mostly caused by repetitive sequences (Fig. 3). Strikingly, this expansion seems to be lineage specific since we do not observe a similar pattern in the two other *Brassica* genome types, i.e. chromosome C05 and C06 in*B. oleracea* and *B. napus* (Belser et al., 2018; Song et al., 2020), and chromosome B05 and B06 in *B. nigra* (Fig. 6A). This lineage specific expansion may play a role in the evolutionary divergence of *Brassica* AA, BB and CC genomes. It is worth noting that such large repetitive regions can only be resolved by long-read sequencing technology. For example, in the previous studies, *B. rapa* Z1 and *B. napus* AA genome assemblies present a similar but relatively weaker pattern than the current assembly(Belser et al., 2018; Song et al., 2020; Zhang et al., 2018) (Fig. S1). However, in the assembly of *B. rapa* (Belser et al., 2018; Song et al., 2020; Zhang et al., 2018) (Figure S1E), sequenced by PacBio Sequel with a N50 of 1.45Mb, does not present the large repetitive regions in its assembly (Supplementary Fig.1E).

The genus *Brassica* contains three basic genomes, *B. rapa*(AA genome), *B. nigra* (BB genome), and *B. oleracea* (CC genome), which further hybridize to give rise to three allopolyploid species, *B. napus* (AACC genome), *B. juncea* (AABB genome), and *B. carinata* (BBCC genome)(Cheng et al., 2016; Sun et al., 2019). In the present study, a phylogenetic tree was constructed to analyze the evolution of the *Brassica* species. Interestingly, the Chinese flowering cabbage shows the closest relationship with the *B. juncea* AA genome but not with two*B. rapa* genomes (Chinese cabbage and yellow sarson)(Fig. 4)(Belser et al., 2018; Zhang et al., 2018). The *B. rapa* species can be further subdivided into six populations: turnips (Chinese and European turnips), sarsons (sarson, rapid cycling and spring/winter oilseed), turnip rapes, taicai and mixed Japanese morphotypes, pak choi (pak choi, wutacai, Chinese flowering cabbage and zicaitai varieties) and heading Chinese cabbages(Cheng et al., 2016). Our results suggested that the donor of the AA genome in*B. juncea* is most likely from the pak choi group (Chinese flowering cabbage) in contrast to other *B. rapa* varieties, such as sarsons and turnips(Belser et al., 2018; Cai et al., 2017). Meanwhile, we found that *B. rapa* Z1 (sarson) was clustered firstly with *B. napus* AA genome and then other AA genomes, implying that it should be the most evolutionary closest donor of the AA genome in *B. napus* . Similarly, the*B. oleracea* can also be subdivided into seven populations such as kohlrabies, Chinese kale, cauliflower, broccoli, Brussels sprouts, kale and cabbages(Cheng et al., 2016). Interestingly, *B. oleracea* var. *capitata*(cabbages) was

clustered firstly with two *B. napus* CC genomes and then with *B. oleracea* var. *italica* (broccoli), implying the donor of CC genome in *B. napus* was probably evolved from*B. oleracea* var. *capitata* (cabbages) (Fig. 4). Thus, we demonstrated that high continuity genome assemblies can aid in the interpretation of evolutionary relationship among *Brassica*species.

Numerous cases of studies found that structural variations can impact larger genomic regions than SNPs. Structural variant (SV) discovery would not only help our understanding of the landscape of genomic variation within and between species but also reveal the functional significance of SVs(Fuentes et al., 2019). In comparison to SVs detection methods that are based on Illumina short reads, the whole assembly-based method can fully recover the SVs in theory but still depend on assembly quality. SVs studies in human(Audano et al., 2019; Huang et al., 2010), and in a wide range of plant species, such as rice(Fuentes et al., 2019), Maize(Mahmoud et al., 2020), tomato(Voichek & Weigel, 2020), and*Arabidopsis* (Voichek & Weigel, 2020) indicate that SVs can affect a large proportion of coding genes. In current study, we detect SVs between the genome assemblies of two *Brassica rapa* lines and identified a total of 27,190 insertions, 26,002 deletions, 1,368 duplications and 46 medium-sized inversions with size from 5.2Kb to 1,431.6 Kb, and 8,565 complex SVs with imprecise breakpoints between them (Fig. 7). This is the first report of SVs that detect between *Brassica* genomes using high contiguity genome assemblies. These SVs may affect coding genes that may further contribute to phenotypic variations, such as morphological and phytochemical characteristics.

In summary, we report a chromosome-level genome assembly of Chinese flowering cabbage and its accurate gene and TE annotation. The phylogenetic analysis indicates this genome has a closer evolutionary relationship with the AA diploid progenitor of *B. juncea* . We also found the lineage specific pericentromeric expansion events on the chromosome 5 and 6 of the *Brassica* AA genome compared to the orthologous genomic regions in the *Brassica* BB and CC genomes. Finally, we report a large amount of structural variations (SVs) between two *B. rapa* lines (Z1 and *parachinensis*) using high continuity genome assemblies. Overall, our high-quality genome assembly of Chinese flowering cabbage provides a valuable genetic resource for deciphering the genome evolution of *Brassica* species and it would serve as the reference genome guiding the molecular breeding practice of*B. rapa* crops.

Acknowledgements

This work was funded by the Science and Technology Program of Guangzhou (202002020007), the Guangdong Basic and Applied Basic Research Foundation (2020A1515011396), the Key-Area Research and Development Program of Guangdong Province (2018B020202010), and Science and Technology Program of Guangzhou (201804010320).

Author contributions

C.-M.C. H.Z. and Y.L.designed the project and wrote the draft manuscript. G.-G.L., Y.L., J.-T.W., and D.J. contributed to the genome assembly, genome evolution analysis, and structural variants analysis. Y.-S.Z., X.-C.D., H.-L.R., J.-J.L., G.-J.C., and B.-H.C., participated in data analysis and substantively revised the manuscript. The final manuscript has been read and approved by all authors.

Data Accessibility

The raw genome and RNA sequencing data were deposited in the China National GeneBank DataBase(CNGBdb) under Bioproject number CNP0001121. The final chromosome assembly was submitted to CNGBdb under the same Bioproject.

References

Audano, P. A., Sulovari, A., Graves-Lindsay, T. A., Cantsilieris, S., Sorensen, M., Welch, A. E., ... Eichler, E. E. (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell*, *176*(3), 663–675.e19.

Bayer, P. E., Hurgobin, B., Golicz, A. A., Chan, C.-K. K., Yuan, Y., Lee, H., ... Edwards, D. (2017). Assembly and comparison of two closely related Brassica napus genomes. *Plant Biotechnology Journal*,

15(12), 1602–1610.

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., ... Aury, J.-M. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants*,*4*(11), 879–887.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.

Cai, C., Wang, X., Liu, B., Wu, J., Liang, J., Cui, Y., ... Wang, X. (2017). Brassica rapa Genome 2.0: A Reference Upgrade through Sequence Re-assembly and Gene Re-annotation. *Molecular Plant*, *10*(4), 649–651.

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., ... Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., ... Wincker, P. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science*,*345*(6199), 950–953.

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., ... Wang, X. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in Brassica rapa and Brassica oleracea.*Nature Genetics*, *48*(10), 1218–1224.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., ... Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*, *356*(6333), 92–95.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98.

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.*Genome Biology*, *16*, 157.

Fuentes, R. R., Chebotarov, D., Duitama, J., Smith, S., De la Hoz, J. F., Mohiyuddin, M., ... Alexandrov, N. (2019). Structural variants in 3000 rice genomes.*Genome Research*, *29*(5), 870–880.

Harris, R. S. (2007).*Improved pairwise Alignmnet of genomic DNA*. Retrieved from https://etda.libraries.psu.edu/catalog/7971

Huang, C. R. L., Schneider, A. M., Lu, Y., Niranjan, T., Shen, P., Robinson, M. A., ... Burns, K. H. (2010). Mobile interspersed repeats are major structural variants in the human genome. *Cell*, *141*(7), 1171–1182.

Kamran, M., Xie, K., Sun, J., Wang, D., Shi, C., & Lu, Y. (2020). Modulation of growth performance and coordinated induction of ascorbate-glutathione and methylglyoxal detoxification systems by salicylic acid mitigates salt toxicity .... *Ecotoxicology* . Retrieved from https://www.sciencedirect.com/science/article/pii/S0147651319312084

Katoh, K., & Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program.*Bioinformatics* , *26*(15), 1899–1900.

Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W., & Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes.*Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11484–11489.

Koch, M. A., Haubold, B., & Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). *Molecular*

Biology and Evolution,*17*(10), 1483–1498.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics.*Genome Research*, *19*(9), 1639–1645.

Liao, Y., Zhang, X., Chakraborty, M., & Emerson, J. J. (2020). Topologically associating domains and their role in the evolution of genome structure and function in Drosophila (p. 2020.05.13.094516). doi: 10.1101/2020.05.13.094516

Liao, Y., Zhang, X., Li, B., Liu, T., Chen, J., Bai, Z., ... Chen, M. (2018). Comparison of Oryza sativa and Oryza brachyantha Genomes Reveals Selection-Driven Gene Escape from the Centromeric Regions. *The Plant Cell*,*30*(8), 1729–1744.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., ... Paterson, A. H. (2014). The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications*, *5*, 3930.

Lysak, M. A., Koch, M. A., Pecinka, A., & Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. *Genome Research*, *15*(4), 516–525.

Mahmoud, M., Gracz-Bernaciak, J., Żywicki, M., Karłowski, W., Twardowski, T., & Tyczewska, A. (2020). Identification of Structural Variants in Two Novel Genomes of Maize Inbred Lines Possibly Related to Glyphosate Tolerance. *Plants*,*9*(4). doi: 10.3390/plants9040523

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., ... Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, *20*(1), 275.

Parkin, I. A. P., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., ... Sharpe, A. G. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biology*, *15*(6), R77.

Schwartz, S., Kent, W. J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R. C., ... Miller, W. (2003). Human-mouse alignments with BLASTZ. *Genome Research*,*13*(1), 103–107.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* , *31*(19), 3210–3212.

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., ... Guo, L. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nature Plants*, *6*(1), 34–45.

Stein, J. C., Yu, Y., Copetti, D., Zwickl, D. J., Zhang, L., Zhang, C., ... Wing, R. A. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza.*Nature Genetics*, *50*(2), 285–296.

Sun, D., Wang, C., Zhang, X., Zhang, W., Jiang, H., Yao, X., ... Shan, X. (2019). Draft genome sequence of cauliflower (Brassica oleracea L. var. botrytis) provides new insights into the C genome in Brassica species. *Horticulture Research*, *6*, 82.

Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., & Tong, C. (2017). The high-quality genome of Brassica napus cultivar "ZS11" reveals the introgression history in semi-winter morphotype. *The Plant*. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.13669

Tan, X. L., Fan, Z., Kuang, J., Lu, W., & Reiter, R. J. (2019). Melatonin delays leaf senescence of Chinese flowering cabbage by suppressing ABFs-mediated abscisic acid biosynthesis and chlorophyll degradation. *Journal of Pineal Research*. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/jpi.12570

Vasimuddin, M., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *2019 IEEE International Parallel and Distributed Processing Symposium*

(IPDPS). doi: 10.1109/ipdps.2019.00041

Voichek, Y., & Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nature Genetics*, *52*(5), 534–540.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., ... Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One*,*9*(11), e112963.

Wang, W., Guan, R., Liu, X., Zhang, H., Song, B., Xu, Q., ... Wang, J. (2019). Chromosome level comparative analysis of Brassica genomes. *Plant Molecular Biology*, *99*(3), 237–249.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., ... Brassica rapa Genome Sequencing Project Consortium. (2011). The genome of the mesopolyploid crop species Brassica rapa. *Nature Genetics*, *43*(10), 1035–1039.

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., ... Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49.

Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., ... Gruning, B. A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Research*, *46*(W1), W11–W16.

Xiao, C.-L., Chen, Y., Xie, S.-Q., Chen, K.-N., Wang, Y., Han, Y., ... Xie, Z. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nature Methods*, *14*(11), 1072–1074.

Xiao, X.-M., Xu, Y.-M., Zeng, Z.-X., Tan, X.-L., Liu, Z.-L., Chen, J.-W., ... Chen, J.-Y. (2019). Activation of the Transcription of BrGA20ox3 by a BrTCP21 Transcription Factor Is Associated with Gibberellin-Delayed Leaf Senescence in Chinese Flowering Cabbage during Storage. *International Journal of Molecular Sciences*, *20*(16). doi: 10.3390/ijms20163860

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., ... Zhang, M. (2016). The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. *Nature Genetics*,*48*(10), 1225–1232.

Yang, X., Liu, D., Liu, F., Wu, J., Zou, J., Xiao, X., ... Zhu, B. (2013). HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*,*14*, 33.

Yang, X., Liu, H., Ma, Z., Zou, Y., Zou, M., Mao, Y., ... Yang, R. (2019). Chromosome-level genome assembly of Triplophysa tibetana , a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. *Molecular Ecology Resources*, *19*(4), 1027–1036.

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., ... Wang, X. (2018). Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Horticulture Research*, *5*, 50.