

# Analysis of Genomic Variation in Different Brassica Napus Synthetic Allopolyploids

**Yunxiao Wei**

CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute

**Guoliang Li**

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

**Fei Li**

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

**Shujiang Zhang**

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

**Shifan Zhang**

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

**Hui Zhang**

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

**Rui Zhang**

CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute

**Rifei Sun** (✉ [yuluoyunxiao@outlook.com](mailto:yuluoyunxiao@outlook.com))

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

---

## Research Article

**Keywords:** Whole-genome sequencing, synthetic Brassica napus, copy number variation (CNV), inter-chromosomal translocation (CTX)

**Posted Date:** May 19th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-516265/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

Allopolyploidy is an evolutionary and mechanistically intriguing process involving the reconciliation of two or more sets of diverged genomes and regulatory interactions, resulting in new phenotypes. In this study, we explored the genomic variation of eight F<sub>2</sub> synthetic *B. napus* using whole-genome sequencing. We found that there was a genetic variation in the F<sub>2</sub> generation. Part of the variation was consistent in the F<sub>2</sub> generation, and a small number of mutations only appeared in a single plant of the F<sub>2</sub> generation. The analysis of copy number variation (CNV) found that most of the AA genome was lost, and most of the CC genome was obtained. In addition, there was inter-chromosomal translocation (CTX) in the F<sub>2</sub> generation and the number of each plant was different. The above results indicate that the F<sub>2</sub> generation showed genetic variation and there was a difference between eight plants, which may lay a molecular basis for the unique field performance of the offspring. It provides a new perspective of genomic variation and trait separation in the early stages of allopolyploid polyploid formation.

## Introduction

Polyploidy refers to the fact that the chromosome numbers of genetically related organisms are in a multiple relationship with each other [1]. The results of the study now suggest that polyploidy is a frequent rather than accidental event, and taxonomists believe that many polyploids are produced through multiple independent polyploidization processes [2]. The phenomenon of high levels of unreduced gamete formation in natural populations of angiosperms [3] provides the basis for the formation of polyploidy, and further forms heteropolyploidy (involving interspecies hybridization) and autopolyploid through hybridization and doubling, both types of polyploid contribute greatly to angiosperm species diversity [4]. The newly formed polyploid undergoes genetic variation and changes in gene expression at an early stage to cope with the impact of multiple genomes in the polyploid. There are many forms of solutions to the impact of multiple genomes, including chromosomal recombination, loss of homologous genes, changes in gene expression, etc [5-11]. Polyploid is not only the addition product of its diploid parent, polyploid is a combination of additive and non-additive expression of the parent [12, 13]. In short, the polyploid phenotypic diversity and wide range of variation can provide species with strong adaptability.

Based on the "U-triangle" (U 1935) theory and Brassica genome sequencing data, it is generally believed that Brassica napus was naturally crossed and doubled from *B. rapa* and *B. oleracea* about 7500 years ago [14]. Moreover, it is easier to obtain hybrid offspring by embryo rescuing of *B. rapa* and *B. oleracea* as parents. The synthetic AACC allotetraploid early generation has abundant trait variations, such as flower size, flowering time, waxy characteristics, leaf shape and size [15, 16]. The role of genetics and epigenetics in heterologous polyploidization leads to changes in gene expression, which in turn leads to new phenotypes [15, 17-20].

Genome shock generated by hybridization and polyploidization can induce hybrid offspring to produce mutations at the genetic level, and the newly generated genetic variation will directly or indirectly cause hybrid offspring to produce new phenotypes and increase the adaptation of hybrid offspring. Rousseau-Gueuti et al. (2017) used Brassica 60K Infinium SNP array to investigate the SNP variation of 30 newly synthesized AACC heterotetraploid hybrids, and found that after heterologous polyploidization, their genomes were completely shuffled, only 8.5% and 3.5% of the C and A genomes are lost in generations. The identified deletions mainly occur in the distal part of the chromosome, and the C genome has a greater degree of variation than the A genome [21]. Wang et al. (2013) re-sequenced hybrid rice introgression lines and found that introgression hybridization caused widespread changes in the rice genome, and some of these mutations led to important new phenotypes [22].

There is little research on the genomic variation of the newly synthesized AACC heterotetraploid hybrids. In this experiment, the Chinese cabbage × Chinese kale and eight F<sub>2</sub> were re-sequenced to analyze the different characters genomic variation of the newly synthesized heteropolyploid. We found that there was a genetic variation in the F<sub>2</sub> generation. Part of the variation was consistent in the F<sub>2</sub> generation, and a small number of mutations only appeared in a single plant of the F<sub>2</sub> generation. The analysis of copy number variation (CNV) found that most of the AA genome was lost, and most of the CC genome was obtained. In addition, there was chromosomal translocation (CTX) in the F<sub>2</sub> generation and the number of each plant was different. The above results indicate that the F<sub>2</sub> generation showed genetic variation and there was a difference between eight plants, which may lay a molecular basis for the unique field performance of the offspring.

## Materials And Methods

### Plant materials

For this study, we used 10 accessions, including the female parent Cai-Xin, male parent Chinese kale, and eight F<sub>2</sub> synthetic allopolyploids (Fig. 1). The materials used are the same as in the previous article [23, 24].

## Whole-genome sequencing

Young leaves next to bud (5 cm in length) were collected, frozen in liquid nitrogen, and stored at -80°C until extraction. DNA was extracted using the CTAB method. Unamplified, high-molecular weight, RNase treated genomic DNA (4–6 µg) was used for WGS. WGS were performed at the Novogene company (Beijing, China) with an Illumina HiSeq 2000. WGS was performed with the TruSeq DNA prep kit. Sequencing was carried out so as to obtain 30× coverage from 2 × 150-bp paired-end reads.

### Analysis of Whole-genome Sequencing Data.

For the raw data, the adapter sequence, undetected bases, and bases with very low sequencing quality are filtered to obtain clean data, and the clean data is used for data analysis. The *B. rapa* genome and the *B. oleracea* genome were combined together to serve as reference genomes for eight F2 generation single plants. Use BWA software to compare the sequencing data of F2 to the merged reference genome. The sequencing data of the P1 was compared to the *B. rapa* genome, and the data of P2 to the *B. oleracea* genome. Use samtools software to detect genomic variation, including SNP and InDel, should be use CNVkit software to detect copy number variation (CNV) [25] and Breakdancer software to detect structural variation (SV) [26].

### Data accessibility statement

The resequencing data we sequenced would be uploaded to genebank database after the article is published.

## Results

### Acquisition of clean reads data

The statistical results of clean reads obtained by filtering out unqualified data with whole genome sequencing are shown in the following table (Table 1). The sequencing data of the parents and 8 F2 single plants are all 30 ×. The sum of the data of the two parents is similar to the 8 F2 single plants, which lays the foundation for the subsequent data analysis. The data of F2 were compared with the data after the parents were combined to analyze the genomic variation of F2.

**Table 1** Number of clean reads from parent and offspring

	AACC1	AACC2	AACC3	AACC4	AACC5	AACC6	AACC7	AACC8	P2	P1
clean reads	3.29E+10	3.19E+10	3.54E+10	3.00E+10	3.66E+10	3.40E+10	4.27E+10	3.60E+10	1.81E+10	1.35E+10

### Sequence Alignment

After comparing the clean reads of the parents and F2 plants with the genome. The results are shown in the table. The experiment found that the comparison rate of the parents and the F2 plants exceeded 90% (Table 2), most of the data can be used.

**Table 2** Comparison ratio of clean reads

	P1	P2	AACC1	AACC2	AACC3	AACC4	AACC5	AACC6	AACC7	AACC8
mapped	98.11%	99.02%	98.63%	98.67%	98.70%	97.92%	90.85%	98.73%	96.26%	98.71%
properly paired	93.16%	94.02%	91.29%	90.68%	90.68%	90.62%	83.03%	91.08%	88.22%	91.36%

### SNP and InDel analysis

Using samtools for SNP and InDel mutation analysis, it was found that the number of SNPs in F2 plants was higher than the sum of the parents, and the number of InDel in F2 plants was similar to the sum of the parents (Fig 2). Most of the F2 plant variants are located in the intergenic region, and a small number are located in the intron and exon regions of the gene. F2 plants had more heterozygous genotypes (ALT), pure and genotypes than their parents (Fig 2). In addition, the experimental analysis of the distribution of mutations found that 8 F2

plants had frequent genomic mutations, some of the mutations were consistent in all progeny (red rectangles), and a few mutations only appeared in one F2 single plant (purple circle) (Fig 3), it is speculated that this phenomenon is related to the differential expression of genes between individual plants, which in turn affects the performance of traits.

### **CTX and CNV analysis**

The experiment used Breakdancer software to detect structural variation (SV). SV includes deletion (DEL), inversion (INV), insertion (INS), intrachromosomal translocation (ITX), and interchromosomal translocation (CTX). Analysis found that there are a large number of interchromosomal translocations (CTX) in offspring, which is consistent with previous observations using chromosomes [27]. The occurrence of inter-chromosomal translocations leads to genomic variation, which in turn leads to the emergence of new phenotypes and trait separation in the field shape of newly synthesized AACC heterotetraploid hybrids. In addition, the experiment found that there are differences between individual plants, the number of 8 plants are: 46782, 45033, 43651, 45546, 44469, 46006, 44705, 46329 (Fig 4). The number of CTX in AACC1 was the highest, and the number of CTX in AACC3 was the least, which was consistent with the results of gene expression and small RNA expression [28]. It is speculated that the unique genetic differences of 8 plants may provide the molecular basis of unique field traits.

In the experiment, CNVkit software was used to detect copy number variation (CNV). The analysis found that in 8 F2 plants, the AA genome was mostly lost, and the CC genome was mostly represented (Fig 4). The results are consistent with the conclusions obtained by transcriptome sequencing analysis of differentially expressed genes and activation / silence genes in previous research (unpublished), indicating that the two genomes responded differently when the synthetic AACC heterotetraploid hybrids experienced WGD.

## **Discussion**

### **Genomic variation of synthetic AACC heterotetraploid**

The genome shock generated by hybridization and polyploidization can induce mutations at the genomic level of the hybrid offspring. Rousseau-Gueutin et al. (2017) used Brassica 60K Infinium SNP array to investigate the SNP variation of the newly synthesized AACC heterotetraploid hybrids and found that the A and C genomes showed significant variation [21]. Wang et al. (2013) re-sequenced hybrid rice introgression lines and found that introgression hybridization caused widespread changes in the rice genome, and some of these mutations led to important new phenotypes [22]. The results of this experiment found that there were SNP and InDel mutations in 8 F2 plants, and the number of F2 single plant mutations was higher than that of the parents, indicating that hybridization and polyploidization caused new mutations in the F2 single plant genome. In addition, the experiment found that some of the mutations were consistent in 8 plants, and a small number of mutations only appeared in a single plant. The unique mutations of each plant were consistent with the results of transcriptome sequencing analysis of non-additively expressed genes in previous research (unpublished), indicating individual plant-specific genetic variation may produce new traits by affecting gene expression. It is speculated that genetic variation lays a molecular foundation for the separation of field traits.

### **Copy number variation and structural variation of synthetic AACC heterotetraploid**

Structural variation (SV) includes deletion (DEL), insertion (INS), inversion (INV), interchromosomal translocation (CTX), and intrachromosomal translocation (ITX). Few people have used sequencing technology to explore the structural variation caused by hybridization and polyploidization. Generally, FISH and GISH are used to investigate the phenomenon of chromosomal translocation in newly synthesized AACC heterotetraploid hybrids [6]. In this experiment, using bioinformatics analysis, it was found that the offspring of the hybridization had more interchromosomal translocations, which was consistent with the previous conclusions. In addition, genome copy number variations (CNVs) refer to complex variations derived from the insertion, amplification, and deletion of DNA fragments  $\geq 1$  kb in the genome aligned with the genome reference sequence. Experimental analysis of copy number variation found that most of the AA genome performance was lost, and the corresponding majority of the CC genome performance was obtained, indicating that the two genomes showed differences in response to WGD. In addition, regardless of the number of CTX or the performance of CNVs, the performance among individual plants is different, which is consistent with the results of transcriptome analysis and small RNA expression analysis in previous studies [28]. Speculate that the specific variation of a single plant lays a molecular foundation for its unique field traits. Although, due to the small number of individual plants, it is difficult to associate genetic variation, small RNA changes, and gene expression with a specific trait, but this study is necessary. The population should be expanded in the future to link the variation with the trait, providing substantial theoretical basis for breeding.

## Conclusion

This study explored the genomic variation of eight F2 synthetic *B. napus* using resequencing. The results have shown that F2 generation showed genetic variation and there was a difference between eight plants, which may lay a molecular basis for the unique field performance of the offspring. It provides a new perspective of genomic variation and trait separation in the early stages of allopolyploid polyploid formation.

## Declarations

### Author Contribution statement

Rifei Sun and Yunxiao Wei designed the experimental design and wrote the article. Fei Li, Shujiang Zhang, Shifan Zhang and Hui Zhang contributed to the interpretation of the results and coordinated the study. Yunxiao Wei performed the experiment. All the authors read and approved the final manuscript.

### Acknowledgments

This research was funded by the National Key Research and Development Program of China (grant number 2017YFD0101802) and the Fundamental Research Funds for Central Non-profit Scientific Institution (grant number IVF-BRF2018003). The experiment was performed at the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, Beijing, China. We would like to thank Accdon for providing linguistic assistance during the preparation of this manuscript.

### Competing interests

The authors declare that they have no conflict of interest.

### Data accessibility statement

The Resequencing data is available at SRA: SRR12113183, SRR12113184, SRR12113185, SRR12113186, SRR12113187, SRR12113188, SRR12113189, SRR12113190, SRR12113191, SRR12113192.

## References

1. Stebbins G: **Genetics, evolution and plant breeding.** *Indian J Genet Plant Breed* 1957.
2. Soltis DE, Soltis PS: **Polyploidy: recurrent formation and genome evolution.** *cell* 1999, **14**.
3. Ramsey J, Schemske DW: **Neopolyploidy in Flowering Plants.** *Annual Review of Ecology, Evolution, and Systematics* 2002, **33**(1):589-639.
4. Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA: **On the relative abundance of autopolyploids and allopolyploids.** *New Phytologist* 2016, **210**(2):391-398.
5. Song K, Lu P, Tang K, Osborn TC: **Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution.** *Proceedings of the National Academy of Sciences* 1995, **92**(17):7719.
6. Xiong Z, Gaeta RT, Pires JC, Wessler SR: **Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid Brassica napus.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(19):7908-7913.
7. Adams KL, Cronn R, Percifield R, Wendel JF: **Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing.** *Proceedings of the National Academy of Sciences* 2003, **100**(8):4649-4654.
8. Adams KL, Wendel JF: **Polyploidy and genome evolution in plants.** *Curr Opin Plant Biol* 2005, **8**(2):135-141.
9. Cui C, Ge X, Zhou Y, Li M, Li Z: **Cytoplasmic and genomic effects on non-meiosis-driven genetic changes in Brassica hybrids and allotetraploids from pairwise crosses of three cultivated diploids.** *PLoS One* 2013, **8**(5):e65078.
10. Ge XH, Ding L, Li ZY: **Nucleolar dominance and different genome behaviors in hybrids and allopolyploids.** *Plant Cell Rep* 2013, **32**(11):1661-1673.
11. Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L *et al*: **Genomewide nonadditive gene regulation in Arabidopsis allotetraploids.** *Genetics* 2006, **172**(1):507-517.

12. Chelaifa H, Monnier A, Ainouche M: **Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (Poaceae).** *New Phytol* 2010, **186**(1):161-174.
13. Yoo MJ, Szadkowski E, Wendel JF: **Homoeolog expression bias and expression level dominance in allopolyploid cotton.** *Heredity (Edinb)* 2013, **110**(2):171-180.
14. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B *et al.*: **Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome.** *Science* 2014, **345**(6199):950-953.
15. Chen ZJ: **Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids.** *Annu Rev Plant Biol* 2007, **58**:377-406.
16. He L-q, Tang R-h, Jiang J, Xiong F-q, Huang Z-p, Wu H-n, Gao Z-k, Zhong R-c, He X-h, Han Z-q: **Rapid gene expression change in a novel synthesized allopolyploid population of cultivated peanut×*Arachis doigoi* cross by cDNA-SCoT and HFO-TAG technique.** *Journal of Integrative Agriculture* 2017, **16**(5):1093-1102.
17. Samans B, Chalhoub B, Snowdon RJ: **Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species *Brassica napus*.** *The Plant Genome* 2017, **10**(3):plantgenome2017.2002.0013.
18. Stein A, Coriton O, Rousseau-Gueutin M, Samans B, Schiessl SV, Obermeier C, Parkin IAP, Chevre AM, Snowdon RJ: **Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*.** *Plant Biotechnol J* 2017, **15**(11):1478-1489.
19. Mason AS, Batley J, Bayer PE, Hayward A, Cowling WA, Nelson MN: **High-resolution molecular karyotyping uncovers pairing between ancestrally related *Brassica* chromosomes.** *New Phytologist* 2014, **202**(3):964-974.
20. Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP: **Detecting &#x26;de Novo&#x27; Homoeologous Recombination Events in Cultivated &#x26;Brassica napus&#x27; Using a Genome-Wide SNP Array.** *G3: Genes/Genomes/Genetics* 2018, **8**(8):2673.
21. Rousseau-Gueutin M, Morice J, Coriton O, Huteau V, Trotoux G, Negre S, Falentin C, Deniot G, Gilet M, Eber F *et al.*: **The Impact of Open Pollination on the Structural Evolutionary Dynamics, Meiotic Behavior, and Fertility of Resynthesized Allotetraploid *Brassica napus* L.** *G3 (Bethesda)* 2017, **7**(2):705-717.
22. Wang ZH, Zhang D, Bai Y, Zhang YH, Liu Y, Wu Y, Lin XY, Wen JW, Xu CM, Li LF *et al.*: **Genomewide variation in an introgression line of rice-*Zizania* revealed by whole-genome re-sequencing.** *PLoS One* 2013, **8**(9):e74479.
23. Wei Y, Li F, Zhang S, Zhang S, Zhang H, Sun R: **Characterization of Interspecific Hybrids between Flowering Chinese Cabbage and Chinese Kale.** *Agronomy* 2018, **8**(11).
24. Wei YX, Li F, Zhang SJ, Zhang SF, Zhang H, Sun RF: **Analysis on Interspecific Hybridization Compatibility and Progeny Characteristics of Flowering Chinese Cabbage and Chinese Kale.** *CHINA VEGETABLES* 2017, **11**.
25. Talevich E, Shain AH, Botton T, Bastian BC: **CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing.** *PLoS Comput Biol* 2016, **12**(4):e1004873.
26. Fan X, Abbott TE, Larson D, Chen K: **BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping.** *Current Protocols in Bioinformatics* 2014, **45**(1):15.16.11-15.16.11.
27. Xiong Z, Gaeta RT, Pires JC: **Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*.** *Proc Natl Acad Sci U S A* 2011, **108**(19):7908-7913.
28. Wei Y, Li F, Zhang S, Zhang S, Zhang H, Sun R: **Analysis of small RNA changes in different *Brassica napus* synthetic allopolyploids.** *PeerJ* 2019, **7**:e7621.

## Figures

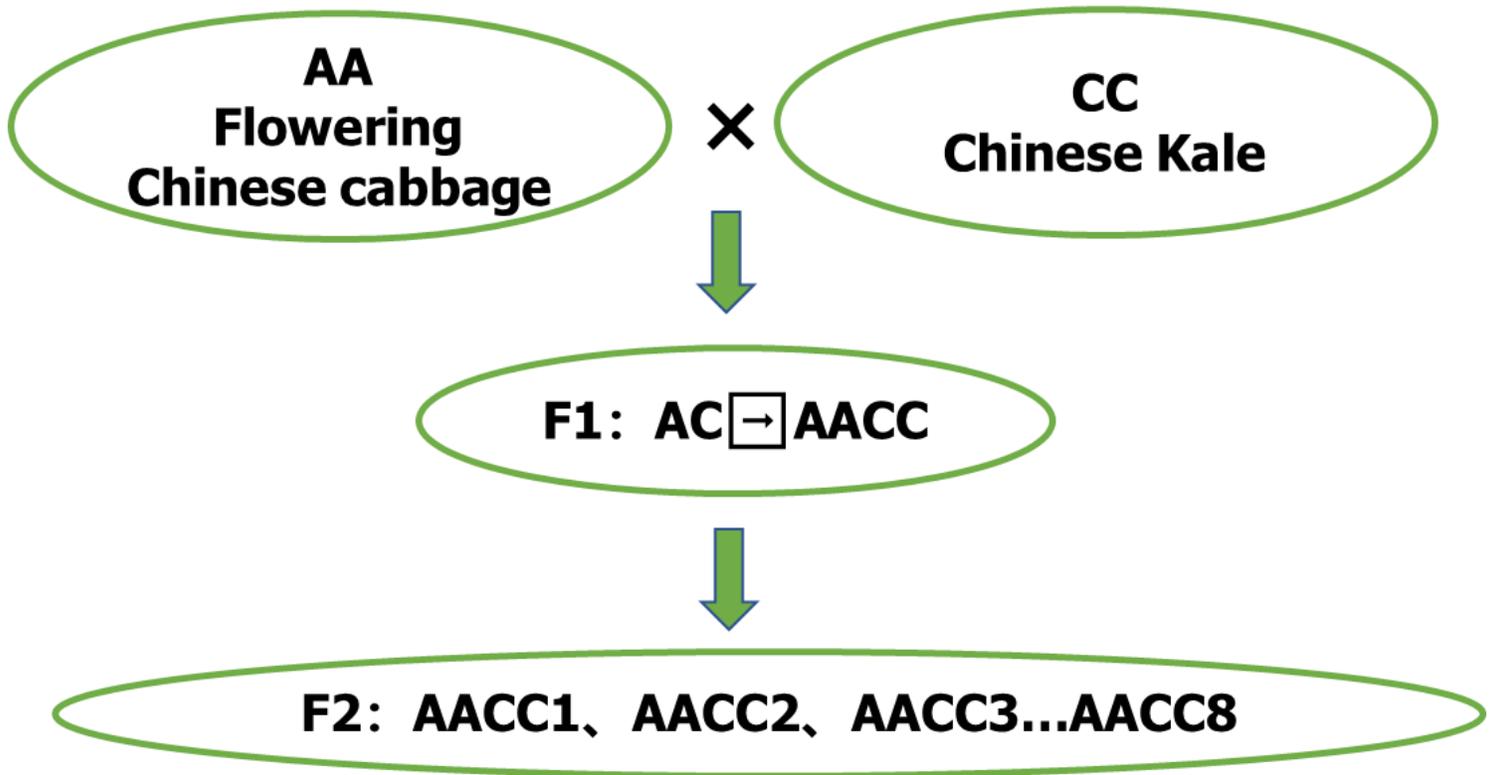


Figure 1

Plant materials

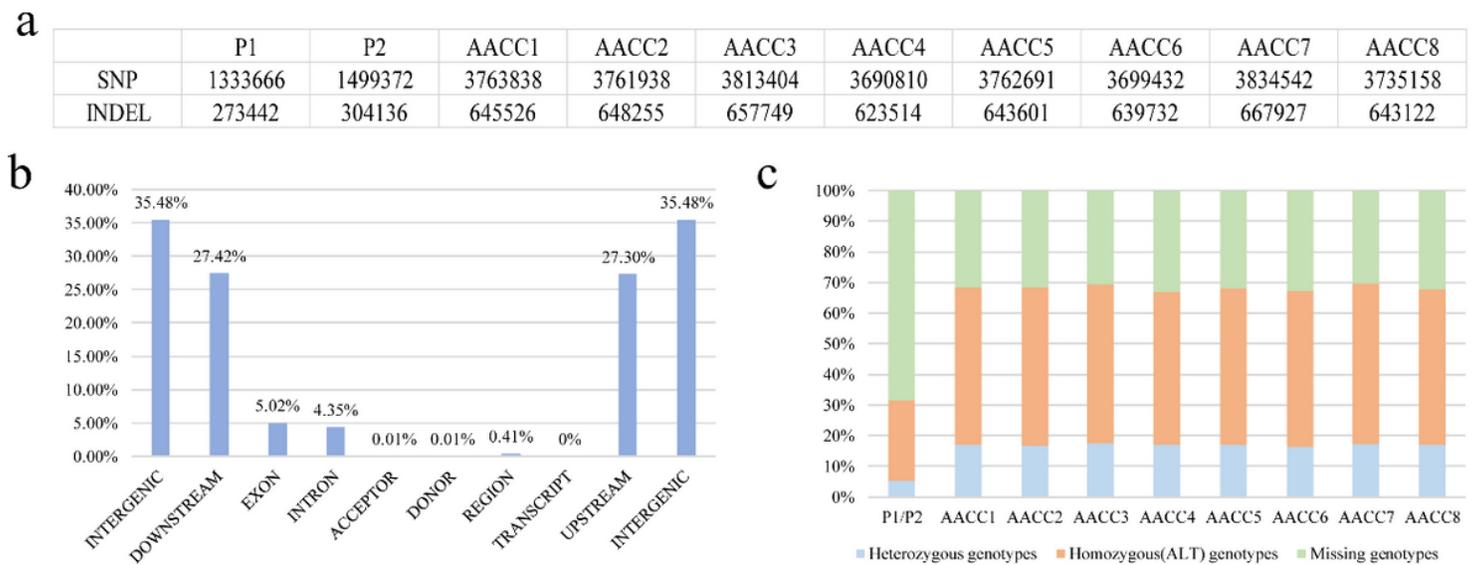
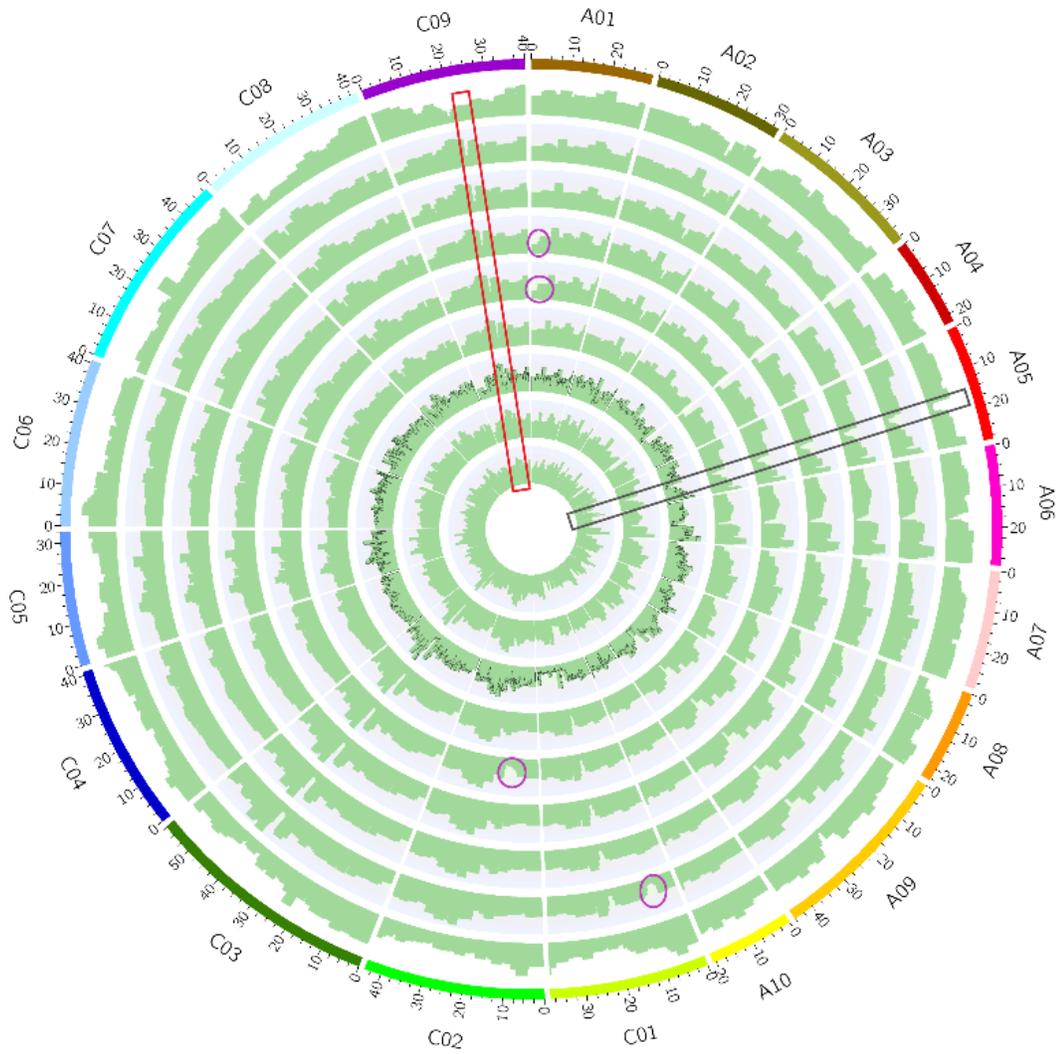


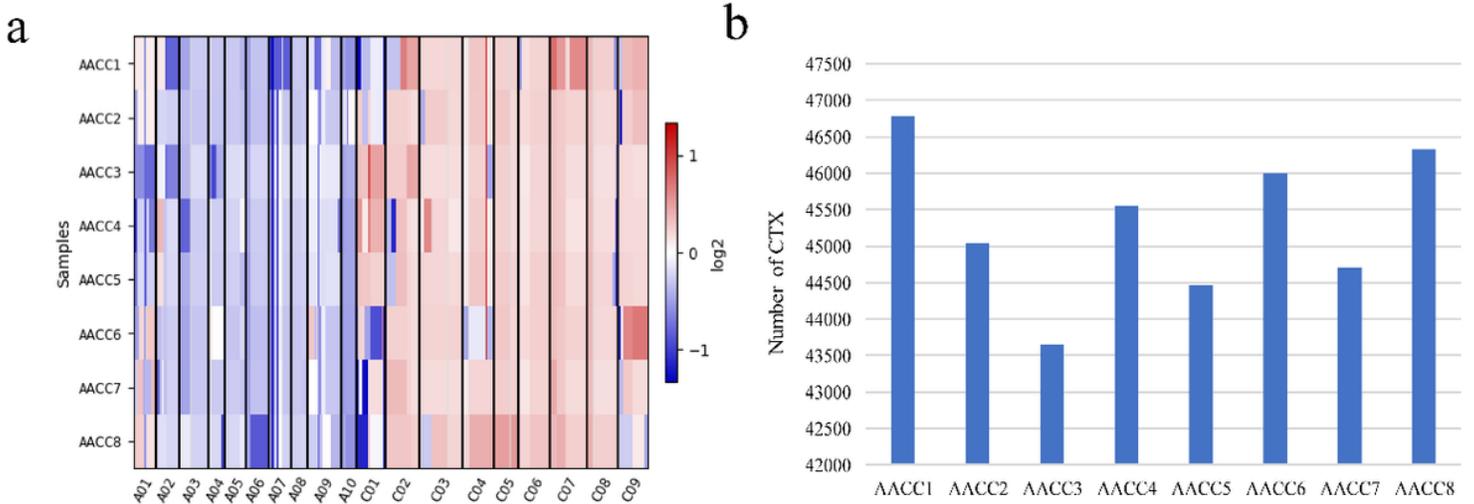
Figure 2

Genetic variation analysis of parents and offspring a: Number of SNP and InDel in parents and offspring; b: Location of genetic variation; c: Number of genotypes in parents and offspring.



**Figure 3**

Variation distribution of parents and offspring The outermost circle is the parent, and the inside is AACC1...AACC8



**Figure 4**

CNV and CTX analysis of parents and offspring a: Distribution of CNV in parents and offspring Blue represents loss, red represents gain; b: Number of CTX in parents and offspring.