Figure S1 Hi-C interaction heatmap of the pak choi genome. The color bar represents the logarithm of contact density from high (red) to low (white). Note that only sequences anchored on chromosomes are shown in the plot.
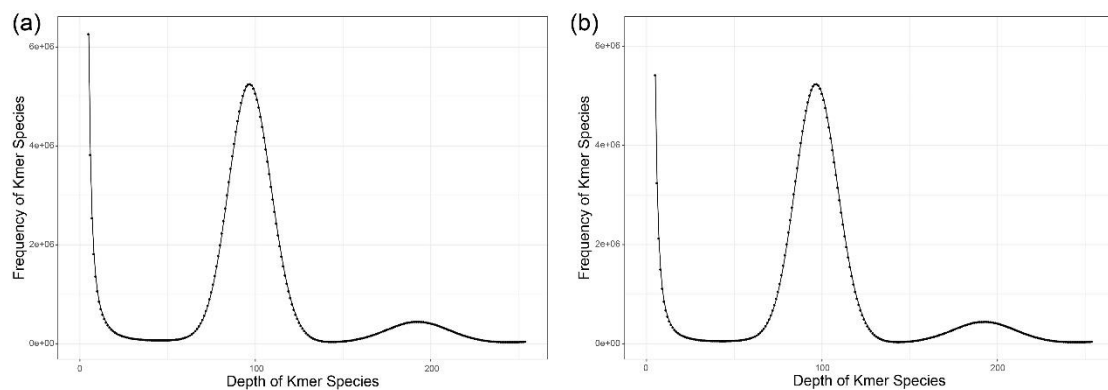

Figure S2 KmerFreq-AR analysis of pak choi genome size. (a) Distribution of Kmer frequencies using all pak choi reads data. (b) Distribution of Kmer frequencies without the organelle reads data. K-mers values were plotted against the frequency. The overall genome size and size without the organelle data were about 463 and 406
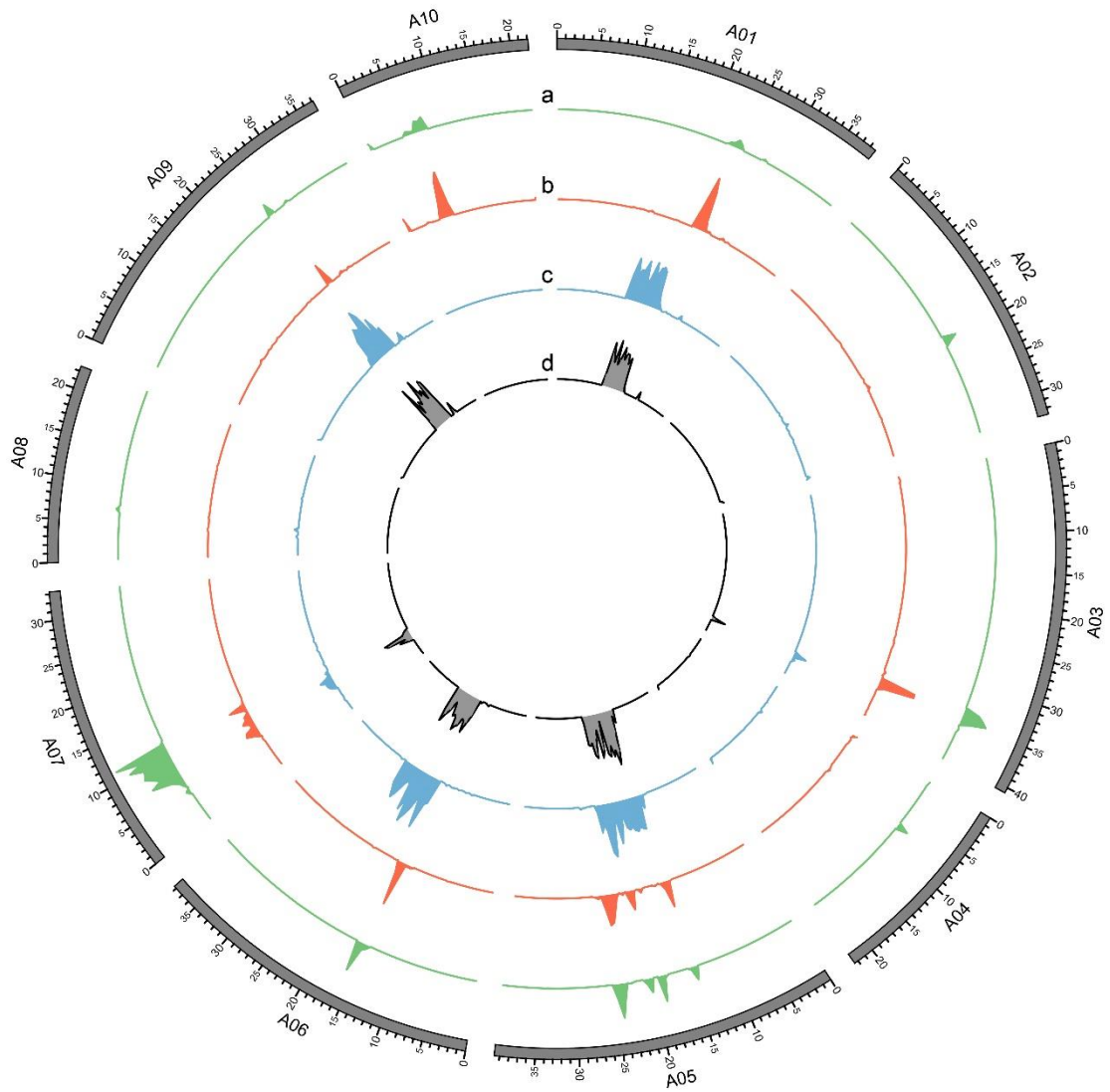
Mb, respectively.

Figure S3 Distribution of centromeric repeats in pak choi genome. Tracks (a) CentBr repeats, (b) CRB repeats, (c) PCRBr repeats, (d) TR238 repeats.
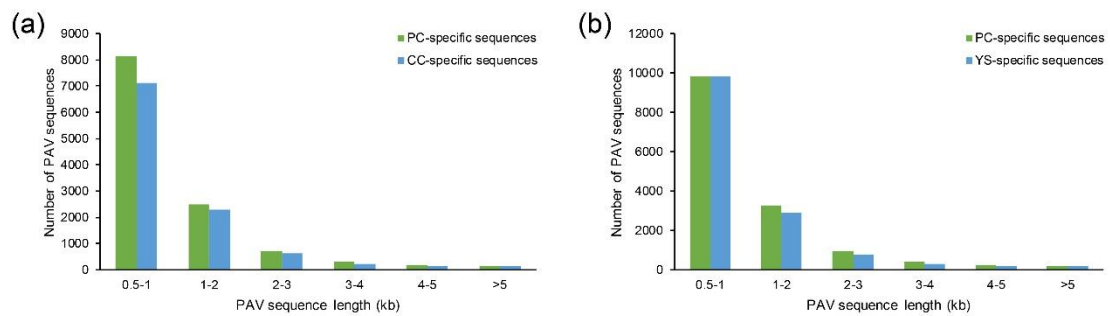

Figure S4 Length distribution of PAV sequences in the Chinese cabbage, yellow

sarson and pak choi genomes. (a) Length distribution of PAV sequences between Chinese cabbage and pak choi genomes. (b) Length distribution of PAV sequences between yellow sarson and pak choi genomes.
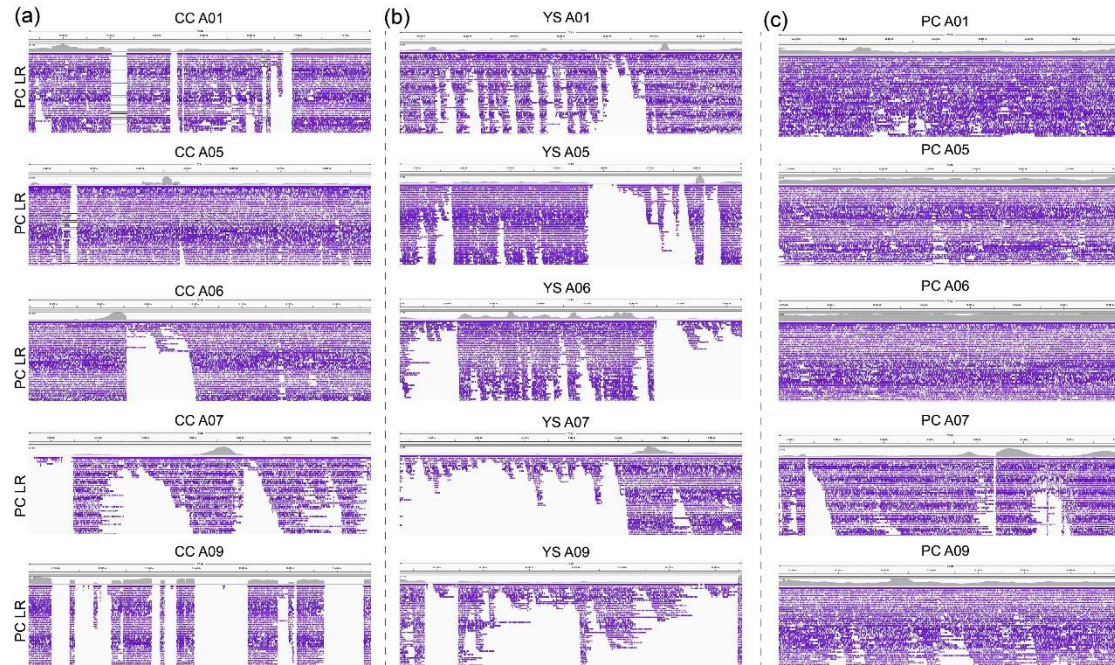

Figure S5 The IGV alignments of centromeric regions using pak choi long reads. The PacBio data of pak choi was mapped to Chinese cabbage (a), yellow sarson (b) and pak choi (c) genomes, respectively.
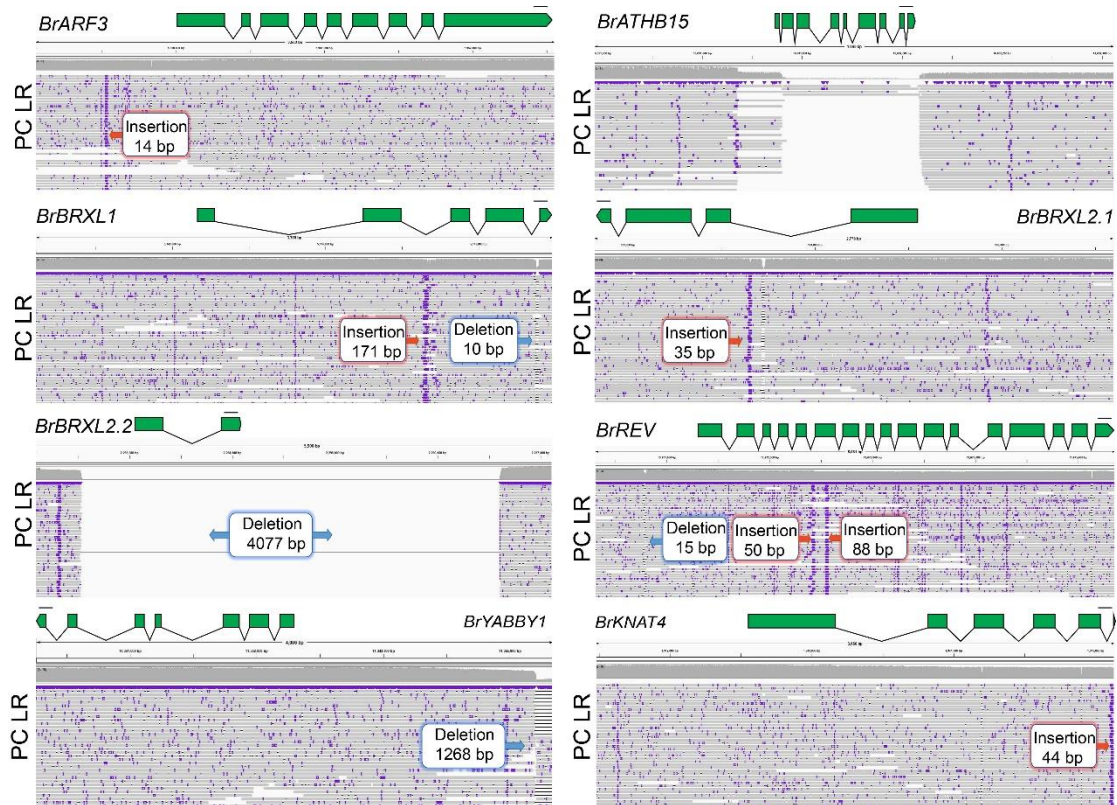
Figure S6 Identification of structural variants of the eight remaining heading genes.
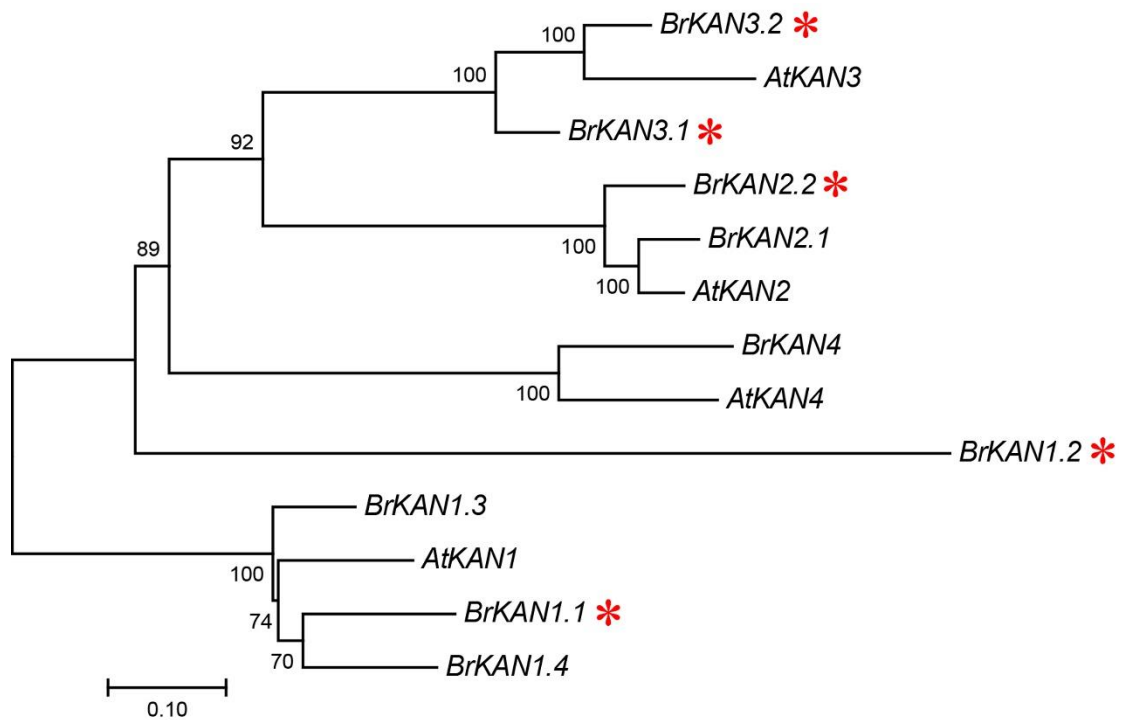

Figure S7 Phylogenies of the *KAN* genes from *A. thaliana* and *B. rapa*

annotations. The annotated genes from *A. thaliana* and *B. rapa* are prefixed with *At* and *Br* respectively. Red star indicates the variant *KAN* genes between pak choi and Chinese cabbage.
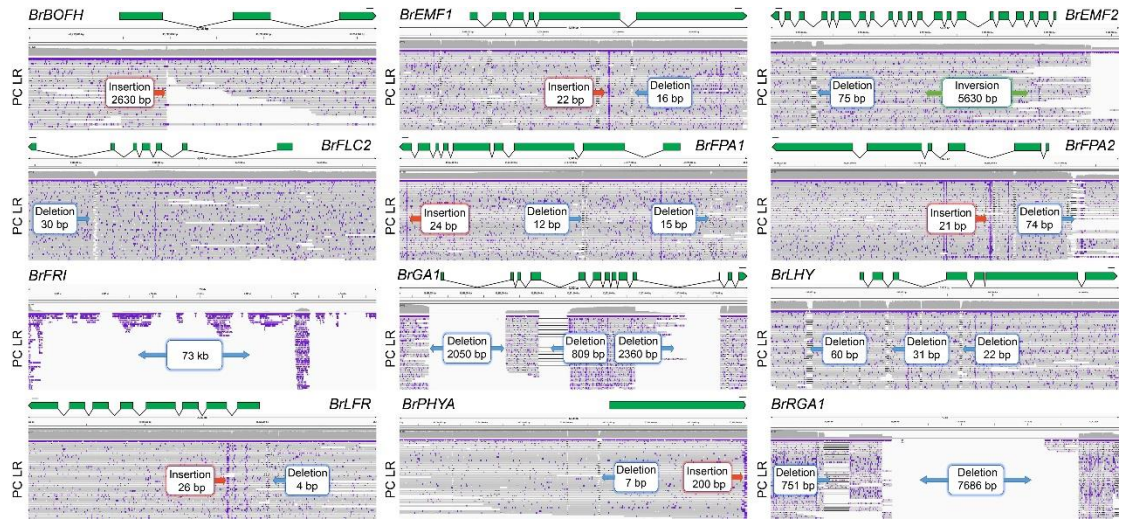

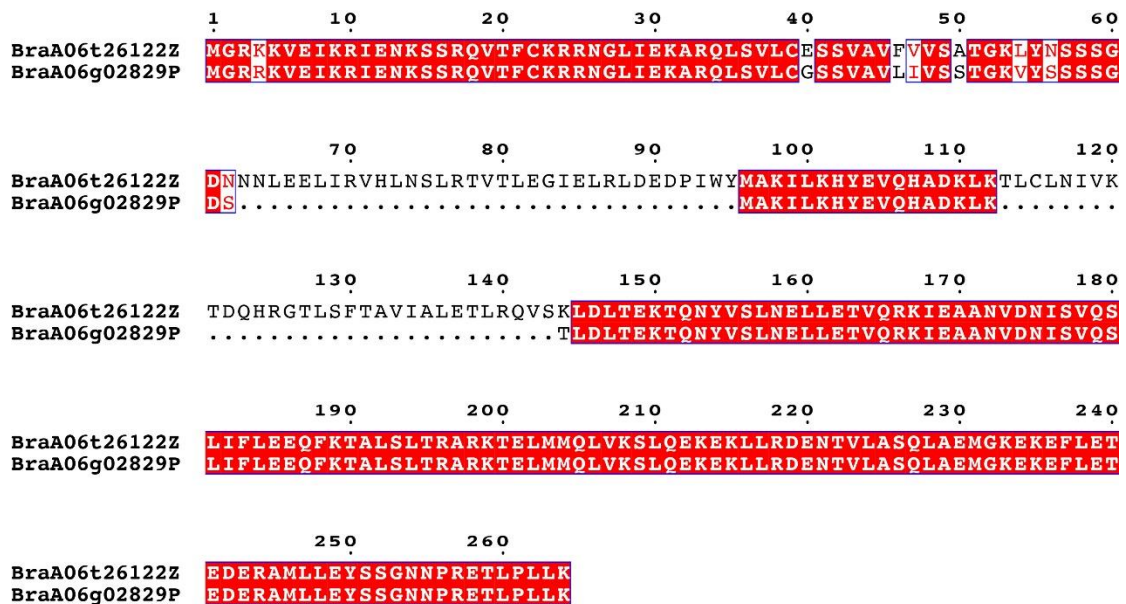**Figure S8 Identification of structural variants of the 12 remaining flowering genes.**

Figure S9 Alignment of BrMAF4-annotated protein sequences between the pak choi and yellow sarson genomes.

Figure S10 The pak choi phenotype (ZYCX).