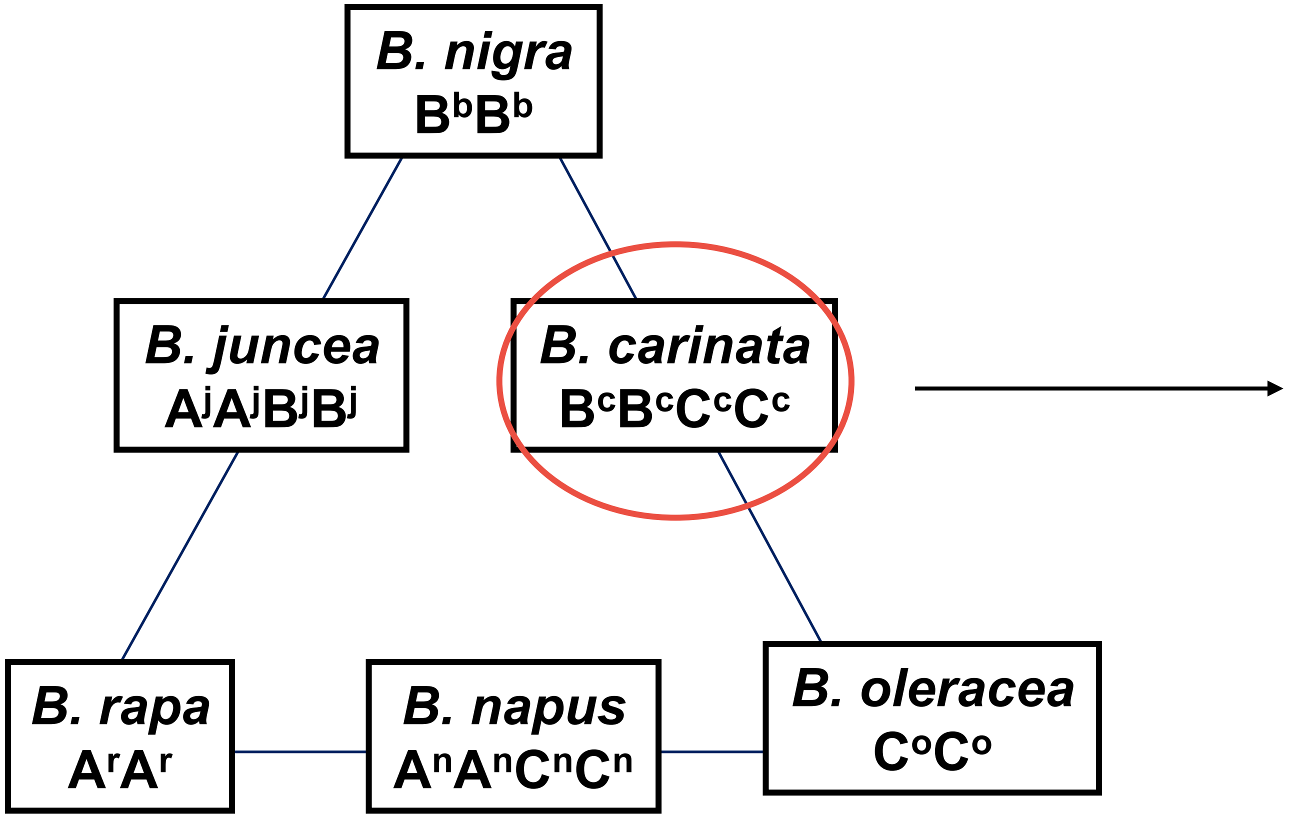
A *Brassica carinata* pan-genome platform for Brassica crop improvement

Yan Niu
Huazhong Agricultural University
WUHAN, CHINA

2023.09.26

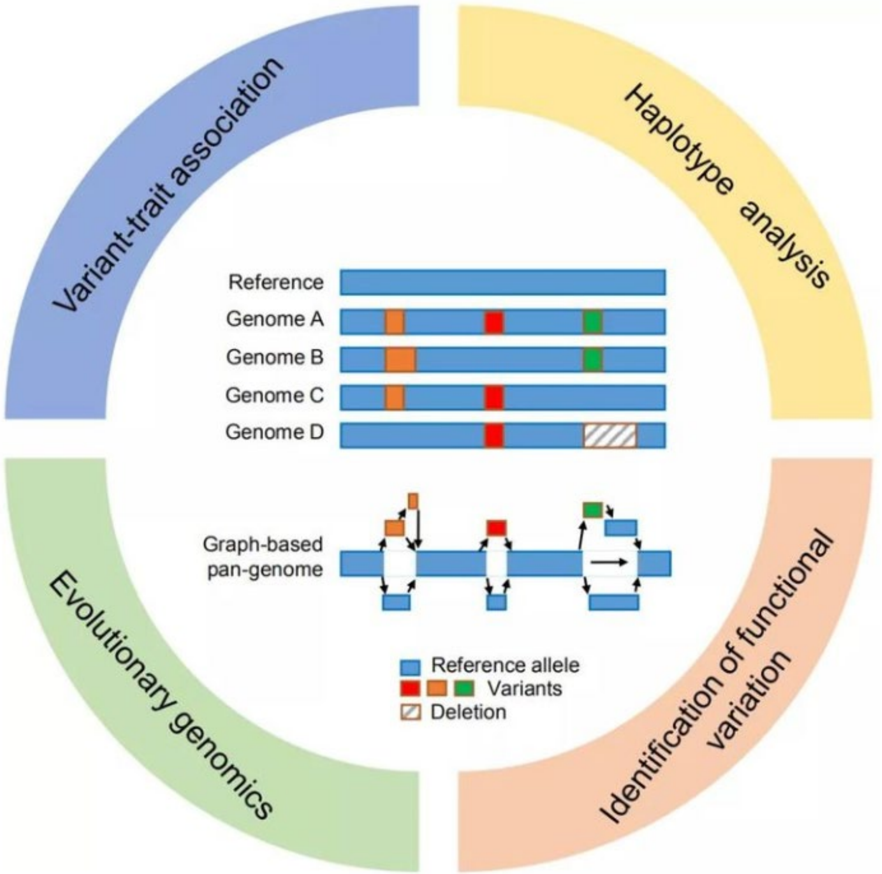
Brassica carinata is an important agricultural crop

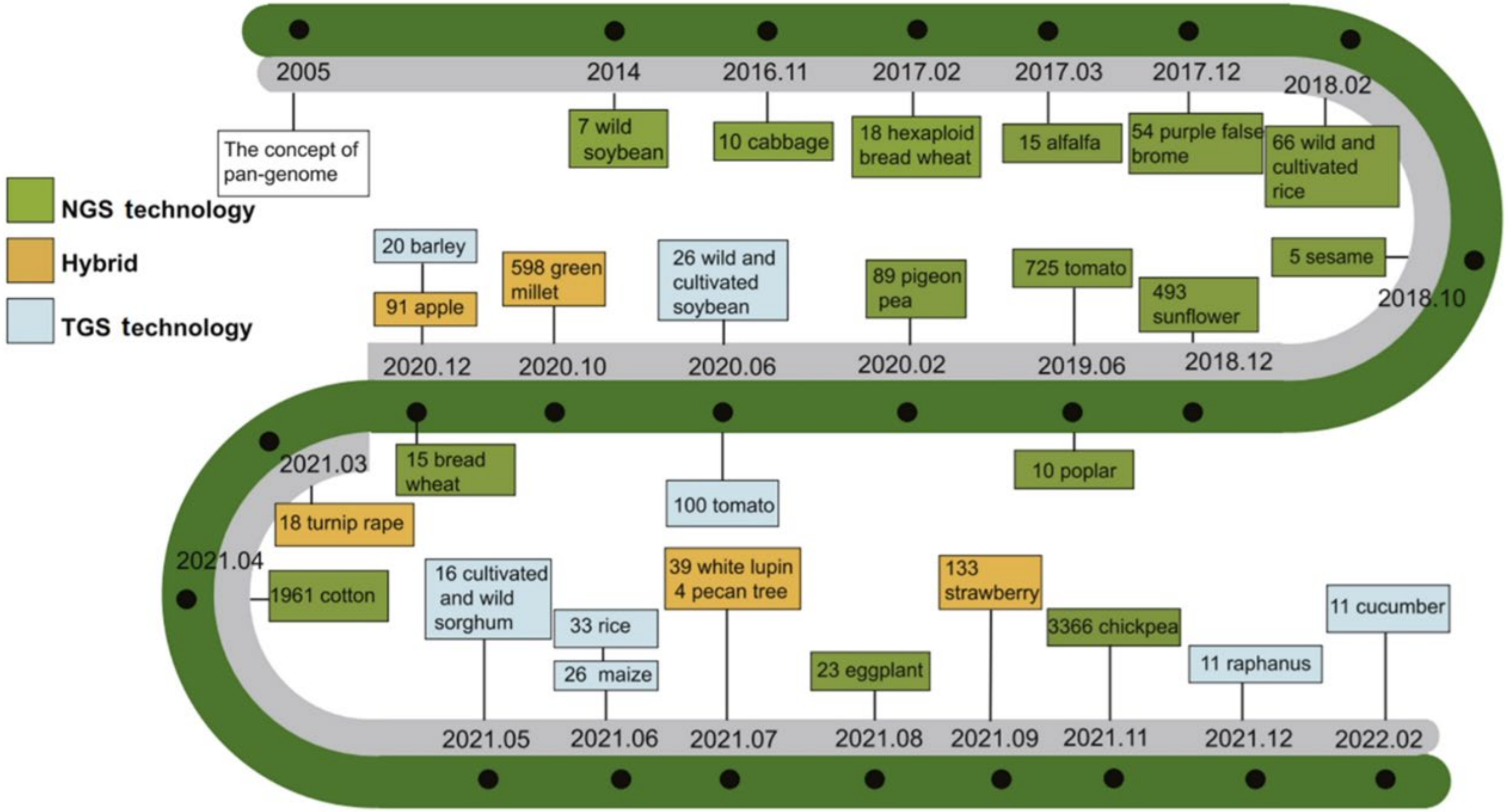


***Brassica* U's triangle**

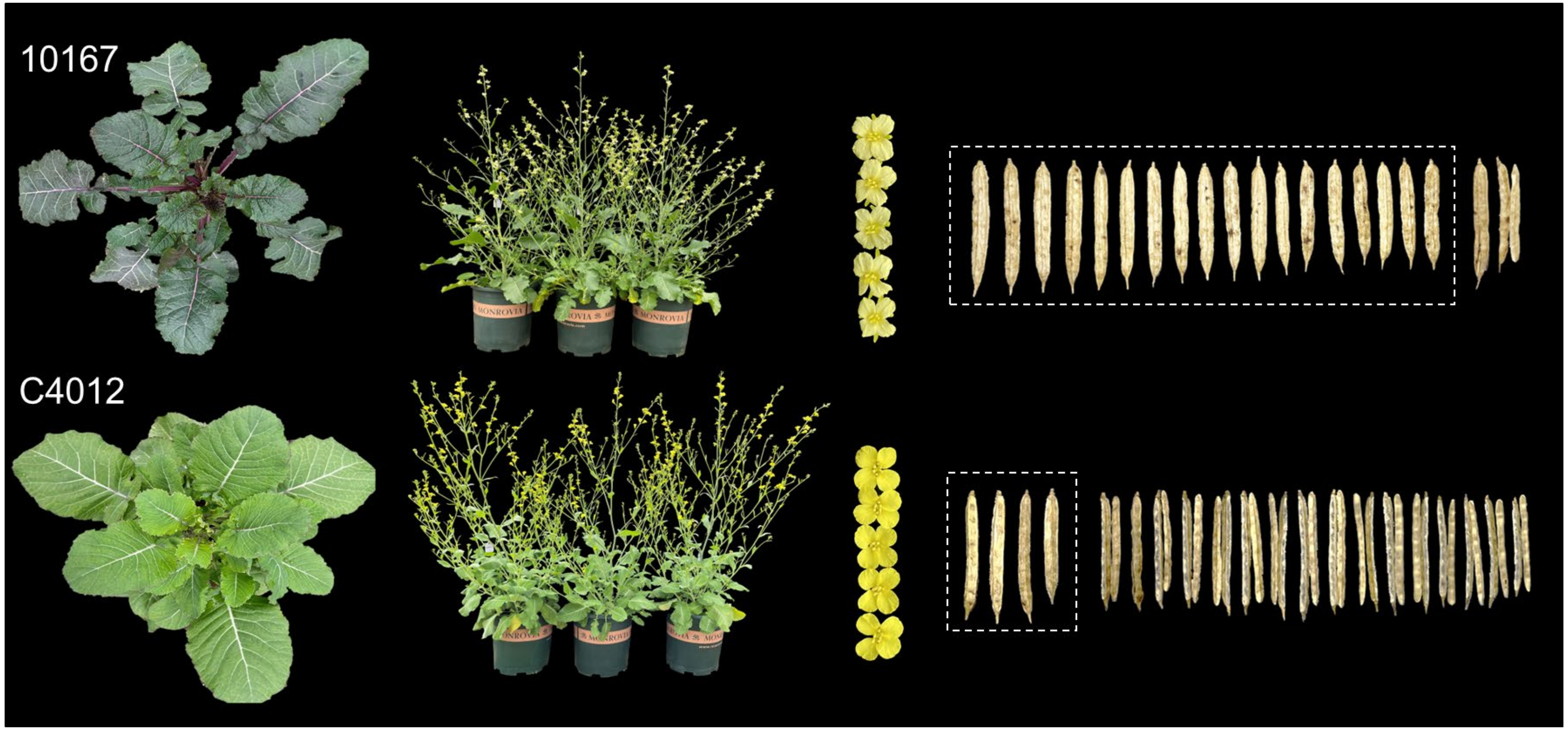
- high glucosinolate
- high erucic acid
- pod shatter resistance
- disease resistance
- drought tolerance
- salt tolerance

No available pan-genome for *B. carinata*





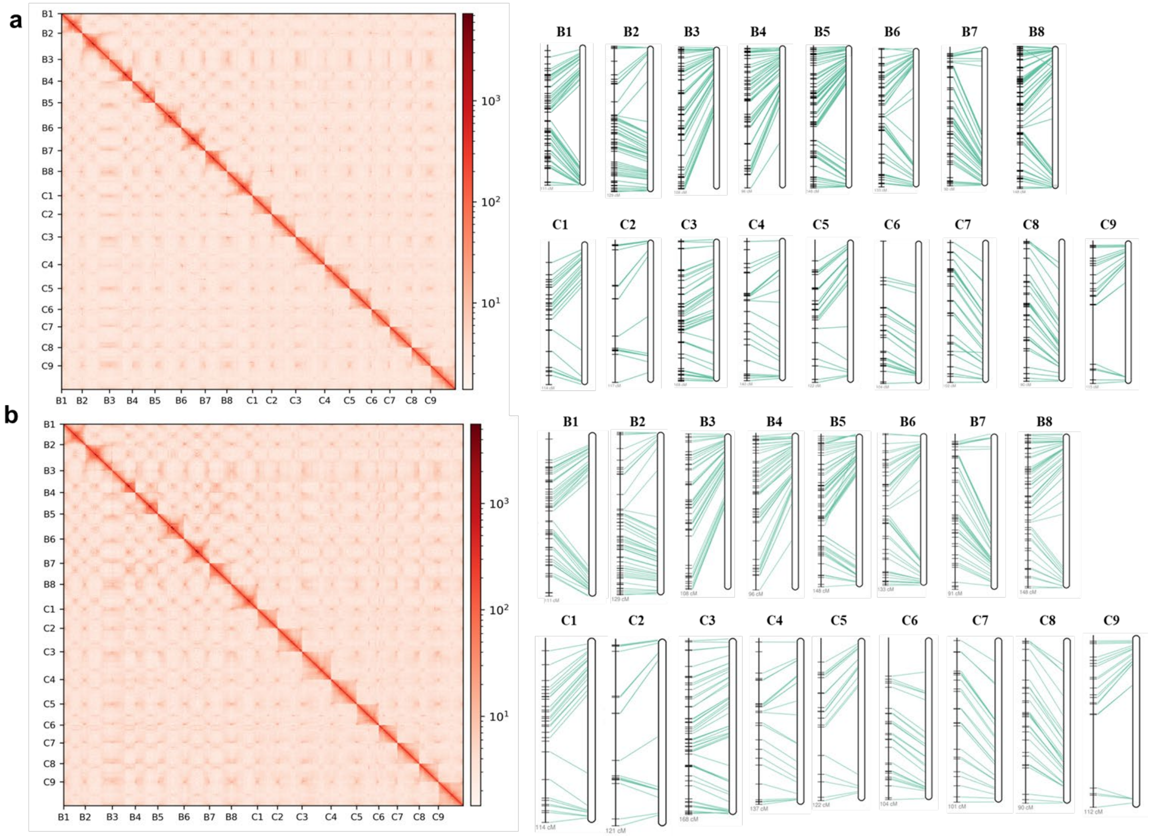
The assembly of two *B. carinata* genomes



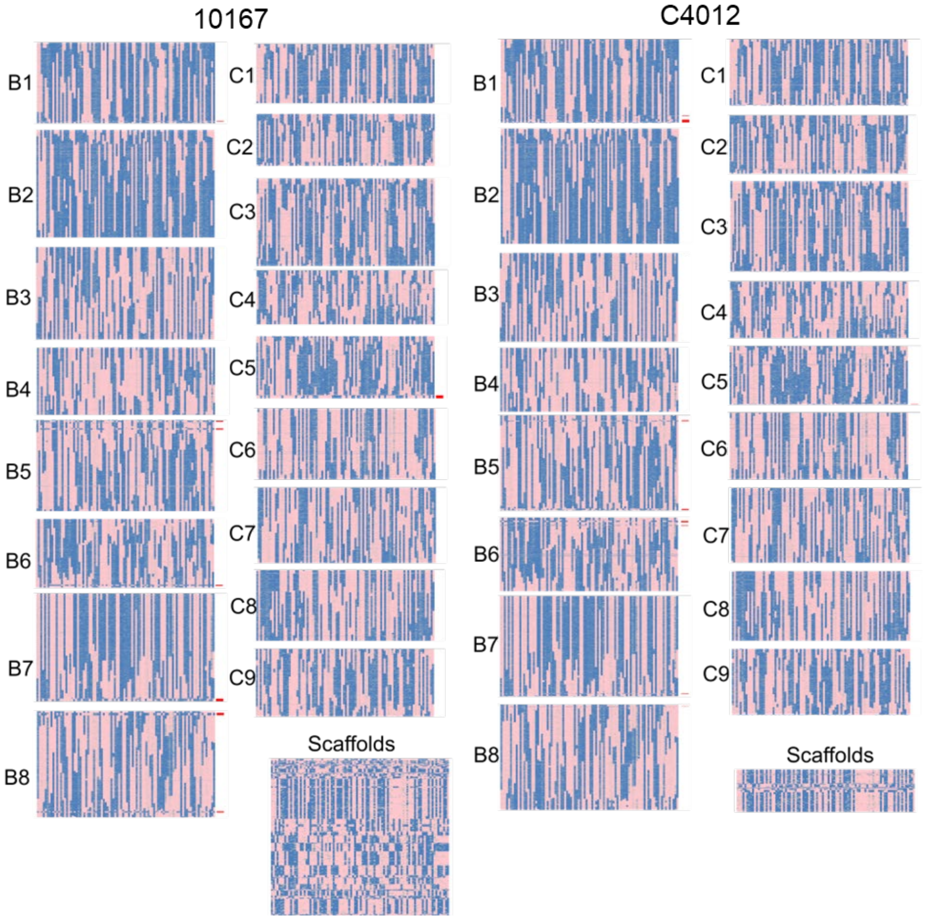
- **10167 (pod shatter-resistant accession):** 50× Illumina + 83× ONT + 121× Hi-C
- **C4012 (shatter-susceptible accession):** 45× Illumina + 75× ONT + 115× Hi-C

The assembly of two *B. carinata* genomes

96.2% and 96.8% contigs of 10167 and C4012 were anchored to 17 pseudochromosomes by **Hi-C**, **linkage mapping** and **GOGGs**.

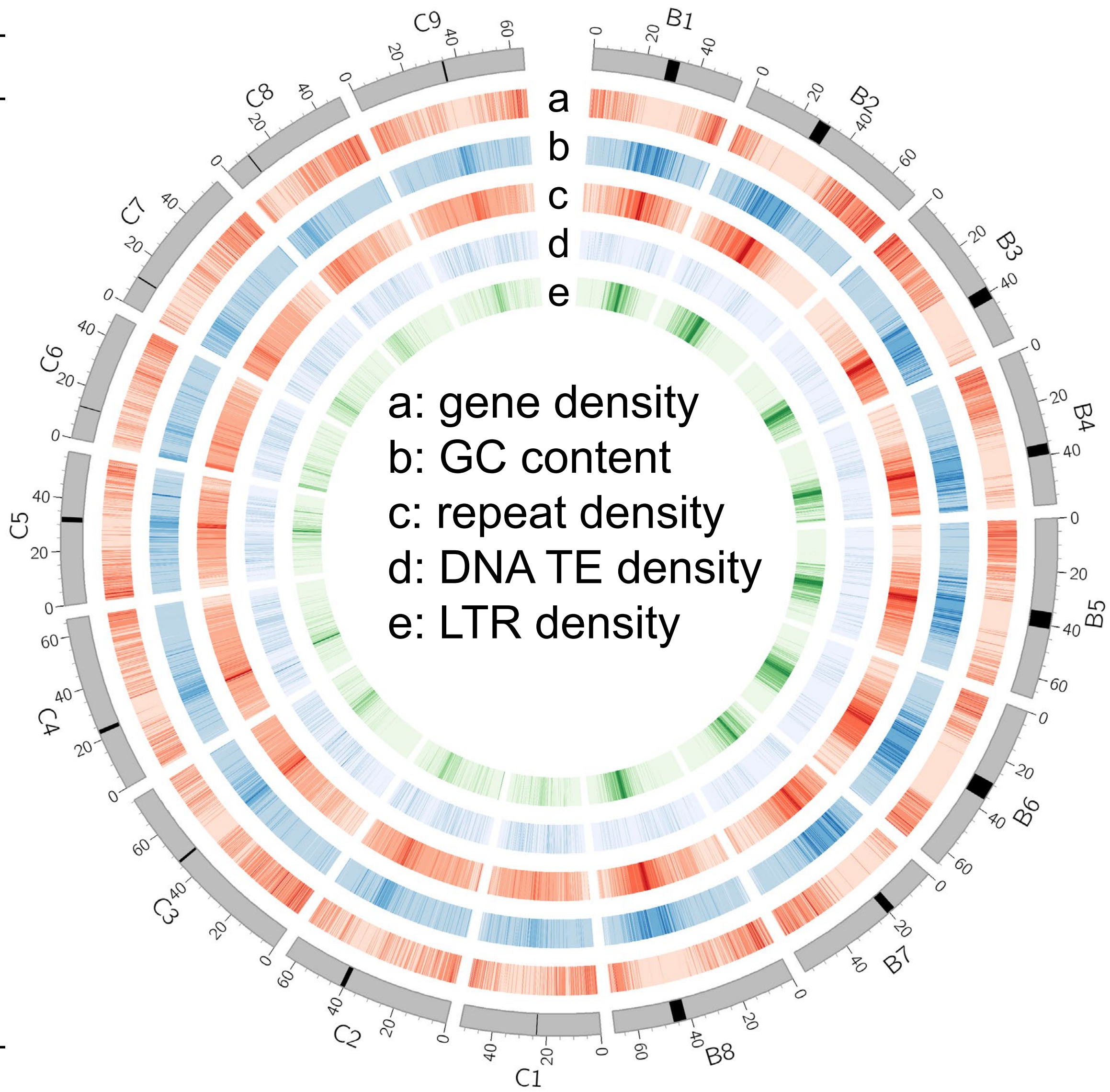


(a-b) The Hi-C map and collinearity with genetic map of 10167 and C4012.



The assembly of two *B. carinata* genomes

Parameters	10167	C4012
Estimated genome size (Mb)	1,156	1,131
Final Genome size (Mb)	**1,071**	**1,074**
Contig number	375	391
Longest contig (Mb)	70.5	45.8
Contig N50 (Mb)	**26.7**	**24.3**
Scaffold number	268	531
Scaffold N50 (Mb)	62.4	64.1
Longest Scaffold (Mb)	74.9	75.4
Anchored to chromosome (%)	96.2	96.8
GC (%)	37.4	37.4
LAI	**15. 5**	**16**
Mapping rate of short reads (%)	99.3	99.4
BUSCO (%)	**99.7**	**99.6**
Number of protein-coding genes	102,334	103,866
Repeat proportion (%)	59.1	58.5



The genomic features of C4012

The construction of *B. carinata* pan-genome

(1) Firstly using *de novo* strategy

Source	Order added	Sub genome	Number	No. At	% At
B. carinata C4012	1	B	46,808	28,726	61.37
B. carinata C4012	1	C	55,593	29,924	53.83
B. carinata 10167	2	B	2,353	717	30.47
B. carinata 10167	2	C	2,350	675	28.72
B. carinata zd1	3	B	2,050	588	28.68
B. carinata zd1	3	C	2,498	716	28.66
B. carinata Gomenzer	4	B	3,077	620	20.15
B. carinata Gomenzer	4	C	2,496	410	16.43
BcaPan		B	54,288	30,651	56.46
		C	62,937	31,725	50.41
		BBCC	117,225	62,376	53.21

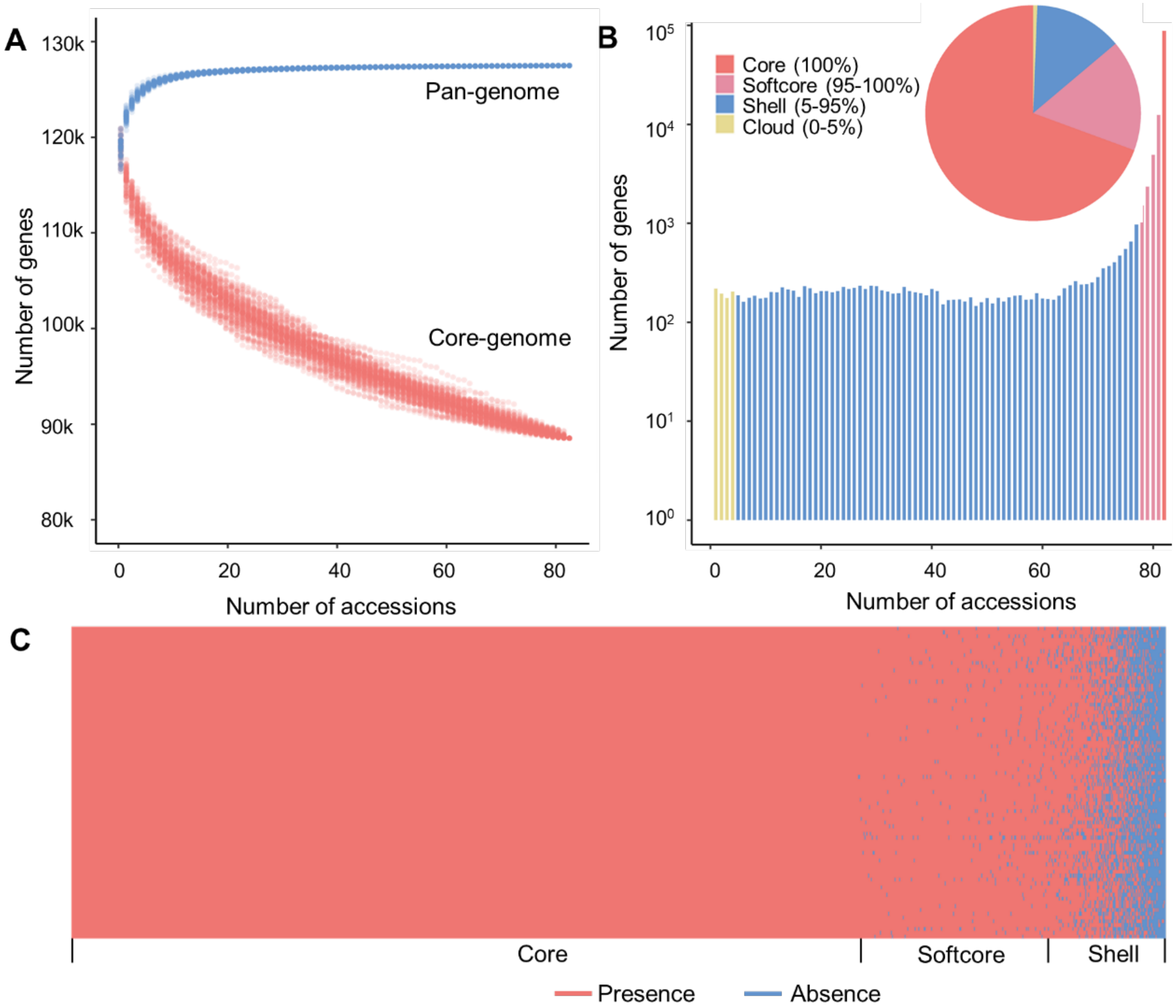
◆ 117,225 gene models in the initial pan-genome

◆ 14,824 (12.65%) gene models were added to the reference genome

The construction of *B. carinata* pan-genome

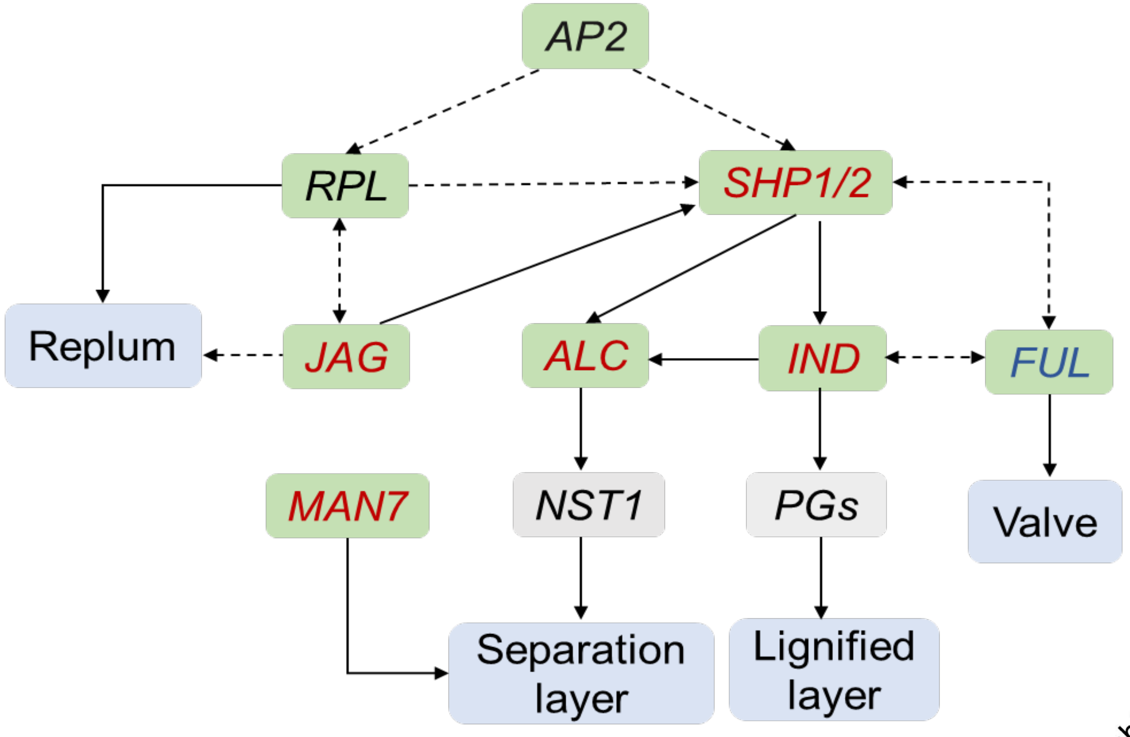
(2) Then using map to pan

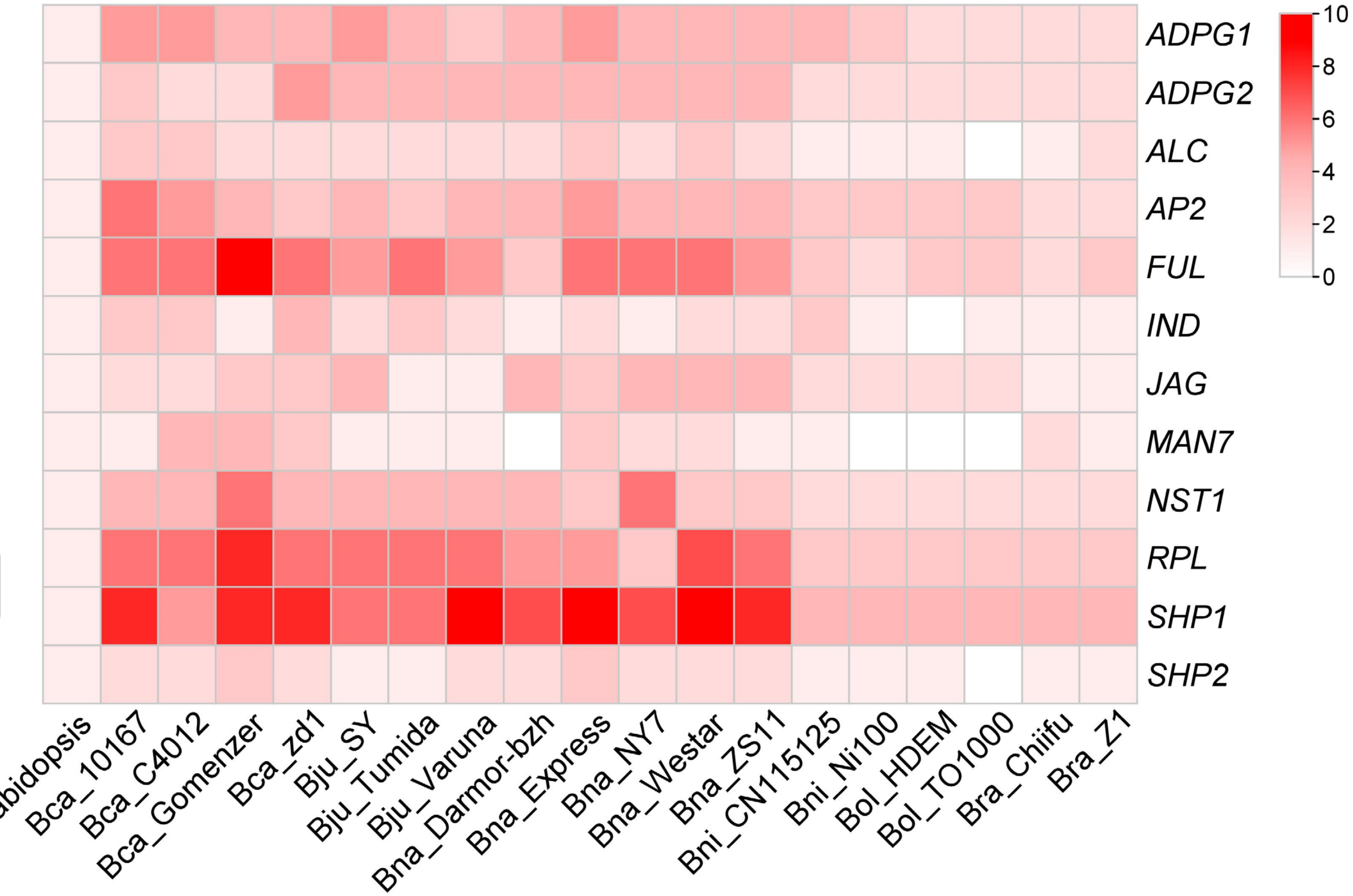
- ➢ The final pan-genome size of the *B. carinata* was ~ **2.52 Gb with 127,421 gene models.**
- ➢ 88,307 (69.42%) core genes were shared by all accessions, and 21,262 softcore, 16,852 shell and 792 cloud genes.



The genomic basis of PSR for *B. carinata*

B. carinata is reported to possess higher levels of **pod shatter resistance (PSR)** than other *Brassica* species.

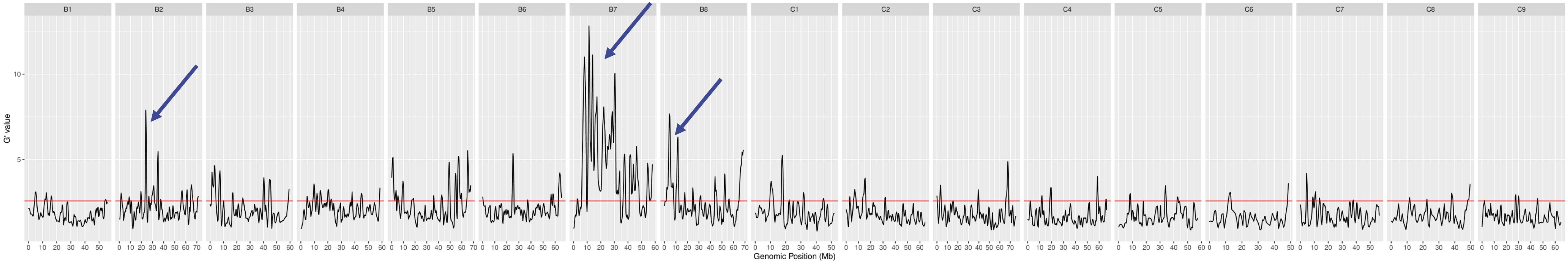


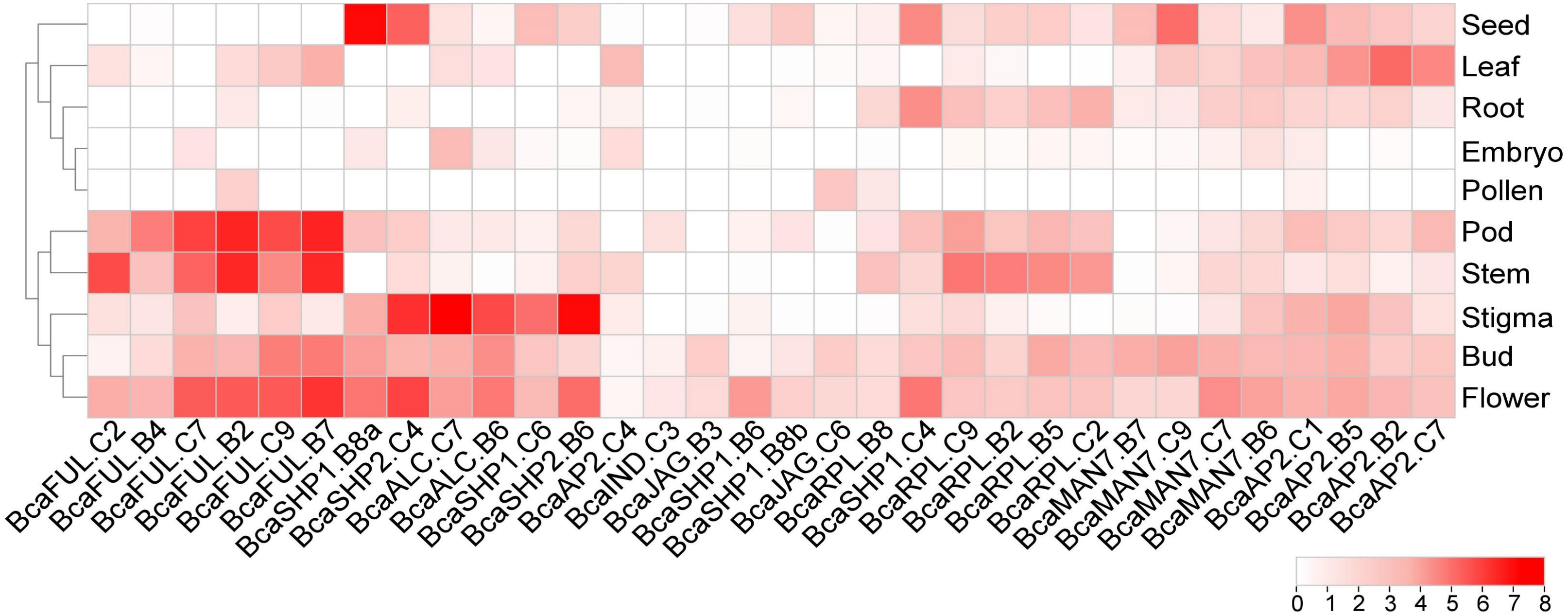


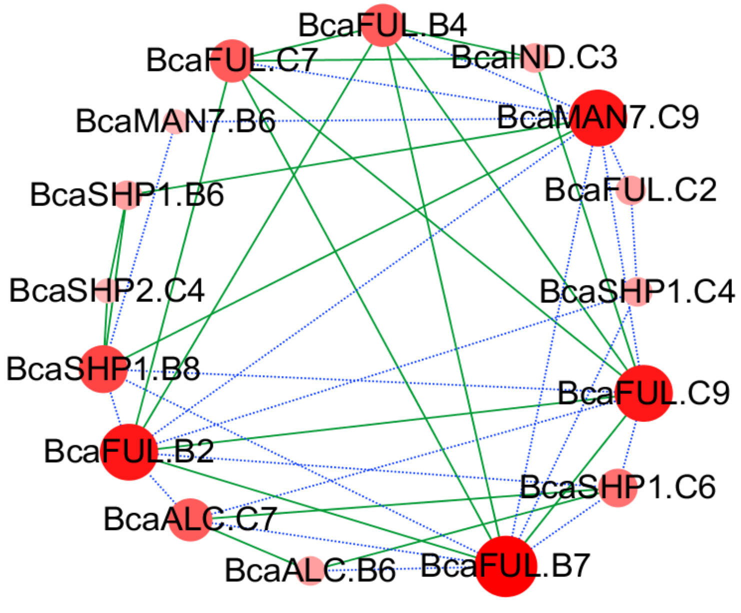
(Raman et al., 2014; Raman *et al.*, 2017)

The genomic basis of PSR for *B. carinata*

BcaFUL.B7 **was the hub gene of PSR**

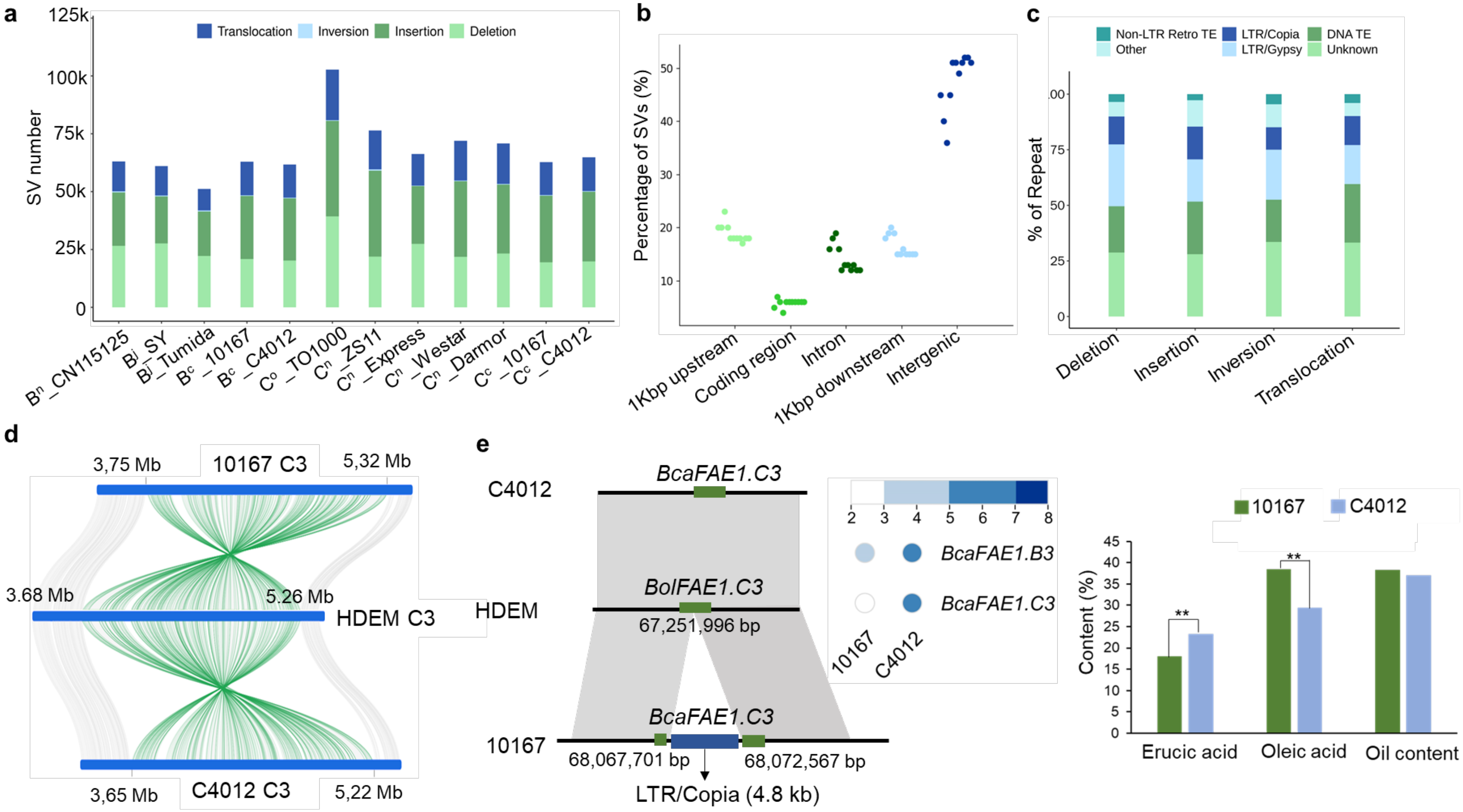






The species specificity of *B. carinata*

- 5,342 SNPs and 9,078 Indels unique to B^c compared to B^j and B^n.
- 4,209 SNPs and 3,981 Indels unique to C^c compared to C^o and C^n.
- Eight large inversions (>1 Mb) unique to *B. carinata.*



Summary

- Two high-quality *B. carinata* reference genomes were assembled with a total size at ~1.07 Gb and **contig N50 length at ~24 Mb**.
- The first *B. carinata* pan-genome **comprising 127,421 gene** models was constructed using 86 accessions.
- **One hub gene *BcaFUL.B7*** ssociated with PSR was detected by co-expression and QTL-seq.
- Species-specific genomic variations for *B. carinata* were identified.

HUAZHONG AGRICULTURAL UNIVERSITY 1898

ACKNOWLEDGEMENTS

CROP GENETIC IMPROVEMENT 1992 NKL

Prof. Jun Zou

Prof. Isobel A. P. Parkin

Prof. Harsh Raman

Prof. Ian Bancroft

Dr. Rosy Raman

Dr. Zhesi He

Prof. Jinling Meng

Dr. Shaoming Guo

NSFC

National Key R&D Program of China

UKRI
Biotechnology and Biological Sciences Research Council

GRDC
GRAINS RESEARCH & DEVELOPMENT CORPORATION