

Supporting tables

Table S1 Assembly statistics of each chromosome in the pak choi (subsp. *chinensis*) genome.

Chromosome	Number of Contigs	Length of Chromosome
A01	190	38,722,542
A02	33	32,748,408
A03	44	40,623,029
A04	24	22,460,690
A05	224	39,612,767
A06	167	38,313,918
A07	130	33,283,850
A08	19	22,567,373
A09	128	51,199,229
A10	43	22,103,602
Total	1,002	341,635,408

Table S2 Statistics of genome completeness in the three *B. rapa* genome assemblies according to BUSCO.

	Pak choi		Chinese cabbage (V3)		Yellow sarson	
	Proteins	Percentage (%)	Proteins	Percentage (%)	Proteins	Percentage (%)
Complete BUSCOs	1,412	98.1	1,410	97.9	1,407	97.7
Complete and single-copy BUSCOs	1,235	85.8	1,235	85.8	1,236	85.8
Complete and duplicated BUSCOs	177	12.3	175	12.1	171	11.9
Fragmented BUSCOs	11	0.8	9	0.6	11	0.8

Missing BUSCOs	17	1.1	21	1.5	22	1.5
Total BUSCO groups searched	1,440	100	1,440	100	1,440	100

Table S3 Statistics of repetitive sequences in the pak choi (subsp. *chinensis*) genome.

Type	RepeatMasker TEs		RepeatProteinMask TEs		De novo		Combined TEs	
	Length (bp)	Percentage (%)	Length (bp)	Percentage (%)	Length (bp)	Percentage (%)	Length (bp)	Percentage (%)
DNA	11,995,663	3.23	1,267,545	0.34	26,981,764	7.27	33,442,663	9.02
LINE	6,210,960	1.67	7,598,542	2.05	12,511,472	3.37	16,031,549	4.32
SINE	13,412	0	0	0	179,293	0.05	192,705	0.05
LTR	32,616,197	8.79	30,611,069	8.25	102,291,456	27.58	108,924,720	29.37
Other	57,585	0.02	0	0	12,224	0	69,809	0.02
Unknown	79,057	0.02	0	0	10,605,485	2.86	10,677,567	2.88
Total TE	50,667,969	13.66	39,475,145	10.64	147,094,599	39.66	166,904,453	45

Table S4 Statistics of centromere information on pak choi chromosomes.

Chr	CentBr		CRB		PCRBr		TR238	
	Start	End	Start	End	Start	End	Start	End
A01	23,267,167	24,628,674	23,247,268	24,639,006	15,704,697	23,158,859	15,571,023	23,313,371
A02	19,809,316	20,033,131	20,664,273	20,845,739	16,898,653	19,867,615	19,849,670	19,909,523
A03	33,848,340	36,152,144	26,079,575	35,103,509	35,968,421	37,245,732	35,957,038	36,148,406
A04	7,051,689	7,189,650	6,852,283	7,184,871	6,845,682	6,856,527	6,989,562	7,042,092
A05	13,690,592	24,796,193	14,008,078	26,401,734	13,222,612	23,578,707	13,698,021	23,561,497
A06	15,468,181	16,449,691	14,100,872	16,253,912	16,554,850	24,796,282	16,419,685	24,771,229
A07	4,894,166	11,526,942	6,331,973	11,577,111	6,464,703	9,655,267	6,441,841	9,666,172
A08	6,876,134	7,709,362	6,742,690	6,801,008	4,070,190	5,809,769	/	/
A09	25,024,368	27,656,897	25,062,604	25,200,299	20,847,237	26,809,796	20,842,115	26,816,206

Table S5 Summary of non-coding RNA (ncRNA) genes in the pak choi (subsp. *chinensis*) genome.

Type	Copy	Average length (bp)	Total length (bp)	Percentage in genome (%)	
miRNA	255	125.4	31,977	0.008622	
tRNA	1,599	75.1	120,140	0.032392	
rRNA	rRNA	4,495	524.2	2,356,562	0.635369
	18S	1,164	1340.1	1,559,873	0.420568
	28S	1,369	197.4	270,190	0.072848
	5.8S	857	449.6	385,282	0.103879
	5S	1,105	127.8	141,217	0.038074
	8S	0	0	0	0
snRNA	snRNA	1,062	110.9	117,813	0.031764
	CD-box	795	102.4	81,369	0.021938
	HACA-box	92	122.8	11,294	0.003045
	splicing	175	143.7	25,150	0.006781
	scaRNA	0	0	0	0

Table S6 Statistics of gene annotation completeness in the three *B. rapa* genome assemblies according to BUSCO.

	Pak choi		Chinese cabbage (V3)		Yellow sarson	
	Proteins	Percentage (%)	Proteins	Percentage (%)	Proteins	Percentage (%)
Complete BUSCOs	1,409	97.9	1,400	97.2	1,391	96.6
Complete and single-copy BUSCOs	1,202	83.5	1,213	84.2	1,204	83.6

Complete and duplicated BUSCOs	207	14.4	187	13	187	13
Fragmented BUSCOs	10	0.7	18	1.2	13	0.9
Missing BUSCOs	21	1.4	22	1.6	36	2.5
Total BUSCO groups searched	1,440	100	1,440	100	1,440	100

Table S7 Pak choi genes with large-effect mutations compared with the Chinese cabbage and yellow sarson genomes.

This table is provided as Microsoft Excel spreadsheet.

Table S8 Pak choi genes with large structure variations compared with the Chinese cabbage and yellow sarson genomes.

This table is provided as Microsoft Excel spreadsheet.

Table S9 Summary of PAV genes in the Chinese cabbage, yellow sarson and pak choi genomes.

This table is provided as Microsoft Excel spreadsheet.

Table S10 SV and PAV genes related to flowering in the pak choi and yellow sarson genomes.

Genes	<i>B. rapa</i> gene ID	Arabidopsis ID	Annotation	Variation Type
<i>BrAP1A</i>	<i>BraA02g01908P</i>	<i>AT1G69120</i>	Floral homeotic protein APETALA 1 A	SV: Premature stop codon
<i>BrBBX19</i>	<i>BraA01t01996Z</i>	<i>AT4G38960</i>	B-box zinc finger protein 19	PAV: PC absent
<i>BrBOFH</i>	<i>BraA06g02518P</i>	<i>AT5G61850</i>	Transcription factor BOFH	PAV: YS absent
<i>BrCSTF77</i>	<i>BraA08g02799P</i>	<i>AT1G17760</i>	Cleavage stimulation factor 77	PAV: PC absent
<i>BrEMF1</i>	<i>BraA03g00457P</i>	<i>AT5G11530</i>	EMBRYONIC FLOWER 1	SV: Premature stop codon, with 3n+-1 indel in CDS

<i>BrEMF2</i>	<i>BraA02g01601P</i>	<i>AT5G51230</i>	EMBRYONIC FLOWER 2	SV: Premature stop codon, Splice acceptor mutation, with 3n+-1 indel in CDS
<i>BrFLC2</i>	<i>BraA02g00340P</i>	<i>AT5G10140</i>	FLOWERING LOCUS C	SV: Premature stop codon
<i>BrFPA1</i>	<i>BraA09g03655P</i>	<i>AT1G27750</i>	Flowering time control protein FPA	SV: Premature stop codon
<i>BrFPA2</i>	<i>BraA09g04862P</i>	<i>AT2G43410</i>	Flowering time control protein FPA	SV: with 3n+-1 indel in CDS
<i>BrFRI</i>	<i>BraA10g00883P</i>	<i>AT4G00650</i>	Protein FRIGIDA	PAV: YS absent
<i>BrGAI</i>	<i>BraA03g02755P</i>	<i>AT4G02780</i>	Ent-copalyl diphosphate synthase	SV: Premature stop codon
<i>BrLFR</i>	<i>BraA05g02429P</i>	<i>AT3G22990</i>	LEAF AND FLOWER RELATED	SV: Premature stop codon
<i>BrLHY</i>	<i>BraA10g00082P</i>	<i>AT1G01060</i>	Protein LHY	SV: Premature stop codon
<i>BrMAF4</i>	<i>BraA06g02829P</i>	<i>AT5G65070</i>	MADS AFFECTING FLOWERING 4	PAV: PC absent
<i>BrPHYA</i>	<i>BraA09g06044P</i>	<i>AT1G09570</i>	Phytochrome A	SV: Premature stop codon, with 3n+-1 indel in CDS
<i>BrRGA1</i>	<i>BraA07g01654P</i>	<i>AT2G01570</i>	DELLA protein RGA1	PAV: YS absent
<i>BrSVP</i>	<i>BraA04g01624P</i>	<i>AT2G22540</i>	MADS-box protein SVP	SV: Premature stop codon
<i>BrTOE2</i>	<i>BraA02g01041P</i>	<i>AT5G60120</i>	Ethylene-responsive transcription factor TOE2	SV: Premature stop codon, with 3n+-1 indel in CDS

Table S11 Primers used for qRT-PCR assays.

Gene	Forward Primer	Reverse Primer	Fragment size
<i>GAPDH</i>	CAGGTTTGGAAATTGTCGAGG	GAGCTGTGGAAGCACCTTTC	175 bp
<i>BrMAF4</i>	TGTTCTCATCGTTTCCTCTACC	GCATGTTGTACTIONCGTAATGCT	93 bp