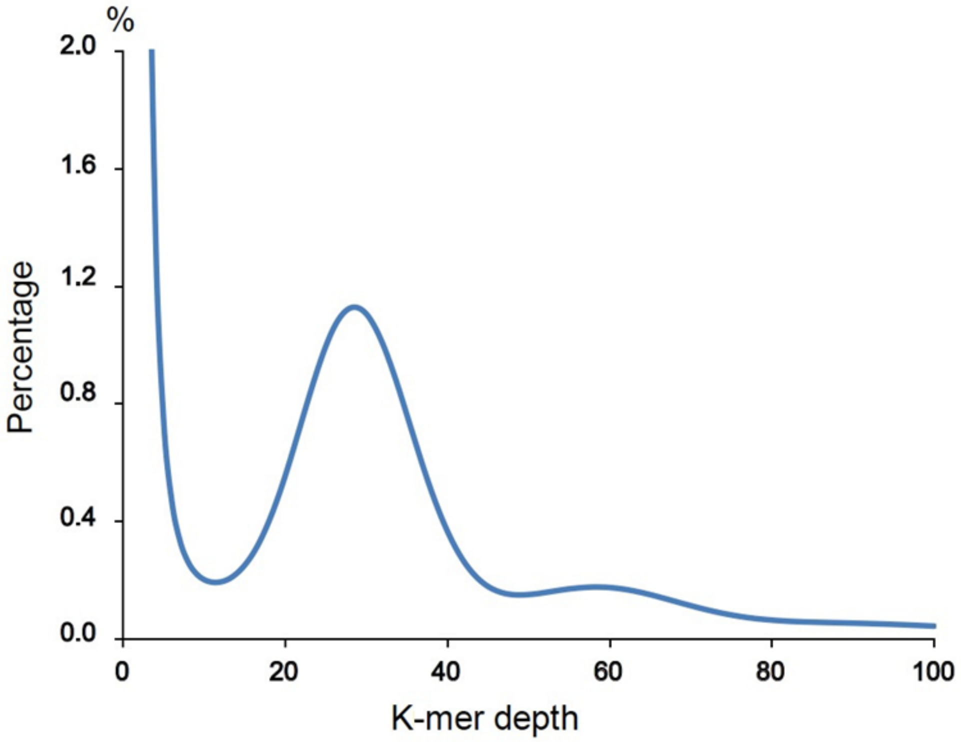
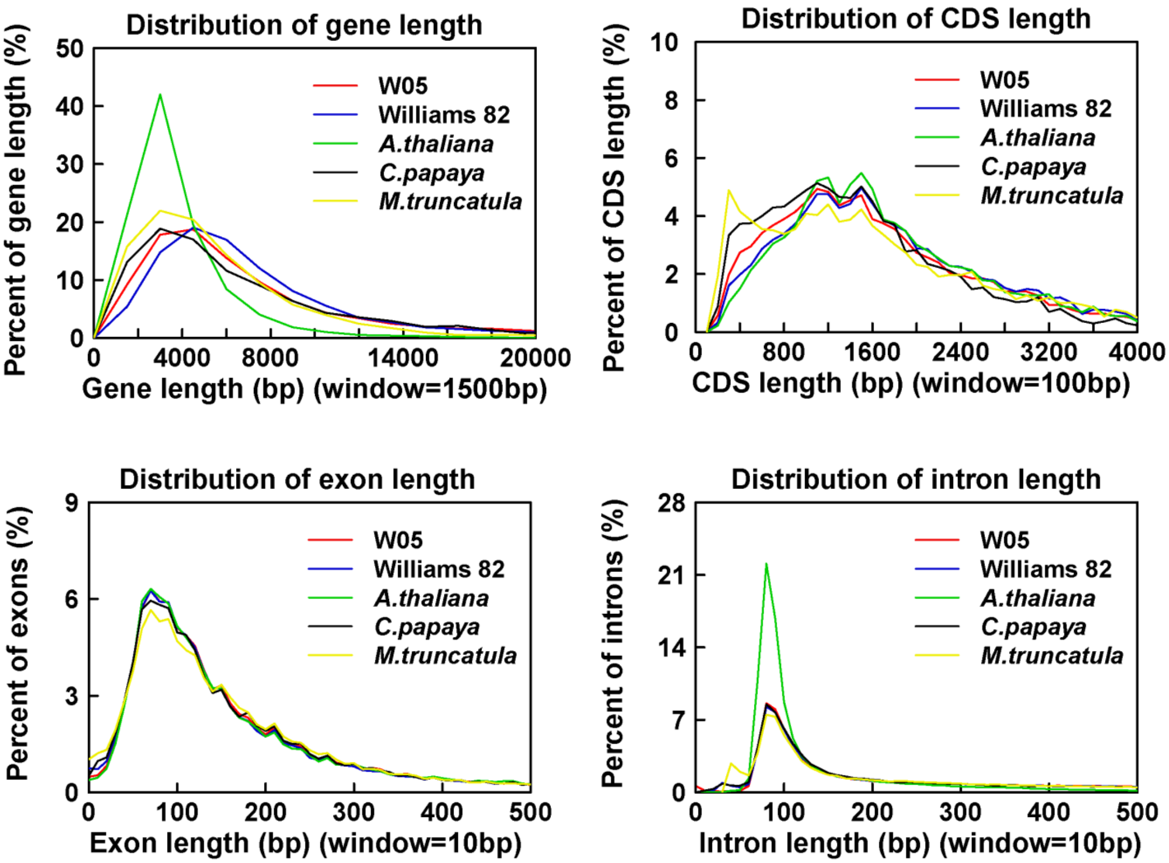
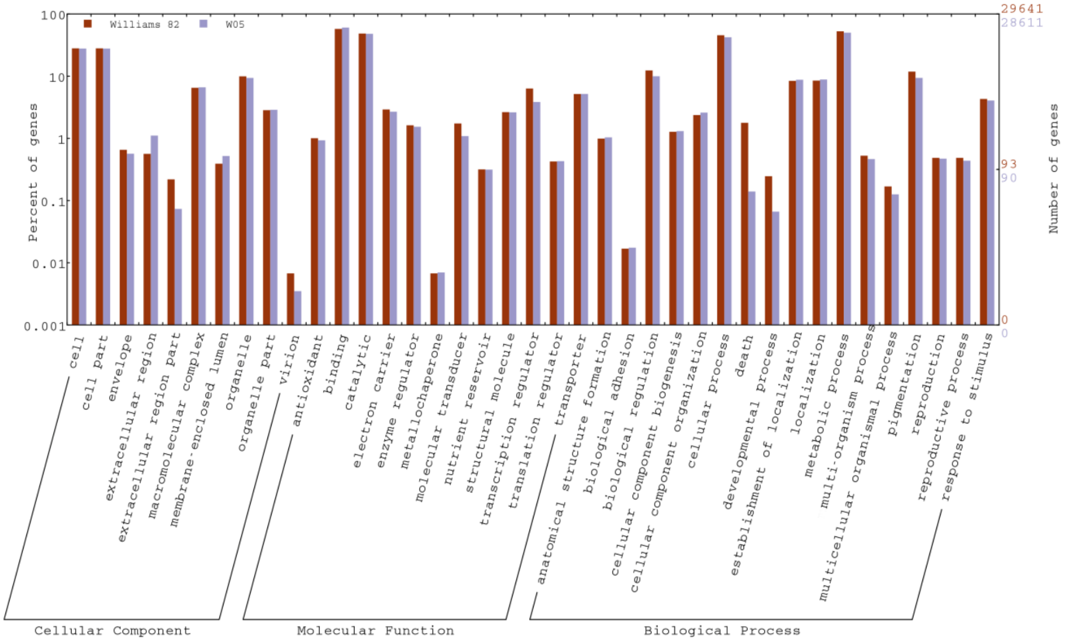
Supplementary Information



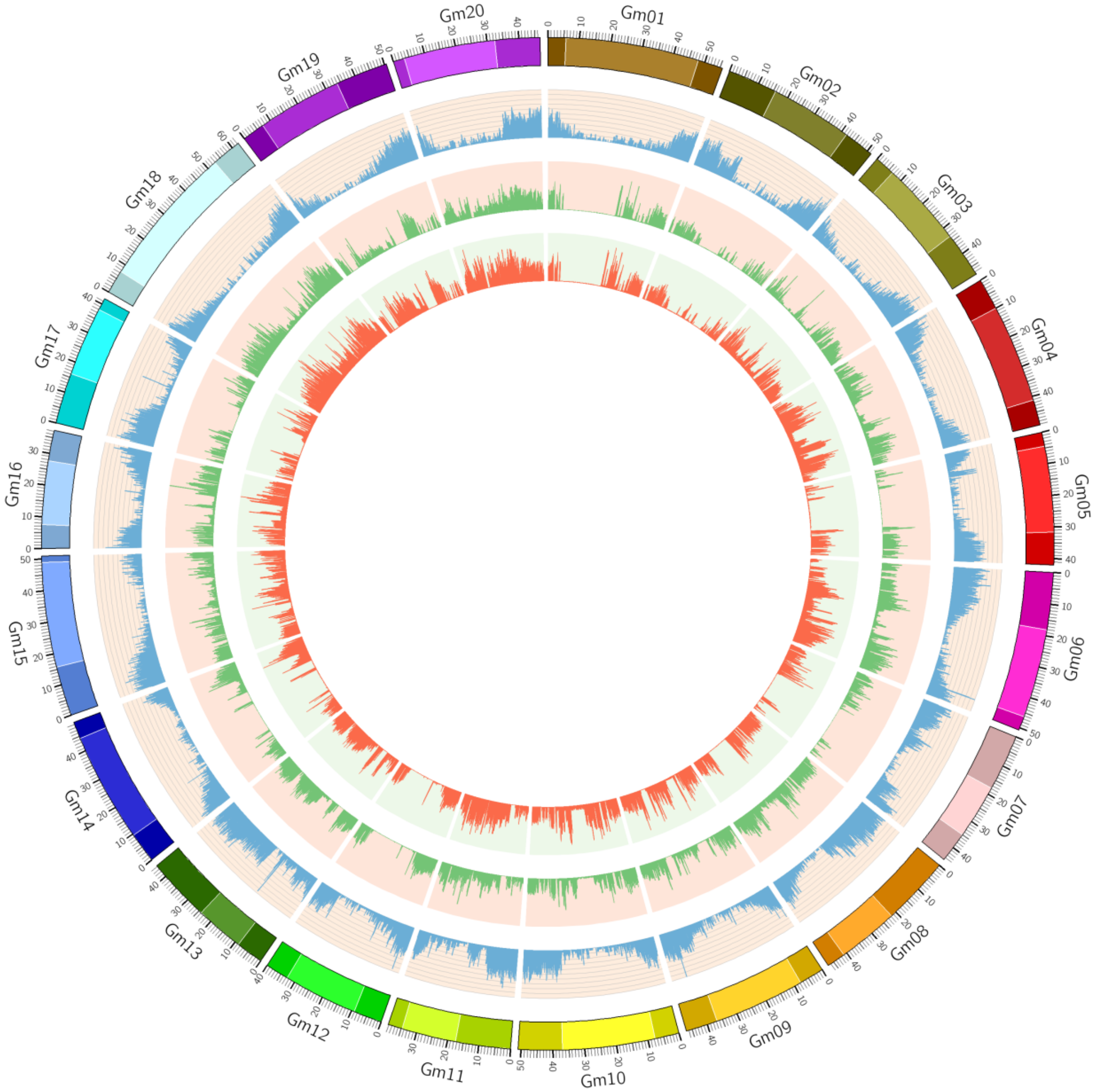
Supplementary Figure 1. K-mer analysis for estimating the genome size. The genome size of W05 is estimated by K-mer analysis using the Lander-Waterman algorithm[1], in which G = N×(L-K+1)/D, where N is the total number of reads, L is the average length of reads, and K is equal to 17 in this case. This method is widely used to estimate the approximate genome size of plant species[2]. To reduce inaccuracy caused by sequencing errors and duplicated regions in the genome, only clean reads from short insert-size libraries are used. The detailed statistics of the K-mer analysis are as follows: 17-Mer number: 33,792,138,754; Peak depth: 29; Estimated genome size: 1,165,246,163 bp; Total base pair used: 44,112,249,666.



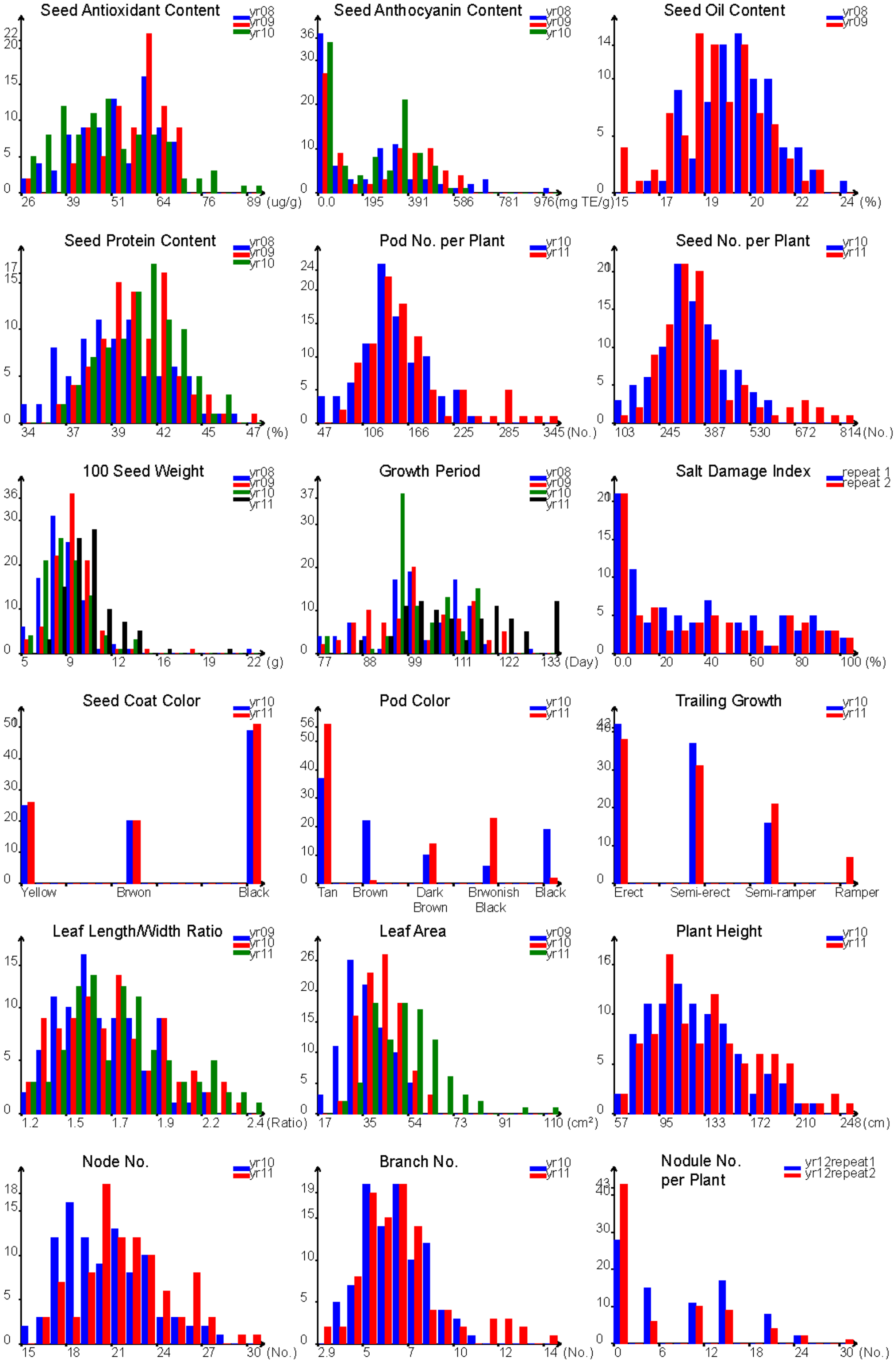
Supplementary Figure 2. Basic information of the gene model. Basic information including predicted gene length, CDS length, exon length, and intron length are collected and compared among the wild soybean accession W05 (this study), the cultivated soybean reference genome Williams 82 (Glyma1.1), *Medicago truncatula* (a close relative of alfalfa (*Medicago sativa*)), *Carica papaya* (papaya), and *Arabidopsis thaliana*.



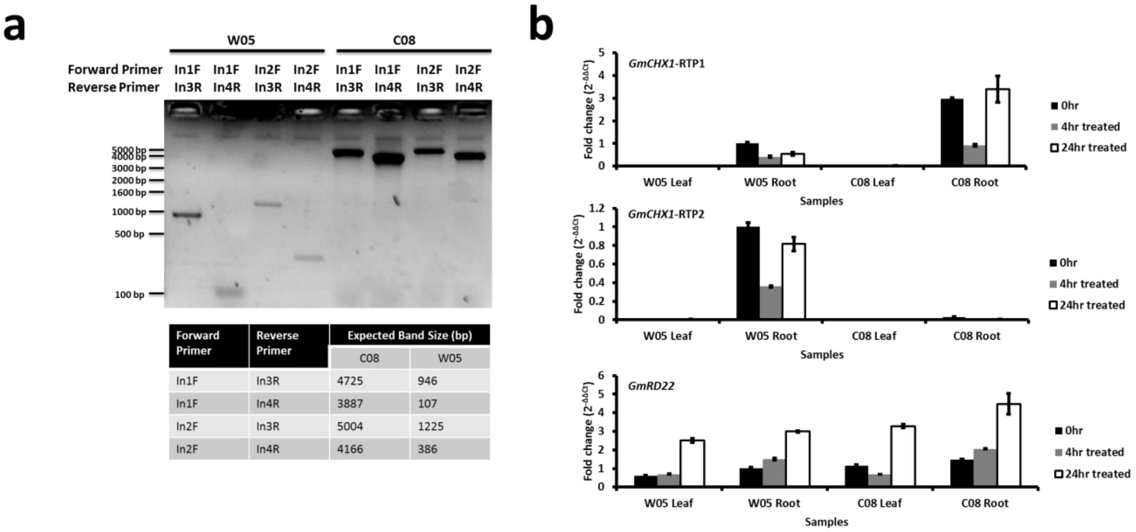
Supplementary Figure 3. Comparison of GO categories between W05 and Williams 82 annotated gene sets.



Supplementary Figure 4. SNP distribution within the soybean genome. SNP distribution is mapped to the reference genome Williams 82. Outermost circle: darker regions indicate the homo-chromatin regions located at the chromosome ends[3]; second circle: gene density; third circle: identified SNPs between the two parental lines (W05 and C08); and innermost circle: identified SNPs in the core panel of RI lines.

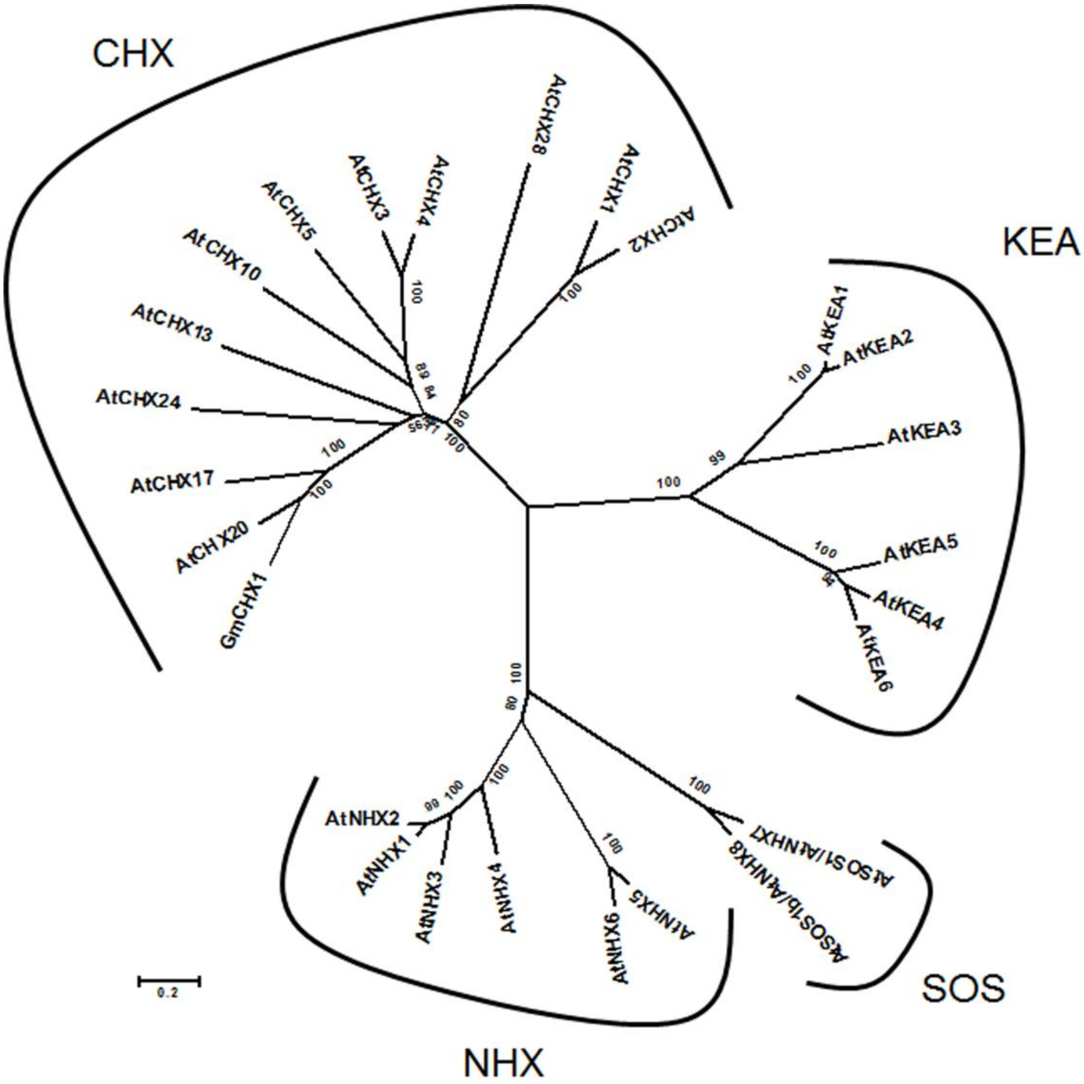


Supplementary Figure 5. Phenotype distributions of 96 core-panel RI lines and the two parents. Different colors indicate records based on data collected in different years (yr08: 2008, yr09: 2009, yr10: 2010, yr11: 2011, yr12: 2012).

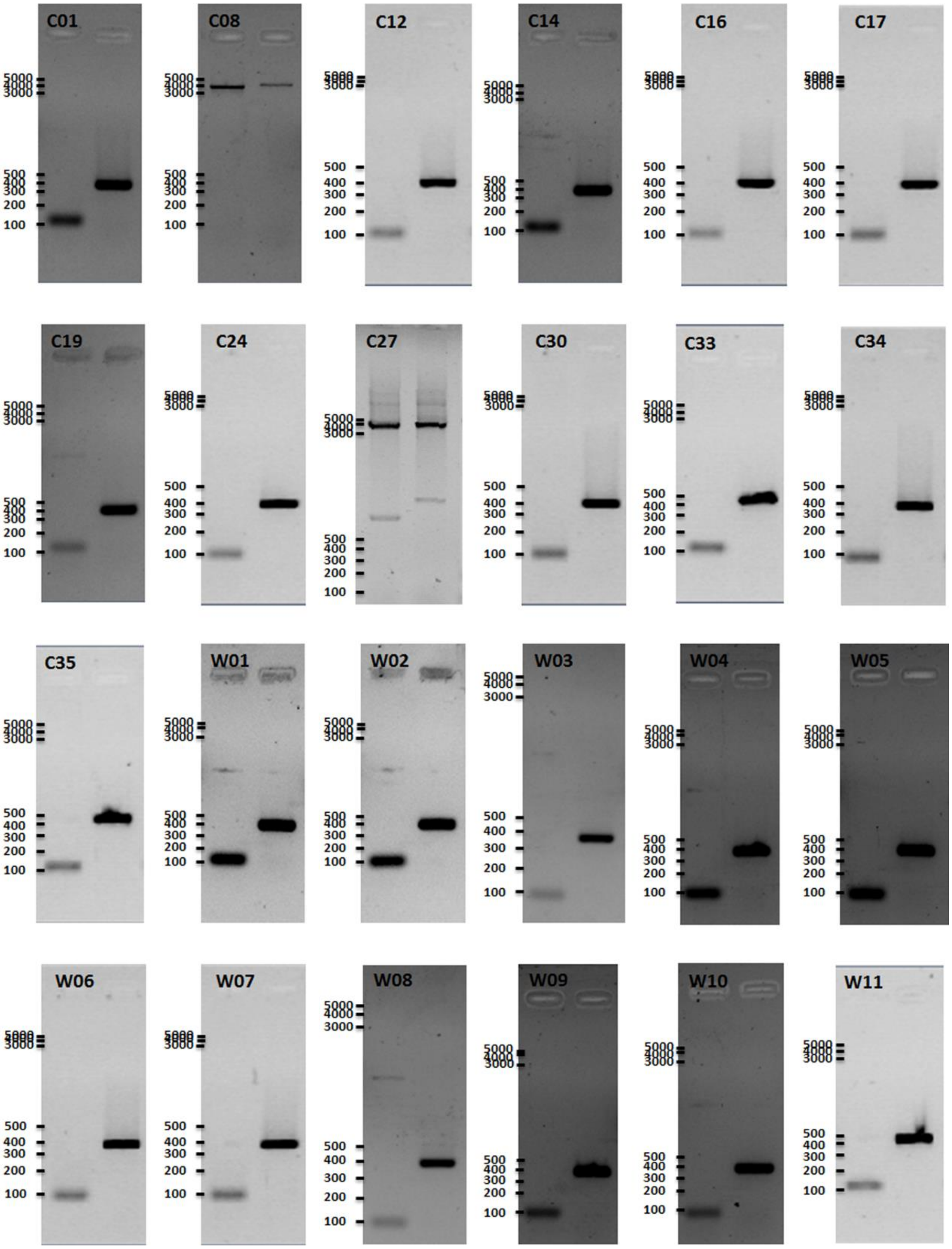


Forward Primer	Reverse Primer	Expected Band Size (bp)	
		C08	W05
In1F	In3R	4725	946
In1F	In4R	3887	107
In2F	In3R	5004	1225
In2F	In4R	4166	386

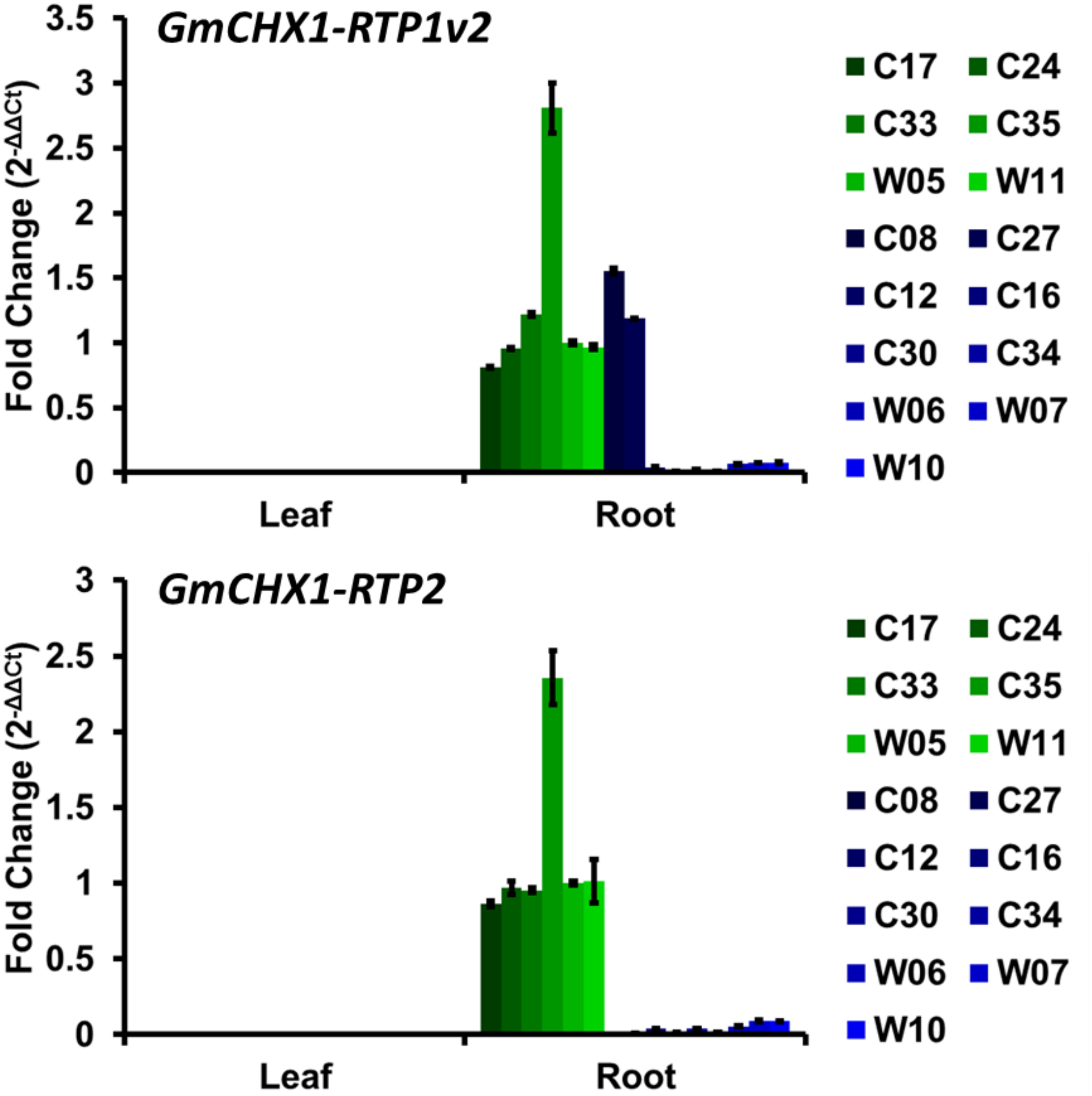
Supplementary Figure 6. An insertion of retrotransposon disrupted the *GmCHX1* gene in C08. (**a**) The insertion in the *GmCHX1* gene of C08 is shown by the larger PCR products in C08 compared to W05. Primer binding sites are marked in **Figure 2d** of the main text. (**b**) Real-time PCR studies of *GmCHX1* using primers specific to regions upstream (upper panel) or downstream (middle panel) of the insertion site depicted in C08, with or without NaCl treatment. Expression of the salt-inducible *GmRD22* (lower panel) was used as the quality control of RNA samples. The $2^{-\Delta\Delta CT}$ method[4] is used to compare gene expression, using the soybean α-tubulin gene (*Glyma20g27280*)[5] for normalization. The expression of *GmCHX1* in the root of W05 at 0 h is set to 1 for comparison. Error bar: standard errors. N=4. The sequences of the primers used are listed in **Supplementary Data File 2**. A lower extension temperature protocol for A-T rich templates is adopted to amplify the retrotransposon insertion[6], with the following modifications. In brief, PCR is conducted in a 25 µL reaction mixture containing 1X buffer, 1.8 mM MgCl$_2$, 0.2 mM dNTPs, 0.1 µM of each primer, 200 ng genomic DNA template and 0.75 U GoTaq® DNA polymerase. The thermal cycler is programmed with an initial 95°C for 2 min, followed by 40 cycles of 95 °C for 30 s and 60 °C for 6 min. PCR products are resolved on 1% agarose gel in 1X TAE buffer.



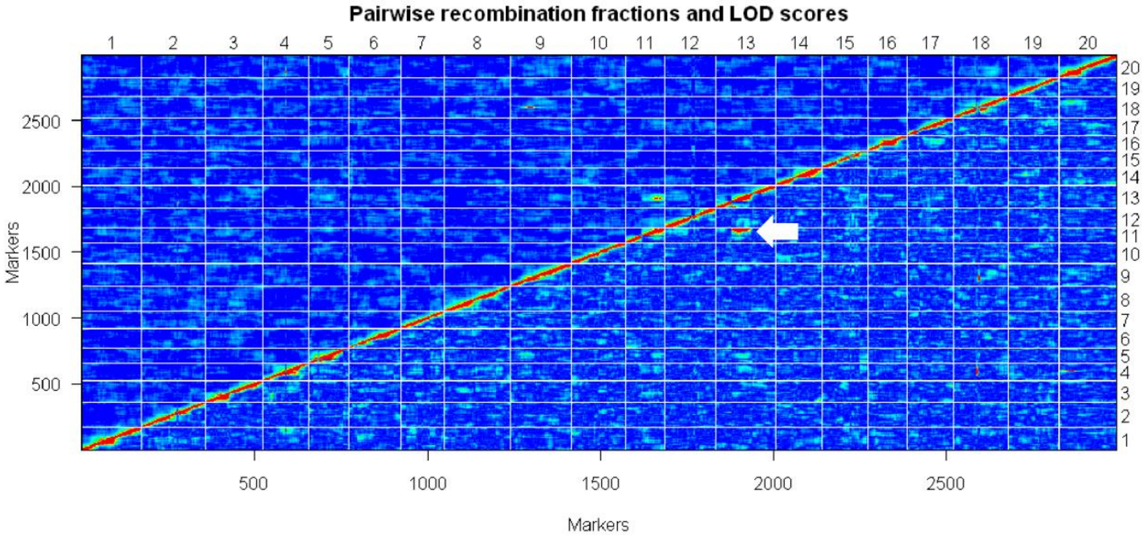
Supplementary Figure 7. Phylogenetic analysis of selected cation proton antiporters in plants. The amino acid sequences of selected proteins are aligned using the MEGA5 program[7]. GenBank/UniProtKB/Swiss-Prot accession numbers are as follows: AtCHX1:Q9SA37.1; AtCHX2:Q9SAK8.1; AtCHX3:Q9FFB8.1; AtCHX4:Q9FYC1.1; AtCHX5:Q3EDG3.1; AtCHX10:Q58P69.2; AtCHX13:O22920.1; AtCHX17:Q9SUQ7.1; AtCHX20:Q9M353.1; AtCHX24:Q1HDT2.2; AtCHX28:Q8L709.2; AtKEA1:Q9ZTZ7.2; AtKEA2:AEE81911.1; AtKEA3:Q9M0Z3.2; AtKEA4:Q9ZUN3.2; AtKEA5:Q8VYR9.1; AtKEA6:B5X0N6.1; AtSOS1/AtNHX8:Q9LKW9.1; AtSOS1b/AtNHX8:Q3YL57.1; AtNHX1:Q68KI4.2; AtNHX2:Q56XP4.2; AtNHX3:Q84WG1.2; AtHNX4:Q8S397.2; AtNHX5:Q8S396.2; AtNHX6:Q8RWU6.3.



Supplementary Figure 8. Detection of Ty1/copia retrotransposon insertion in *GmCHX1* of 24 soybean germplasms using PCR. The primers In1F and In2F are used as forward primers and In8F is used as a reverse primer as shown in **Figure 2d** in the main text. The specific sizes of bands with and without insertion can be found in **Supplementary Figure 6**.



Supplementary Figure 9. Additional expression studies of *GmCHX1* in characterized soybean germplasms. Real-time PCR of the *GmCHX1* gene using primers specific to regions upstream or downstream of the retrotransposon insertion site in C08, with RNA from salt-tolerant (green bars) and salt-sensitive germplasms (blue bars). N=3. Error bar: standard errors.



Supplementary Figure 10. Recombination fractions and LOD scores of the whole soybean genome. Upper left triangle: estimated recombination fractions for all markers; lower right triangle: LOD scores. Red indicates an LOD score ≥ 10 or a small recombination fraction; blue indicates an LOD score < 10 or a large recombination fraction. The white arrow indicates a recombination hot spot which may involve translocation events[8].

Supplementary Table 1. Sequencing libraries and statistics of clean reads.

Insertion size (bp)	**Total data (Gb)**	**Read length (bp)**	**Sequencing depth (X)**
180	4.6	75	3.95
260	27.4	75	23.52
326	11.4	75	9.79
817	8.0	75/44	6.87
2k	4.1	44	3.52
2.3k	9.5	44	8.15
6k	2.3	44	1.97
10k	6.2	44	5.32
Total	73.5	-	63.09

Supplementary Table 2. Statistics of SOAPdenovo-assembled soybean genome.

	Contig		Scaffold	
	Size(bp)	**Number**	**Size(bp)**	**Number**
N90	3,427	40,973	43,843	3,281
N80	7,996	25,959	88,676	1,910
N70	12,865	18,043	154,429	1,171
N60	18,132	12,762	243,754	720
N50	24,167	8,897	401,309	442
Longest (bp)	295,587		7,596,757	
Total Length (bp)	808,678,523		868,000,813	
Total Number (>200 bp)	129,013		34,137	
Total Number (>2k bp)	51,178		9,008	
Length of Ns (Mb)	-		59	

Supplementary Table 3. Statistics of short reads successfully mapped back to the assembled scaffolds.

Insertion Size (bp)	Reads Number	Mapped (%)	Properly paired (%)
180	61,187,966	94.16	89.29
260	364,899,960	93.07	82.58
326	160,013,874	94.44	82.19
817	156,556,972	95.42	79.25

Supplementary Table 4. Statistics of the gene model prediction.

Gene set		Number	Average gene length (bp)	Average CDS length (bp)	Average no. of exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	**Augustus**	63,034	2485.51	1035.03	4.80	215.73	381.93
	GlimmerHMM	69,296	2144.68	847.65	3.91	216.75	445.61
Homology -based	***A. thaliana***	36,183	3602.32	1245.63	5.65	220.46	506.80
	C. papaya	48,928	2526.80	971.26	4.23	229.56	481.44
	M. truncatula	58,762	2457.57	946.45	3.98	237.77	507.00
	G. max	65,599	2882.28	1056.39	4.92	214.63	465.57
EST(NCBI)		62,638	1435.87	632.29	3.11	203.18	380.48
RNA-Seq		252,055	3031.92	1270.02	4.83	262.75	459.60
Final gene set		52,395	3398.43	1083.93	4.87	222.59	503.55

Integrated gene model strategy including *de novo* prediction (programs: Augustus, Glimmer-HMM), homologue searching (*A. thaliana*, *C. papaya*, *M. truncatula* and *G. max*), RNA-seq (W05) data supporting and EST (*G. soja* and *G. max* from NCBI database) data supporting. All gene models are combined by GLEAN to give a final gene set.

Supplementary Table 5. Functional annotation of W05 genes.

	Number	**Percentage**
Total	52,395	100
Annotated	49,560	94.59
Swissprot	36,827	70.29
TrEMBL	49,475	94.43
InterPro	38,505	73.49
KEGG	28,781	54.93
GO	28,611	54.61
Un-annotated	2,835	5.41

Supplementary Table 6. Rate of Trinity-assembled contigs mapping back to the annotated genome.

Total no. of best CDS candidates	**Mapped to W05 annotated genome (No. (%))**
32,102	28,973 (90%)

Supplementary Table 7. Core Eukaryotic Genes Mapping Approach (CEGMA) analysis.

		No. of Core Proteins Identified	**Completeness**
Williams 82	**Complete**	224	90.32%
	Partial	243	97.98%
W05	**Complete**	215	86.69%
	Partial	240	96.77%

Supplementary Table 8. Statistics of ncRNA annotated in the W05 genome.

Type		Copy	Average length (bp)	Total length (bp)	Percentage in genome
miRNA		**376**	**122.22**	**45,953**	**0.00529**
tRNA		**864**	**75.11**	**64,898**	**0.00748**
rRNA	**Total**	**100**	**133.12**	**13,312**	**0.00153**
	18S	28	192.96	5,403	0.00062
	28S	16	119.63	1,914	0.00022
	5.8S	3	113.33	340	0.00004
	5S	53	106.70	5,655	0.00065
snRNA	**Total**	**437**	**106.17**	**46,398**	**0.00535**
	CD-box	301	94.08	28,318	0.00326
	HACA-box	45	120.53	5,424	0.00062
	splicing	91	139.08	12,656	0.00146

Supplementary Table 9. Statistics of repeat contents in the W05 genome.

Repeat Class		**Repeat length**	**Percentage**
DNA transposon	**Total**	**57,532,948**	**6.63**
	CMC	9,224,152	1.06
	hAT	5,479,340	0.63
	MULE	8,726,061	1.01
	TcMar	388,955	0.04
	CACTA	3,063,982	0.35
	Mutator	12,347,579	1.42
	Helitron	2,763,885	0.32
	other	15,538,994	1.79
LTR	**Total**	**268,111,653**	**30.89**
	Gypsy	163,812,079	18.87
	Copia	83,752,475	9.65
	Caulimovirus	1,992,703	0.23
	ERV	121,687	0.01
	other	18,432,709	2.12
Non-LTR	**LINE**	**16,640,627**	**1.92**
	L1	13,605,914	1.57
	L2	84,907	0.01
	RTE	2,573,153	0.30
	other	376,653	0.04
	SINE	**1,204,272**	**0.14**
	tRNA-Like	888,579	0.10
	other	315,693	0.04
Low complexity		**3,218,693**	**0.37**
Simple repeat		**12,984,986**	**1.50**
Satellite		**1,875,157**	**0.22**
Other		**346,855**	**0.04**
Novel/Unknown		**14,841,785**	**1.71**
Total		**376,756,976**	**43.41**

Supplementary Table 10. Measurement methods for different agronomic traits.

Agronomic Traits [a]		**Wild Parent (W05)**	**Cultivated Parent (C08)**	**Phenotypes recorded for 96 RI lines**
Seed Nutritional Quality	Seed Antioxidant Content [b]	Higher	Lower	Trolox equivalent antioxidant capacity (TEAC) assay was used. In brief, antioxidant from 0.5 g soybean seed powder was extracted using HCl-acidified methanol. Antioxidant activities were determined by measuring the decrease in absorbance at 734 nm upon the reduction of $ABTS^+$ radicals by antioxidants in the sample. An antioxidation activity of trolox was used as the standard for quantitative measurement. Each data point represents the average value of 3 random samples from pooled seeds of each line; 3-year record.
	Seed Anthocyanin Content [c]	Higher	Lower	Total anthocyanin assay. In brief, anthocyanin was extracted from 0.5 g soybean seed powder using 80% methanol with 0.1 N HCl. Monomeric anthocyanin pigment contents were determined with the differential A_{520nm}-A_{700nm} at pH 1.0 (KCl buffer) and pH 4.5 (NaOAc buffer). Anthocyanin content was calculated as cyanidin-3-glucoside-equivalence using the formula: $(A \times MW \times DF \times 1000) \div (\varepsilon \times l)$ where A is (A_{520nm}-A_{700nm}) at pH 1.0 – (A_{520nm}-A_{700nm}) at pH 4.5, MW is the molecular weight of cyanidin-3-glucoside (449.2 g mol^{-1}), DF is the dilution factor and ε is the molar extinction coefficient of cyanidin-3-glucoside (26,900) and l is the length of light path. Each data point represents the average value of 3 random samples from pooled seeds of each line; 3-year record.
	Seed Oil Content [d]	Lower	Higher	AOAC Official Method was used. In brief, 2g of soybean seed powder was digested

				with HCl in 80^{o}C water bath. Fat/oil was extracted with ethanol:ether:petroleum ether mixture (10 ml:25 ml:25 ml). Fat/Oil content was determined by weight after the solvent was completely evaporated. Each data point represents the average value of 3 random samples from pooled seeds of each line; 2-year record.
	Seed Protein Content [e]	Higher	Lower	AOAC Official Method was used. In brief, 2g of soybean seed powder was digested with H_2SO_4 at 420 ^{o}C. Proteins were converted to NH_3 by Kjeldahl catalyst. NH_3 was distilled and titrated using standardized HCl to estimate the protein content using the equation: $\frac{(V_{sample}-V_{blank})\times M\times A}{W\times 10}\times F$, where V_{sample} is the volume of standardized acid used to titrate a sample, V_{blank} is the volume of standardized acid to titrate the blank, M is the molarity of standardized acid, A is the atomic mass of N (14.01), W is the weight of sample, 10 is the percentage conversion factor, and F is the conversion factor from percentage nitrogen to protein content (6.25). Each data point represents the average value of 3 random samples from pooled seeds of each line; 3-year record.
Yield Related	Pod Number Per Plant	Higher	Lower	Number of pod per plant was counted at growth stage R8. Each data point represents the average value of 8 to 10 randomly selected plants for each line; 2-year record.
	Seed Number Per Plant	Higher	Lower	Seeds were harvested and counted at growth stage R8. Each data point represents the average value of 8 to 10 randomly selected plants for each line; 2-year record.
	100-Seed Weight	Lower	Higher	Seeds were harvested, air-dried and weighed at growth stage R8. Each data point represents the average value of 3 batches of 100 seeds randomly picked from a seed pool of 20 to 40 plants of each line; 3-year record.

Growth Period	Growth Period	Longer	Shorter	From the first day after sowing (June) to growth stage R8 (September to October); 3-year record.
Stress Tolerance	Salt Tolerance [f]	Tolerant	Sensitive	Salt damage index = $\frac{\sum CiNi}{SN} \times 100\%$; Ci: degree of damage, Ni: no. of plants with the corresponding degree of damage (scaled from 0 to 5, where 0 refers no observable damage and 5 refers to lethal damage to the terminal bud), N: Total no. of plants, S: highest degree of damage (= 5 in this scale). Each data point represents the average value of triplicates from 6 randomly selected young seedlings (with 2 primary leaves opened); two independent experiments were performed.
Coloration	Seed Coat Color	Black	Yellow	The major color of dried seeds: 1: yellow; 2: brown; 3: black; 2-year record.
	Pod Color	Black	Tan	Recorded at growth stage R8: 1: tan; 2: grayish brown; 3: brown; 4: dark brown; 5: black; 2-year record.
Plant Morphology	Trailing Growth	Trailing	Erect	Recorded at growth stage R4-R6: 1: erect; 2: semi-erect; 3: semi-trailing; 4: trailing; 2-year record.
	Leaf Length/Width Ratio	Larger	Smaller	The middle leaflets of a mature trifoliate located in the middle node were collected at growth stage R3-R5. The leaf length and leaf width were measured by Image-pro-plus (version 6.0.0); 3-year record.
	Leaf Area	Smaller	Larger	The middle leaflets of a mature trifoliate located in the middle node were collected at growth stage R3-R5. The leaf area was measured by Image-pro-plus (version 6.0.0); 3-year record.
	Plant Height	Longer	Shorter	Plant height was measured from first node (cotyledon) to the apex of plant at growth stage R8. Each data point represents the average of 8 to 10 randomly selected plants

				for each line; 2-year record.
	Node Number	Larger	Smaller	Each data point represents the average of the number of nodes on the main stem from 8 to 10 randomly selected plants at growth stage R8 for each line; 2-year record.
	Branch Number	Larger	Smaller	Each data point represents the average number of effective branches (branches with more than one mature pods and more than two nodes) from 8 to 10 randomly selected plants at growth stage R8 for each line; 2-year record.
Nodulation by *S. fredii* CCBAU25509	Nodule No. Per Plant	Nodule formed	No nodule	The Rhizobium strain *Sinorhizobium fredii* CCBAU25509 was cultured on yeast extract mannitol broth [9] for 48 h and then diluted to OD600 = 0.6. Surface-sterilized [10] seeds were grown in autoclaved vermiculite mixed with B&D medium [11]. Each soybean plant was inoculated with 1 ml of Rhizobium suspension. The number of root nodules formed on each plant was counted 4 weeks after inoculation.

[a] The plants for salt damage index recording and the nodulation experiment were performed in a greenhouse setting in the Chinese University of Hong Kong with the temperature ranging from 22-30 °C, and the humidity at 75-90% (2010 to 2012). For the other measurements, the 96 core RI lines and their parents were planted in the field at the experimental farm of Qing County (38° 35′ 0″ N, 116° 48′ 0″ E) in Hebei Province, China, during the growth season - June to October (2009 to 2011). For nutritional analyses, whole soybean seeds were powdered with a sample mill using a 0.5 mm screen (Foss, Model: 1093 Cyclotec).

[b] Methods were according to Miller et al (1996) [12].

[c] Methods were according to AOAC Official Method 2005.02.

[d] Methods were according to AOAC Official Method 922.06.

[e] Methods were according to AOAC Official Method 2001.11.

[f] Soybean seeds were germinated on vermiculite in a greenhouse. Seedlings were irrigated with half-strength Hoagland's solution with 1.8% (w/v) NaCl at the primary-leaf stage. The salt damage symptoms were recorded every 3 days. Salt damage index scores were assessed as previously described[13].

Supplementary Table 11. Summary of the re-sequencing data of 96 core RI lines.

RI line ID	Read No.	Mapping Rate (%)	Properly paired and uniquely mapped (bases)	Cov (%)	Heterozygosity
R01	13081398	86.46	502251125	52.87	0.002
R02	12032834	86.92	477793784	50.29	0.002
R03	12124934	86.85	481759841	50.71	0.002
R04	12211424	87.43	490979601	51.68	0.002
R05	12721766	87.09	493656995	51.96	0.002
R06	11463280	88.14	476766703	50.19	0.002
R07	14399146	86.25	529384433	55.72	0.002
R08	10689498	86.42	441926678	46.52	0.002
R09	14027912	86.95	525066179	55.27	0.002
R10	12966572	86.47	503569819	53.01	0.002
R11	13899200	87.37	524201664	55.18	0.003
R12	13056478	88.74	522338401	54.98	0.002
R13	14318894	87.9	536323427	56.46	0.003
R14	14309310	85.16	527464482	55.52	0.003
R15	14551542	86.35	538368852	56.67	0.004
R16	13161244	84.75	501212930	52.76	0.002
R17	14492334	88.18	546707953	57.55	0.002
R18	13867474	89.02	540222518	56.87	0.002
R19	12038962	86.96	479826084	50.51	0.002
R20	12535010	87.52	497469312	52.37	0.002
R21	13522782	86.97	512531573	53.95	0.002
R22	12574538	87.73	505145541	53.17	0.003
R23	13921542	87.45	525937850	55.36	0.002
R24	14260342	88.57	545494549	57.42	0.002
R25	16945956	86.3	582537377	61.32	0.003
R26	15237508	88.66	568480986	59.84	0.003
R27	13527038	86.71	517157929	54.44	0.003
R28	14076302	87.34	531267252	55.92	0.002
R29	15395946	88.4	575978017	60.63	0.003
R30	12477462	86.91	493878421	51.99	0.002
R31	12756208	82.75	477569910	50.27	0.002
R32	11531900	82.81	449357645	47.30	0.002

R33	12047300	82.84	456561294	48.06	0.002
R34	13900976	81.65	491809505	51.77	0.002
R35	13235946	82.51	488970546	51.47	0.002
R36	13095498	82.63	478033767	50.32	0.002
R37	15532164	83.03	542551279	57.11	0.002
R38	14669920	81.44	502826223	52.93	0.002
R39	12654884	82.07	464612987	48.91	0.002
R40	10913832	82.2	423653267	44.60	0.001
R41	10874032	81.64	421733649	44.39	0.002
R42	12102384	85.1	475530241	50.06	0.002
R43	12743300	85.96	500517894	52.69	0.002
R44	11719326	85.39	465418251	48.99	0.002
R45	13152174	85.81	506058932	53.27	0.002
R46	13851874	85.25	515828769	54.30	0.002
R47	12237926	85.22	475259015	50.03	0.002
R48	12609218	86.09	494708848	52.07	0.002
R49	13140184	85.06	496896443	52.30	0.002
R50	10700396	85.64	443208396	46.65	0.002
R51	11033998	85.04	441796028	46.50	0.002
R52	10161538	84.73	419344613	44.14	0.001
R53	11480034	84.99	458847929	48.30	0.002
R54	11411480	84.32	452098328	47.59	0.002
R55	11380134	86.48	471068218	49.59	0.002
R56	11072392	85.01	442359512	46.56	0.002
R57	13989886	84.78	515650083	54.28	0.002
R58	10907500	84.56	441932379	46.52	0.002
R59	10509924	84.54	435965241	45.89	0.002
R60	11011364	84.43	439088785	46.22	0.002
R61	11978824	86.01	481303735	50.66	0.002
R62	10862624	84.02	433807767	45.66	0.002
R63	8829392	83.63	369726381	38.92	0.001
R64	10972392	88.1	433037037	45.58	0.003
R65	9098954	89.11	383045512	40.32	0.002
R66	10444410	87.85	417797290	43.98	0.003
R67	9388098	89.69	397561539	41.85	0.002
R68	9658042	88.65	396518128	41.74	0.002
R69	10464038	88.21	415369040	43.72	0.003

R70	11575688	88.32	447032299	47.06	0.003
R71	10811610	87.44	431850195	45.46	0.003
R72	9189578	87.51	380202879	40.02	0.002
R73	13784458	89.35	495433511	52.15	0.004
R74	10582092	88.59	422671774	44.49	0.003
R75	14771240	84.92	531721929	55.97	0.002
R76	15780440	84.5	557476672	58.68	0.003
R77	14570446	85.83	538314766	56.66	0.002
R78	13599320	85.77	516344266	54.35	0.002
R79	15694070	85.76	553450220	58.26	0.003
R80	15198202	85.18	534976048	56.31	0.003
R81	15253650	86.3	553568836	58.27	0.002
R82	13886792	85.63	514718356	54.18	0.002
R83	14411672	86.62	537563494	56.59	0.002
R84	12572850	86.57	491660680	51.75	0.002
R85	12349628	85.13	478613566	50.38	0.002
R86	14141942	87.03	533875413	56.20	0.002
R87	13215086	86.59	506111255	53.27	0.002
R88	15879770	87.23	570823223	60.09	0.002
R89	15855932	86.34	560243008	58.97	0.003
R90	15545988	86.93	567973403	59.79	0.002
R91	14926364	87.06	550483665	57.95	0.002
R92	16771988	85.99	576087083	60.64	0.003
R93	12738494	86.95	504889676	53.15	0.002
R94	16239624	86.12	574249789	60.45	0.003
R95	13770322	85.74	511555412	53.85	0.002
R96	15404446	86.06	555113606	58.43	0.003

Supplementary Table 12. Detailed information on the previously identified QTLs and their putative causal genes

Agronomic Traits	**QTL Locus Identified in Previous Studies**	**Putative causal gene and its functions**	**References**
Growth Period	*E1 (E1)*	*E1* contains an intron-free gene that suppresses flowering and delays maturity.	14
	E2 (GmGIa)	A *GmGIa* gene, which is a homolog of *GIGANTEA* in *A. thaliana*, is identified as a candidate for the *E2* locus.	15
Salt Tolerance	*Ncl*	Not reported	16, 17, 18, 19
Seed Coat Color	*I* (*CHS* gene cluster)	The *I* locus on Chr08 contains several *CHS* genes (encoding chalcone synthase, a key biosynthetic enzyme for anthocyanin and proanthocyanidins). It was reported that the yellow seed coat color of cultivated soybeans was due to tissue-specific RNAi silencing of *CHS* genes.	20
Leaf Length/Width Ratio	*Ln* (*Gm-JAGGED1*)	*ln* locus on Chr20 contains the *Gm-JAGGED1* gene (homologous to the *JAG* gene in *Arabidopsis thaliana* which encodes a regulator of lateral organ development and fruit patterning).	21, 22
Nodule No. Per Plant	*Rj2* (*Rj2*)	*Rj2* locus that harbors a resistance gene involved in rhizobia infection specificity.	23

Supplementary Table 13. Correlation between selected traits. [a]

Year	**Trait**		**Kendall's tau_b**	**Spearman's rho**
2010	Seed Coat Color	Seed Anthocyanin Content	.662**	.823**
	Pod Color	Seed Antioxidant Content	.329**	.423**
	Pod Number Per Plant	Trailing Growth	.483**	.614**
	Pod Number Per Plant	Seed Number Per Plant	.822**	.954**
	Seed Number Per Plant	Trailing Growth	.487**	.621 **
2011	Pod Number Per Plant	Trailing Growth	.431**	.544**
	Pod Number Per Plant	Seed Number Per Plant	.861**	.969**
	Seed Number Per Plant	Trailing Growth	.447**	.562**

[a] Kendall’s tau_b and Spearman’s rho values were calculated using the Predictive Analytics Suite Workstation Statistics (PASW Statistics 18.0.0). ** $p < 0.001$.

Supplementary References

1. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* **2**, 231-239 (1988).
2. Zhang G, *et al.* Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnol* **30**, 549-554 (2012).
3. Du J, *et al.* Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. *Plant Cell* **24**, 21-32 (2012).
4. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-408 (2001).
5. Phang TH, Shao G, Liao H, Yan X, Lam HM. High external phosphate (Pi) increases sodium ion uptake and reduces salt tolerance of 'Pi-tolerant' soybean. *Physiol Plant* **135**, 412-425 (2009).
6. Su X-Z, Wu Y, Sifri CD, Wellems TE. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res* **24**, 1574-1575 (1996).
7. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739 (2011).
8. Findley SD, *et al.* A fluorescence *in situ* hybridization system for karyotyping soybean. *Genetics* **185**, 727-744 (2010).
9. JP V. A manual for the practical study of the root-nodule bacteria. *A manual for the practical study of the root-nodule bacteria*, 164 (1970).
10. Kereszt A, *et al.* *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nature Protoc* **2**, 948-952 (2007).
11. Imaizumi-Anraku H, Kawaguchi M, Koiwa H, Akao S, Syōno K. Two ineffective-nodulating mutants of *Lotus japonicus*—different phenotypes caused by the blockage of endocytotic bacterial release and nodule maturation. *Plant Cell Physiol* **38**, 871-881 (1997).
12. Miller NJ, Sampson J, Candeias LP, Bramley PM, Rice-Evans CA. Antioxidant activities of carotenes and xanthophylls. *FEBS Letters* **384**, 240-242 (1996).
13. Shao G, Song J. Preliminary studies on the evaluation of salt tolerance in soybean varieties. *Agr Sci China* **6**, 30-35 (1986).
14. Xia Z, *et al.* Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc Natl Acad Sci USA* **109**, E2155–E2164 (2012).

15. Watanabe S, *et al.* A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* **188**, 395-407 (2011).
16. Hall D, Evans AR, Newbury HJ, Pritchard J. Functional analysis of CHX21: a putative sodium transporter in Arabidopsis. *J Exp Bot* **57**, 1201-1210 (2006).
17. Hamwieh A, Tuyen DD, Cong H, Benitez ER, Takahashi R, Xu DH. Identification and validation of a major QTL for salt tolerance in soybean. *Euphytica* **179**, 451-459 (2011).
18. Hamwieh A, Xu D. Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. *Breeding Sci* **58**, 355-359 (2008).
19. Lee GJ, *et al.* A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. *Theor Appl Genet* **109**, 1610-1619 (2004).
20. Tuteja JH, Vodkin LO. Structural features of the endogenous silencing and target loci in the soybean genome. *Crop Sci* **48**, S-49-S-68 (2008).
21. Jeong N, Moon J-K, Kim H, Kim C-G, Jeong S-C. Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean. *Theor Appl Genet* **122**, 865-874 (2010).
22. Jeong N, *et al. Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* **24**, 4807-4818 (2012).
23. Yang S, Tang F, Gao M, Krishnan HB, Zhu H. R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proc Natl Acad Sci USA* **107**, 18735-18740 (2010).