Article title: The AP2/ERF transcription factor TOE4b regulates photoperiodic flowering and grain yield per plant in soybean

Haiyang Li[+], Haiping Du[+], Zerong Huang[+], Milan He, Lingping Kong, Chao Fang, Liyu Chen, Hui Yang, Yuhang Zhang, Baohui Liu*, Fanjiang Kong*, Xiaohui Zhao*

Supporting Information:

Supporting figures:

Figure S1 Genome-wide averaged distance of linkage disequilibrium decayed. Calculation via squared correlations of allele frequencies (r^2) against the distance between polymorphic sites in wild soybeans, landraces and cultivars.

Figure S2 Expression of 15 genes in the linkage disequilibrium block in the leaves and stem tips. (a) Data are from transcriptomic analysis in wild-type W82. The values represent RPKM value of genes in transcriptome. Data are presented as means ± SD of two biological replicates. (b) Data are from Phytozome 13 (https://phytozome-next.jgi.doe.gov/). FPKM (Fragments Per Kilobase of exon model per Million mapped reads).

Figure S3 Natural variation of *TOE4b*. (a) Summary of *TOE4b* haplotypes across a diversity panel comprising 1,744 soybean accessions. (b) Origin tree of the *TOE4b* haplotypes. The data are from 1,744 accessions, consisting of 221 wild soybeans, 516 landraces, and 874 improved cultivars. The red arrows indicate the direction of evolution.

Figure S4 *TOE4b*^{H4} variation affects one of the conserved AP2 domain. (a) Alignment of the amino acid sequences of TOE4b with its orthologs and homologs. Amino acid sequences from *Glycine max*, *Phaseolus vulgaris*, *Medicago truncatula*, *Arabidopsis thaliana* and *Oryza sativa* were aligned by Clustal W. The red asterisk indicates the amino acid in TOE4b^{H4} that changes from the conserved G residue to S. (b) CAPS marker for the identification of the *TOE4b*^{H1} and *TOE4b*^{H4} alleles. PCR products were digested with the restriction enzyme *Hae*III. Hetero, heterozygous plant.

Figure S5 Effects of combination of *TOE4b* and *E2* alleles on soybean flowering time.

(a, b) Flowering time of six allelic combinations between *TOE4b* and *E2* across 449 soybean accessions grown in Harbin in 2017 (a) and 2020 (b). The number above the *x*-axis represents the number of samples. Different lowercase letters indicate significant differences. The phenotype with 2 samples was not statistically analyzed with other samples.

Figure S6 Generation *TOE4b* overexpression lines in soybean. (a) Relative transcript levels of *TOE4b* in wild-type W82 and T1 overexpression transgenic plants. *β-Tubulin* (*TUB*) was used as a normalization transcript. (b) Abundance of TOE4b-3Flag in the leaves of T_1 transgenic plants, as determined by immunoblotting using anti-Flag antibodies. The non-specific band (NS) was used as a loading control.

Figure S7 Phenotypes of *TOE4b* mutant and overexpression lines. Agronomic traits of plant height (a, c, e), measured from the cotyledonary node of the main stem to the apex in centimeters; number of nodes (b, d, f). The plants were grown in an incubator under LD (16-h light/8-h dark) (a-b) or SD (12-h light/12-h dark) conditions with an average photon flux of 300 μmol m^{-2} s^{-1} (c-d). For the mutant plants were grown in an incubator under LD (16-h light/8-h dark) with an average photon flux of 450 μmol m^{-2} s^{-1} (e-f). W82, wild-type Williams 82. Five plants were scored for each phenotype; data are means ± SD. $P < 0.05$, as determined by multiple comparison testing by one-way ANOVA. Different lowercase letters indicate significant differences.

Figure S8 Agronomic traits associated with the major haplotypes of *TOE4b*. Agronomic trait of plant height (a), measured from the cotyledonary node of the main stem to the apex in centimeters; number of nodes (b); average of internode length (c); branch number per plant (d); pod number per plant (e); and grain weight per plant (f). $TOE4b^{H1}$, n = 68; $TOE4b^{H4}$, n = 83. The plants were grown in a standard field in Harbin under natural conditions in 2017. At least 6 plants were scored for each phenotype; data are means ± SD. $P < 0.05$, two-tailed Student's *t*-test. Different lowercase letters indicate significant differences.

Supporting tables:

Table S1 Significant loci associated with flowering time identified through GWAS.

Table S2 The $-\log_{10}(p)$ values of significant loci associated with flowering time for GWAS.

Table S3 Information of candidate genes within LD block.

Table S4 Transcription levels of the 15 genes in the leaves and stem tips.

Table S5 List of soybean accessions used for GWAS.

Table S6 Primers used in this study.

References

Bailey M, Ivanauskaite A, Grimmer J, Akintewe O, Payne AC, Osborne R, Labandera AM, Etherington RD, Rantala M, Baginsky S *et al*. 2021. The *Arabidopsis NOT4A* E3 ligase promotes *PGR3* expression and regulates chloroplast translation. *Nat Commun* **12**: 251.

Guan X, Okazaki Y, Zhang R, Saito K, Nikolau BJ. 2020. Dual-localized enzymatic components constitute the fatty acid synthase systems in mitochondria and plastids. *Plant Physiol* **183**: 517-529.

Hasi Q, Kakimoto T. 2022. ROP interactive partners are involved in the control of cell division patterns in *Arabidopsis* leaves. *Plant Cell Physiol* **63**: 1130-1139.

Lee TA, Bailey-Serres J. 2019. Integrative analysis from the epigenome to translatome uncovers patterns of dominant nuclear regulation during transient stress. *Plant Cell* **31**: 2573-2595.

Li B, Byrt C, Qiu J, Baumann U, Hrmova M, Evrard A, Johnson AA, Birnbaum KD, Mayo GM, Jha D *et al*. 2016. Identification of a stelar-localized transport protein that facilitates root-to-shoot transfer of chloride in arabidopsis. *Plant Physiol* **170**: 1014-29.

Lupanga U, Röhrich R, Askani J, Hilmer S, Kiefer C, Krebs M, Kanazawa T, Ueda T, Schumacher K. 2020. The *Arabidopsis* V-ATPase is localized to the TGN/EE via a seed plant-specific motif. *Elife* **9**: e60568.

Tan YF, O'Toole N, Taylor NL, Millar AH. 2010. Divalent metal ions in plant mitochondria and their role in interactions with proteins and oxidative stress-

induced damage to respiratory function. *Plant Physiol* **152**: 747-61.

Xu C, Min JR. 2011. Structure and function of WD40 domain proteins. *Protein Cell* **2**: 202-214.

Zhang A, Matsuoka K, Kareem A, Robert M, Roszak P, Blob B, Bisht A, De Veylder L, Voiniciuc C, Asahina M *et al*. 2022. Cell-wall damage activates DOF transcription factors to promote wound healing and tissue regeneration in *Arabidopsis thaliana*. *Curr Biol* **32**: 1883-1894.e7.

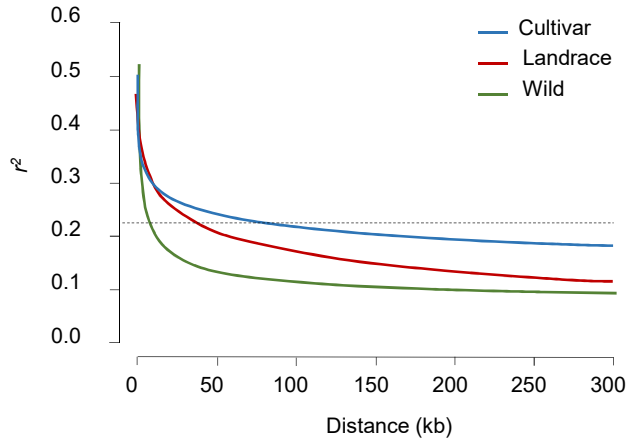

Figure S1 Genome-wide averaged distance of linkage disequilibrium decayed. Calculation via squared correlations of allele frequencies (r^2) against the distance between polymorphic sites in wild soybeans, landraces and cultivars.

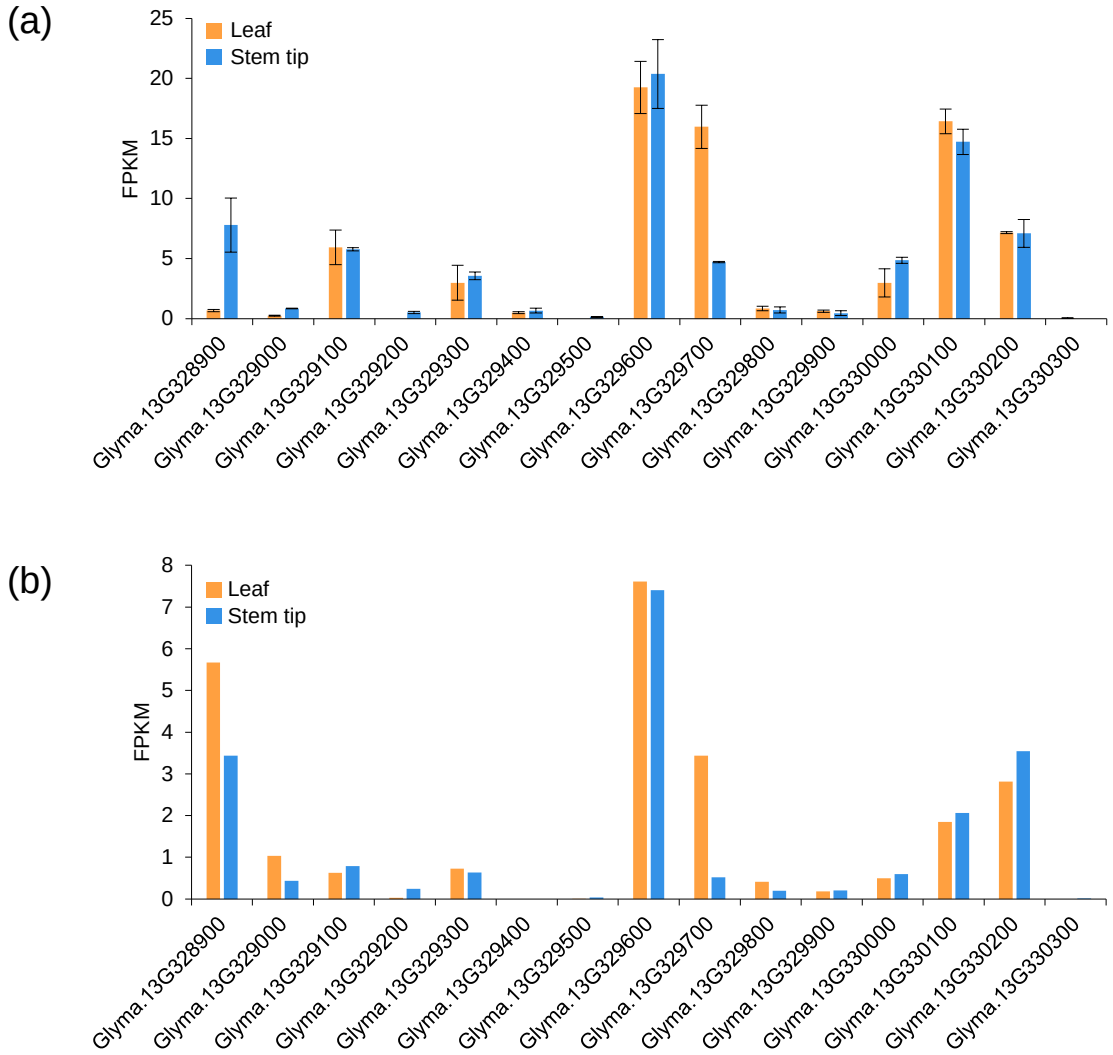

Figure S2 Expression of 15 genes in the linkage disequilibrium block in the leaves and stem tips. (a) Data are from transcriptomic analysis in wild-type W82. The values represent RPKM value of genes in transcriptome. Data are presented as means ± SD of two biological replicates. (b) Data are from Phytozome 13 (https://phytozome-next.jgi.doe.gov/). FPKM (Fragments Per Kilobase of exon model per Million mapped reads).

(a)

Reference Genome W82		Haplotype									Number
TOE4b	*H1*	G	T	G	C	G	C	C	A	C	1189
	H2	G	A	G	A	G	C	C	A	C	69
	H3	G	T	A	C	G	C	C	A	C	52
	H4	G	T	G	C	A	C	C	A	C	283
	H5	C	T	G	A	G	C	C	A	C	4
	H6	G	A	G	A	G	C	C	C	C	1
	H7	G	T	G	C	G	A	C	A	C	1
	H8	G	T	G	C	G	C	A	A	C	1
	H9	G	T	G	C	G	C	C	C	C	6
	H10	G	T	G	C	G	C	C	A	A	5
CDS nt # in W82		245	283	306	328	715	841	848	958	1086	
AA # in W82		82	95	102	110	239	281	283	320	362	
AA identified in W82		S	F	M	H	G	P	A	N	F	
AA change		T	I	I	N	S	T	D	H	L	

(b)

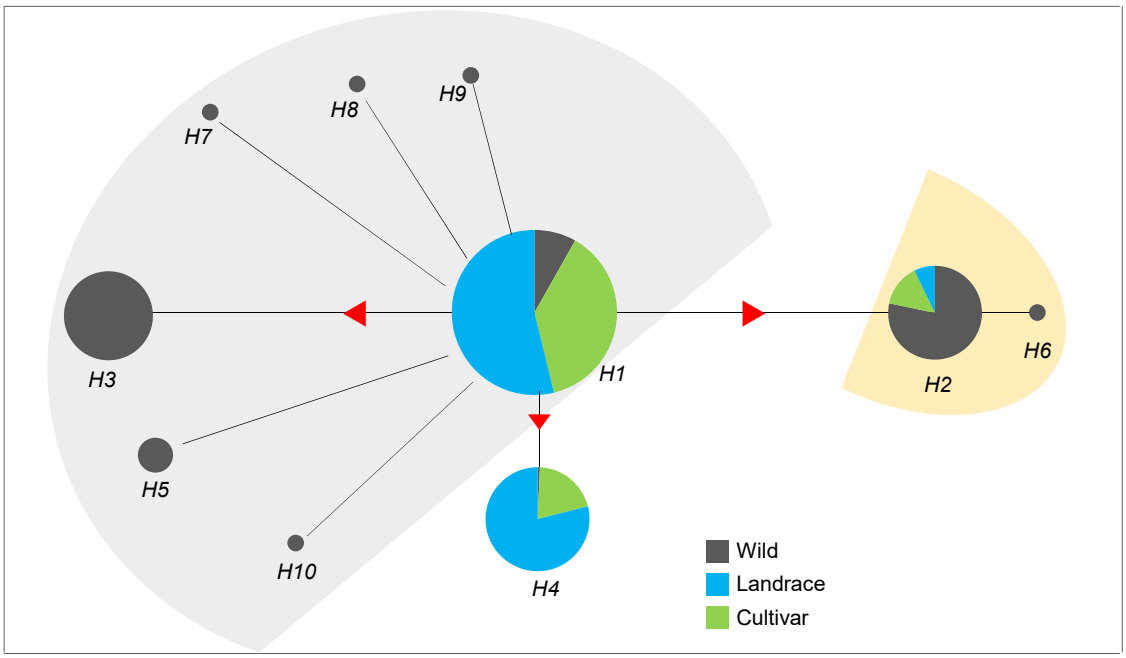

Figure S3 Natural variation of *TOE4b*. (a) Summary of *TOE4b* haplotypes across a diversity panel comprising 1,744 soybean accessions. (b) Origin tree of the *TOE4b* haplotypes. The data are from 1,744 accessions, consisting of 221 wild soybeans, 516 landraces, and 874 improved cultivars. The red arrows indicate the direction of evolution.

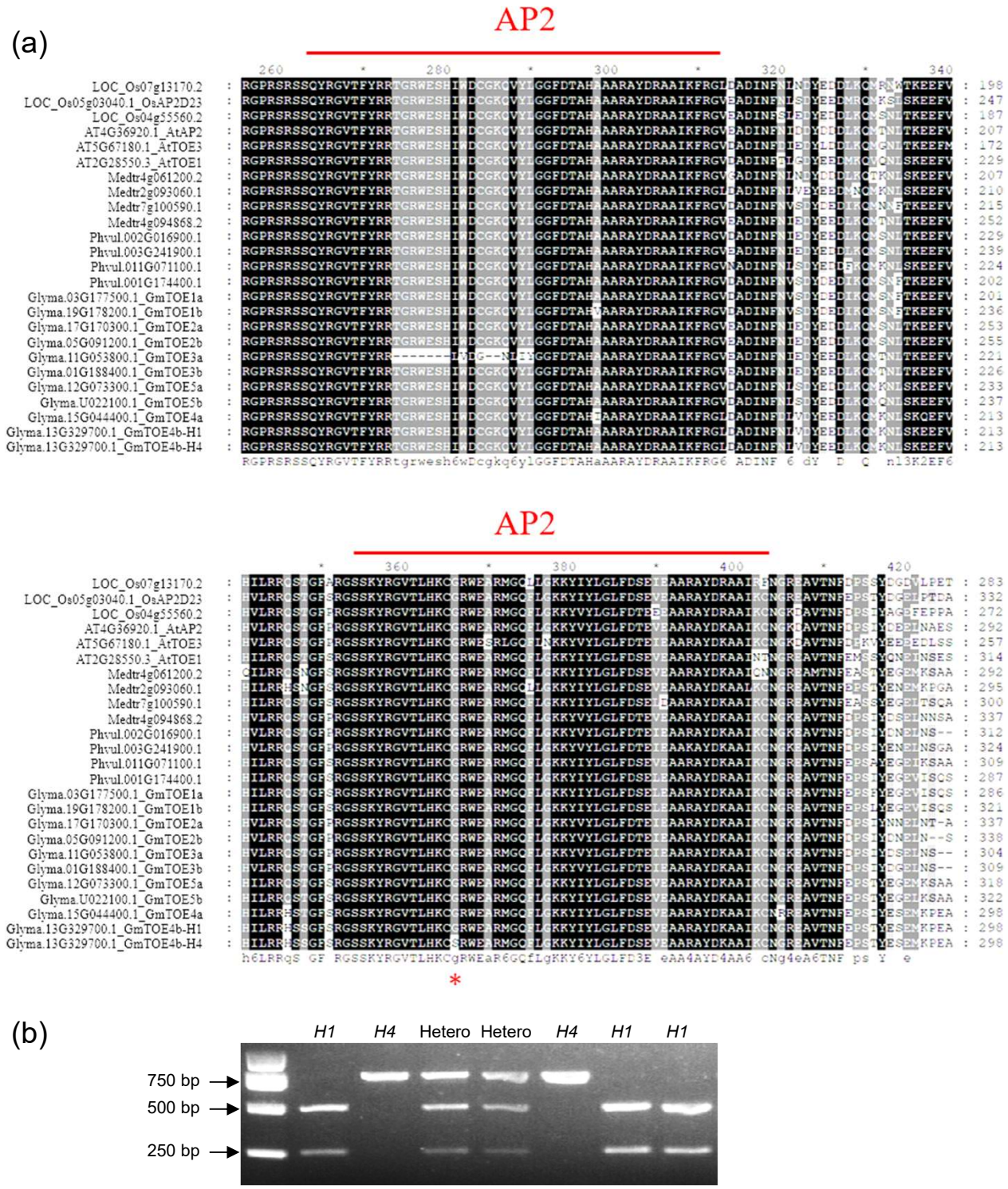

Figure S4 *TOE4b*^{H4} variation affects one of the conserved AP2 domain. (a) Alignment of the amino acid sequences of TOE4b with its orthologs and homologs. Amino acid sequences from *Glycine max*, *Phaseolus vulgaris*, *Medicago truncatula*, *Arabidopsis thaliana* and *Oryza sativa* were aligned by Clustal W. The red asterisk indicates the amino acid in TOE4b^{H4} that changes from the conserved G residue to S. (b) CAPS marker for the identification of the *TOE4b*^{H1} and *TOE4b*^{H4} alleles. PCR products were digested with the restriction enzyme *Hae*III. Hetero, heterozygous plant.

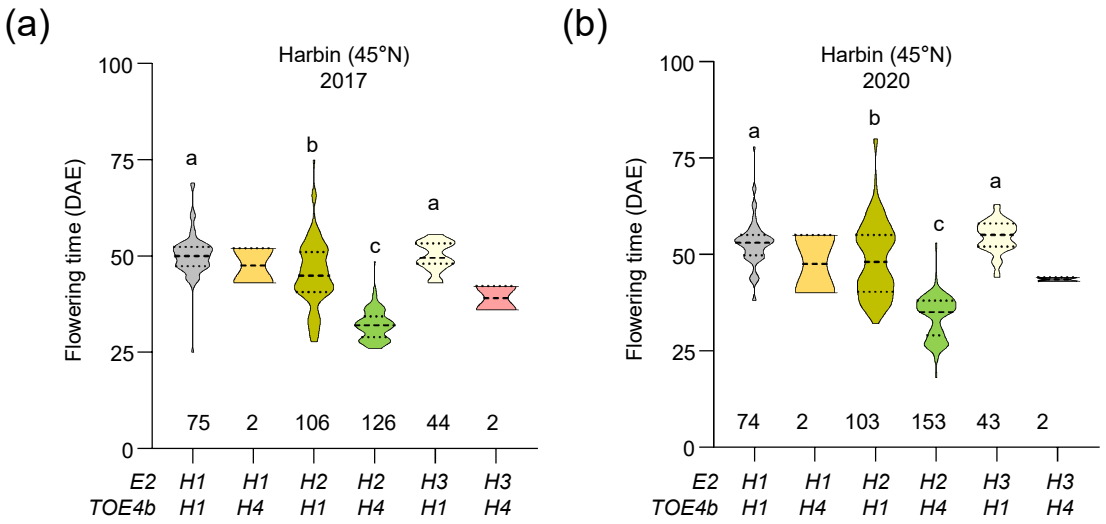

Figure S5 Effects of combination of *TOE4b* and *E2* alleles on soybean flowering time. (a, b) Flowering time of six allelic combinations between *TOE4b* and *E2* across 449 soybean accessions grown in Harbin in 2017 (a) and 2020 (b). The number above the *x*-axis represents the number of samples. Different lowercase letters indicate significant differences. The phenotype with 2 samples was not statistically analyzed with other samples.

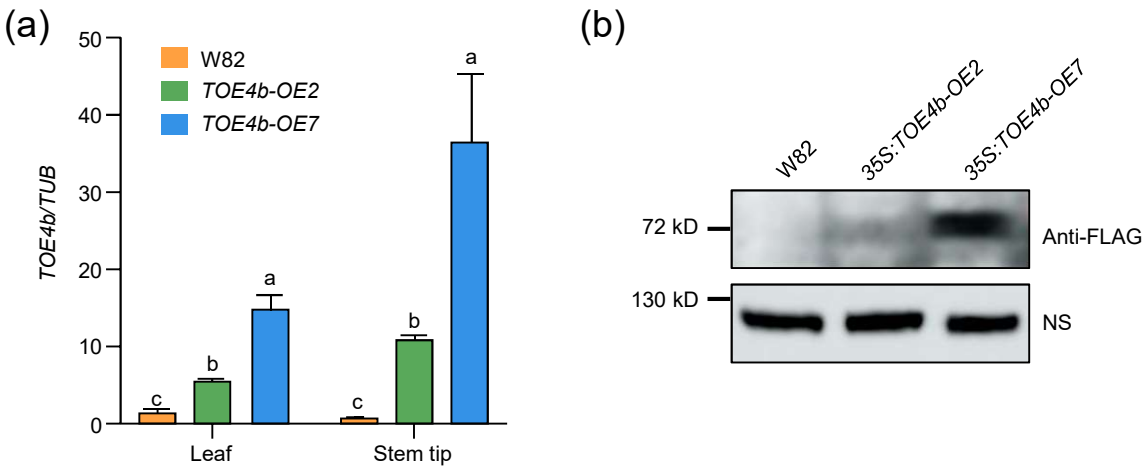

Figure S6 Generation *TOE4b* overexpression lines in soybean. (a) Relative transcript levels of *TOE4b* in wild-type W82 and T1 overexpression transgenic plants. *β-Tubulin* (*TUB*) was used as a normalization transcript. (b) Abundance of TOE4b-3Flag in the leaves of T_1 transgenic plants, as determined by immunoblotting using anti-Flag antibodies. The non-specific band (NS) was used as a loading control.

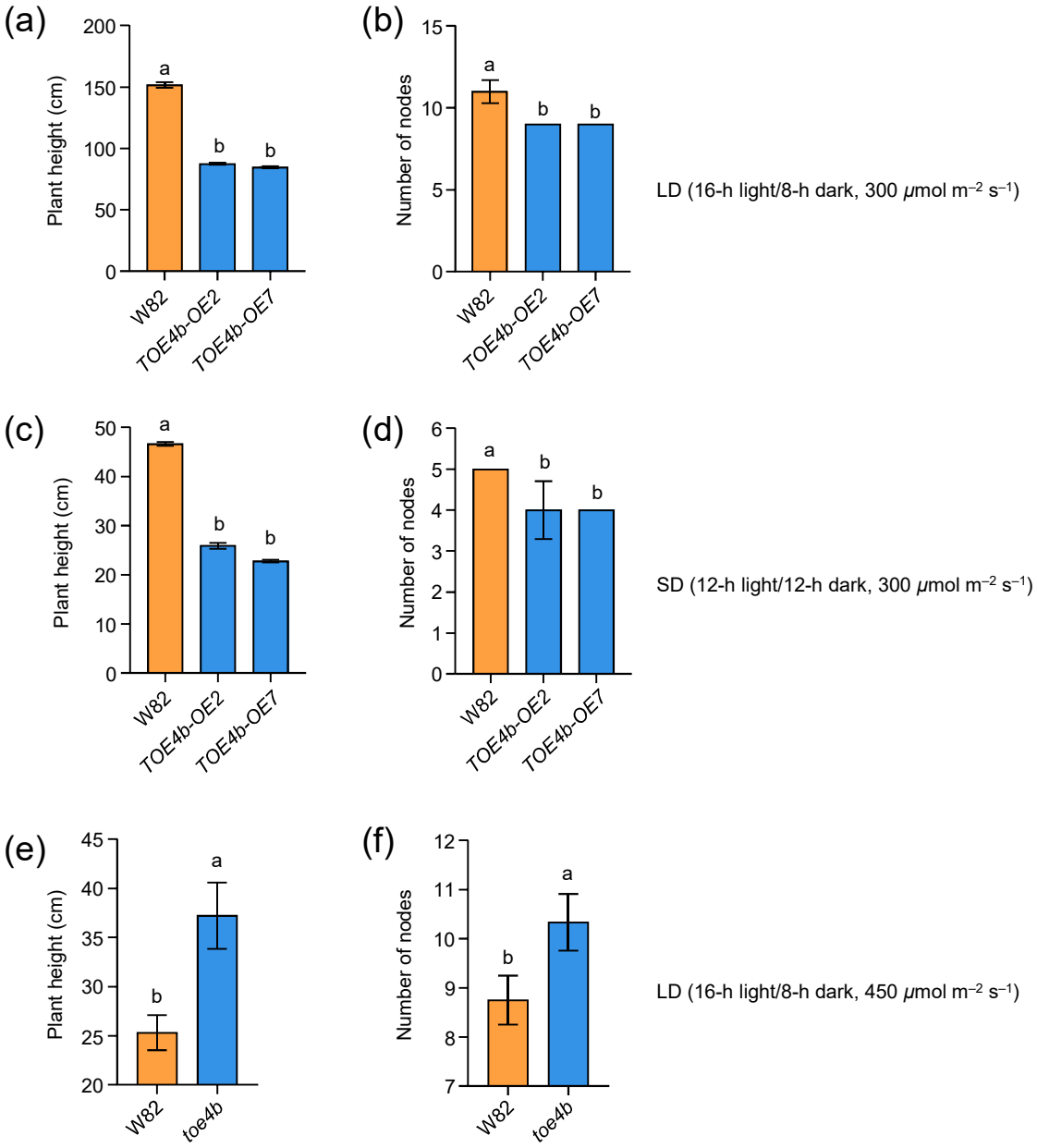

Figure S7 Phenotypes of *TOE4b* mutant and overexpression lines. Agronomic traits of plant height (a, c, e), measured from the cotyledonary node of the main stem to the apex in centimeters; number of nodes (b, d, f). The plants were grown in an incubator under LD (16-h light/8-h dark) (a-b) or SD (12-h light/12-h dark) conditions with an average photon flux of 300 μmol m^{-2} s^{-1} (c-d). For the mutant plants were grown in an incubator under LD (16-h light/8-h dark) with an average photon flux of 450 μmol m^{-2} s^{-1} (e-f). W82, wild-type Williams 82. Five plants were scored for each phenotype; data are means ± SD. $P < 0.05$, as determined by multiple comparison testing by one-way ANOVA. Different lowercase letters indicate significant differences.

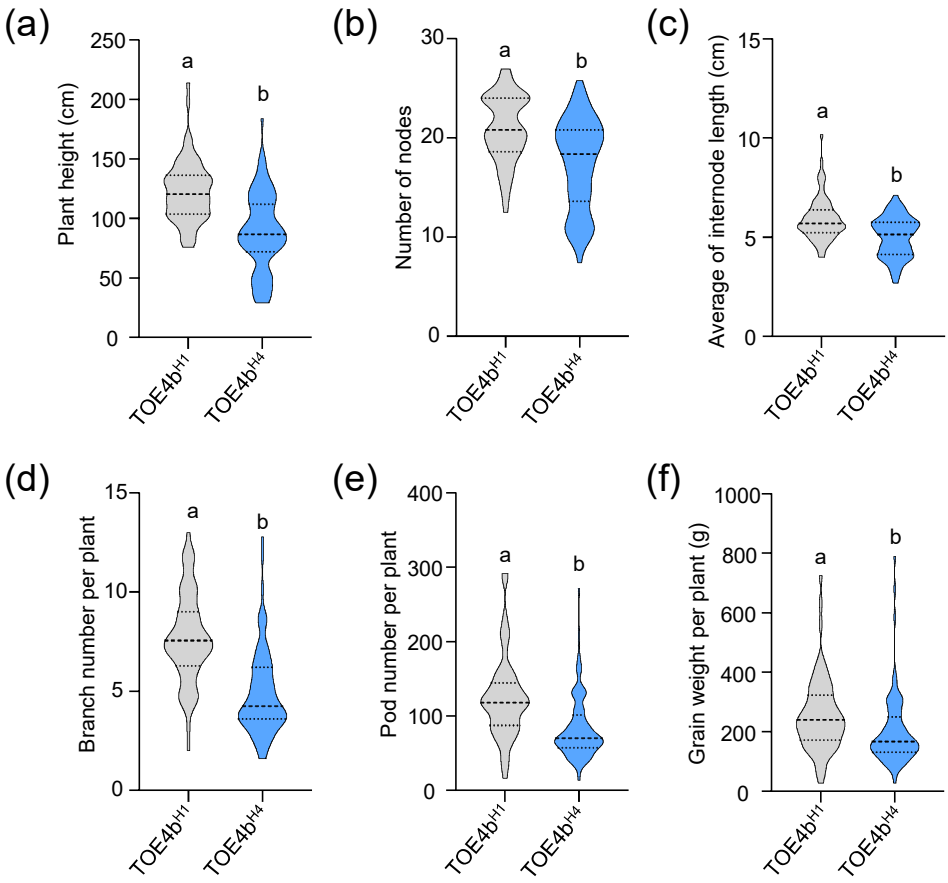

Figure S8 Agronomic traits associated with the major haplotypes of *TOE4b*. Agronomic trait of plant height (a), measured from the cotyledonary node of the main stem to the apex in centimeters; number of nodes (b); average of internode length (c); branch number per plant (d); pod number per plant (e); and grain weight per plant (f). *TOE4b*^{H1}, n = 68; *TOE4b*^{H4}, n = 83. The plants were grown in a standard field in Harbin under natural conditions in 2017. At least 6 plants were scored for each phenotype; data are means ± SD. $P < 0.05$, two-tailed Student's *t*-test. Different lowercase letters indicate significant differences.