

Genome-wide DNA methylation analysis of soybean curled-cotyledons mutant and functional evaluation of a homeodomain-leucine zipper (HD-Zip) I gene *GmHDZ20*

Hui Yang^{†,*}, Zhongyi Yang[†], Zhuozhuo Mao, Yali Li, Dezhou Hu, Xiao Li, Guixia Shi, Fang Huang, Baohui Liu, Fanjiang Kong and Deyue Yu^{*}

*** Correspondence:**

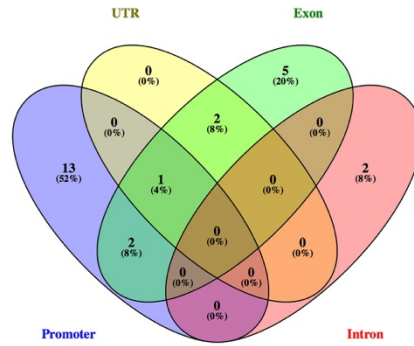
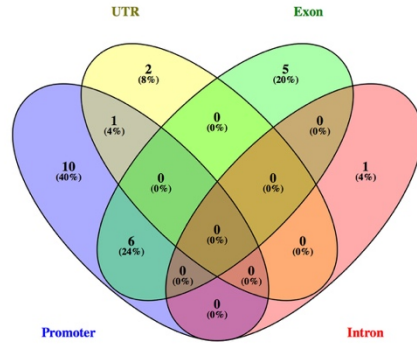
Hui Yang: smashing_hui@163.com

Deyue Yu: dyyu@njau.edu.cn

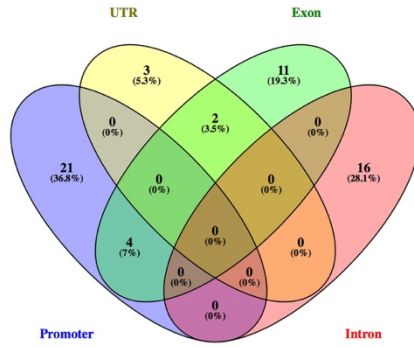
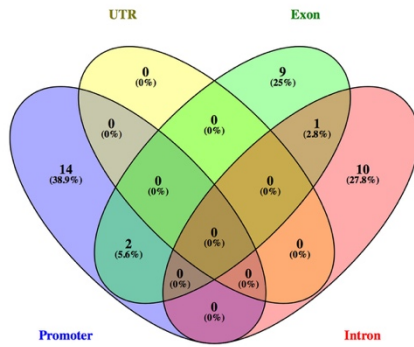
Hypermethylated-DMGs

Hypomethylated-DMGs

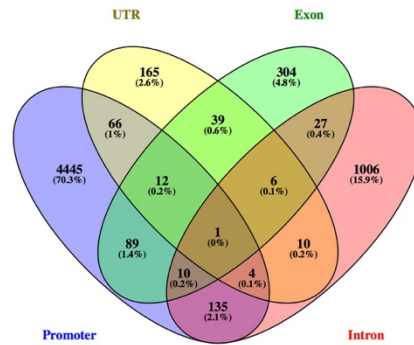
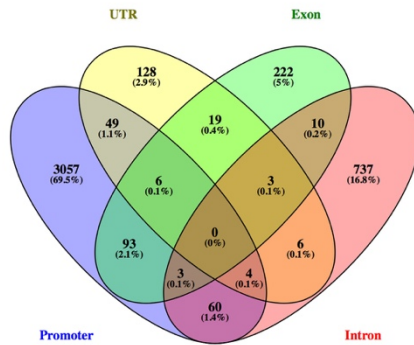
A



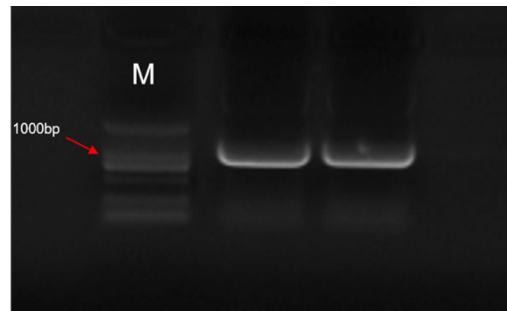
B



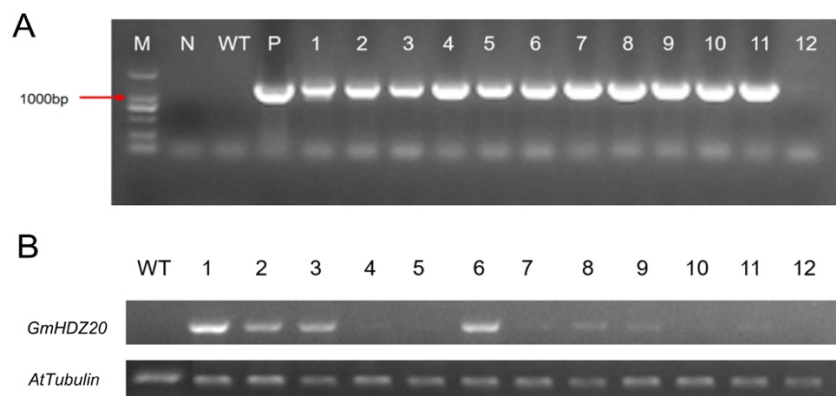
C



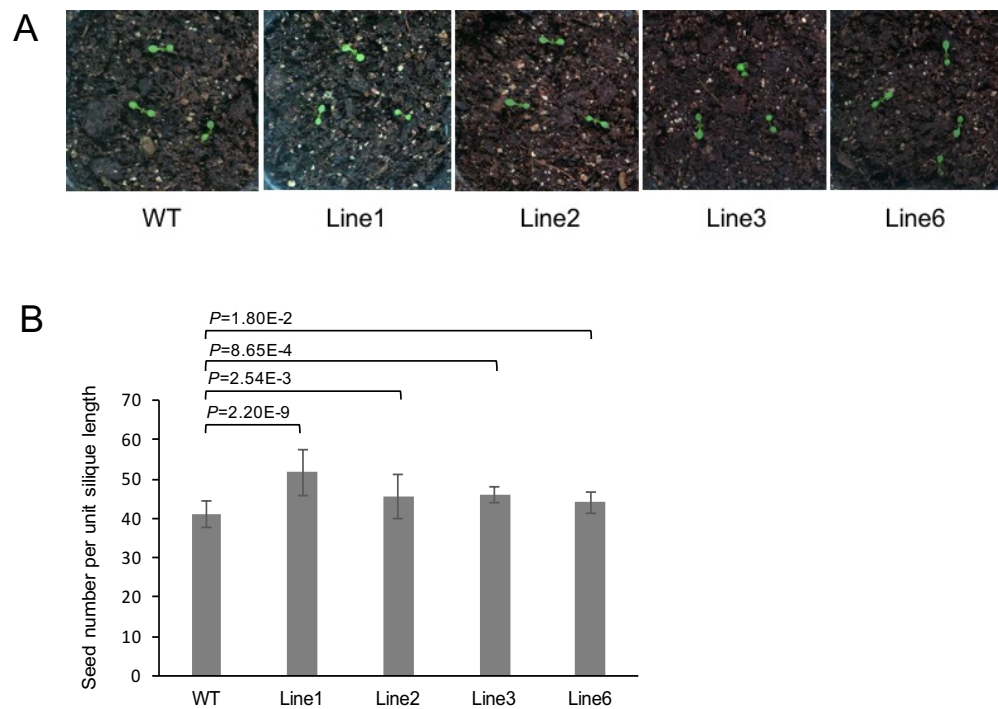
Supplementary Figure S1. Venn diagram of hyper- and hypo-DMGs in gene functional regions of the CG (**A**), CHG (**B**), and CHH (**C**) contexts.



Supplementary Figure S2. Amplification of the *GmHDZ20* coding sequence. M: Trans2K DNA marker.



Supplementary Figure S3. Identification of T₂ transgenic *Arabidopsis* plants heterologously expressing *GmHDZ20*. **(A)** Transgenic plants were identified by amplifying *GmHDZ20* from the genomic DNA. **(B)** Semi-quantitative RT-PCR analysis of *GmHDZ20* expression in different transgenic lines. M: Trans2K DNA marker; P: positive control (recombinant plasmid); N: negative control (water); WT: Col-0 ecotype; Numerical code: different transgenic lines.



Supplementary Figure S4. Comparison of cotyledon morphology and seed number per unit silique length between *Arabidopsis* WT and *GmHDZ20* transgenic plants. **(A)** The cotyledons of 10-day-old seedlings. **(B)** Statistics of seed number per unit silique length. Data are given as means $\pm$ SD ($n = 10$ seedlings). Two-tailed, paired-sample t -test were used to generate the p values.

Supplementary Table S1 Primer pairs used for experiments

Gene Names	Sense primers (5'-3')	Antisense primers (5'-3')	Function
<i>GmHDZ20</i>	ACAACCCTCTTCTCAACCTCT	GGTCTTTGGCTAGTTTGGTCT	qRT-PCR
<i>GmTubulin</i>	GGAGTTCACAGAGGCAGA	CACTTACGCATCACATAGCA	qRT-PCR
<i>AtTubulin</i>	CTCAAGAGGTTCTCAGCAGTA	TCACCTTCTTCATCCGCAGTT	qRT-PCR
<i>GmHDZ20</i>	<u>AAGTCCGGAGCTAGCTCTAGATGGCGG</u> GTAGTGGAAGTGC	<u>GCCCTTGCTCACCATGGATCCG</u> TAGTAA GCCCAGGTCCAAAGG	Subcellular localization
<i>GmHDZ20</i>	<u>GGGGACAAGTTTGTACAAAAAAGCAG</u> <u>GCTATGGCGGGTAGTGGAAGTGC</u>	<u>GGGGACCACTTTGTACAAGAAAGCTGG</u> <u>GTTTAGTAGTAAGCCCAGGTCCAAA</u>	Transformation
<i>GmHDZ20-FL</i>	<u>TCTCAGAGGAGGACCTGCATATGGCGG</u> GTAGTGGAAGTGC	<u>CCGCTGCAGGTCGACGGATC</u> TTAGTAG TAAGCCCAGGTCCAAA	Transcriptional activation activity analysis
N243	<u>TCTCAGAGGAGGACCTGCATATGGCGG</u> GTAGTGGAAGTG	<u>CCGCTGCAGGTCGACGGATC</u> GGGTGA TGAAAGTACTCGTC	Transcriptional activation activity analysis
N411	<u>TCTCAGAGGAGGACCTGCATATGGCGG</u> GTAGTGGAAGTG	<u>CCGCTGCAGGTCGACGGATC</u> GTCTTC CACCGAGCTCTACG	Transcriptional activation activity analysis
N549	<u>TCTCAGAGGAGGACCTGCATATGGCGG</u> GTAGTGGAAGTGC	<u>CCGCTGCAGGTCGACGGATC</u> CCCTCTT GCAAGCACCTTCTCAG	Transcriptional activation activity analysis
H1H2	<u>TCTCAGAGGAGGACCTGCATGAGAAG</u> AAGCGACGTCTCTCTGCG	<u>CCGCTGCAGGTCGACGGATC</u> CCCTCTT GCAAGCACCTTCTCA	Transcriptional activation activity analysis
HD	<u>TCTCAGAGGAGGACCTGCATGAGAAG</u> AAGCGACGTCTCTCTG	<u>CCGCTGCAGGTCGACGGATC</u> GTCTTC CACCGAGCTCTACG	Transcriptional activation activity analysis
HALZ	<u>TCTCAGAGGAGGACCTGCATAAACAGC</u> TGGAGAAGGATTACGAGA	<u>CCGCTGCAGGTCGACGGATC</u> CCCTCTT GCAAGCACCTTCTCA	Transcriptional activation activity analysis
C447	<u>TCTCAGAGGAGGACCTGCATAAACAA</u> GAGGGGCACATGAAGC	<u>CCGCTGCAGGTCGACGGATC</u> TTAGTAG TAAGCCCAGGTCCAAAGG	Transcriptional activation activity analysis
ΔH1H2-1	<u>TCTCAGAGGAGGACCTGCATATGGCGG</u> GTAGTGGAAGTG	GCTTCATGTGCCCTCTTGTGTTGGGTTG ATGAAAGTACTCGTC	Transcriptional activation activity analysis
ΔH1H2-2	GACGAGTACTTTCATCAACCCAAACAA GAGGGGCACATGAAGC	<u>CCGCTGCAGGTCGACGGATC</u> TTAGTAG TAAGCCCAGGTCCAAAGG	Transcriptional activation activity analysis