Phylogeny, gene structures and expression patterns of the ERF gene family in soybean (*Glycine max* L.)

Authors: Gaiyun Zhang [1,2, §], Ming Chen [2, §], Xueping Chen [1, *], Zhaoshi Xu [2], Shan Guan [2], Lian-Cheng Li [2], Aili Li [2], Jiaming Guo [1], Long Mao [2], and Youzhi Ma [2, *]

Affiliation:

[1] Department of Chemistry, University of Science and Technology of China, Hefei 230026, China;

[2] The National Key Facility for Crop Genetic Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, 100081, China

* **Corresponding author:**

Youzhi Ma
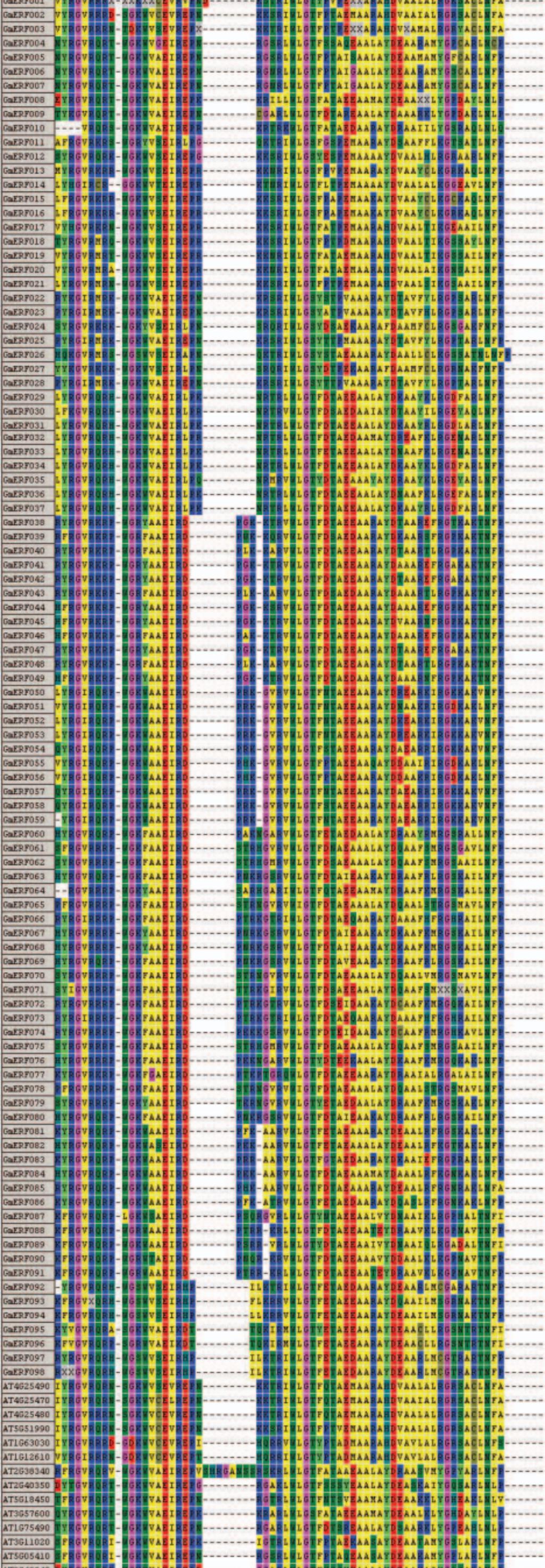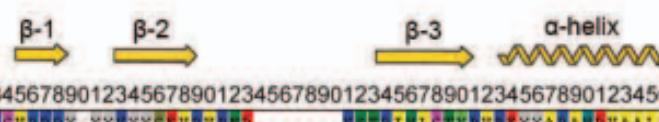
Phone : +86 10 82108789
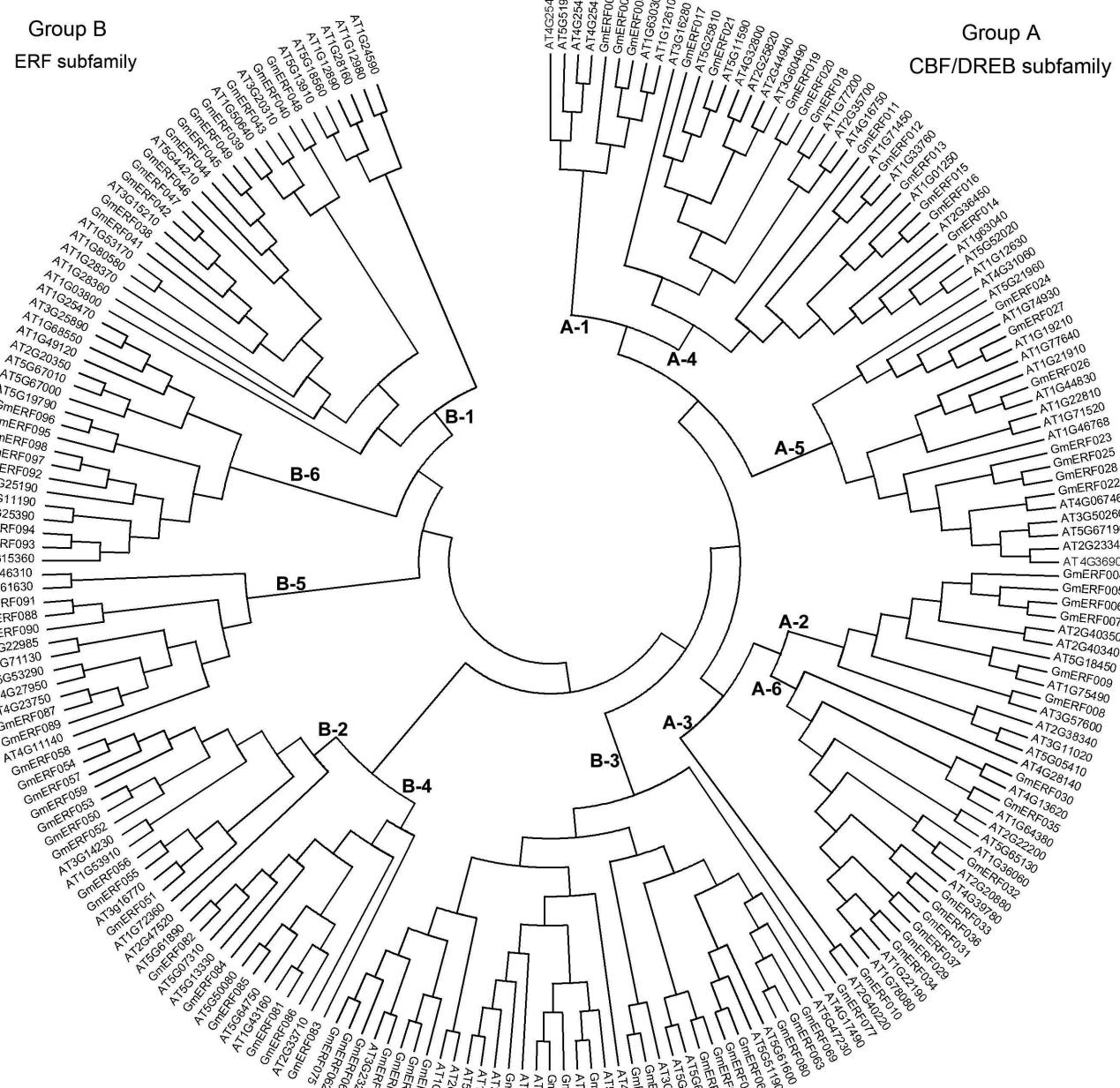
Fax : +86 10 82108789

E-mail: mayzh@mail.caas.net.cn

Institute of Crop Science, Chinese Academy of Agricultural Sciences, 12[#] Zhongguancun South Street, Beijing 100081, China; Institute of Natural Resources and Regional Planning, CAAS, Beijing 100081, China
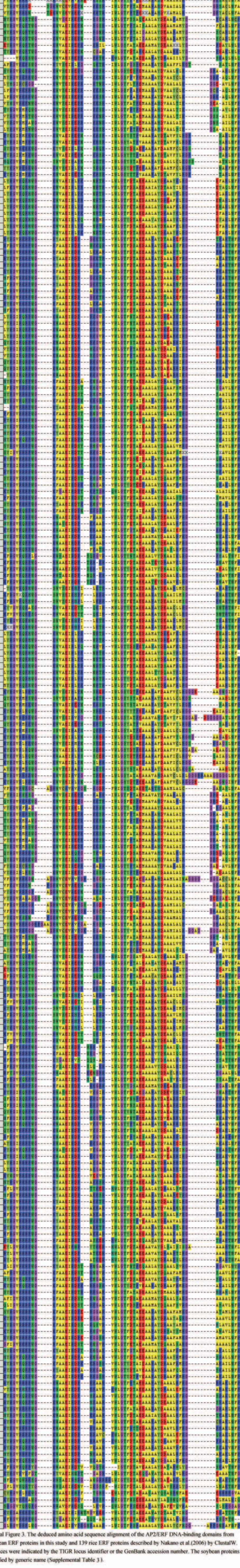
§ These authors contributed equally to this work.

Supplemental Figure1. The deduced amino acid sequence alignment of the AP2/ERF DNA-binding domains from the 98 soybean ERF proteins in this study and 122 Arabidopsis ERF proteins described by Nakano et al. (2006) by ClustalW. Arabidopsis sequences had AGI identifiers; the soybean proteins were identified by generic name (Supplemental Table 3).
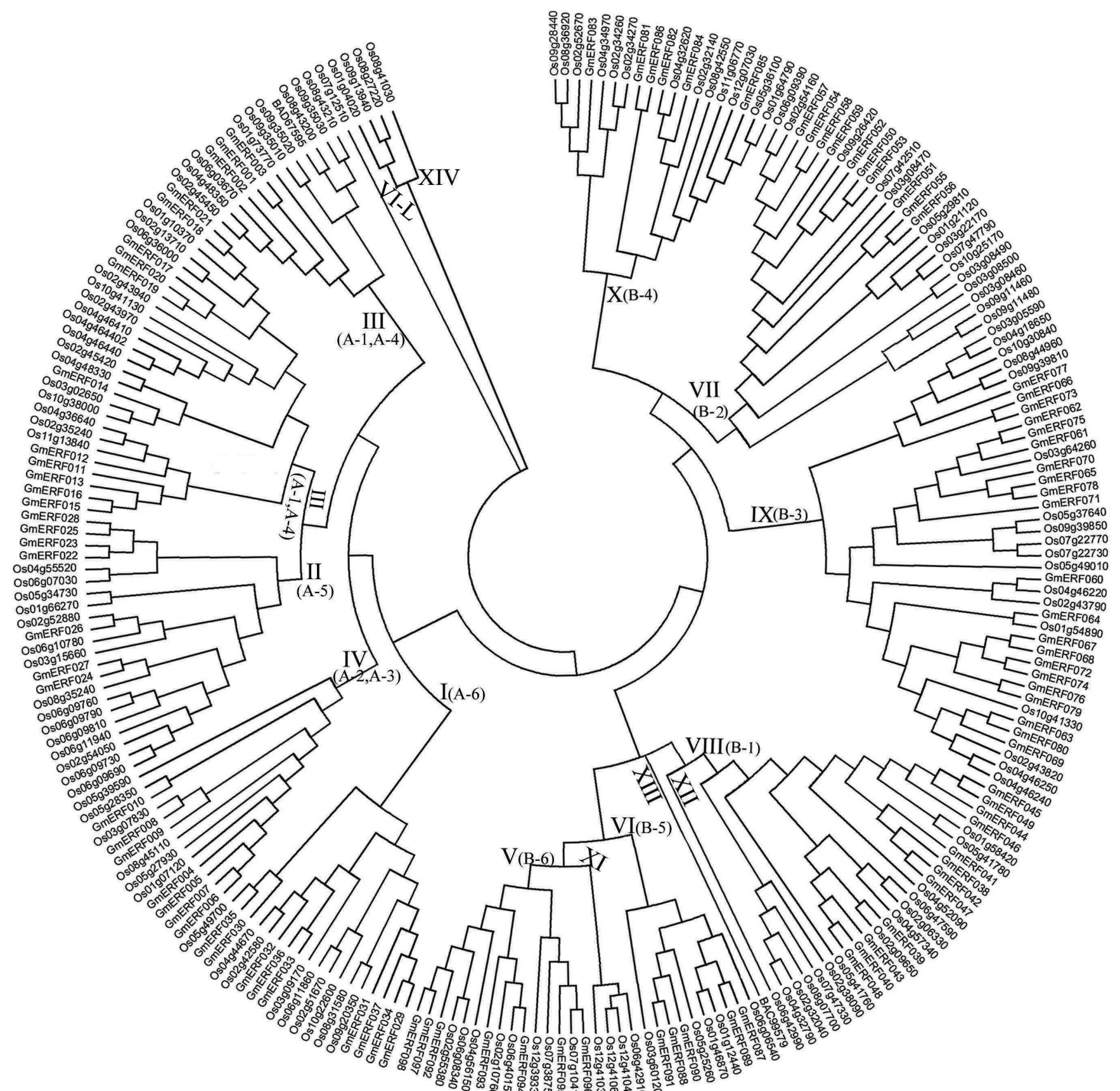
Supplemental Figure 2. An unrooted phylogenetic tree of the ERF family of the soybean and Arabidopsis. The amino acid sequences of the AP2/ERF domain of 98 soybean ERF family proteins and 122 Arabidopsis ERF proteins were aligned by ClustalW and the phylogenetic tree was constructed using MEGA 4.0 and the NJ method. The soybean proteins were named according to GmERF numbers (see Supplemental Table 3); Arabidopsis sequences had AGI identifiers.

Supplemental Figure 3. The deduced amino acid sequence alignment of the AP2/ERF DNA-binding domains from the 98 soybean ERF proteins in this study and 139 rice ERF proteins described by Nakano et al.(2006) by ClustalW. Rice sequences were indicated by the TIGR locus identifier or the GenBank accession number. The soybean proteins were identified by generic name (Supplemental Table 3).

Supplemental Figure 4. An unrooted phylogenetic tree of the ERF family of the soybean and rice. The amino acid sequences of the AP2/ERF domain of 98 soybean ERF family proteins and 139 rice ERF proteins were aligned by ClustalW and the phylogenetic tree was constructed using MEGA 4.0 and the NJ method. The soybean proteins were named according to GmERF numbers (see Supplemental Table 3). Rice sequences were indicated by the TIGR locus identifier or the GenBank accession number. Classification was indicated according to the report of Nakano et al. (2006), and soybean classification by comparative analysis with Arabidopsis was indicated in parentheses.

Supplemental Table 1.

The sequence of primers used for RT- PCR amplification of selected 9 target unigenes under different stress treatments

TC number	Primer
GmERF039	Forward: 5-GTGCTGTCTCCTTGTTCTCCATCGT-3
	Reverse: 5-AGTGGTGGTCCGCCTCTGCC-3
GmERF056	Forward: 5- CGAGACCCATAGTGCTGCCA -3
	Reverse: 5- CCAACCCTAGAAACCGTTCAAGT -3
GmERF057	Forward:5- GCCTGACCGCCGATTACCTGT-3
	Reverse: 5- AGCACTGAACTTGTGACCCAGATTG-3
GmERF061	Forward: 5-AGGCTCTGTTTACTCATCTGGTGT-3
	Reverse: 5-CATCTTCCTTCTTCCTCTTAGCAG-3
GmERF069	Forward: 5-ACCGTCGCTGCTAATCGCTGTT-3
	Reverse: 5-GGCGACAACGGCGGAACG-3
GmERF079	Forward: 5-TGAAGGAGAACGACACCAATGAC-3
	Reverse: 5-ACTATCCTCACACTTCTGCAATGC-3
GmERF081	Forward: 5-CATCTACTTATGTTGTTGCCACCTC-3
	Reverse: 5-GCAACTGGGCAGAATACAAGG-3
GmERF089	Forward: 5- AGCCCTCCATGTCCTCCAC- 3
	Reverse: 5- TCACACTCTCTATGCCAACTCGT-3
GmERF098	Forward: 5-AGGCGTCCGACAAAGGCATT-3
	Reverse: 5-GTGGTGGAATCACAGAGCATAGC-3