A GATA Transcription Factor from Soybean (Glycine max) Regulates Chlorophyll Biosynthesis and Suppresses Growth in the Transgenic Arabidopsis thaliana

Chanjuan Zhang 1, Yi Huang 1, Zhiyuan Xiao 1, Hongli Yang 1, Qingnan Hao 1, Songli Yuan 1, Haifeng Chen 1, Limiao Chen 1, Shuilian Chen 1, Xinan Zhou 1,* and Wenjun Huang2,3,*

Supplementary Materials:

Figure S1. Multiple alignment of deduced amino acids of GmGATA58 with other three GATA factors. The highly conserved GATA domain, conserved leucine-leucine-methionine (LLM) domain and nuclear localization signal (NLS) were indicated. The protein alignment was generated based on the full-length protein sequences using the Clustal Omega algorithm at the EMBL-EBI website (http://www.ebi.ac.uk/Tools/msa/clustalo/) with the default settings. Protein sequences were retrieved from GenBank and their accession numbers were as follows: Glycine max GATA58 (XP_003550634.1) and GATA44 (XP_003543725.1), Arabidopsis thaliana GNC (NP_200497.1) and GNL (NP_194345.1).
Figure S2. Restoration of greening defect of Arabidopsis thaliana gnc mutant by GmGATA58 overexpression. Phenotypic comparison of wild-type (WT), gnc mutant and overexpression of GmGATA58 in the gnc mutant background (gnc-OX7).

Table S1. Putative cis-acting regulatory elements predicted in the GmGATA58 promoter sequence.

| No. | Cis-Element Name                      | Sequence         | Function                        | Element Number |
|-----|--------------------------------------|------------------|---------------------------------|----------------|
| 1   | ABRE (ACGT-containing ABA Response Element) | CACGTG          | abscisic acid responsiveness    | 1              |
| 2   | ACE (ACGT-Containing Element)          | AAAACGTTTA       | light responsiveness            | 1              |
| 3   | AE-box (Activating Element)            | AGAAACCTT        | light response                   | 1              |
| 4   | ARE (Anaerobic Response Element)       | TGGTTT           | anaerobic induction              | 2              |
| 5   | ATI-motif Box 4                        | ATTAATTTTACA     | light responsive module          | 1              |
| 6   | Box 4                                 | ATTAAT           | light responsiveness            | 3              |
| 7   | CATT-motif                            | GCACTTC          | light responsive element         | 2              |
| 8   | (Chs Conserved DNA Module Array 2)     | TTACTTTAA        | light responsive element         | 1              |
| 9   | Circadian (Circadian Control)          | CAAAGATATC       | circadian control               | 1              |
| 10  | G-Box                                 | CACGTG           | light responsiveness            | 1              |
| 11  | G-box                                 | GACATGGTGT       | light responsiveness            | 1              |
| 12  | GA-motif                              | AAGGAAGA         | light responsive element         | 1              |
| 13  | GATA-motif                            | AAGATAAGATT      | light responsive element         | 1              |
| 14  | GCN4_motif                            | TGACTCA          | endosperm expression            | 1              |
| 15  | HSE (Heat Shock Element)              | AAAAAATTTTC      | heat stress responsiveness       | 2              |
| 16  | I-box                                 | AGATAAGG         | light responsive element         | 1              |
| 17  | MBS (MYB Binding Site)                | T(C)AACTG        | drought-inducibility            | 4              |
| 18  | MRE (MYB Recognition Element)         | AACCTAA          | light responsiveness            | 2              |
| 19  | O2-site (Opaque-2 site)               | GT(A)TGACG(A)TGA | zein metabolism regulation      | 2              |
| 20  | P-box (Pyrimidine Box)                | CCTTTTG          | gibberellin-responsive element   | 1              |
| 21  | Ski-1_motif                           | GTCAT            | endosperm expression            | 5              |
| 22  | Sp1                                   | CC(G/A)CCC       | light responsive element         | 1              |
| 23  | TATC-box                              | TATCCCC          | gibberellin-responsive          | 1              |
| 24  | TCA-element                           | CCATCTTTTTT      | salicylic acid responsiveness   | 1              |
| 25  | TCCC-motif                            | TCTCCCT          | light responsive element         | 2              |
| 26  | TCT-motif                             | TCTTAC           | light responsive element         | 1              |
| 27  | TGA-element                           | AACGAC           | auxin-responsive element         | 1              |
### Table 2. Primers used in this study for cloning, vector construction, semi-quantitative RT-PCR and qPCR assay.

| Gene Name | Accession No. | Primer Sequence (Forward/Reverse) | Note |
|-----------|---------------|-----------------------------------|------|
| GmGATA58  | Glyma.17G055200 | GTGTTTTTATATGTGTGTTCCTCCA; ATGGAAGTGGTGGTACCTCTAGCC | Full-length cDNA cloning |
|           |               | TCCCCGGGAGATTCCAGCTATGGCCCA; CGGATGATCACTGGAACAAGCCCTATAAGATTAA | Subcellular localization assay |
|           |               | TGATGTTAGGTTAGCTTTACATGTC; GATGGAGGAACACACATATAAAAC | Promoter isolation |
| GmGATA58  | Glyma.17G055200 | TCCCCCGGGAGATTCCAGCTATGGCCCA; CGGATGATCACTGGAACAAGCCCTATAAGATTAA | Subcellular localization assay |
|           | Glyma.17G055200 | ATGATCAGAGCCTGTGCAG; TCAATGAACAAGCCCTATAAGATTAA | Semi-quantitative RT-PCR assay |
|           | Glyma.17G055200 | ATGATCAGAGCCTGTGCAG; TCAATGAACAAGCCCTATAAGATTAA | Semi-quantitative RT-PCR assay |
| AtUBQ10   | At4g05320 | GGCTGATTACAATATCCAGTCG; GCCACCATTGGAGGAGAGT | Semi-quantitative RT-PCR assay |
| GmCHLH1   | Glyma.03G137000 | ACGGTATCGGATAAGTCGGTCCTACCCCTACCTCGT; AGAATCTGTGTTGATCTTTAAATATGTAAGTAAAC | promoter isolation for dual-luciferase reporter assay |
| GmCHLH3   | Glyma.19G139300 | ACGGTATCGGATAAGTCGGTCCTACCCCTACCTCGT; AGAATCTGTGTTGATCTTTAAATATGTAAGTAAAC | promoter isolation for dual-luciferase reporter assay |
| GmCHLII1  | Glyma.13G232500 | ACGGTATCGGATAAGTCGGTCCTACCCCTACCTCGT; AGAATCTGTGTTGATCTTTAAATATGTAAGTAAAC | promoter isolation for dual-luciferase reporter assay |
| GmACT11   | Glyma.18G290800 | ATCTTGACCTGACGCTGTTATCC; GCTTGCTCCTGCTGCTCC | qPCR assay |
| AtDXS     | At4g15560 | AACTTACTTGTGGGAGGACATTAG; CATCTTGGTGGAGAGCG | qPCR assay |
| AtDXR     | At5g62790 | GCCGTATGCTGTATAGCCTGTGCTTACA; AGAATCTGTGTTGATCTTTAAATATGTAAGTAAAC | qPCR assay |
| AtHEMA1   | At1g58290 | CACCGGTTTACATGTGGACGG; CACCCATCGTCTCAACACTGTG | qPCR assay |
| AtHEMA3   | At2g31250 | GACAAAAAGACGAGGAAGCAG; | qPCR assay |
| Gene   | Accession | Primer Forward | Primer Reverse |
|--------|-----------|----------------|----------------|
| AtGSA1 | At5g63570 | GCTGCTACCAGTCCCATCGTA | ATACCTTTGGAAATCACGCCTGATC; ATCTCCAATAATATCAGCCTACC |
| AtGSA2 | At3g48730 | GACCGATGTATCAAGCTGGTACG; AAGTATCATATGTCCCTGGGTGAC | ATCTCCATAATATCAGCCTACC |
| AtPPX I1 | At4g01690 | TGAGCATGAAACGCAGGTTTG; AAGTCCCTTCTGAAAAGAACC | AGACATAGTGACTAAACAGACAGCAGC |
| AtPPX I2 | At5g14220 | CAACCTTTATTGGTGGGAGTAGGAA; TGGTTGACAGAACCGGTTCA | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHLI1 | At4g18480 | AGACCATAGTGACTAAACAGACAGCAGC | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHLI2 | At5g45930 | GACCGGAGCTGAGTAGTAAGA; CTTCCTTGCAGTAAATCTCGCTCC | ATGTGGTAGGTAGG |
| AtCHLD | At1g08520 | CCGATGGCTAGCCACCATTA; TTGGACGCTGGGCTAGGAGCA | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHLH | At5g13630 | ACACCAATCCCAACTCTCTTCG; TCAGCGGAAGTCCCTGAGTAG | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHLM | At4g25080 | TGAGATGAAGGCAAGACCCAGCACA; GTCTGCTTGGTGTGGGTA | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHL27 | At3g56940 | GATTTCGCCGAGTTTGAGCC; AAATCAAGAACCAGCCGTAANACTC | AGACATAGTGACTAAACAGACAGCAGC |
| AtPORA | At5g54190 | TGATTTGGAGCTGGAACACACAGCACA; CCAAAACGTGAAACACCCGAGG | AGACATAGTGACTAAACAGACAGCAGC |
| AtPORB | At4g27440 | AAGGCTCTGAAAGTGTGGGAGA; TCGATTGGTACCAGGAGGTTGC | AGACATAGTGACTAAACAGACAGCAGC |
| AtPORC | At1g03630 | ACACATACGCTGTTTCCG; CAATACACTCTGACTCTCAAGAAG | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHLP | At1g74470 | GATGCTCTGTGATTTAGACTACG; CACTTACCGGAAACCACACCATAG | AGACATAGTGACTAAACAGACAGCAGC |
| AtCHLG | At3g51820 | CTCTTACGCCAGATGTTGTTTC; TGCCAAAGCTACTGGGAGAGA | AGACATAGTGACTAAACAGACAGCAGC |
| AtGAPDH | At3g26650 | CTTGGAAGAGCCTAGGATGGAAC| AGACATAGTGACTAAACAGACAGCAGC |
