Figure S1. Conserved motifs in the GmNTL N terminal region.

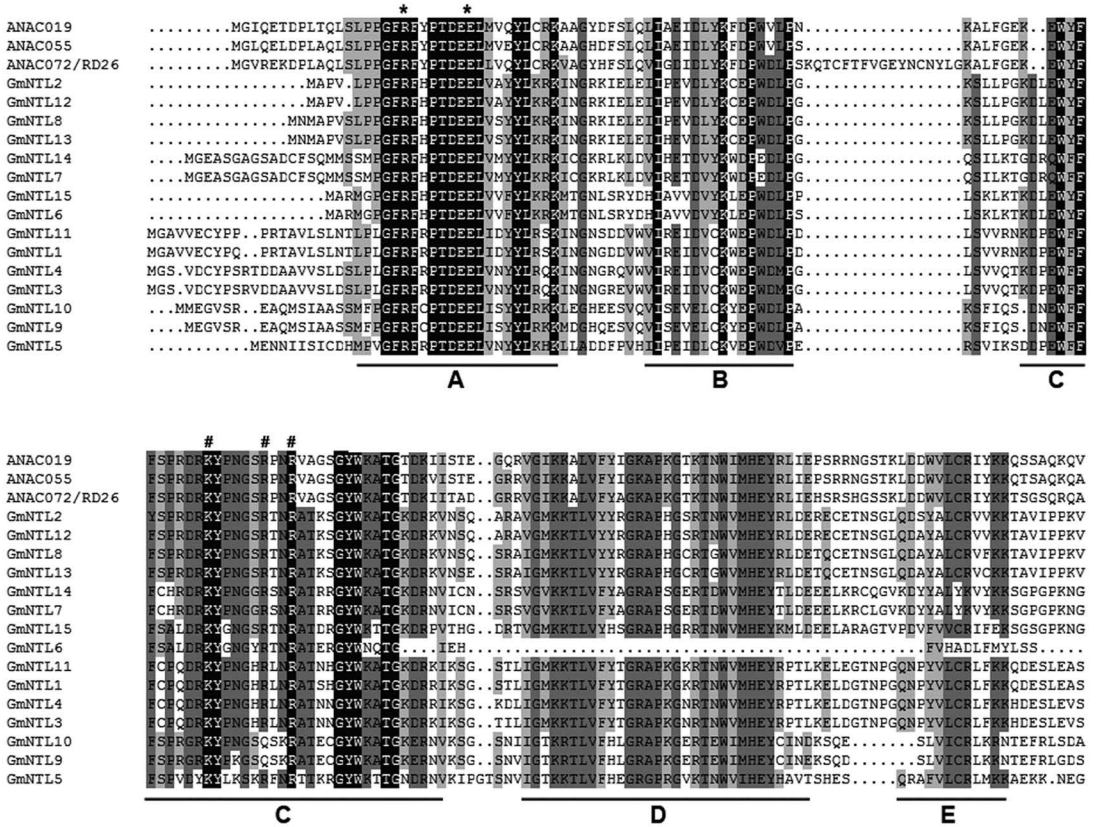

Figure S1. Conserved motifs in the GmNTL N-terminal region.

A comparison of the polypeptide sequences of the GmNTLs, ANAC013, ANAC055 and ANAC072 is shown. Identical residues indicated by white letters on a black background. The consensus NAC sub-domains A-E are indicated by underlining. ∗: residues forming a salinity bridge stabilizing the dimerization interface (Smyczynski et al., 2006). #: residues in a region containing several highly conserved residues of importance to DNA binding (Duval et al., 2002). Sequences used in the analysis can be found in as Supplementary Data Set 2.

Figure S2. The cds sequences tree of NTL genes.

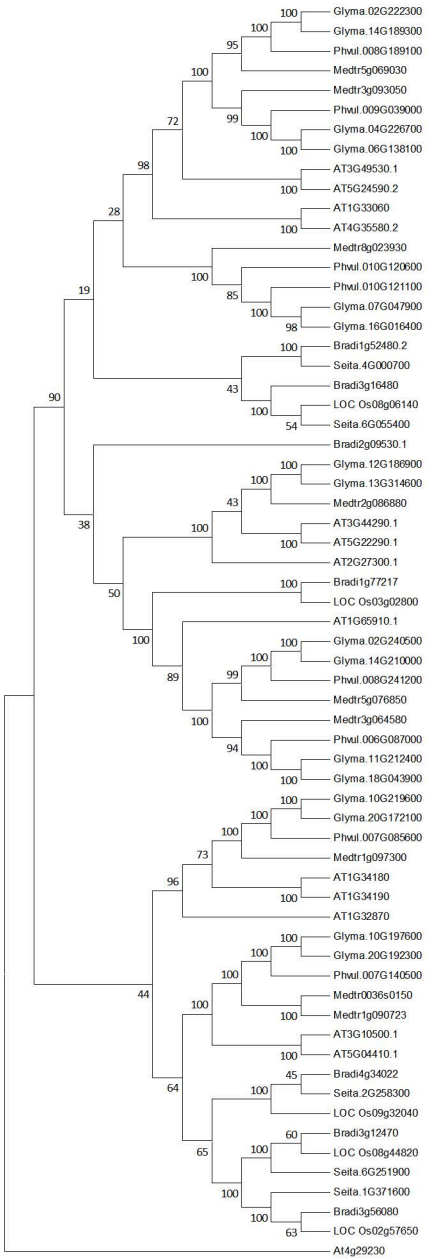

Figure S2. The cds sequences tree of NTL genes.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Jukes-Cantor model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed . The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach.Evolutionary analyses were conducted in MEGA6.The sequences used to construct the tree are given in Data Set S3.The ClustalW-based alignment of *GmNTL* cds sequences employed a gap open penalty of 10 and a gap extension penalty of 0.1, as implemented within MEGA6 software (Tamura et al., 2013).

Figure S3. Gene duplication for the *GmNTLs*.

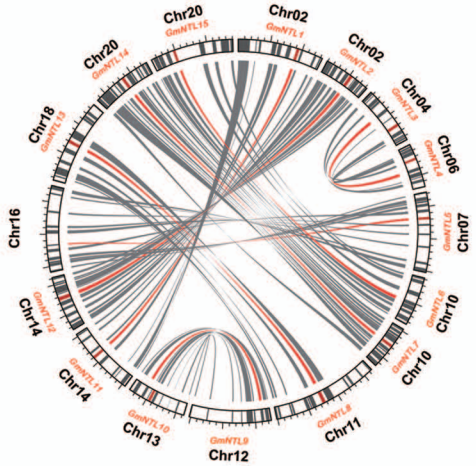

Figure S3. Gene duplication for the *GmNTLs*.

Segmentally duplicated genes located on 11 soybean chromosomes. Gray lines indicate the known synteny blocks in the soybean genome, and the red lines suggest duplicated pairs of *GmNTL/NAC* genes.

Figure S4

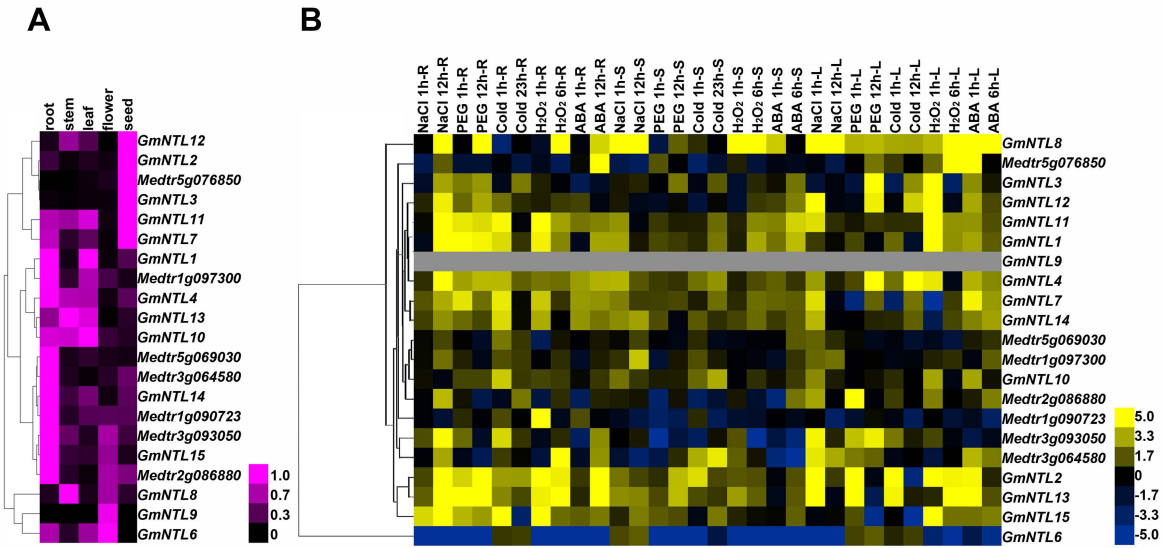

Figure S4. Hierarchical clustering of *GmNTL* duplicated pairs and their homologs in *Medicago truncatula*..

The expression data used are the same as in Figure 4. Hierachical clustering was performed using software Cluster 3.0 by Average linkage method. A, Organ specificity of transcription. Transcript abundances normalized to the most abundant transcript. B, The transcriptional response to salinity (200 mM NaCl), moisture stress (20% w/v PEG 6000), low temperature (4 °C), oxidative stress (10 mM H_2O_2) and 100 μM ABA treatment. For organ specifity assay, the levels were normalized against the transcript abundance in the most strongly transcribing organ. For the stress response assay, fold changes in abundance were $\log_2$ transformed. R: root, S: stem, L: leaf, h: hours.

Figure S5. Substitution rate analysis using free ratio model.

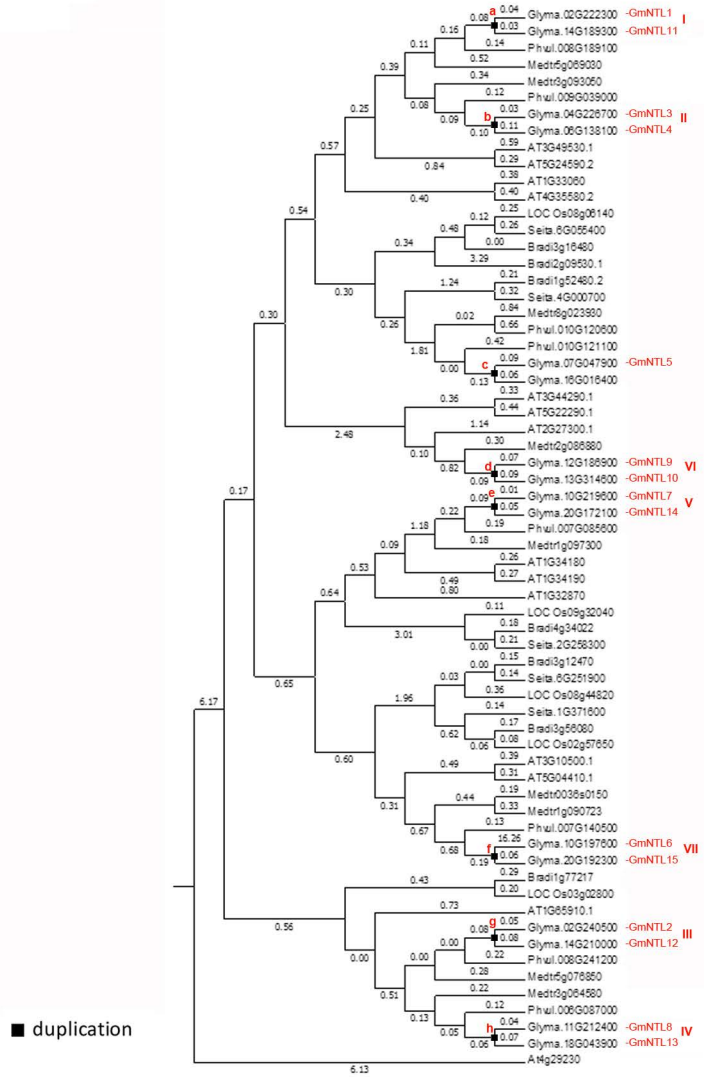

Figure S5. Substitution rate analysis using free ratio model.

The dN/dS ratios (ω) of each branch was estimated using free ratio model. GmNTL genes were lighted in red font. Both of branch was estimated using free ratio model. GmNTL above analysis led to key extant genes (Table 1).

Fig. S6 The transcriptional activation activity assay of 15 GmNTL proteins.

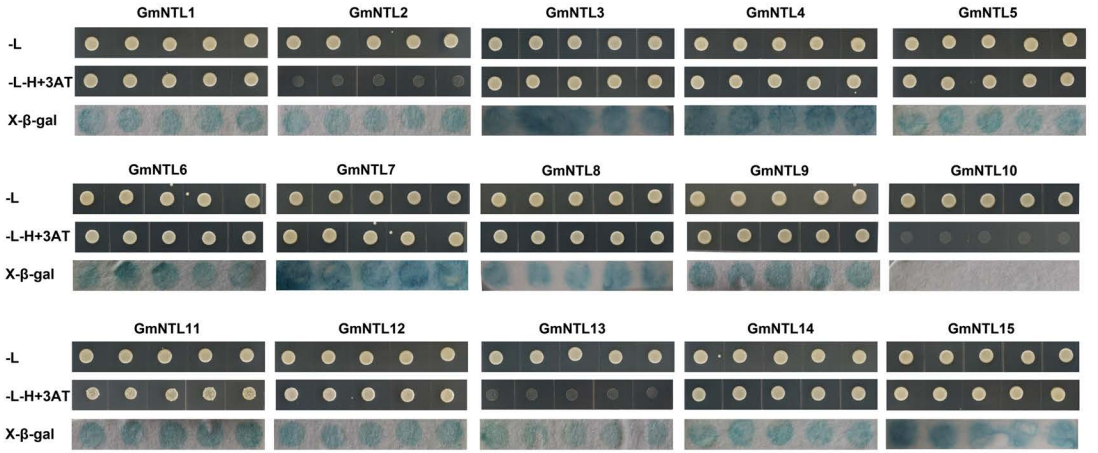

Figure S6. The transcriptional activation activity assay of 15 GmNTL proteins.

Transcriptional activation activity assay. The ability of yeast transformants to grow on a medium lacking histidine, leucine and supplemented with 3AT (10 mM) and the formation of colour in the X-gal assay indicate transcriptional activity. This picture showed the result of 5 random selected transformants of each construct.

Duval M, Hsieh TF, Kim SY, Thomas TL (2002) Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. Plant Mol Biol **50:** 237-248

Smyczynski C, Roudier F, Gissot L, Vaillant E, Grandjean O, Morin H, Masson T, Bellec Y, Geelen D, Faure JD (2006) The C terminus of the immunophilin PASTICCINO1 is required for plant development and for interaction with a NAC-like transcription factor. J Biol Chem **281:** 25475-25484

Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution **16:**111-120.

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution **30**: 2725-2729

Figure S7. The promoter tree of *GmNTLs*

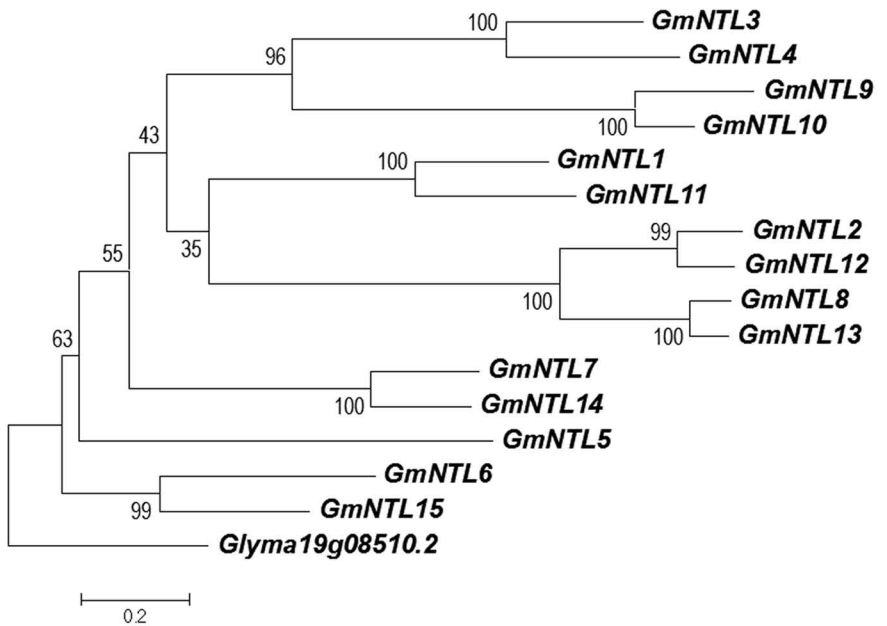

Figure S7. The promoter tree of *GmNTLs*.

The tree was constructed using the -1 to -1500 bp promoter sequences of *GmNTLs* with promoter of *Gm19g08510.2* as out-group. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura M et al., 1980). The tree with the highest log likelihood (-13041.8555) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 731 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).