Transcriptome dynamics of developing maize leaves and genomewide prediction of *cis* elements and their cognate transcription factors

Chun-Ping Yu[a,1], Sean Chun-Chang Chen[b,1], Yao-Ming Chang[b,1], Wen-Yu Liu[b,1], Hsin-Hung Lin[b,1], Jinn-Jy Lin[b,c,d], Hsiang June Chen[b], Yu-Ju Lu[b], Yi-Hsuan Wu[b], Mei-Yeh Jade Lu[b], Chen-Hua Lu[e], Arthur Chun-Chieh Shih[e], Maurice Sun-Ben Ku[f,g], Shin-Han Shiu[h,2], Shu-Hsing Wu[i,2], and Wen-Hsiung Li[a,b,j,2]

[a]Biotechnology Center, National Chung-Hsing University, Taichung, Taiwan 40227; [b]Biodiversity Research Center, [c]Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Science, [e]Institute of Information Science, and [i]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan 115; [d]Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan 300; [f]Institute of Bioagricultural Science, National Chiayi University, Chiayi, Taiwan 600; [g]School of Biological Sciences, Washington State University, Pullman, WA 99164; [h]Department of Plant Biology, Michigan State University, East Lansing, MI 48824; and [j]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637

Contributed by Wen-Hsiung Li, January 15, 2015 (sent for review November 25, 2014; reviewed by Justin O. Borevitz)

Maize is a major crop and a model plant for studying C4 photosynthesis and leaf development. However, a genomewide regulatory network of leaf development is not yet available. This knowledge is useful for developing C3 crops to perform C4 photosynthesis for enhanced yields. Here, using 22 transcriptomes of developing maize leaves from dry seeds to 192 h post imbibition, we studied gene up- and down-regulation and functional transition during leaf development and inferred sets of strongly coexpressed genes. More significantly, we developed a method to predict transcription factor binding sites (TFBSs) and their cognate transcription factors (TFs) using genomic sequence and transcriptomic data. The method requires not only evolutionary conservation of candidate TFBSs and sets of strongly coexpressed genes but also that the genes in a gene set share the same Gene Ontology term so that they are involved in the same biological function. In addition, we developed another method to predict maize TF–TFBS pairs using known TF–TFBS pairs in *Arabidopsis* or rice. From these efforts, we predicted 1,340 novel TFBSs and 253 new TF–TFBS pairs in the maize genome, far exceeding the 30 TF–TFBS pairs currently known in maize. In most cases studied by both methods, the two methods gave similar predictions. In vitro tests of 12 predicted TF–TFBS interactions showed that our methods perform well. Our study has significantly expanded our knowledge on the regulatory network involved in maize leaf development.

maize transcriptomes | coexpressed genes | *cis* binding site

Maize (*Zea mays*) is a major crop and a model plant for studying C4 photosynthesis and leaf development. However, the regulatory network that controls maize leaf development is still not well understood. In fact, the number of known maize transcription factor (TF)-binding sites (TFBSs) is far smaller than that for *Arabidopsis thaliana* (1–3).

To understand better the regulation of maize leaf development, several recent studies used next-generation sequencing (NGS) technologies to survey transcriptomic differences among maize leaf cell types and developmental stages. The first large-scale study was by Li et al. (4), who studied the transcriptomes of the base of the blade, the tip, and two middle regions of the third maize leaves in 9-d-old maize plants. The four regions represent different leaf developmental stages, with the base being the youngest and the tip the oldest. Li et al. (4) also obtained the transcriptomes of mesophyll (M) and bundle sheath (BS) cells to study the regulatory and functional differentiation between M and BS cells. A subsequent study by Chang et al. (5) led to the identification of genes differentially expressed between M and BS cells, including two metabolite transporter genes and genes for the specialized BS cell wall structure. Liu et al. (6) obtained the time-course transcriptomes of embryonic leaves every 6 h, starting from dry seeds to hour 72 post imbibition. This set of data provided a global picture of the transcriptional dynamics of genes for early leaf development during maize seed germination and shed light on the succession of biological processes during this period. Wang et al. (7) investigated the transcriptomes of Kranz (i.e., the foliar leaf blade) and non-Kranz (the husk leaf sheath) maize leaves to identify cohorts of genes associated with procambium initiation and vascular patterning. Recently, Wang et al. (8) conducted comparative transcriptomic and metabolomic analyses of developing leaves in maize and rice and identified putative structural and regulatory components important for C4 and C3 photosynthesis. These studies provided insights into the regulatory mechanisms underlying the development of Kranz leaf anatomy in maize.

In the present study, we obtained nine transcriptomes of the second leaf from 84–192 h post imbibition at 12-h (6:00 AM and 6:00 PM) or 24-h (6:00 PM only) intervals. Together with the 13

Significance

Maize is a major crop and a model plant for studying C4 leaf development. However, its regulatory network of leaf development is poorly understood. We used transcriptomes of developing leaves to study gene-expression dynamics and coexpression to reveal functional transition during maize leaf development. More significantly, we developed methods to predict transcription factor-binding sites (TFBSs) and their cognate transcription factors (TFs) or to use the known *Arabidopsis* TF–TFBS pairs to predict the maize TF–TFBS pairs. In total, we predicted 1,340 novel TFBSs and 253 new TF–TFBS pairs in maize. Twelve predicted TF–TFBS interactions were validated by functional tests, suggesting that our methods perform well. Our study has significantly expanded our knowledge of the regulatory network of maize leaf development.

Author contributions: C.-P.Y., S.C.-C.C., Y.-M.C., W.-Y.L., H.-H.L., M.S.-B.K., S.-H.S., S.-H.W., and W.-H.L. designed research; C.-P.Y., S.C.-C.C., Y.-M.C., W.-Y.L., H.-H.L., J.-J.L., H.J.C., Y.-J.L., Y.-H.W., M.-Y.J.L., C.-H.L., A.C.-C.S., M.S.-B.K., S.-H.S., S.-H.W., and W.-H.L. performed research; C.-P.Y., S.C.-C.C., Y.-M.C., J.-J.L., A.C.-C.S., and S.-H.S. analyzed data; and C.-P.Y., S.C.-C.C., Y.-M.C., W.-Y.L., H.-H.L., M.S.-B.K., S.-H.S., S.-H.W., and W.-H.L. wrote the paper.

Reviewers included: J.O.B., Australian National University.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the Sequence Read Archive, www.ncbi.nlm.nih.gov/sra (accession no. SRS654357).

[1]C.-P.Y., S.C.-C.C., Y.-M.C., W.-Y.L., and H.-H.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: shius@msu.edu, shuwu@gate.sinica.edu.tw, or whli@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1500605112/-/DCSupplemental.

transcriptomes of Liu et al. (6), we now have 22 transcriptomes with a time span from hour 0 (dry seeds, T000) to 192 h post imbibition (T192). These time series data are excellent materials for studying the transcriptional dynamics of genes, for revealing which TF genes are up- or down-regulated during the early development of maize embryonic leaves, and particularly for inferring strongly coexpressed genes with potentially common regulatory basis. Moreover, this dataset can be used to predict TFBSs and TF–TFBS interactions, information that is highly valuable because our current knowledge of the maize leaf regulatory network is very limited. To this end, we developed methods for predicting TFBSs and for inferring the cognate TF of a predicted TFBS, and we conducted experimental tests of some predicted TF–TFBS pairs. In addition, when a TF–TFBS pair is known in one species (e.g., *Arabidopsis thaliana*), we developed a method to predict its homologous TF–TFBS pair in maize. Finally, we found that the two approaches of predicting TF–TFBS pairs gave similar results in most of the cases studied.

Results and Discussion

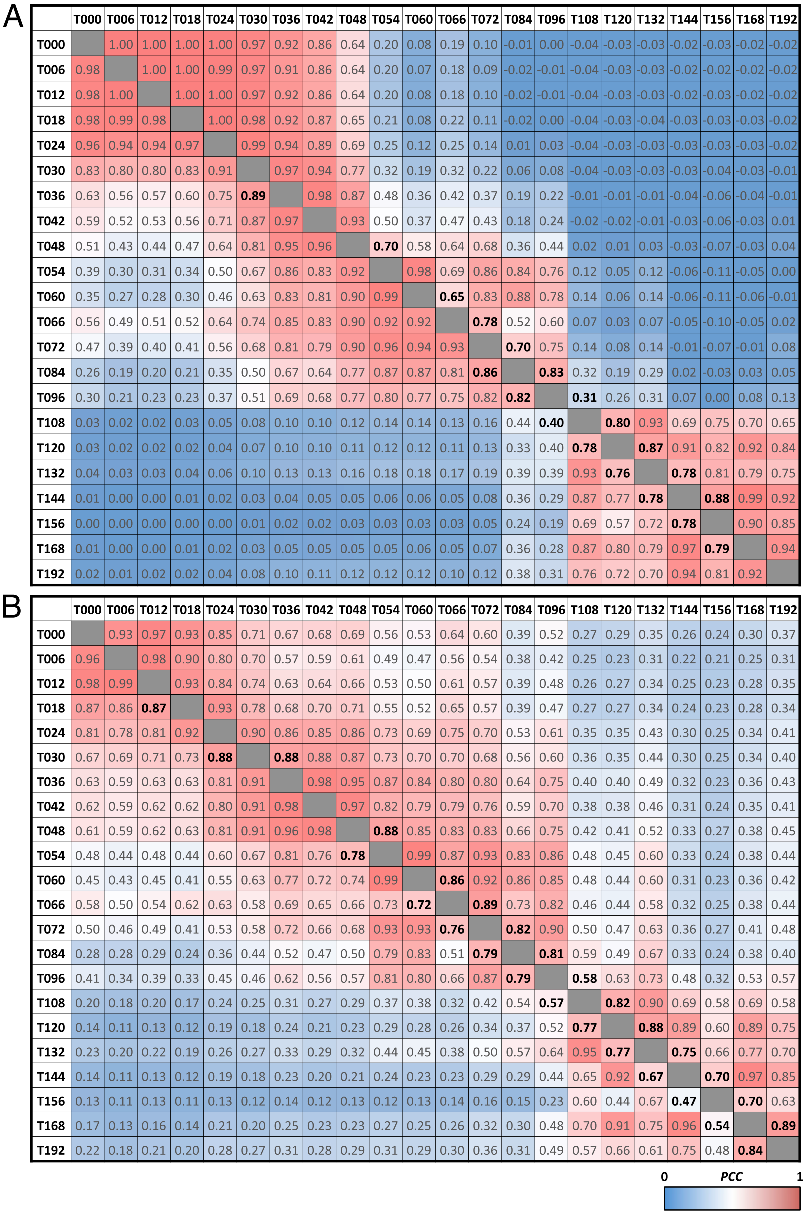
Transcriptome Profiling of Maize Leaf Development. *SI Appendix*, Fig. S1 shows the sampling time points for all 22 transcriptomes and corresponding phenotypes of the germinating seeds and seedlings, and *SI Appendix*, Table S1 shows the numbers of raw reads and mapped reads of the nine newly collected transcriptomes of developing second maize leaves from 84 h post imbibition (T084) to T192. These nine transcriptomes plus the 13 transcriptomes of embryonic leaves from Liu et al. (6) provide an extensive dataset for studying the transcriptome dynamics of genes in developing maize leaves. A gene is defined as expressed if its RPKM (reads per kilobase per million mapped reads) value is ≥1 in at least 2 of the 22 transcriptomes. In total, 32,777 genes, including 1,574 TF genes, were found to be expressed in the 192-h time period.

It should be noted that the cultivar we used for our transcriptomic study is *Zea mays cv.* White Crystal, and the maize reference genome was from B73. This difference would tend to reduce the read mappable rate and thus would tend to underestimate the expression level of a gene if there are sequence differences in the transcribed regions between the two maize genotypes. However, the difference may not significantly affect the pattern of gene-expression dynamics or the expression correlation between genes, which are the major gene-expression characteristics considered in this study. Our remapping procedure allowed two mismatches for a read. This allowance should compensate to some extent for the difference in sequences between the two cultivars. In this study, the mappable rate at each time point (except 77% at T108) was higher than 80% (*SI Appendix*, Table S1), similar to the mappable rates reported in the B73 maize transcriptomes by Chen et al. (9).

For the top 1% highly expressed genes at the 22 time points (1,256 genes), the Pearson correlation coefficient (PCC) in expression levels between two neighboring time points decreased slowly until T048 (PCC >0.9, Fig. 1*A*, above the diagonal). Before T072, two sharp transitions occurred. The first one was from T048 to T054 (PCC = 0.70), and the second was from T060 to T066 (PCC = 0.65). Over the entire time course, the lowest correlation occurred from T096 to T108 (PCC = 0.31), mainly because of the up-regulation of genes related to cell wall and light reaction in photosynthesis and the down-regulation of genes related to cell organization and division, signifying a major developmental transition. In summary, the top 1% of genes showed a similar expression profile from T000 to T048, a low expression correlation from T048 to T108, and a moderate expression correlation from T108 to T192. In contrast, no obvious transition of the bottom 1% of genes (1,523) occurred until T96 (Fig. 1*A*, below the diagonal). The expression of the middle 98% of genes (30,333) (Fig. 1*B*, above the diagonal) was more dynamic over the time course than that of the top 1% of genes. In addition, TF genes showed more dynamic expression changes than the other genes from T012 to T018 (PCC = 0.87), from T024 to T030 (PCC = 0.88), and from T048 to T054 (PCC = 0.78; Fig. 1*B*, below the diagonal), suggesting that TF genes have important roles in germination and early leaf development. In addition, clear transitions were observed for genes in all groups after T072 (Fig. 1), possibly because of larger alternations in the expression of diurnal- or circadian-regulated genes at dawn/dusk junctions.

Coexpression Modules. The developmental time series data are useful for identifying genes that are potentially coregulated and/or involved in the same biological processes. Therefore, we classified the expressed genes into 30 coexpression modules, each containing genes with similar expression patterns (Fig. 2*A*). In each module, overrepresented functional categories were identified based on the MapMan annotation of 16,657 expressed genes (Fig. 2*B*; see also *SI Appendix*, Fig. S2). According to the timing of peak expression, the 30 modules were divided into the first (C1–9), second (C10–23), and third (C24–30) stages (Fig. 2*A* and Dataset S1). The first stage coincided with the transition from seed dormancy to germination, involving extensive physiological changes. C1 genes were highly expressed in dry seeds (T000) and are related to RNA processing and cell vesicle transport. The expression of C2–C7 genes peaked at T006 but was down-regulated gradually after T012. C2–C7 genes include genes related to RNA binding, protein degradation and synthesis, stress, abscisic acid (ABA) and ethylene hormones, and late embryogenesis abundant and storage proteins. Functions related to stress, mitochondrial electron transport, and lipid degradation were overrepresented in C8 (Fig. 2*B*). Many hormone-related TFs that regulate the transition from seed dormancy to germination also belong to C2–C9 (Fig. 2*A*), such as *ABA-INSENSITIVE3*/*VIVIPAROUS1* (*ABI3*/*VP1*), *ABI4*, and *ABI5*, *GIBBERELLIC ACID-INSENSITIVE* (*GAI*), and *ETHYLENE-INSENSITIVE3* (*EIN3*), as is consistent with their antagonistic functions in the ABA and GA/ethylene pathways at this stage (10–12).

The expression peaks at the second stage occurred from T030 to T084. Auxin-related genes, e.g., the auxin efflux carrier *PIN-FORMED 1A* and *1C* (*PIN1A* and *PIN1C*), were overrepresented in C10 and C11, respectively. C12–C15 genes showed an expression peak at T036 and then were expressed in an oscillatory manner after T072, likely because of the influence of day/night cycles. Genes overrepresented in C12–C15 included light signaling in C12, G protein signaling in C13, and signaling of MAP kinases and phosphoinositides in C15, suggesting that maize embryonic leaves became more responsive to light and external or internal stimuli starting at T036. C16–C20 genes started to be gradually up-regulated at T024 and down-regulated after T084 or T096. Genes involved in DNA synthesis and repair, cell organization, cell division, cell cycle, cell wall, and secondary metabolism were overrepresented in these modules (Fig. 2*B*), indicating that the expansion of the second leaf (*SI Appendix*, Fig. S1*B*) likely is accompanied by an increased rate of cell division during this stage. In addition, vascular tissue and BS development are apparent during this period because relevant TFs such as *MONOPTEROS* (*MP*), *HOMEOBOX GENE 8* (*HB8*), *PHABULOSA*/ *PHAVOLUTA* (*PHB*/*PHV*), *REVOLUTA* (*REV*), *SHORTROOT* (*SHR*), and *SCARECROW* (*SCR*) (6, 13, 14) were all found in modules C16–C20 (Fig. 2*A*). *Arabidopsis* SHR–SCR complex has been suggested to control the development of vasculature in all tissues, including the Kranz anatomy in C4 leaves (15–17). Consistent with this hypothesis, the maize SHR and SCR genes not only were highly expressed during the development of Kranz anatomy in embryonic and foliar leaves (7) but also were preferentially expressed in the developing BS strands that contain both BS and vasculature cells (5). Finally, genes in C21 and C22 showed an expression peak at T084 and were overrepresented by many metabolic pathways, including amino acid, lipid, nucleotide,



A

	T000	T006	T012	T018	T024	T030	T036	T042	T048	T054	T060	T066	T072	T084	T096	T108	T120	T132	T144	T156	T168	T192
T000		1.00	1.00	1.00	1.00	0.97	0.92	0.86	0.64	0.20	0.08	0.19	0.10	-0.01	0.00	-0.04	-0.03	-0.03	-0.02	-0.03	-0.02	-0.02
T006	0.98		1.00	1.00	0.99	0.97	0.91	0.86	0.64	0.20	0.07	0.18	0.09	-0.02	-0.01	-0.04	-0.03	-0.03	-0.02	-0.02	-0.02	-0.02
T012	0.98	1.00		1.00	1.00	0.97	0.92	0.86	0.64	0.20	0.08	0.18	0.10	-0.02	-0.01	-0.04	-0.03	-0.03	-0.02	-0.03	-0.02	-0.02
T018	0.98	0.99	0.98		1.00	0.98	0.92	0.87	0.65	0.21	0.08	0.22	0.11	-0.02	0.00	-0.04	-0.03	-0.04	-0.03	-0.03	-0.03	-0.02
T024	0.96	0.94	0.94	0.97		0.99	0.94	0.89	0.69	0.25	0.12	0.25	0.14	0.01	0.03	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.02
T030	0.83	0.80	0.80	0.83	0.91		0.97	0.94	0.77	0.32	0.19	0.32	0.22	0.06	0.08	-0.04	-0.03	-0.03	-0.03	-0.04	-0.03	-0.01
T036	0.63	0.56	0.57	0.60	0.75	**0.89**		0.98	0.87	0.48	0.36	0.42	0.37	0.19	0.22	-0.01	-0.01	-0.01	-0.04	-0.06	-0.04	-0.01
T042	0.59	0.52	0.53	0.56	0.71	0.87	0.97		0.93	0.50	0.37	0.47	0.43	0.18	0.24	-0.02	-0.02	-0.01	-0.03	-0.06	-0.03	0.01
T048	0.51	0.43	0.44	0.47	0.64	0.81	0.95	0.96		**0.70**	0.58	0.64	0.68	0.36	0.44	0.02	0.01	0.03	-0.03	-0.07	-0.03	0.04
T054	0.39	0.30	0.31	0.34	0.50	0.67	0.86	0.83	0.92		0.98	0.69	0.86	0.84	0.76	0.12	0.05	0.12	-0.06	-0.11	-0.05	0.00
T060	0.35	0.27	0.28	0.30	0.46	0.63	0.83	0.81	0.90	0.99		**0.65**	0.83	0.88	0.78	0.14	0.06	0.14	-0.06	-0.11	-0.06	-0.01
T066	0.56	0.49	0.51	0.52	0.64	0.74	0.85	0.83	0.90	0.92	0.92		**0.78**	0.52	0.60	0.07	0.03	0.07	-0.05	-0.10	-0.05	0.02
T072	0.47	0.39	0.40	0.41	0.56	0.68	0.81	0.79	0.90	0.96	0.94	0.93		**0.70**	0.75	0.14	0.08	0.14	-0.01	-0.07	-0.01	0.08
T084	0.26	0.19	0.20	0.21	0.35	0.50	0.67	0.64	0.77	0.87	0.87	0.81	**0.86**		**0.83**	0.32	0.19	0.29	0.02	-0.03	0.03	0.05
T096	0.30	0.21	0.23	0.23	0.37	0.51	0.69	0.68	0.77	0.80	0.77	0.75	0.82	**0.82**		**0.31**	0.26	0.31	0.07	0.00	0.08	0.13
T108	0.03	0.02	0.02	0.03	0.05	0.08	0.10	0.10	0.12	0.14	0.14	0.13	0.16	0.44	**0.40**		**0.80**	0.93	0.69	0.75	0.70	0.65
T120	0.03	0.02	0.02	0.02	0.04	0.07	0.10	0.10	0.11	0.12	0.12	0.11	0.13	0.33	0.40	**0.78**		**0.87**	0.91	0.82	0.92	0.84
T132	0.04	0.03	0.03	0.04	0.06	0.10	0.13	0.13	0.16	0.18	0.18	0.17	0.19	0.39	0.39	0.93	**0.76**		**0.78**	0.81	0.79	0.75
T144	0.01	0.00	0.00	0.01	0.02	0.03	0.04	0.05	0.05	0.06	0.06	0.05	0.08	0.36	0.29	0.87	0.77	**0.78**		**0.88**	0.99	0.92
T156	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.02	0.03	0.03	0.03	0.03	0.05	0.24	0.19	0.69	0.57	0.72	**0.78**		0.90	0.85
T168	0.01	0.00	0.00	0.01	0.02	0.03	0.05	0.05	0.05	0.05	0.06	0.05	0.07	0.36	0.28	0.87	0.80	0.79	0.97	**0.79**		0.94
T192	0.02	0.01	0.02	0.02	0.04	0.08	0.10	0.11	0.12	0.12	0.12	0.10	0.12	0.38	0.31	0.76	0.72	0.70	0.94	0.81	0.92	

B

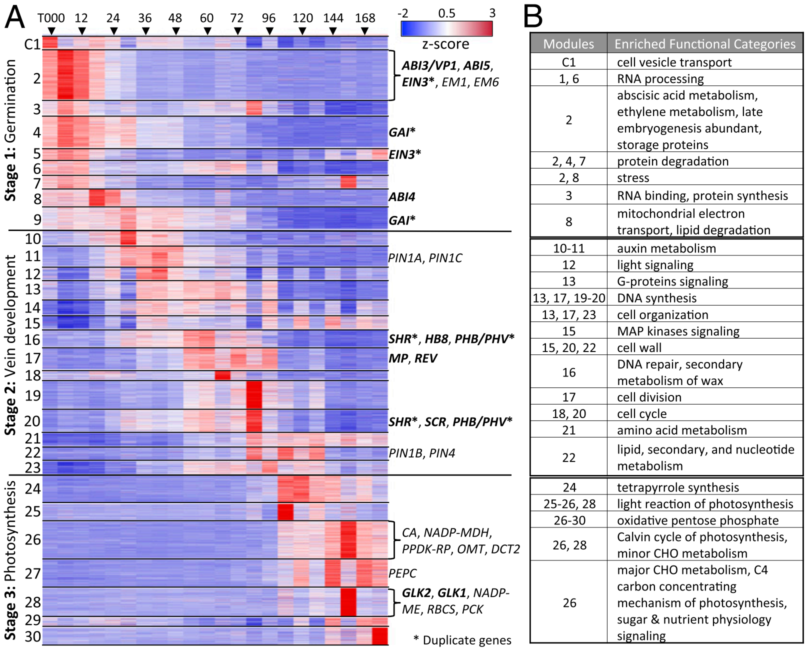
	T000	T006	T012	T018	T024	T030	T036	T042	T048	T054	T060	T066	T072	T084	T096	T108	T120	T132	T144	T156	T168	T192
T000		0.93	0.97	0.93	0.85	0.71	0.67	0.68	0.69	0.56	0.53	0.64	0.60	0.39	0.52	0.27	0.29	0.35	0.26	0.24	0.30	0.37
T006	0.96		0.98	0.90	0.80	0.70	0.57	0.59	0.61	0.49	0.47	0.56	0.54	0.38	0.42	0.25	0.23	0.31	0.22	0.21	0.25	0.31
T012	0.98	0.99		0.93	0.84	0.74	0.63	0.64	0.66	0.53	0.50	0.61	0.57	0.39	0.48	0.26	0.27	0.34	0.25	0.23	0.28	0.35
T018	0.87	0.86	**0.87**		0.93	0.78	0.68	0.70	0.71	0.55	0.52	0.65	0.57	0.39	0.47	0.27	0.27	0.34	0.24	0.23	0.28	0.34
T024	0.81	0.78	0.81	0.92		0.90	0.86	0.85	0.86	0.73	0.69	0.75	0.70	0.53	0.61	0.35	0.35	0.43	0.30	0.25	0.34	0.41
T030	0.67	0.69	0.71	0.73	**0.88**		**0.88**	0.88	0.87	0.73	0.70	0.70	0.68	0.56	0.60	0.36	0.35	0.44	0.30	0.25	0.34	0.40
T036	0.63	0.59	0.63	0.63	0.81	0.91		0.98	0.95	0.87	0.84	0.80	0.80	0.64	0.75	0.40	0.40	0.49	0.32	0.23	0.36	0.43
T042	0.62	0.59	0.62	0.62	0.80	0.91	0.98		0.97	0.82	0.79	0.79	0.76	0.59	0.70	0.38	0.38	0.46	0.31	0.24	0.35	0.41
T048	0.61	0.59	0.62	0.63	0.81	0.91	0.96	0.98		**0.88**	0.85	0.83	0.83	0.66	0.75	0.42	0.41	0.52	0.33	0.27	0.38	0.45
T054	0.48	0.44	0.48	0.44	0.60	0.67	0.81	0.76	**0.78**		0.99	0.87	0.93	0.83	0.86	0.48	0.45	0.60	0.33	0.24	0.38	0.44
T060	0.45	0.43	0.45	0.41	0.55	0.63	0.77	0.72	0.74	0.99		**0.86**	0.92	0.86	0.85	0.48	0.44	0.60	0.31	0.23	0.36	0.42
T066	0.58	0.50	0.54	0.62	0.63	0.58	0.69	0.65	0.66	0.73	**0.72**		**0.89**	0.73	0.82	0.46	0.44	0.58	0.32	0.25	0.38	0.44
T072	0.50	0.46	0.49	0.41	0.53	0.58	0.72	0.66	0.68	0.93	0.93	**0.76**		**0.82**	0.90	0.50	0.47	0.63	0.36	0.27	0.41	0.48
T084	0.28	0.28	0.29	0.24	0.36	0.44	0.52	0.47	0.50	0.79	0.83	0.51	**0.79**		**0.81**	0.59	0.49	0.67	0.33	0.24	0.38	0.40
T096	0.41	0.34	0.39	0.33	0.45	0.46	0.62	0.56	0.57	0.81	0.80	0.66	0.87	**0.79**		**0.58**	0.63	0.73	0.48	0.32	0.53	0.57
T108	0.20	0.18	0.20	0.17	0.24	0.25	0.31	0.27	0.29	0.37	0.38	0.32	0.42	0.54	**0.57**		**0.82**	0.90	0.69	0.58	0.69	0.58
T120	0.14	0.11	0.13	0.12	0.19	0.18	0.24	0.21	0.23	0.29	0.28	0.26	0.34	0.37	0.52	**0.77**		**0.88**	0.89	0.60	0.89	0.75
T132	0.23	0.20	0.22	0.19	0.26	0.27	0.33	0.29	0.32	0.44	0.45	0.38	0.50	0.57	0.64	0.95	**0.77**		**0.75**	0.66	0.77	0.70
T144	0.14	0.11	0.13	0.12	0.19	0.18	0.23	0.20	0.21	0.24	0.23	0.23	0.29	0.29	0.44	0.65	0.92	**0.67**		**0.70**	0.97	0.85
T156	0.13	0.11	0.13	0.11	0.13	0.12	0.14	0.12	0.13	0.12	0.13	0.14	0.16	0.15	0.23	0.60	0.44	0.67	**0.47**		**0.70**	0.63
T168	0.17	0.13	0.16	0.14	0.21	0.20	0.25	0.23	0.23	0.27	0.25	0.26	0.32	0.30	0.48	0.70	0.91	0.75	0.96	**0.54**		**0.89**
T192	0.22	0.18	0.21	0.20	0.28	0.27	0.31	0.28	0.29	0.31	0.29	0.30	0.36	0.31	0.49	0.57	0.66	0.61	0.75	0.48	**0.84**	

Fig. 1. Expression correlations between time points. (*A*) Heatmaps of PCCs between time points for the genes at the top 1% expression level (above the diagonal) and for the genes at the bottom 1% expression level (below the diagonal). (*B*) Heatmap for all genes excluding those with the top and bottom 1% expression level (above the diagonal) and heatmap for the TF genes (below the diagonal). High PCC values are shown in red, and low PCC values are shown in blue (see the color bar at the bottom of the figure). PCC values <0.9 between neighboring time points are highlighted in black boldface type.

and secondary metabolisms. Thus, at the second stage, leaf development seems to be regulated mainly by light and auxin signaling (Fig. 2*B*) (18, 19).

The third stage includes modules C24–C30 with the expression peak at T108 or later. The genes in C24 and C25 that contributed most to this sharp transition tend to be involved in tetrapyrrole synthesis and light reaction of photosynthesis. Genes in C26–C30 were overrepresented by oxidative pentose phosphate, carbohydrate metabolism, light reaction, Calvin cycle, and C4 carbon-concentrating mechanism in photosynthesis, signaling active carbon fixation, and energy metabolism (Fig. 2*B*). The overrepresentation of photosynthesis-related genes during this third stage indicates high photosynthetic activity at T108 and later. Furthermore, key C4 enzymes, e.g., *NADP-DEPENDENT MALIC ENZYME* (*NADP-ME*), *PPDK REGULATORY PROTEIN* (*PPDK-RP*), and *PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PEPC*), and transporters *2-OXOGLUTARATE/MALATE TRANSPORTER* (*OMT*) and *GENERAL DICARBOXYLATE TRANSPORTER2* (*DCT2*) were found in C25, C26, and C27 (Fig. 2*A*), suggesting that C4 photosynthesis fully turns on during the third stage.

Differential Gene Expression Between Time Points. To study functional transitions further, we followed Liu et al. (6) to identify two types of differentially expressed genes (DEGs) between time points (Fig. 3). Type 1 DEGs are differentially expressed in a time point and the preceding time point, signifying a significant transition in gene expression in 6 h (T000–T072), 12 h (T072–T168), or 24 h (T168–T192). Notably, type 1 DEG numbers at later time points, especially T084, T096, T108, T144, T156, and T168, tend to be larger than those at earlier time points (Fig. 3 *A* and *C*). One reason for this phenomenon is that the time difference between two time points is larger at later than at earlier



Modules	Enriched Functional Categories
C1	cell vesicle transport
1, 6	RNA processing
2	abscisic acid metabolism, ethylene metabolism, late embryogenesis abundant, storage proteins
2, 4, 7	protein degradation
2, 8	stress
3	RNA binding, protein synthesis
8	mitochondrial electron transport, lipid degradation
10-11	auxin metabolism
12	light signaling
13	G-proteins signaling
13, 17, 19-20	DNA synthesis
13, 17, 23	cell organization
15	MAP kinases signaling
15, 20, 22	cell wall
16	DNA repair, secondary metabolism of wax
17	cell division
18, 20	cell cycle
21	amino acid metabolism
22	lipid, secondary, and nucleotide metabolism
24	tetrapyrrole synthesis
25-26, 28	light reaction of photosynthesis
26-30	oxidative pentose phosphate
26, 28	Calvin cycle of photosynthesis, minor CHO metabolism
26	major CHO metabolism, C4 carbon concentrating mechanism of photosynthesis, sugar & nutrient physiology signaling

Fig. 2. Heatmap of gene-expression levels and enriched functional categories for coexpression modules. (*A*) Each gene (row) in the 30 coexpression modules (C1–C30) is sorted according to the time point (column) at which peak expression occurred. All gene expression levels were transformed to z-scores and are colored blue, white, or red to represent low, moderate, or high expression levels, respectively. The 30 modules can be divided into three groups that correspond to three stages represented by genes in three key processes: germination, vein development, and photosynthesis. Selected TFs (in bold face) or genes related to the three stages are listed to the right of the corresponding modules. (*B*) Each entry of table represents the enriched functions (*Right*) of the corresponding modules (*Left*). Gene names: *ABI3*, *4*, *5*: *ABA-INSENSITIVE3*, *4*, *5*; *CA*: *CARBONIC ANHYDRASE*; *DCT2*: *GENERAL DICARBOXYLATE TRANSPORTER2*; *EIN3*: *ETHYLENE-INSENSITIVE3*; *EM1*, *6*: *LATE EMBRYOGENESIS ABUNDANT1*, *6*; *GAI*: *GA-INSENSITIVE*; *GLK1*, *2*: *GOLDEN2-LIKE 1*, *2*; *HB8*: *HOMEOBOX GENE 8*; *MP*: *MONOPTEROS*; *NADP-MDH*: *NADP-MALATE DEHYDROGENASE*; *NADP-ME*: *NADP-DEPENDENT MALIC ENZYME*; *OMT*: *2-OXOGLUTARATE/MALATE TRANSPORTER*; *PCK*: *PHOSPHOENOLPYRUVATE CARBOXYKINASE*; *PEPC*: *PHOSPHOENOLPYRUVATE CARBOXYLASE*; *PHB/PHV*: *PHABULOSA/ PHAVOLUTA*; *PIN4*, *1A*, *1B*, *1C*: *PIN-FORMED4*, *1A*, *1B*, *1C*; *PPDK-RP*: *PPDK REGULATORY PROTEIN*; *RBCS*: *RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL SUBUNIT*; *REV*: *REVOLUTA*; *SCR*: *SCARECROW*; *SHR*: *SHORTROOT*; *VP1*: *VIVIPAROUS1*.

time points (12 h vs. 6 h). Also, it is possible that the expression levels of many genes were more strongly affected at later time points by the day/night cycle and/or by photosynthesis-related processes (Fig. 2 and *SI Appendix*, Fig. S3*A*). For example, genes involved in photosystem I were up-regulated in 24-h cycles from T084 (6:00 AM) to T156, and those in photosystem II were down-regulated from T096 (6:00 PM) to T168. In addition, the genes encoding Pseudo ARABIDOPSIS RESPONSE REGULATOR TYPE B (ARR-B)/PSEUDO RESPONSE REGULATOR (PRR), the Rubisco large subunit, and the CONSTANT (CO)-Like zinc finger family also showed periodic expression from T096 or T108 on. *Pseudo ARR-B*/*PRR* genes, such as *TIMING OF CAB EXPRESSION1* (*TOC1*/*PRR1*), *PRR9*, *PRR7*, and *PRR5*, are key regulators of the circadian clock (20). Thus, the circadian regulation in early leaf development was prominent after being entrained by 12-h light/12-h darkness for a few days.

Type 2 DEGs are genes differentially expressed between a time point and T000 (dry seeds), representing genes that are up- or down-regulated for the first time after imbibition (Fig. 3 *B* and *D*). These genes potentially signify the onset/termination of developmental and/or physiological processes at a particular stage. Several genes encoding ethylene-responsive APETALA2 (AP2) and ETHYLENE RESPONSE FACTOR (ERF) TF families were up-regulated by T018, suggesting their involvement in germination. Also maize orthologs of rice genes that are induced by GA treatment, including those encoding early nodulin (GRMZM2G131421/GRMZM2G147399) and cationic peroxidase (GRMZM2G108207) (21), were up-regulated by T024 or T036. Orthologs of *Arabidopsis* TF genes known to modulate cell proliferation in organ growth, including *AINTEGUMENTA* (*ANT*) (22) and *GROWTH REGULATING FACTOR5* (*GRF5*) (23), also were up-regulated by T024. Importantly, TF genes implicated in auxin responses and vein initiation/differentiation were up-regulated between T018 and T054, including many *AUXIN RESPONSE FACTORs* (*ARFs*) and orthologs of *Arabidopsis LONESOME HIGHWAY* (*LWH*), *ATHB8*, *AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN3* (*AHL3*), and *AHL4* (24–26). Likewise, non-TF genes known to be involved in auxin signal transduction and responses were up-regulated at the early developmental stage, including *AUXIN BINDING PROTEIN 1* (*ABP1*) and auxin transporters (e.g., *PIN1* and *PIN4*) (27–30). These genes are likely involved in active cell division and in the formation of new cell types, such as vasculatures and BS cells.

For down-regulation, genes involved in ABA biosynthesis or response were down-regulated early, from T024 or later, including *VP1* and maize orthologs of *Arabidopsis ABI5*. The decreased expression of these genes leads to the breaking of seed dormancy. Genes that promote the transition from vegetative to reproductive development are expected to be repressed at this stage. For example, the ortholog of *Arabidopsis SQUAMOSA PROMOTER BINDING PROTEIN* (*SBP*)*-Like2* (*SBL2*), which controls lateral organ development in association with shoot maturation in the reproductive phase (31), was down-regulated at T024. The down-regulation of a negative regulator for xylem cell specification, the ortholog (GRMZM2G083347) of *VASCULAR-RELATED NAC-DOMAIN INTERACTING2* (*VNI2*), at T036 likely initiates xylem development.

Expression Dynamics of TF Gene Families During Germination and Early Leaf Development. Because genes in a TF family may serve similar functions, when a TF gene is down-regulated, its regulatory role may be assumed by another TF gene in the same family. Thus, it is interesting to consider the total expression level (total

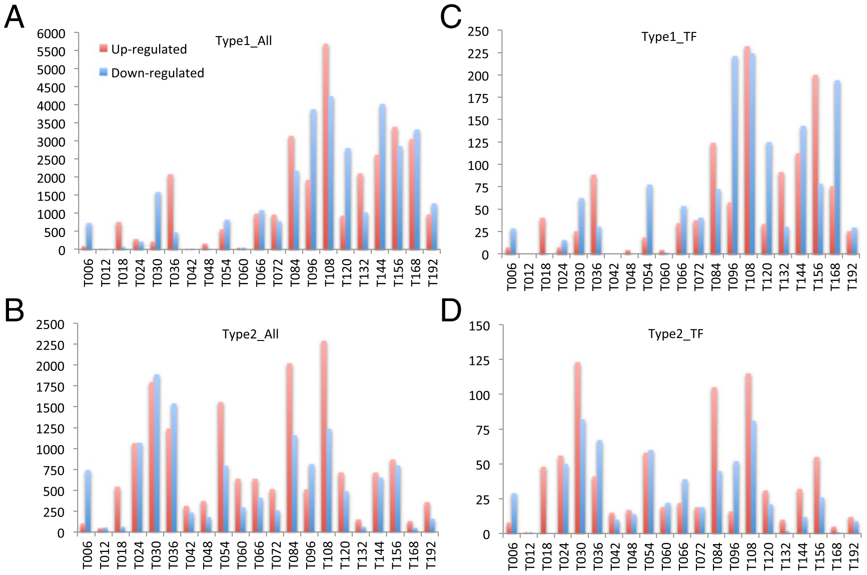


Fig. 3. Numbers of genes differentially expressed at two time points. (*A*) Type 1 DEGs. Type 1 DEGs at time T_i were the genes that showed differential expression between T_i and T_{i-1}, where T_i denotes time point I and T_{i-1} denotes the preceding time point. The *y* axis denotes the number of genes. (*B*) Type 2 DEGs. Type 2 DEGs at T_i were the genes that were not differentially expressed until time T_i; that is, T_i was the first time since time 0 that these genes became significantly up- or down-regulated. (*C*) Type 1 TF DEGs. (*D*) Type 2 TF DEGs.

RPKM) of the TF genes in a family over the time course. To assess which TF gene families are more highly expressed and thus are likely play a more prominent role than other families at a specific stage of development, we calculated the proportion (P) of the total RPKM of each TF gene family relative to the total RPKM of all TF genes in a transcriptome (the red-white-green line in each row in Fig. 4 and *SI Appendix*, Table S2). Because P is family-size dependent, we also presented the most highly expressed gene (i.e., the gene with the highest average RPKM over the 22 transcriptomes) in each TF gene family (the yellow-white-black line in each row in Fig. 4); the IDs of the most highly expressed genes are listed in the figure legend.

Highly expressed TF gene families. Five TF gene families [ERF, BASIC LEUCINE ZIPPER (bZIP), NAM, ATAF1/2, and CUC2 (NAC), CCCH ZINC FINGER (C3H), and HEAT STRESS TRANSCRIPTION FACTOR (HSF)] were highly expressed ($P > 5\%$) in dry seeds and at the early stage of germination (T000–T024) (Fig. 4 and *SI Appendix*, Table S2). The ERF, bZIP, and C3H gene families were highly expressed over the entire time course, suggesting that they play diverse roles during seed germination and early leaf development. ERFs are involved in embryo development and in stress and ethylene responses (32). About half (104) of the 205 ERF genes were expressed over the entire time course (Fig. 4). They showed a total *P* value >10% from T000 to T048, perhaps reflecting the requirement for ethylene action in germination. bZIP proteins control seed and leaf development, nitrogen/carbon balance, and photomorphogenesis (33–35), whereas C3H proteins are implicated in embryogenesis (36) and stress response (37). The NAC gene family maintained high expression ($P > 5\%$) until T030. NAC TFs modulate stress response, meristem boundary formation (38), auxin signaling (39), and xylem development (40). The HSF gene family showed high expression from T000 to T018 and at T066, suggesting involvement in other functions in addition to heat stress response (41).

As leaf development proceeded, several TF gene families became highly expressed ($P > 5\%$), including C2H2 ZINC FINGER (C2H2) (T036–T096), BASIC/HELIX–LOOP–HELIX (bHLH) (T036–T192), HIGH MOBILITY GROUP BOX (HMG) (T054–T192), and MYELOBLASTOSIS (MYB)-related (T108–T192). The high expression of *C2H2* genes, including the ortholog of the *Arabidopsis DEFECTIVELY ORGANIZED TRIBUTARIES5* (*DOT5*), which is involved in vascular development (42), during T036–T096 coincides with the timing of vascular tissue development (6). bHLH proteins regulate stomatal development, leaf expansion, epidermal cell fate determination, and light responses (43). HMG proteins are involved in germination, seedling growth, and stress response (44, 45). The top *HMG* gene was highly expressed through the entire 192-h time course, suggesting that it regulates multiple processes. MYB-related proteins regulate seed/leaf development and circadian rhythm (46, 47), and their up- and down-regulation time points were mostly synchronized with those of other photosynthesis genes (*SI Appendix*, Fig. S3*A*).

Moderately highly expressed TF gene families. Some TF families were moderately expressed ($1\% < P < 5\%$) from T000 through T192, including B3, GATA, Trihelix, ARF, GLABROUS1 ENHANCER-BINDING PROTEIN (GeBP), MYB, WRKY, *GAI*, *REPRESSOR OF GAI* and *SCR* (GRAS), HOMEODOMAIN- LEUCINE ZIPPER (HD-ZIP), METHYL-CPG-BINDING DOMAIN (MBD), GOLDEN2 (G2)-like, and MITOCHONDRIAL TRANSCRIPTION TERMINATION FACTOR (mTERF). B3 proteins interact with auxin or ABA signaling pathways to regulate embryogenesis and the development of shoot meristem or leaf shape (48, 49). The most highly expressed *B3* gene, coding for *VP1*, regulates seed maturation and germination (50). MYB proteins are involved in primary and secondary metabolism, cell fate and identity, developmental processes, and biotic and abiotic stress responses (51). GRAS proteins mediate GA-responsive plant growth and Kranz anatomy development in maize (15, 52). HD-ZIP proteins are involved in leaf and vascular development or meristem maintenance (53); the most highly expressed gene is orthologous to *ATHB5*, which mediates ABA responsiveness in developing seedlings (54). TFs in this category may play a key role in the differentiation of vascular cells and the development of Kranz anatomy.

Many families were moderately highly expressed only at specific developmental stages. Five of these families [GRF, LYSINE-SPECIFIC DEMETHYLASE (LSD), NODULE INCEPTION (Nin)-like, M-Type and EIN3-LIKE (EIL)] exhibited moderately high expression in dry seeds, but their expression level decreased as development proceeded. Two TF gene families exhibited moderately high expression during seed germination and in photosynthesis: AP2 proteins (at T000–T054 and T108–T192), which control embryo/flower development (32), and B-box proteins, including the DOUBLE B-BOX (DBB) and CO-like families (at T000–T024 and T108–T192), which are output genes of circadian rhythm (55, 56) and indeed showed periodicity in expression.

Eight families became moderately highly expressed soon after imbibition or at a later stage. Many of them are implicated in development, such as LATERAL ORGAN BOUNDARIES (LBD) (T018–T048) in organ boundary determination (57), SBP (T030–T066) in vegetative phase change, branching, and leaf initiation rate (58), HMGI/HMGY (T030–T096, T132) in vascular tissue patterning (26), YABBY (T054–T072) in the regulation of abaxial–adaxial polarity (59), TEOSINTE BRANCHED1, CYCLOIDEA and PROLIFERATING CELL FACTORS (TCP) (T030–T132) in cell proliferation (60), and THREE-AMINO ACID-LOOP-EXTENSION (TALE) (T036–T048 and T096–T192) in meristem formation or maintenance and secondary cell wall biosynthesis (61, 62). The most highly expressed *TALE* gene is orthologous to *KNOTTED-LIKE FROM ARABIDOPSIS3* (*KNAT3*), which is light- and cytokinin-regulated and which modulates ABA response during germination and early seedling development (61, 63).

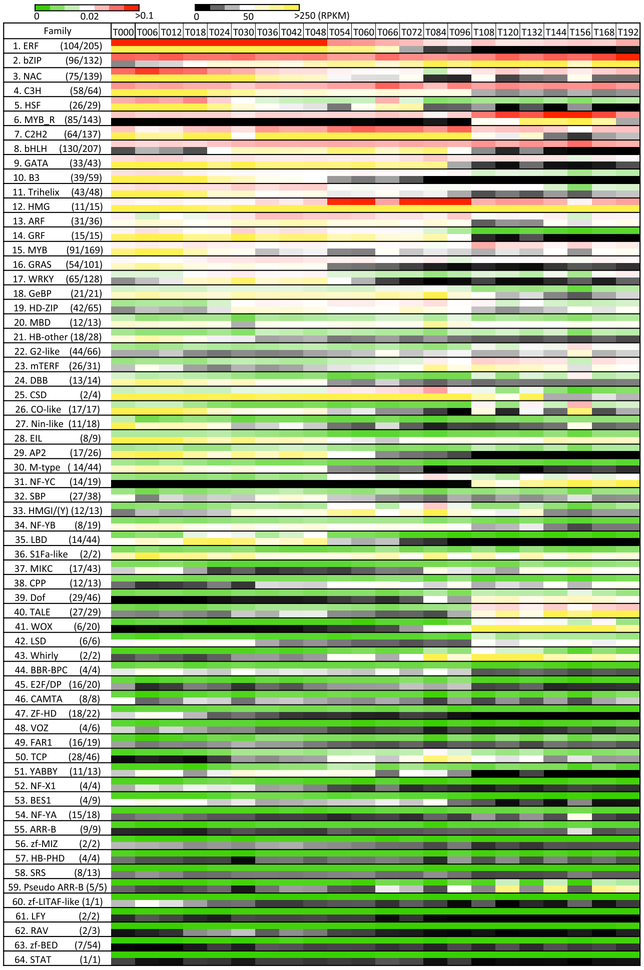


Fig. 4. Expression dynamics of TF families. The expression dynamics of each TF gene family are shown in two lines. The upper line (green-white-red) shows the proportion of the total RPKM of all TF genes contributed by the members of a particular TF gene family in a transcriptome. The lower line (black-white-yellow) shows the RPKM of the highest expressed TF gene in the TF family. In each row, the first number in parentheses refers to the number of expressed genes in a TF family, and the second number refers to the total number of genes in the TF gene family in question. MYB_R: MYB-related. The gene IDs of the highest expressed genes are (with the prefix GRMZM2G except for family numbers 6, 12, 25, 39, and 53) 1: 053503; 2: 479885; 3: 347043; 4: 071034; 5: 105348; 6: AC204212.4_FG001; 7: 159032; 8: 128807; 9: 114775; 10: 133398 (*VP1*); 11: 481163; 12: GRMZM5G834758; 13: 378580; 14: 178261; 15: 149958; 16: 431309; 17: 120320; 18: 083886; 19: 056600 (*ATHB5*); 20: 157470; 21: 111204; 22: 173943; 23: 113181; 24: 028594; 25: GRMZM5G895313; 26: 038783; 27: 105004; 28: 033570; 29: 086573; 30: 320549; 31: 089812; 32: 169270; 33: 106133; 34: 064426; 35: 386674; 36: 147424; 37: 059102; 38: 701689; 39: AC233935.1_FG005; 40: 055243 (*KNAT3*); 41: 010929; 42: 173425; 43: 155662; 44: 179366; 45: 060000; 46: 341747; 47: 089619; 48: 111696; 49: 155980; 50: 120151; 51: 529859; 52: 087787; 53: GRMZM5G812774; 54: 361842; 55: 479110; 56: 075582; 57: 314546; 58: 136700; 59: 095727 (*PRR7*); 60: 010235; 61: 180190; 62: 169654; 63: 317835; and 64: 453001.

Finally, the TF families that respond to light signals or stress became moderately highly expressed from T084 or T108 on. These include DNA BINDING WITH ONE FINGER (DOF) proteins (T084–T168) in light responses (64) and LSD and WUS HOMEOBOX-CONTAINING (WOX) proteins (T108–T168) in cell death under oxidative stress (65) and in embryonic patterning, stem-cell maintenance, and organ formation (66), respectively. The *Pseudo ARR-B*/*PRR* genes (T120, T144, T168, and T192) exhibit circadian expression. The top *Pseudo ARR-B*/*PRR* gene is orthologous to *Arabidopsis PPR7*, which negatively regulates the expression of key morning genes *CLOCK ASSOCIATED1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*) (67).

In summary, TF gene families involved in early germination, stress responses, or photosynthesis tend to be highly expressed, whereas those involved in development tend to be moderately expressed. Additionally, our data revealed distinct circadian expression patterns of some TF genes; for example, a few *CO*-like, *DBB*, and *MYB*-related TF genes (e.g., *LHY*) showed expression peak at 6:00 AM, whereas *Pseudo ARR-B*/*PRR* TF genes showed peak expression at 6:00 PM. Also, both *G2-LIKE1* (*GLK1*) and *GLK2*, which are involved in maize M and BS chloroplast development (5, 6), showed a strict light/darkness modulation of expression.

Predicted TFBSs and Their Cognate TFs. The 22 developing leaf transcriptomes provide a sufficiently large number of time points for inferring sets of strongly coexpressed genes that can be used to predict TFBSs (*Materials and Methods*). For each of these gene sets, a subset of genes that share the same Gene Ontology (GO) term are selected, so that the genes in each subset not only are coexpressed but also are involved in the same biological function. To uncover TFBSs, for each subset, we identified sequence motifs that are overrepresented among the genes of the subset and are mapped to locations well conserved among maize, *Sorghum bicolor*, *Setaria italica*, *Oryza sativa*, and *Brachypodium distachyon* (*Materials and Methods*). We predicted 1,459 maize motifs, which are called "putative TFBSs" (pTFBSs) (*SI Appendix*, Fig. S4 and Datasets S2 and S3). We checked the overlaps of our pTFBSs with the 33 known maize TFBSs (1–3), which actually represent only 20 nonredundant TFBSs because some of them are recognized by the same TFs. Among these 20 TFBSs, 14 share significant similarity with 119 pTFBSs (*P* value <0.001) (*SI Appendix*, Table S3). Many of the remaining 1,340 pTFBSs are likely novel maize TFBSs, significantly expanding the potential *cis*-regulatory landscape in maize.

We next asked which TFs likely bind these pTFBSs. Assuming that TF binding specificity is largely conserved across species, we used each maize pTFBS to find the best-matching TFBS in TRANSFAC (2) (www.gene-regulation.com/pub/databases.html), JASPAR (3) (jaspar.genereg.net/), and AthaMap (1) (www.athamap.de/) or in a collection of plant protein-binding microarray datasets (*Materials and Methods* and *SI Appendix*, Fig. S5) (68, 69). If a significantly similar TFBS existed in the TF–TFBS interaction datasets, its corresponding TF was used to find the maize homologous TFs. Each of these maize TFs then was used to identify a set of coexpressed maize genes and to predict its TFBSs. Thus, for a given maize pTFBS, we might obtain a set of potential maize cognate TFs with predicted TFBSs. We chose the TF with the predicted TFBS best matching the given pTFBS as the cognate TF of the given pTFBS. In total, we predicted 176 cognate TFs for the pTFBSs in Dataset S3. Hereafter, we refer to this prediction method as "Method 1."

Maize TF–TFBS Pairs Inferred from Known TF–TFBS Pairs in *Arabidopsis* and Rice. A second approach to identifying putative TF–TFBS interactions in maize is to start with interaction data from another species. We found 122 TF–TFBS pairs from the *Arabidopsis* TF databases (TRANSFAC, JASPAR, and AthaMap), 108 TFBSs for 63 *Arabidopsis* TFs obtained by the protein-binding microarray technique (68), and 254 TFBSs for 240 *Arabidopsis* TFs from the *Catalog of Inferred Sequence Binding Proteins* (CIS-BP; cisbp.ccbr.utoronto.ca/) (69). By combining the three sources, a total of 353 different *Arabidopsis* TF–TFBS pairs were obtained. In addition, we included 36 TF–TFBS pairs of *O. sativa* from the TF databases and CIS-BP. These TF–TFBS pairs were used to find the corresponding TF–TFBS pairs in maize as follows.

The DNA-binding domains (DBDs) in each *Arabidopsis* or rice TF protein were used to find maize TFs that contain the DBDs. For the 353 *Arabidopsis* TFs, 287 have homologous TFs in maize (BLAST E-value $<10^{-20}$), 286 of which were considered expressed in our time series data. For each of these 286 maize TF genes, the genes coexpressed with the TF gene (PCC >0.8) were enriched with a GO term and were selected to form a strongly coexpressed gene set (*Materials and Methods*). Using these maize gene sets and the new method we developed in this study (*Materials and Methods* and *SI Appendix*, Fig. S6), we predicted 219 TF–TFBS pairs in maize (Datasets S2 and S4). Moreover, using the 36 rice TFs, we found 30 homologous maize TFs that were expressed in our transcriptomes and predicted 20 TF–TFBS pairs in maize (Dataset S5). Thus, we predicted 239 maize TF–TFBS pairs for the 316 (286 + 30) homologous maze TFs, allowing a prediction rate of 76%. These pairs represent 135 nonredundant maize TFs.

There were 57 overlaps between the 135 TFs identified using the above method and the 176 TFs identified using Method 1 detailed in the previous section. In one case (an NAC TF), the TFBS models disagreed between AthaMap and CIS-BP, and this TF was not considered further. Eight of the remaining 56 cases had a *P* value between 0.01 and 0.2 (*SI Appendix*, Table S4), using a motif comparisons tool (70). Five of these eight cases were MYB TFs that have two known core sequences with either GTGGT or GTAGGT, and three were ERF TFs that have diverse TFBSs with a low-complexity C/G core sequence, so that in each case the prediction of TFBS was difficult. In the remaining 48 cases, the maize TFBSs predicted for a TF by the two methods were similar (*P* value <0.01). Thus, in general, the two methods give consistent predictions. Removing the overlaps between the two methods, we obtained 253 new TF–TFBS pairs in maize, which represent 29 of the 64 maize TF families (71). Hereafter, we refer to this method as "Method 2."

Experimental Validation of Predicted TF–TFBS Interactions. To verify the authenticity of predicted TF–TFBS interactions, we selected 12 cases predicted by Method 1 for experimental validation with EMSA (Fig. 5). In each case, in the absence of the predicted cognate TF protein, only the fast-migrating, free biotin-labeled TFBS probe was observed. When the purified TF protein was included in the binding reaction, a strong TF–TFBS complex was observed as a slowly migrating complex, indicating an interaction between the TF protein and the corresponding promoter sequence. To examine binding specificity, we performed competition experiments with a 20-fold excess of non–biotin-labeled (i.e., unlabeled) TFBS probes or with TFBS probes carrying mutations. As shown in Fig. 5, only unlabeled wild-type TFBS probes, but not the mutated TFBS probe, could reduce the formation of TF–TFBS complex efficiently, supporting the binding specificity of the TF–TFBS pairs predicted. Because the selected TFs represent eight very different TF families, our EMSA experiments demonstrated the reliability and breadth of our prediction of TF–TFBS interactions.

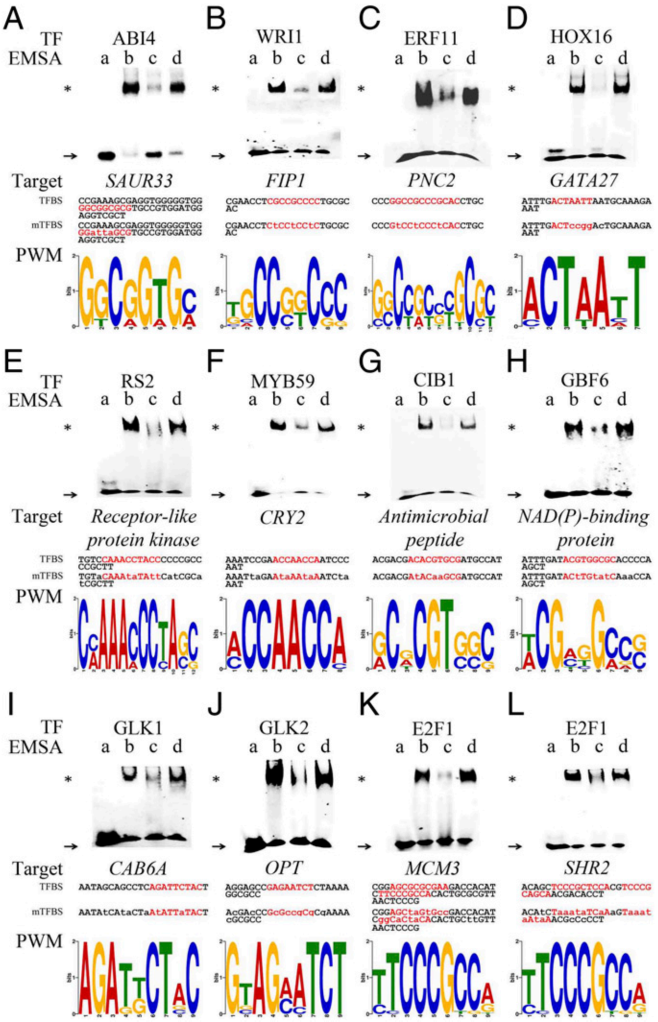


Fig. 5. Assays of the binding of transcription factors to the *cis* elements in target genes by EMSA. (*A*–*L*) TF proteins purified from *E. coli* were incubated with *cis* element probes corresponding to different TFBSs and their mutated sequences (mTFBS) for EMSA. EMSA was performed with labeled probe alone (lane a) or with combined purified TF and labeled probe (lane b). DNA-binding specificity was tested by adding a 20-fold excess of unlabeled probe (lane c) or a 20-fold excess of unlabeled mutated probe (lane d). "Target" denotes the gene with the predicted TFBS in its promoter. PWM denotes the sequence logo of the position weight matrix for the binding-site motifs. Specific DNA–protein complexes and free probes are indicated by asterisks and arrows, respectively. The core binding sequences are shown in red.

Concluding Remarks

We studied the dynamics of gene expression during the first 192 h of maize embryonic leaf development. Through coexpression and differential gene expression analysis we were able to uncover molecular signatures associated with dominant developmental and physiological processes at different time points. Combined with analysis of TF expression, our study not only provides a general picture of transcriptional regulation during this period but also highlights the key TF families involved, indicating future directions in elucidating their biological functions in maize leaf development. Given the dearth of both functional and regulatory details of these TFs, we used a combination of computational approaches to predict more than 1,000 TFBSs and 253 TF–TFBS pairs in maize, far exceeding the 30 TF–TFBS pairs currently known in maize. There were three reasons that we were able to predict so many TF–TFBS pairs. First, the 22 transcriptomes, which were used repeatedly in our analysis, provided a sufficiently large number of time points for constructing tightly coexpressed gene sets. We note the added requirement that the genes in a gene set share the same GO term. Second, the number of good-quality TF–TFBS pairs in *Arabidopsis* and rice in TF databases and literature recently has become fairly large. Third, the two methods we have developed for predicting TF–TFBS pairs seem reliable. The exceptions are for TFBSs with low-complexity motifs and for "gapped TFBSs," i.e., TFBSs with one or more highly degenerate sites within the TFBS sequence. Because predicting TF–TFBS interactions is a challenging task, further improvement in methodology is needed. Also, many more maize TF–TFBS pairs remain to be identified. Nonetheless, our study provides a detailed look at the transcriptomic changes highly relevant to germination and early leaf development in a major C4 crop. Our findings significantly expand the transcriptional regulatory network currently known in maize, providing a number of testable hypotheses of TF and TFBS interactions for experimental verification. This approach should contribute to our understanding of the regulatory circuit underlying maize leaf development.

Materials and Methods

Collection of Gene Sets and Gene Set Enrichment Analysis. The gene clusters obtained by the preceding procedure (i.e., gene coexpression modules) are usually too loose and too few for inferring TFBSs. For this purpose, we identified sets of strongly coexpressed genes as described below. Because coexpression does not always imply coregulation, we added the condition that the genes in a set should belong to the same functional category, which was retrieved from GO (AGPv2, Ensembl Plants; plants.ensembl.org/index.html), MaizeCyc (v2.0.2; maizecyc.maizegdb.org/), or MapMan (v0.9) (72). Because GO terms are formed hierarchically so that a child term is a subset of its parental term, a gene's GO term was assigned to all its parental GO terms (Matlab function *getancestors*). Because MapMan assigns a transcript's ID to a pathway (represented by a bincode), we converted transcript IDs to their gene IDs and then assigned them to all corresponding parental bincodes. Finally, MaizeCyc gene sets, which include mostly metabolic and transport pathway genes, were compiled and integrated into our database of gene sets. To have sufficient statistical power while keeping computational analysis feasible, only gene sets with ≥5 and ≤5,000 expressed genes were selected—a total of 3,783 gene sets including 21,441 expressed genes (65% of all expressed genes; see Dataset S3).

Next, we conducted gene set enrichment analysis (GSEA) (73). We refined each gene set by focusing on the subset of genes with a high PCC. The enrichment score (ES) of each gene set was calculated based on two types of rank lists, H1 and H2. To construct the H1 list for each expressed gene, its expression similarities (PCCs) to all other expressed genes were calculated, and the top 20 PCCs were averaged and taken as the representative PCC for the gene in question. Then a rank list was constructed by sorting the representative PCCs of all expressed genes in descending order. For the H2 list, the PCCs were calculated between each gene in the gene set of interest and every expressed gene, and the average PCC of each gene was computed from the top 5% genes in the gene set (selecting at least five genes) to generate the rank list. Note that an H2 list was constructed for each of the selected 3,783 gene sets. The H2 list was added to make sure that the genes in a strongly coexpressed gene set are highly coexpressed with each other.

Based on the two rank lists, the ES of a gene set was calculated with the modified Kolmogorov–Smirnov statistic (73). By walking down each rank list, a running-sum statistic was increased by $|r|/m$ if the running rank was a gene within the set or decreased by $1/n$ if the gene was outside the set (r: the representative PCC of the gene; m: the gene set size; n: the number of expressed genes not in the gene set). The ES is the maximum deviation of the running sum from 0 and is normalized to be between −1 and 1. To select the core genes of a strongly coexpressed gene set, a leading-edge subset (LES) was identified consisting of genes that are at the upper ranks before reaching the maximum deviation of the running-sum statistic.

To test if the ES of a gene set is significantly higher than random expectation, we used the permutation method. For H1, all gene labels and their representative PCCs were randomly permuted, and the ES obtained was taken as the null ES for each gene set. The *P* value of a gene set was computed as the probability of the observed ES ≥ the null ES by repeating the procedure 1,000 times. For H2, we randomly selected genes from all expressed genes to form a randomized test set of the same size as the gene set of interest. A null ES was computed from the rank list of H2, and this process was repeated 1,000 times to estimate the *P* value for each gene set. To combine the *P* values derived from H1 and H2 for each gene set (74), we calculated $T_{composite} = 0.5T_{H1} + 0.5T_{H2}$ [$T_H = \Phi^{-1}(1 - p)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution function]. By using the statistic $T_{composite}$, the combined *P* value of a gene set has the normal distribution with the mean($T_{composite}$) and std($T_{composite}$), which are the mean and the SD of the composite Ts of the gene set. According to the combined *P* values, false-discovery rates (FDRs) were estimated using the procedure of Benjamini and Hochberg (75). Finally, for each significant gene set, the genes that were common in the two LESs identified based on H1 and H2 were selected, and the gene set thus obtained may be clustered further into strongly coexpressed gene sets, using *k*-means with automatic detection of k ($1 \leqq k \leqq 10$). The strongly coexpressed gene sets were analyzed further (Dataset S3).

Discovering Overrepresented Motifs in Promoter Sequences and Identifying TFBSs. The putative promoter sequence of a gene was defined as the region from −1,000 bp to +200 bp relative to the transcription start site (TSS) of the gene. Simple repeats and low-complexity DNA sequences then were masked using RepeatMasker (version open-4.0.0; www.repeatmasker.org), resulting in ∼1% masked sequences.

We used MEME (76) to detect overrepresented motifs in the promoter sequences of a set of strongly coexpressed genes by considering motifs (5–12 bp) that were located on either strand of a promoter sequence, occurring (the best hit in terms of maximum likelihood) in >50% of the promoters. In this study, the promoter of a gene was defined as the region from −1,000 to +200 bp of the TSS because most TFBSs are concentrated within 250 bp upstream of the TSS and also may occur in the downstream region of the TSS (77, 78). The background model was the second-order Markov model of 5,000 randomly selected promoter sequences. Each motif was reported as a position weight matrix (PWM) (Dataset S2), and we selected the top 10 motifs for each gene set to determine if their mapped sites were conserved between orthologous promoters of maize, *S. bicolor* (Sorbi1), *S. italica* (JGIv2.0), *O. sativa* (MSU6), and *B. distachyon* (v1.0).

The conservation of a motif was assessed in two steps (*SI Appendix*, Fig. S4C). First, orthologous relationships among these five species were defined. The one-to-one orthologous relationships among these five species from Ensembl Plants were selected by the criterion of synteny (79). This criterion allowed us to identify syntenic orthologs. If no syntenic ortholog was found, the Ensembl Plants orthologous definition was determined by two criteria: (*i*) the sequence identity between the target and the query is >50% and (*ii*) among the potential orthologs it has the highest average sequence identity with the maize gene. If neither synteny nor the two criteria held for any of the four grasses, we discarded the gene from our analysis. In the second step, we examined whether the sites of a motif, mapped using FIMO (80), were also present in orthologous promoters (*P* value $<1 \times 10^{-4}$) based on alignments generated by MUSCLE (v3.8.31) (81). A motif site in maize was considered conserved if the motif appeared in an orthologous promoter on the same strand and was located within 100 bp of the maize motif in the alignment. This pair of the maize site and an orthologous site in another species was considered a "conserved pair." For each gene set, the total number of conserved pairs (*N*) was counted for calculating the conservation score (*N* divided by the total number of orthologous pairs of the maize genes in the same gene set that have the overrepresented motif). Then a *P* value to assess the significance of conservation of a motif was computed according to the binomial distribution with the success probability of 0.04, under the assumption that a motif occurs uniformly in the region of 100 bp in the same strand within a 1.2-kb promoter, i.e., 100/(1,200 × 2) = ∼0.04. Significantly conserved motifs (*P* value $<10^{-10}$) were regarded as pTFBSs.

Inferring the Cognate TFs of Maize pTFBSs. Once a pTFBS is obtained, its cognate TF may be inferred as follows (*SI Appendix*, Fig. S5): The pTFBS was used to find similar known TFBSs, using the motif comparison tool TOMTOM (82) (*P* value $<10^{-4}$). We used three sources of known TF–TFBS interactions: (*i*) The TF databases TRANSFAC (2), JASPAR (3), and AthaMap (1); (*ii*) the set of 63 *Arabidopsis* TFs and their TFBSs that were identified using the protein-binding microarray (PBM) technique (68); and (*iii*) the set of 240 *Arabidopsis* TFs (254 TFBSs) and 23 rice TFs (25 TFBSs) in CIS-BP (69), in which the selected TF–TFBS pairs were determined directly by PBM or Systematic Evolution of Ligands by Exponential (SELEX) enrichment approaches. Because for each TFBS in these sources the cognate TF was known, we obtained 8,432 cognate TF sequences and used them to search for their maize homologs in a set of 5,246 maize TF sequences (71) with BLAST E-value $<10^{-7}$. To reduce the false identification rate, the gene-expression profile of maize putative cognate TF(s) of a pTFBS was required to be correlated (PCC >0.85) with the mean profile of the gene set from which the pTFBS was obtained. For each candidate cognate TF, we identified a set of genes strongly coexpressed with the TF gene and predicted its putative TFBSs, as we did in the prediction of pTFBSs (see above). We then selected the TF that had the putative TFBS best matching the given pTFBS.

Finding the Maize TF–TFBS Pair Using a Known TF–TFBS Pair in Another Species. The TF–TFBS pair in maize can be inferred if the TF–TFBS pair is known in another species, such as *A. thaliana*, and if maize gene-expression data are available for inferring coexpressed genes of the TF gene (*SI Appendix*, Fig. S6). Note that in Method 1 described above, we started with a predicted TFBS, whereas in the Method 2 we started with a known TF–TFBS pair. We collected the known TF–TFBS pair information for 353 *A. thaliana* and 36 *O. sativa* (rice) TFs from AthaMap, TRANSFAC, and JASPAR (1–3) and from Franco-Zorrilla et al. (68) and Weirauch et al. (69).

Our analysis was done in three steps. First, the maize homolog(s) of each *Arabidopsis* or rice TF in the collected dataset was identified by DBD sequence similarity. With a threshold E-value $<10^{-20}$, 287 *Arabidopsis* and 31 rice TFs were found to have maize homologs. However, we excluded one *Arabidopsis* TF because its maize homologous TF genes were not expressed in the 22 maize transcriptomes, and we also excluded one rice TF for the same reason. Second, for a given *Arabidopsis*/rice TF gene, each of the homologous maize TF genes was used to find a set of coexpressed genes (PCC >0.8) in maize, which then was subjected to GSEA using Fisher's exact test and multiple testing correction (75). We required that the number of coexpressed genes in the gene set be >5 but ≦ 100; if the number of genes was >100, we selected the top 100 coexpressed genes. The coexpressed genes that were enriched with a GO term (FDR $<10^{-3}$) were selected for detecting overrepresented motifs in their promoter sequences. If an overrepresented motif passed the conservation test described above (*P* value $<10^{-5}$), it was selected as a pTFBS of the maize TF, leading to a potential maize TF–TFBS pair. Finally, all the maize TF–pTFBS pairs derived from an *Arabidopsis*/rice TF–TFBS pair were tested one by one, starting from the top homologous TF (the lowest BLASTp E-value). The test was to determine if the TFBS of a maize TF is the one most similar to the TFBS of the *Arabidopsis*/rice TF used to identify the maize TF in question. If the similarity between the two motifs was significant (TOMTOM, *P* value <0.005), this most-similar pTFBS and its corresponding maize TF were selected as a putative TF–TFBS pair.

EMSA to Validate TF–TFBS Interactions. EMSA was conducted to test whether an inferred cognate TF indeed binds the predicted TFBS sequence. For the production of a recombinant TF protein, the full-length cDNA of the TF in maize was cloned into a pPET42a vector to create an in-frame fusion with the histidine (His) tag with primer pairs in Dataset S6. The construct then was transformed into *Escherichia coli* Rosetta (DE3) for TF protein expression and purification according to the manufacturer's suggestions (GE). Briefly, the protein was induced to express with 0.5 mM isopropyl β-d-1-thiogalactopyranoside at 37 °C for 3 h. The cell pellet was resuspended with the 1× Gibco PBS (Life Technologies) and 1× SigmaFAST proteinase inhibitor mixture (Sigma) and was homogenized further by microfluidizer (Hyland Scientific). The cleared lysate was subjected to affinity chromatography by incubation with Ni Sepharose (GE) for 14 h at 4 °C followed by elution with 250 mM imidazole in 500 mM sodium chloride and 20 mM sodium phosphate, pH 7.4. For each TF, the purified recombinant protein and probes containing its corresponding predicted TFBS sequence were used for the EMSA experiment. The synthetic double-stranded oligonucleotide probe (Dataset S6) was biotin-labeled with Klenow fragment (Thermo Scientific), 0.1 mM biotin-dUTP (Thermo Scientific), 0.1 mM dTTP, and 0.2 mM dATP, dCTP, and dGTP at 37 °C for 30 min and was purified by the QIAquick PCR purification kit (Qiagen). The EMSA reaction mixture containing binding buffer (250 mM KCl, 0.1 mM DTT, 0.1 mM EDTA, 10 mM Tris, pH 7.4), 100 ng poly (dI-dC), and biotin-labeled probes was incubated for 20 min at 22 °C. Competition experiments were performed with excess unlabeled probes or unlabeled probes with mutations in the pTFBS site as competitors. The EMSA mixture was separated by a 5% polyacrylamide native gel and transferred to a Hybond N^+ membrane (GE) by semidry transfer cell (Bio-Rad). The biotin-labeled probe and the TF–probe complexes were detected by streptavidin-HRP conjugates (Life Technologies) with substrates from ECL plus (GE). The chemiluminescent signals were visualized by the UVP BioSpectrum imaging system (UVP).

ACKNOWLEDGMENTS. We thank Chien-Yu Chen for suggestions. This study was supported by Academia Sinica, Taiwan Grant AS-102-SS-A13, the Innovative Translational Agricultural Research Program, and National Science Foundation Grant 1119778 (to S.-H.S.).

1. Bülow L, Steffens NO, Galuschka C, Schindler M, Hehl R (2006) AthaMap: From in silico data to real transcription factor binding sites. *In Silico Biol* 6(3):243–252.
2. Matys V, et al. (2006) TRANSFAC and its module TRANSCompel: Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34(Database issue):D108–D110.
3. Mathelier A, et al. (2014) JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 42(Database issue):D142–D147.
4. Li P, et al. (2010) The developmental dynamics of the maize leaf transcriptome. *Nat Genet* 42(12):1060–1067.
5. Chang YM, et al. (2012) Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiol* 160(1):165–177.
6. Liu WY, et al. (2013) Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination. *Proc Natl Acad Sci USA* 110(10):3979–3984.
7. Wang P, Kelly S, Fouracre JP, Langdale JA (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *Plant J* 75(4):656–670.
8. Wang L, et al. (2014) Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat Biotechnol* 32(11):1158–1165.
9. Chen J, et al. (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol* 166(1):252–264.
10. Bewley JD (1997) Seed Germination and Dormancy. *Plant Cell* 9(7):1055–1066.
11. Kucera B, Cohn MA, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. *Seed Sci Res* 15(4):281–307.
12. Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: Early seed germination. *J Exp Bot* 62(10):3289–3309.
13. Ohashi-Ito K, Fukuda H (2010) Transcriptional regulation of vascular cell fates. *Curr Opin Plant Biol* 13(6):670–676.
14. Cui H, Kong D, Liu X, Hao Y (2014) SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in Arabidopsis thaliana. *Plant J* 78(2): 319–327.
15. Slewinski TL, Anderson AA, Zhang C, Turgeon R (2012) Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol* 53(12):2030–2037.
16. Slewinski TL, Zhang C, Turgeon R (2013) Structural and functional heterogeneity in phloem loading and transport. *Front Plant Sci* 4:244.
17. Fouracre JP, Ando S, Langdale JA (2014) Cracking the Kranz enigma with systems biology. *J Exp Bot* 65(13):3327–3339.
18. Halliday KJ, Martínez-García JF, Josse EM (2009) Integration of light and auxin signaling. *Cold Spring Harb Perspect Biol* 1(6):a001586.
19. Sassi M, Wang J, Ruberti I, Vernoux T, Xu J (2013) Shedding light on auxin movement: Light-regulation of polar auxin transport in the photocontrol of plant development. *Plant Signal Behav* 8(3):e23355.
20. Strayer C, et al. (2000) Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* 289(5480):768–771.
21. Yazaki J, et al. (2004) Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: Phenotyping and comparative analysis between rice and Arabidopsis. *Physiol Genomics* 17(2):87–100.
22. Elliott RC, et al. (1996) AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 8(2): 155–168.
23. Horiguchi G, Kim GT, Tsukaya H (2005) The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J* 43(1):68–78.
24. Baima S, et al. (1995) The expression of the Athb-8 homeobox gene is restricted to provascular cells in Arabidopsis thaliana. *Development* 121(12):4171–4182.
25. Ohashi-Ito K, Oguchi M, Kojima M, Sakakibara H, Fukuda H (2013) Auxin-associated initiation of vascular cell differentiation by LONESOME HIGHWAY. *Development* 140(4):765–769.
26. Zhou J, Wang X, Lee JY, Lee JY (2013) Cell-to-cell movement of two interacting AT-hook factors in Arabidopsis root vascular tissue patterning. *Plant Cell* 25(1):187–201.
27. Okada K, Ueda J, Komaki MK, Bell CJ, Shimura Y (1991) Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. *Plant Cell* 3(7):677–684.
28. Jones AM, et al. (1998) Auxin-dependent cell expansion mediated by overexpressed auxin-binding protein 1. *Science* 282(5391):1114–1117.
29. Blakeslee JJ, et al. (2007) Interactions among PIN-FORMED and P-glycoprotein auxin transporters in Arabidopsis. *Plant Cell* 19(1):131–147.

30. Guo X, Lu W, Ma Y, Qin Q, Hou S (2013) The BIG gene is required for auxin-mediated organ growth in Arabidopsis. *Planta* 237(4):1135–1147.
31. Shikata M, Koyama T, Mitsuda N, Ohme-Takagi M (2009) Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. *Plant Cell Physiol* 50(12):2133–2145.
32. Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiol* 140(2):411–432.
33. Walsh J, Waters CA, Freeling M (1998) The maize gene liguleless2 encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. *Genes Dev* 12(2):208–218.
34. Silveira AB, et al. (2007) The Arabidopsis AtbZIP9 protein fused to the VP16 transcriptional activation domain alters leaf and vascular development. *Plant Sci* 172(6): 1148–1156.
35. Corrêa LG, et al. (2008) The role of bZIP transcription factors in green plant evolution: Adaptive features emerging from four founder genes. *PLoS ONE* 3(8):e2944.
36. Li Z, Thomas TL (1998) PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. *Plant Cell* 10(3):383–398.
37. Wang D, et al. (2008) Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. *BMC Genomics* 9:44.
38. Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in Arabidopsis: An analysis of the cup-shaped cotyledon mutant. *Plant Cell* 9(6):841–857.
39. Xie Q, Frugis G, Colgan D, Chua NH (2000) Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev* 14(23): 3024–3036.
40. Xu B, et al. (2014) Contribution of NAC transcription factors to plant adaptation to land. *Science* 343(6178):1505–1508.
41. Scharf KD, Berberich T, Ebersberger I, Nover L (2012) The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochim Biophys Acta* 1819(2):104–119.
42. Petricka JJ, Clay NK, Nelson TM (2008) Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. *Plant J* 56(2):251–263.
43. Castilhos G, Lazzarotto F, Spagnolo-Fonini L, Bodanese-Zanettini MH, Margis-Pinheiro M (2014) Possible roles of basic helix-loop-helix transcription factors in adaptation to drought. *Plant Sci* 223:1–7.
44. Kwak KJ, Kim JY, Kim YO, Kang H (2007) Characterization of transgenic Arabidopsis plants overexpressing high mobility group B proteins under high salinity, drought or cold stress. *Plant Cell Physiol* 48(2):221–231.
45. Lildballe DL, et al. (2008) The expression level of the chromatin-associated HMGB1 protein influences growth, stress tolerance, and transcriptome in Arabidopsis. *J Mol Biol* 384(1):9–21.
46. Wang ZY, Tobin EM (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* 93(7):1207–1217.
47. Du H, et al. (2013) Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res* 20(5):437–448.
48. Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. *Trends Plant Sci* 13(12):647–655.
49. Brooks L, 3rd, et al. (2009) Microdissection of shoot meristem functional domains. *PLoS Genet* 5(5):e1000476.
50. Hoecker U, Vasil IK, McCarty DR (1995) Integrated control of seed maturation and germination programs by activator and repressor functions of Viviparous-1 of maize. *Genes Dev* 9(20):2459–2469.
51. Dubos C, et al. (2010) MYB transcription factors in Arabidopsis. *Trends Plant Sci* 15(10):573–581.
52. Cao D, Hussain A, Cheng H, Peng J (2005) Loss of function of four DELLA genes leads to light- and gibberellin-independent seed germination in Arabidopsis. *Planta* 223(1): 105–113.
53. Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. *Trends Plant Sci* 12(9):419–426.
54. Johannesson H, Wang Y, Hanson J, Engström P (2003) The Arabidopsis thaliana homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings. *Plant Mol Biol* 51(5):719–729.
55. Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. *Plant Physiol* 131(4):1855–1867.
56. Kumagai T, et al. (2008) The common function of a novel subfamily of B-Box zinc finger proteins with reference to circadian-associated events in Arabidopsis thaliana. *Biosci Biotechnol Biochem* 72(6):1539–1549.
57. Husbands A, Bell EM, Shuai B, Smith HM, Springer PS (2007) LATERAL ORGAN BOUNDARIES defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic Acids Res* 35(19):6663–6671.
58. Preston JC, Hileman LC (2013) Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Front Plant Sci* 4:80.
59. Bowman JL (2000) The YABBY gene family and abaxial cell fate. *Curr Opin Plant Biol* 3(1):17–22.
60. Martín-Trillo M, Cubas P (2010) TCP genes: A family snapshot ten years later. *Trends Plant Sci* 15(1):31–39.
61. Serikawa KA, Martinez-Laborda A, Kim HS, Zambryski PC (1997) Localization of expression of KNAT3, a class 2 knotted1-like gene. *Plant J* 11(4):853–861.
62. Hamant O, Pautot V (2010) Plant development: A TALE story. *C R Biol* 333(4):371–381.
63. Kim D, et al. (2013) BLH1 and KNAT3 modulate ABA responses during germination and early seedling development in Arabidopsis. *Plant J* 75(5):755–766.
64. Lijavetzky D, Carbonero P, Vicente-Carbajosa J (2003) Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evol Biol* 3:17.
65. Epple P, Mack AA, Morris VR, Dangl JL (2003) Antagonistic control of oxidative stress-induced cell death in Arabidopsis by two related, plant-specific zinc finger proteins. *Proc Natl Acad Sci USA* 100(11):6831–6836.
66. van der Graaff E, Laux T, Rensing SA (2009) The WUS homeobox-containing (WOX) protein family. *Genome Biol* 10(12):248.1–248.9.
67. Nakamichi N, et al. (2010) PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the Arabidopsis circadian clock. *Plant Cell* 22(3):594–605.
68. Franco-Zorrilla JM, et al. (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci USA* 111(6):2367–2372.
69. Weirauch MT, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158(6):1431–1443.
70. Mahony S, Auron PE, Benos PV (2007) DNA familial binding profiles made easy: Comparison of various motif alignment and clustering strategies. *PLOS Comput Biol* 3(3):e61.
71. Lin JJ, Yu CP, Chang YM, Chen SC, Li WH (2014) Maize and millet transcription factors annotated using comparative genomic and transcriptomic data. *BMC Genomics* 15:818.
72. Thimm O, et al. (2004) MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37(6): 914–939.
73. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545–15550.
74. Tsang JS, Ebert MS, van Oudenaarden A (2010) Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. *Mol Cell* 38(1): 140–153.
75. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57(1):289–300.
76. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
77. Heyndrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *Plant Cell* 26(10):3894–3910.
78. Lin Z, Wu WS, Liang H, Woo Y, Li WH (2010) The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. *BMC Genomics* 11:581.
79. Schnable JC, Freeling M, Lyons E (2012) Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol Evol* 4(3):265–277.
80. Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.
81. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
82. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS (2007) Quantifying similarity between motifs. *Genome Biol* 8(2):R24.1–R24.9.