Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination

Wen-Yu Liu, Yao-Ming Chang, Sean Chun-Chang Chen, Chen-Hua Lu, Yeh-Hwa Wu, Mei-Yeh Jade Lu, Di-Rong Chen, Arthur Chun-Chieh Shih, Chiou-Rong Sheue, Hsuan-Cheng Huang, Chun-Ping Yu, Hsin-Hung Lin, Shin-Han Shiue, Maurice Sun-Ben Ku, and Wen-Hsiung Li

Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination

Our anatomical analysis revealed that a dry maize seed contains four to five embryonic leaves at different developmental stages. Rudimentary kranz structure (KS) is apparent in the first leaf with a substantial density, but its density decreases toward younger leaves. Upon imbibition, leaf expansion occurs rapidly with new KSs initiated from the palisade-like ground meristem cells in the middle of the leaf. In parallel to the anatomical analysis, we obtained the time course transcriptomes for the embryonic leaves in dry and imbibed seeds every 6 h up to hour 72. Over this time course, the embryonic leaves exhibit transcripts of 30,255 genes at a level that can be regarded as “expressed.” In dry seeds, ∼25,500 genes are expressed, showing functional enrichment in transcription, RNA processing, protein synthesis, primary metabolic pathways, and calcium transport. During the 72-h time course, ∼13,900 genes, including 590 transcription factor genes, are differentially expressed. Indeed, by 30 h postimbibition, ∼2,200 genes expressed in dry seeds are already down-regulated, and ∼2,000 are up-regulated. Moreover, the top 1% expressed genes at 54 h or later are very different from those before 30 h, reflecting important developmental and physiological transitions. Interestingly, clusters of genes involved in hormone metabolism, signaling, and responses are differentially expressed at various time points and TF gene expression is also modular and stage specific. Our dataset provides an opportunity for hypothesizing the timing of regulatory actions, particularly in the context of KS development.

Maize, a well-studied crop, has been used as a model plant for C4 photosynthesis study, as its leaves possess the kranz structure (KS) for efficient photosynthesis. However, how its leaves develop from seed following imbibition has not been well studied. In particular, it is unclear how KS forms during leaf development. Using the next generation sequencing technology, Li et al. (1) studied the leaf transcriptomes of four regions of 9-d-old third maize leaves; the base, the tip, and two middle regions of the leaf, representing different leaf developmental stages, with the base being the youngest. The data revealed a dynamic transcriptome profile, showing different transcripts enriched in different regions and providing a preliminary view of molecular changes during maize leaf development. However, as the leaf base already exhibits distinct KS, it is not early enough to represent the early leaf development when KS begins to form. Indeed, our anatomical study reveals that KS already exists in a rudimentary form in the first two embryonic leaves of maize dry seeds, and the embryonic leaves develop rapidly after seed imbibition (see below). To correlate the transcriptomic dynamics with the KS development during seed germination, we have obtained the time course transcriptomes of embryonic leaves at every 6 h, starting from dry seeds to hour 72 postimbibition. This set of data provides a clear picture of the transcriptional dynamics of genes for the early leaf development during maize seed germination. Moreover, we use the data to infer the succession of biological processes during this period. Finally, we examine what hormone-related genes and transcriptional regulators are differentially expressed in the embryonic leaves over time, providing a global view of the molecular changes underlying their anatomical and physiological transitions during seed germination.

Results and Discussion

Anatomical Changes During Early Leaf Development. Fig. 1A shows the morphological changes in maize seeds during the first 72 h of imbibition; the tissue swells over time until the appearance of coleoptile at the 72nd hour (T72). A dry maize seed contains four to five embryonic leaves (Fig. 1B). The first leaf already has many well-differentiated rudimentary KS units separated by one to two ground meristem (GM) cells in the middle (median) layer of leaf primordium. The second leaf is less developed, but has recognizable KS units. Upon imbibition, all embryonic leaves expand within 12 h (Fig. 1C). Compared with the first embryonic leaf, the younger leaves contain fewer veins (KS units), frequently separated by more than five palisade-like median ground meristem (mGM) cells. During the next 2 d, the vein density rapidly increases (Fig. 1 D and E).

The most significant feature of the maize embryonic leaves, except for the main vein regions, is that they consist of five layers of distinct cell types in transverse section (Fig. 1F): two layers of protoderm (P) and three layers of GM cells, with the mGM cell layer located in the middle of the leaf and sandwiched by two layers of round-shaped ground meristem (sGM) cells. The two layers of protoderm form the adaxial and abaxial epidermal cell layers, whereas contiguous mGM cells differentiate into vascular tissue, bundle sheath (BS) cells, and mesophyll (M) cells. Clearly, all KS units of the minor veins rise from the mGM cells, initially with a simple formation of three to four contiguous cells with two cells sandwiching one or two smaller cells. Fig. 1 G–K shows the progressive changes in KS or vein development. The sGM cells next to the protodermal cell layers give rise to the subprotodermal mesophyll (sM) cells (Fig. 1K) separating the veins and the epidermal cells in mature leaves. KS units at different developmental stages can be found in all embryonic leaves, especially in the younger leaves. In initiating new KS units the progenitor mGM cells begin to divide and...
study of the base of the third through fifth leaves from 2-wk-old maize plants (2), showing that both main and minor veins have a similar pattern in initiation and development. Also clear is that the two M cells in the middle layer (mM in Fig. 1K) of the maize leaf intervening BSs are ontogenetically more closely related to the BS cells than are the subprotodermal M cells (sM in Fig. 1K).

Transcriptome Profiling of Early Leaf Development. The RNA samples were taken from the embryonic leaves of dry seeds (denoted T00) and imbibed seeds every 6 h up to T72. The 13 RNA samples were pair-end sequenced with Illumina HiSeq2000 (Table S1). The reads were mapped to the maize genome and the mappable reads were used to estimate the transcription level with the reads per kilobase per million (RPKM) mapped reads measure. The RPKM values were then normalized for comparison between time points.

The current maize annotation contains a “working gene set” (WGS) and a “filtered gene set” (FGS, a subset of WGS). After excluding repetitive elements and pseudogene entries, 63,363 WGS and 39,441 FGS genes remain. To reduce the effect of background transcription, a WGS gene is included in our analysis only if its RPKM value is ≥1 at more than one time point (Dataset S1). With this criterion, 30,255 (48%) WGS genes, including 23,900 (61%) FGS genes, are selected for our analysis (Table S2). For the 2,070 maize transcription factors (TFs) (3), 1,238 genes (60%) are included in our analysis. For the selected gene set, a gene is considered expressed at a time point if its normalized RPKM value is ≥1 at that time point (Dataset S1).

We now ask how similar the expression patterns between time points are. For the collection of all top 1% expressed genes at each time point (693 genes), the correlation (r) in expression level between two consecutive time points decreases very slowly until T42, but then a sharp transition occurs between T48 and T54 (r = 0.65; Fig. 2A, Upper half). This transition is mostly due to the up-regulation of ribosomal protein genes and down-regulation of heat stress and glycolysis genes at T54. Another sharp transition occurs between T60 and T66 (r = 0.51; Fig. 2A, Upper half), due to the up-regulation of heat stress genes and down-regulation of ribosomal protein genes and genes involved in DNA synthesis and glycolysis at T66 (Functional Transition Postimbibition). Some of these changes, such as those attributed to heat stress, may be related to response to sunlight and rising temperature during daytime; T54 was midnight, whereas T66 was noon. At any rate, the top 1% of genes at T54 and later are very different from the top 1% of genes before T30 because all r values are <0.3 (Fig. 2A, Lower half). The bottom 1% (1,415) of genes show a very similar transition pattern to the top 1% but their r values decrease somewhat faster, especially after T48, with the exception of T66 (Fig. 2A, Lower half). The middle 98% (28,152) of genes show

---

**Fig. 1.** Changes in maize seed morphology and anatomy during germination. (A) Changes in seed morphology from 0 h (dry seed) to 72 h after imbibition. (i) Within the first 36 h: seed expands but no change is seen on the phenotype. (ii) At the 48th hour: primary root breaks the skin. (iii) At the 60th hour: epicotyl is now visible. (iv) At the 66th hour: primary root hairs and lateral roots are now visible. (v) At the 72nd hour: epicotyl becomes green. (B–E) Cross-sections of maize embryonic leaves at (B) 0 h (dry seed; arrows indicate midribs), (C) 12 h, (D) 48 h, and (E) 72 h after imbibition (arrowheads indicate lignified cells in major veins). (F–K) Progressive changes in kranz structure development in embryonic leaves during seed germination. 1–4, first to fourth embryonic leaves; BS, bundle sheath cells; KS, kranz structure (vein + BS); LS, leaf sheath; mM, median ground meristem cells; mM, median layer mesophyll cells; P, protoderm; rKS, rudimentary kranz structure; sM, round-shaped subprotodermal ground meristem cells; sM, subprotodermal mesophyll cells; V, vein. [Scale bars, (A) 1 mm; (B and C) 50 μm; (D and E) 100 μm; and (F–K) 10 μm.]

---

**Fig. 2.** Expression correlations between time points. (A) Heatmaps of Pearson’s correlation coefficients (PCCs) between time points calculated for the genes at the top 1% expression levels (Upper half) and for the genes at the bottom 1% expression level (Lower half). (B) PCC heatmaps for all genes with middle 98% expression level (Upper half) and for TF genes (Lower half). Scale on the Lower Left shows the night (dark) and day (light) cycles.
a slightly faster decrease in \( r \) than the top 1% of genes until T48 but a slower rate of decrease after T48 (Fig. 2B, Upper half). TF genes show a slower rate of decrease in \( r \) initially than other genes but a faster rate after T54 (Fig. 2B, Lower half).

As mentioned above, an earlier study obtained the transcriptomes of four regions of 9-d-old third maize leaves (1). No significant correlation exists between their transcriptomes and ours, except for the leaf base (the youngest region), and the maximum correlation is only 0.64, which occurs at T60 (Fig. S1). Thus, the base of a 9-d-old maize leaf may not represent the very early leaf development and our time course data provide a unique global view of gene expression over maize early leaf development.

**Genes Expressed in the Embryonic Leaves of Dry Seeds.** In dry seeds, transcripts for 20,603 (52%) FGS and 25,401 (40%) WGS genes are already present with RPKM \( \geq 1 \). These transcripts may jump-start the maize leaf development upon imbibition. To assess the roles of these stored transcripts, we examine MapMan functional categories containing an over- or underrepresented number of expressed genes at T00 (Fig. S2A).

Three types of genes are overrepresented at T00. The first type includes genes involved in transcription, RNA processing, protein synthesis, and primary metabolic pathways, including carbohydrate biosynthesis, the tricarboxylic acid (TCA) cycle, mitochon-drial electron transport, and fatty acid, nucleotide, and amino acid biosynthesis. These are likely transcripts stored in the embryonic leaves for meeting the demands for rapid growth upon imbibition. Similarly, cell division genes are overrepresented, potentially contributing to rapid cell proliferation. The second type includes genes involved in calcium transport, likely reflecting the role of calcium in removing the inhibitory effect of abscisic acid (ABA) during germination (4). The third type involves stress genes, whose expression may be caused by dehydration and osmotic stress during seed maturation. Interestingly, genes involved in abiotic stress are overrepresented but those in biotic stress are underrepresented in the quiescent leaves of the dormant seeds.

Some of the above categories are enriched not only in dry seeds but also throughout the time course (Fig. S2A). These categories include primary metabolic pathways, such as the TCA cycle, some components of mitochondrial electron transport, and biosynthesis of amino acids, proteins, and nucleotides, likely for the need of energy and building blocks for early leaf development. On the other hand, genes in some functional categories are underrepresented, including biotic stress, cell wall modification and degradation, and secondary metabolism.

**Functional Transition Postimbibition.** Although some functional categories contain consistently over- or underrepresented numbers of expressed genes over the entire time course, only 852 (2.8%) genes are expressed at similar levels (coefficient of variation <0.2) during the first 72 h of imbibition and 71 of the 852 genes show more than twofold changes in RPKM value during the 72 h. Thus, most genes are up- or down-regulated to various degrees over the 72-h period (Fig. S4A). In addition, some biological processes are enriched in a time-point-specific manner (Fig. S2A). For example, genes involved in cell division and calcium transport are overrepresented until T30 (Fig. S2A). Genes of E3 ligases involved in ubiquitin-mediated protein degradation are particularly prominent at T12–T18, indicating the elevated need for protein turnover. In addition, chromatin structural genes tend to be expressed from T30 to T72. Moreover, many categories have underrepresented numbers of expressed genes in a time-point-specific manner. For example, ethylene signal transduction genes exhibit reduced expression during T54–T72, presumably reflecting a reduced influence of ethylene at this stage.

To assess functional transition due to highly expressed genes, we conducted enrichment tests on genes in the top 10% or 1% (Fig. S2B and C). Consistent with the above analysis using all genes, categories containing, for example, ATP and protein synthesis or abiotic stress genes are enriched. However, the analysis of the top 10% of genes reveals more categories that are time-point specific. For example, a major class of plant cell wall proteins, hydroxyproline-rich glycoproteins (HRGPs), which are involved in growth and cell wall development, are enriched at T00 and T12. One HRGP (GRMZM2G168651) has been shown to be expressed in maize meristem and vascular tip (5), consistent with the appearance of vascular tissues in embryonic leaves (Fig. 1). Another example is that, at T54, categories containing ribosomal proteins, ATPase transporters, and peroxiredoxins are overrepresented. Peroxiredoxins are targeted to chloroplasts to protect the photosynthetic membrane (6) by scavenging toxic peroxides produced during electron transport. Finally, at T72 genes involved in brassinosteroid (BR) metabolism are overrepresented, suggesting the rapid formation of vascular tissues at this stage, as BR promotes cell expansion and vascular differentiation (7, 8).

**Differential Gene Expression Between Time Points.** We identify three types of differentially expressed genes (DEGs) between time points (Table 1). Type 1 DEGs are identified by comparing the transcriptome at a time point against its preceding one, signifying the transition in gene function in 6 h. From T00 to T06, the transcript levels of 229 genes (including 7 TF genes) are decreased, whereas from T12 to T18, 481 genes (including 22 TF genes) are up-regulated. Notably, a sharp increase occurs in down-regulated DEGs from T24 to T30 (1,136 genes, including 28 TF genes) and in up-regulated DEGs from T30 to T36 (5,85 genes, including 50 TF genes). These and other large changes in DEG numbers suggest major developmental or physiological transitions. Interestingly, a group of genes down-regulated at T30 is up-regulated at T36, including RNA regulation of transcription (including many TFs), chromatin remodeling factors, histone acetyltransferases, DNA synthesis/chromatin structure, and cell cycle/division (Fig. S3 A, D, and E).
chloroplast targeting are up-regulated at T30, fatty acid metabolism at T36, BR signal transduction are enriched in type 2 DEGs at various time points, including 18 h after imbibition and many functional categories are DEGs overlap with type 1 DEGs. Many genes become up-regulated postimbibition (Table 1 and Fig. S3B); some type 2 DEGs overlap with type 1 DEGs. Many genes are up- and/or down-regulated postimbibition (Table 1 and Fig. S3B); some type 2 DEGs overlap with type 1 DEGs. Many genes become up-regulated 18 h after imbibition and many functional categories are enriched in type 2 DEGs at various time points, including arabinogalactan proteins (AGPs) and cell wall proteins at T18, DNA synthesis/chromatin structure at T24, cell wall cellulose synthesis at T30, fatty acid metabolism at T36, BR signal transduction at T42, the YABBY TF family at T48, and nucleotide metabolism at T54. At T72 genes for nitrate metabolism and chloroplast targeting are up-regulated (Fig. S3B). In Atabadosis, YABBY members are responsible for specifying the leaf adaxial identity (9). On the other hand, genes in a number of categories are down-regulated such as RNA helicase at T06, late embryogenesis abundant (LEA) proteins at T24, biodegradation of xenobiotics at T30 and T36, ATP synthesis at T36, and many TF families [AP2 (apetala 2)/ERF (ethylene-responsive factor), CO (constans)-like, EIN3 (ethylene insensitive 3)-like and bZIP (basic leucine zipper)] and ubiquitin E3 at T48.

The third type includes genes not differentially expressed until the indicated time points; they are a subset of type 2 DEGs. Type 3 DEGs potentially signify the onset/termination of developmental and/or physiological processes at a particular stage (Fig. S3C). An example is the differential regulation of cell wall proteins (Dataset S1). AGPs and leucine-rich repeat proteins (LRRs) are up-regulated at T18, T30, or T60; cellulose synthases are up-regulated at T18 or T66; cell wall modification proteins are up-regulated at T18, T24, or T54; pectin esterases, which are involved in cell wall modification and breakage, are up-regulated at T24 or T30; and cell wall degradation proteins (e.g., cellulases and pectate lyases) are up-regulated at T24, T30, or T66, which may be important for subsequent cell division (Dataset S1). Differential regulation of genes of the same function/process reveals how distinct components are used at different stages during leaf development.

Overall, the total number of DEGs over the entire time course is estimated to be 13,907 if all pairwise comparisons are included. Thus, close to half (46%) of the 30,255 expressed WGS genes may be regarded as differentially expressed during the 72-h time course. The corresponding proportion for expressed TF genes is 47% (590/1,238).

**Coexpression Modules.** Using hierarchical clustering, we classify the expressed WGS genes into 30 coexpression modules, each of which contains genes with highly similar expression patterns (Fig. 3A and Fig. S4A). In the following analysis, we select 20 modules (Fig. S4C) that together contain ~83% (12,716) of the 15,372 expressed genes with MapMan annotation (Fig. S4 B and D). With some exceptions, the clusters can be sorted according to the time point the peak expression occurs. This clustering analysis allows us to aggregate genes over multiple time points, providing another view of functional transitions along the early maize leaf development.

The 20 selected modules can be roughly divided into early (T00–T12), middle (T18–T48), and late (T54–T72) stages (Fig. 3A and Fig. S4 C and D). The early stage, best represented by modules C01, C04–C06, and C08, coincides with the transition from seed dormancy to germination, which involves extensive physiological changes (10). In C01, the expression level is high at T00 and T12 but becomes lower after T18. C01 includes genes related to lipid metabolism (Fig. 3A and Fig. S4D). These transcripts are likely mRNAs stored for the very early stage of leaf development during germination. In C04–C06, and C08, gene expression peaks at T06 but different modules have distinct profiles after the peak. ABA and ethylene-related genes belong to these modules (Fig. S4 D and E), reflecting the roles of these hormones during early leaf development (11). Genes related to protein degradation and synthesis are also enriched in these modules, suggesting that protein metabolism is essential for the development of embryonic leaves. For example, functional categories related to ubiquitin E3 SCF (Skp1-Cullin-F-box protein) and RING (really interesting new gene) domain proteins are overrepresented in C05 (Fig. S4B), perhaps because several E3 ligases are major components of hormone signaling (12).

The middle stage (T18–T48), represented by C13, C14, C16–C21, and C23, is typified by the up-regulation of auxin and gibberellic acid (GA) related genes, which are involved in cell division and differentiation, and genes in transcriptional regulation (Fig. S4 D and E). This stage coincides with the rapid embryonic leaf and KS development in the first 2 d of imbibition (Fig. 1 C and D). Also, genes in mitochondrial electron transport, ATP synthesis and glycolysis and a putative gene encoding a transcriptional activator are up-regulated. This suggests a heightened energy consumption, presumably taking place following digestion of endosperm. Meanwhile, the elevated transcription levels of transcriptional regulatory genes indicate the onset of multiple regulatory cascades responsible for various aspects of leaf development.

At the late stage (T54–T72), which includes C23–C25, and C30, jasmonate (JA) metabolic genes are up-regulated (Fig. S4E). C30 includes genes involved in photosynthesis, e.g., the Rubisco ribulose-1,5-bisphosphate carboxylase/oxygenase) small subunit gene. Note that at T66 the entire shoot apex covered by sheath could already be seen outside the seed and became light green (Fig. L4). Transcripts for photosynthesis genes start to accumulate at T48–T54 and become highly enriched at T72, likely signifying the onset of heightened photosynthetic activities. This is consistent with the analysis of type 1 DEGs, where genes involved in photosynthesis start to be up-regulated at T48 (Fig. S3A).

**Developmental Transition of Hormonal Functions.** During the first 72 h of leaf development, 431 expressed genes are relevant to the action of eight hormones: ABA, auxin, BR, cytokinin (CK), ethylene, GA, JA, and salicylic acid (SA) (Fig. 4). For each hormone class, genes are classified into three types: (i) synthesis-degradation, (ii) signal transduction, and (iii) induced-regulated-responsive-activated (Fig. 4 and Dataset S2). Most of these hormone-related genes are expressed in a time-point/range-specific manner (Fig. 4). A relatively small number of hormone-related genes have transcripts in the embryonic leaves of dry seeds (T00), including aldehyde oxidase 1 and 4 involved in seed ABA biosynthesis and maize orthologs of GA3 (gibberellin A3) and GA20X1 (GA-2
Expression of these genes in the embryonic leaves at the later stage of germination reflects the active differentiation of vascular tissues and the onset of plant defense mechanisms.

Type 2 genes include those relevant to hormone perception and downstream signaling, e.g., protein kinases and transcriptional regulators. For ABA, GA, BR, and ethylene, there are distinct coexpression modules among the type 2 genes (Fig. 4, type 2). For example, for both type 1 and type 2 ABA genes, there are three distinct modules: from T00 to T24, from T24 to T48, and from T48 to T72 (Fig. 4A, dotted rectangles). The T00–T24 module includes maize VPI (viviparous1) and the maize ortholog of ABI5 (ABA-insensitive 5), regulating the transition from germination to vegetative growth (Dataset S2), and AFB2 (ABRE-binding factor 2) up-regulated for ABA signaling under water stress (14). The T48–T72 module includes the maize ortholog of LEC2 (leafy cotyledon2) that functions in embryonic development (Dataset S2). Type 2 modules also include GA genes (Fig. 4B, type 2). The T00–T12 module includes a GA receptor and a GA signaling repressor, the maize orthologs of GID1c (GA insensitive dwarf 1c) and GAI (GA-insensitive), which are involved in seed germination and stem elongation (Dataset S2). The second type 2 GA module (T18–T72) contains another two maize GA1 and SPY (spindly) orthologs with a peak expression at T36–T42 (Dataset S2). SPY is related to suppression of GA response, while promoting CK signaling (15). Thus, T18 may mark a transition point between GA and CK action in germination and from cell elongation to differentiation, coinciding with the rapid maturation of existing KSs and initiation of new KSs after T18 (Fig. 1C). Another type 2 gene that signifies interactions between hormones is the maize orthologs of the ethylene receptor ETR1 (ethylene receptor1), known to inhibit ABA signaling (Dataset S2). There are seven putative orthologs of CK receptors including AHK2 (arabidopsis histidine kinase 2), AHK3, and AHK4, but only one putative ortholog of auxin receptor TIR1 (transport inhibitor response 1) in maize. These AHK genes are important regulators of shoot vascular tissue development in A. thaliana (16) and most of their transcripts start to accumulate after T12. In contrast, TIR1 expression peaks at T06, followed by a reduction, and peaks again between T30 and T72. Interestingly, the second peak expression coincides with the maize orthologs of BIG (formerly DOGC1/TIR3/UMBI/ASA1), PIN1 (pin-formed 1), and PIN4, all of which play important roles in auxin transport and vein patterning (17). These observations are consistent with the vein density increase after T24 (Fig. 1C and D) and suggest that leaf vein or KS development may be mediated by relocation of the hormones via the actions of these receptors and transport genes before T24.

One may expect that the type 3 genes that are induced by the same hormones as those for types 1 and 2 genes would have similar ranges of expression (Fig. 4, type 3). Indeed, two maize orthologs of ABA-responsive genes Em1 (late embryogenesis abundant 1) and EM6 required for buffering water loss during normal seed development (18) exhibit the highest expression levels from T00 to T24 (Dataset S2). Another example is the maize orthologs of ABA-induced HV1/A22 (HVA22 domain protein), which regulates seed germination by inhibiting GA-mediated vacuolation, are highly expressed at T06, but much lower thereafter (Dataset S2).

Expression Dynamics of TFs. The 1,238 expressed TF genes are classified into 16 coexpression modules or clusters, which are ordered by their expression peaks (Fig. 3B). Almost all TF genes have relatively narrow windows of expression (Fig. 3B), suggesting regulatory successions that underlie physiological and developmental transitions.

C1, the largest cluster, contains 328 TFs that have high expression levels during the first 24 h, followed by a gradual decrease to a very low level at T72 (Fig. 3B). Most TF families have members in C1 (Fig. 3C), including VP1, which is responsible for the establishment of seed dormancy (19), and Opaque2 (O2) heterodimerizing protein 1, which together with O2 regulates the expression of zein storage protein genes (20). Similar to C1, C8–C11 and C13–C16 have single expression peaks at different time points, indicative of regulatory succession. For example, MWP1...
(milkweed pod1), which is in C10, is responsible for adaxial-abaxial leaf surface patterning (21). Because C10 has a peak at T30, the polarity of newly emerged leaves may be established at the second day postimbibition. Another example is *Golden 2* (*G2*), which is involved in maize leaf BS chloroplast development (22). *G2* (*Zmghk2*) transcript is present in dry seeds and is in C14, which peaks at T54, indicating the onset of BS chloroplast development after T45, consistent with the inference from the DEG (Fig. S3 A–C) and cluster (Fig. S3 D and E) analyses. In contrast, the transcript of *ZmGk1*, which is responsible for M chloroplast development in maize leaves (23), remains low until T54 (C15). Thus, BS chloroplasts may develop before M chloroplasts. There are other TFs with known functions in these clusters but their roles in early leaf development are not clear. For example, *ZFL2* (*zea floricula leaf2*) in C10 is known to regulate floral transition and flower development (24) but it is not clear why it is expressed so early postimbibition.

Although the role of most TFs in early leaf development is not known, their orthologs in *A. thaliana* provide hints on their functions. Take vascular development as an example. Several key *A. thaliana* regulators have roles in establishing and maintaining procambial and cambial cell populations and for further specification and differentiation into distinct cell types within xylem or phloem (25). For example, *A. thaliana* *BDL* (bodenlos), an AUX/IAA family protein with suppression of MP (monopolaris), which is important for procambial development and procambium differentiation in embryo (26). Expression of *BDL* ortholog GRMZM2G142768 peaks at T06 and gradually declines thereafter (Fig. S5). In contrast, the expression levels of maize MP orthologs (GRMZM2G034840 and GRMZM2G086949) start arising at T24 when *BDL* starts to drop. This contrasting pattern of expression suggests that the inhibition for vascular development is released at ∼T24 because *BDL* is down-regulated by an elevated level of auxin (27). This is consistent with elevated levels of the maize orthologs of *TIR1* and *PIN1* as discussed above. In addition, the levels of genes involved in vein formation begin to rise at T24 (Fig. S5). Thus, our time course data allows hypotheses to be generated and tested regarding when these TFs may function over the course of early maize leaf development.

**Concluding Remarks**

Our tissue samples for transcriptomic analysis contained all of the embryonic leaves in the seeds. As the first leaf is the largest, it should have the largest contribution to the transcriptomes. However, as new KS units or veins develop in all of these leaves, our transcriptomes are useful for hypothesizing the regulatory genes involved in KS development. The samples also contained the shoot apical meristem (SAM), but as SAM is small compared with the leaves, its contribution to the transcriptomes may be small. In any event, our transcriptomes reveal a dynamic profile of transcriptional transition in parallel to anatomical transition in maize early leaf development during the first 72 h of seed germination.

**Materials and Methods**

A full description for Materials and Methods is provided in *SI Text*.

**Plant Growth and Sample Collection.** Seeds of Zea mays cv. White Crystal were imbibed. The embryonic leaves were taken every 6 h for 3 d for transcriptomic analysis. Details for plant growth conditions and anatomical studies are provided in *SI Text*, *Plant Growth Conditions and Sample Collection* and *SI Text*, *Anatomical Studies*, respectively.

**RNA Sequencing and Data Processing.** Total RNA was extracted, purified, quantified, and assessed for quality before illumina sequencing. The library preparation steps and sequencing details are outlined in *SI Text*, *RNA Extraction and Sequencing*. Low quality reads were filtered based on quality scores. Read alignments to the B73 genome as well as gene expression level quantification were carried out following the "alternative protocol B" (28) as detailed in *SI Text*, *Data Processing and Analysis*.

**Expression Profile Analyses.** Genes were classified into coexpression modules based on the Pearson correlation coefficients of expression profiles between genes using hierarchical clustering (*SI Text*, *Expression Profile Correlation and Clustering*). We used the nonparametric method of Tarazona et al. (29) to identify DEGs between two samples (*SI Text*, *Identification of Differentially Expressed Genes*). Fisher’s exact tests were conducted to assess functional enrichment based on MapMan annotations (30).

**Acknowledgments.** We thank Jen Sheen, Michael Feeley, and Thomas Okita for comments. This study was supported by Academia Sinica (AS-101-SS-A08 and AS-102-SS-A13) and National Science Council (Taiwan) Fellowship NSC-106-2917-I-166-053 (to S.C.-C.) and NSC 99-2321-B-001-041-MT2.

Supporting Information

Liu et al. 10.1073/pnas.1301009110

SI Text

Plant Growth Conditions and Sample Collection. Seeds of Zea mays cv. White Crystal, a gluttonous maize cultivar, were purchased from a local supplier and stored at room temperature. For germination, seeds were imbibed in distilled water at 6:00 PM by first shaking for 10 min at 200 rpm and then germinating on wet filter paper on Petri dishes in the greenhouse under natural sunlight conditions (approximately 6:00 AM to approximately 7:00 PM) during the summer. The first set of seeds was imbibed on 6/2/2011 and leaf tissue for RNA preparation was taken every 6 h for 2 d [up to the 48th hour (T48)]. A second set of seeds was imbibed at 6:00 PM on 7/24/2011 and leaf tissue for RNA preparation was taken at T54, T60, T66, and T72. The epicotyls were collected, and then coleoptiles were removed before the embryonic leaves were isolated. These isolated leaves were then stored in liquid nitrogen. The whole procedure was done in an hour. The maximum photosynthetic photon flux density around noon was \( \sim1,600 \text{ \mu mol m}^{-2} \text{ s}^{-1} \), and the day/night temperatures were 30–35 °C/25–28 °C, with a relative humidity of 60–70%. Samples of the embryonic leaves and the shoot apical meristem (SAM) were collected from dry seeds (T00) and germinating seeds every 6 h up to 72 h for leaf anatomical examination and for RNA isolation and transcriptomic analysis. For RNA isolation, \( \sim200 \) plumules were dissected from the seeds and after removing the sheath tissue, the entire embryonic leaves and the SAM were stored in liquid \( \text{N}_2 \) until RNA extraction. Our tissue samples for transcriptomic analysis contained all of the four to five embryonic leaves in the maize seeds. As the first leaf is the largest, it should have the largest contribution to the transcriptomes. However, as new kranz structures (KSS) develop in these leaves, the transcriptomes are useful for hypothesizing the regulatory genes involved in KSS development. The samples also contained SAM, but because SAM is small, its contribution to the transcriptomes may not be significant.

Anatomical Studies. For anatomical examination, samples of the embryonic leaves and SAM were fixed in 2.5% (wt/vol) glutaraldehyde and postfixed in 1% (wt/vol) OsO4, both in 0.1 M sodium phosphate buffer (1). After dehydration through an ethanol series, the samples were infiltrated and embedded in Spurr’s resin. Semithin cross-sections (0.9 \( \mu \)m) were cut from the base of embryonic leaves with an Ultracut E Microtome (Reichert-Jung) and stained with 0.1% (wt/vol) toluidine blue and 0.1% (wt/vol) borax for observation with a light microscope (Olympus).

RNA Extraction and Sequencing. Total RNA was extracted using TRIzol reagent (Invitrogen), and 1 \( \mu \)L TURBO DNase (Ambion) per 10 \( \mu \)g RNA was added and incubated for 30 min at 37 °C to remove traces of contaminating DNA; and RNA was subsequently purified by the phenol/chloroform procedure. The RNAs were quantified using Qubit RNA Pico reagent (Invitrogen) and assessed for purity and quality using NanoDrop (Thermo Scientific) and BioAnalyzer (Agilent). RNA-seq libraries were prepared using the Illumina Standard mRNA-seq library preparation kit (Illumina) based on the manufacture’s protocol with the following modifications. Briefly, 15 \( \mu \)g of total RNA was subjected to two rounds of oligo-dT bead purification, fragmented for 2.75 min, and primed for first strand cDNA using 0.8x amount of random primer. After double strand cDNA synthesis and adaptor ligation, each sample was fractionated on 2% (wt/vol) Low Range Ultra Agarose (Bio-Rad) gels, from which the libraries of two different size ranges (\(~300 \text{ bp and } \sim400 \text{ bp}\)) were excised. The purified libraries were then independently amplified by 15 cycles of PCR and cleaned up using AMPure XP beads (Beckman Agencourt). RNA-seq libraries were subjected to Illumina sequencing by paired-end 2*101 nt (both ends with 101 nucleotides long) on a HiSeq2000 at the NGS Core Facility of Biodiversity Research Center in Academia Sinica, Taiwan. A total of 195–292 million pairs of raw reads was obtained from the two libraries (two lanes) for each time point sample; each pair of reads was counted as two reads. For each time point, \( \sim390–584 \) million raw reads and 240–351 million mappable reads were obtained.

Data Processing and Analysis. Low-quality bases and reads were removed by three criteria: (i) the consecutive bases from the end of a read with a default low-quality score of 2 (phred score of 2 or Q2 (2)), (ii) the bases from the beginning of a read until all of the scores of the first 20 remaining bases were at least Q20 (the base call error rate of \( \sim1\% \)), and (iii) the trimmed reads with less than 60 remaining bases. (Phred score is a general metric for the accuracy of a sequencing platform (2). The Q2 indicator does not give a specific error rate, but rather indicates a specific portion of the read that should not be used in further analyses.) Each pair of reads was treated as two single reads. Then, all alignment and quantification processes followed the “alternative protocol B” in ref. 3). The processed reads were mapped to the maize genome (ZmB73_RefGen_v2) and its embedded aligner Bowtie version 0.12.7 (http://bowtie-bio.sourceforge.net). Each read was aligned by the “-n” policy, and at most 10 hits were allowed. The expression level of each working gene set (WGS) gene was estimated in the reads per kilobase per million (RPKM) (4), using Cufflinks version 2.0.2 (http://cufflinks.cbcb.umd.edu/). We selected only genes with RPKM \( \geq 2 \) in two or more time points for further analysis. To compare the expression levels of the selected genes across all time points, we applied the upper quartile normalization procedure (5), which reduces the bias in RPKM values due to highly expressed genes. We used the transcriptome at T42 as the reference because its RPKMs were most evenly distributed among the 13 transcriptomes. The RPKM values at a time point were scaled by a constant factor to make the 75th percentile of the RPKM values for all expressed genes at that time point equal to the one at T42.

Expression Profile Correlation and Clustering. The clustering tool gCLUTO (Graphical Clustering Toolkit) was downloaded from http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/overview. It includes several hierarchical clustering methods and we selected the repeated bisection method, which is a top-down hierarchical clustering algorithm that separates samples with the largest distance. We calculated the Pearson correlation coefficient (PCC) between each pair of expression profiles of the genes under study and used the PCC values to cluster the expression profiles. A scalar quantity, called the figure of merit (FOM) (6), was used to assess the quality of the clustering algorithm and to determine the optimal number of clusters. To determine the functional enrichment for each cluster, we used the MapMan functional annotation (7). The annotated entries with the functional category name of “not assigned.unknown” were excluded in this study. We then used Fisher’s exact test to examine whether a function was significantly over- or un-


[^2]: "[http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/overview](http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/overview)"
derrepresented in a selected cluster of genes against the set of all expressed genes with MapMan annotations.

**Identification of Differentially Expressed Genes.** We used the non-parametric method of Tarazona et al. (8) to identify differentially expressed genes (DEGs) between two samples. Here, we set the \( q \) value (differentially expression probability) in the method to be 0.7, because in our dataset genes with \( q > 0.7 \) have at least a 2.3-fold change in RPKM between the two samples. The test is stringent. For example, the number of DEGs between T06 and T12 was only 39, and because 30,255 genes (tests) were used, the probability of passing a test was only \( \sim 39/30,000 = 0.0013 \), which is much smaller than 0.01 (the 1% significance level). Thus, the test is conservative and can be used to identify DEGs between two time points.

The above analysis was between two time points. To find all genes differentially expressed during leaf development, we have developed the following method to make all pairwise comparisons between time points, which takes care of the problem of multiple testing. Suppose the probability of “success” in a single test is \( p \) and a total of \( N \) tests are made. We use the binomial expansion to calculate the probability of at least \( X \) successes in \( N \) tests. In the above, we have seen that \( p \) is only \( \sim 0.0013 \). To be conservative, let us take \( p = 0.005 \). Among the 13 time points there are 78 possible pairwise comparisons, that is \( n = 78 \). Then the probability to have at least \( X = 4 \) “successes” is only \( \sim 0.0001 \). Under this criterion, we find that 13,907 (46%) of the 30,255 expressed genes and 590 (47%) of the 1,238 expressed TF genes are differentially expressed during the 72-h time course.


<table>
<thead>
<tr>
<th></th>
<th>T00</th>
<th>T06</th>
<th>T12</th>
<th>T18</th>
<th>T24</th>
<th>T30</th>
<th>T36</th>
<th>T42</th>
<th>T48</th>
<th>T54</th>
<th>T60</th>
<th>T66</th>
<th>T72</th>
</tr>
</thead>
<tbody>
<tr>
<td>base</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
<td>0.22</td>
<td>0.33</td>
<td>0.39</td>
<td>0.51</td>
<td>0.44</td>
<td>0.46</td>
<td>0.63</td>
<td>0.64</td>
<td>0.49</td>
<td>0.52</td>
</tr>
<tr>
<td>-1 cm</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.10</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>+4 cm</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>tip</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Fig. S1.** Comparison between mature leaf and embryonic leaf time course transcriptomes. Pearson’s correlation coefficients (PCCs) between the four mature leaf transcriptomes of Li et al. (1), as indicated in the first column, and our time course transcriptomes, as indicated in the first row.

Liu et al. (1) Blue and red represent under- and overrepresented categories, respectively.

Fig. S2. MapMan functional categories enriched in expressed genes at each time point. Enrichment tests conducted using (A) all expressed genes and (B) genes with top 10% or (C) top 1% RPKM values in each time point. The heatmap is color-coded based on the Z transformation of adjusted P values (1). Blue and red represent under- and overrepresented categories, respectively.

Fig. S3. (Continued)
Fig. S3. (Continued)
Fig. S3. (Continued)
Fig. S3. MapMan functional categories enriched in differentially expressed genes (DEGs). Enrichment tests with (A) type 1 DEGs, (B) type 2 DEGs, and (C) type 3 DEGs. (D) Clusters of all three types of DEGs. (E) Enrichment tests on each DEG cluster shown in D.
**Fig. S4. (Continued)**

| Mapman Functional Categories | 00 | 05 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
|-----------------------------|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| C01 | | | | | | | | | | | | | | | | | | | | | |
| C02 | | | | | | | | | | | | | | | | | | | | | |
| C03 | | | | | | | | | | | | | | | | | | | | | |
| C04 | | | | | | | | | | | | | | | | | | | | | |
| C05 | | | | | | | | | | | | | | | | | | | | | |
| C06 | | | | | | | | | | | | | | | | | | | | | |
| C07 | | | | | | | | | | | | | | | | | | | | | |
| C08 | | | | | | | | | | | | | | | | | | | | | |
| C09 | | | | | | | | | | | | | | | | | | | | | |
| C10 | | | | | | | | | | | | | | | | | | | | | |
| C11 | | | | | | | | | | | | | | | | | | | | | |
| C12 | | | | | | | | | | | | | | | | | | | | | |
| C13 | | | | | | | | | | | | | | | | | | | | | |
| C14 | | | | | | | | | | | | | | | | | | | | | |
| C15 | | | | | | | | | | | | | | | | | | | | | |
| C16 | | | | | | | | | | | | | | | | | | | | | |
| C17 | | | | | | | | | | | | | | | | | | | | | |
| C18 | | | | | | | | | | | | | | | | | | | | | |
| C19 | | | | | | | | | | | | | | | | | | | | | |
| C20 | | | | | | | | | | | | | | | | | | | | | |
| C21 | | | | | | | | | | | | | | | | | | | | | |
| C22 | | | | | | | | | | | | | | | | | | | | | |
| C23 | | | | | | | | | | | | | | | | | | | | | |
| C24 | | | | | | | | | | | | | | | | | | | | | |
| C25 | | | | | | | | | | | | | | | | | | | | | |
| C26 | | | | | | | | | | | | | | | | | | | | | |
| C27 | | | | | | | | | | | | | | | | | | | | | |
| C28 | | | | | | | | | | | | | | | | | | | | | |
| C29 | | | | | | | | | | | | | | | | | | | | | |
| C30 | | | | | | | | | | | | | | | | | | | | | |
Fig. S4. MapMan functional categories enriched in genes of various coexpression gene clusters. (A) Expression profiles of genes sorted according to cluster orders. (B) Enrichment tests of detailed overrepresented functional categories on each cluster as shown in A. (C) Expression profiles of genes in selected clusters. (D) Enrichment tests of both over- and underrepresented high-level categories on each cluster as shown in C. (E) Enrichment tests of both over- and underrepresented hormone-related genes on each cluster as shown in C.
Fig. S5. Maize transcription factors likely regulating vascular development. (A) The current known components and relationships of vascular development genes in Arabidopsis thaliana. (B) Expression patterns of putative maize orthologs of A. thaliana vascular development genes.
Table S1. Read count statistics of the deep sequencing data for the 13 time points studied

<table>
<thead>
<tr>
<th>Sample</th>
<th>Library length, bp</th>
<th>Quality filtering</th>
<th>Mappable reads</th>
<th>Rate, %</th>
<th>Total reads</th>
<th>Rate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total reads</td>
<td>Rate, %</td>
<td></td>
<td>Total</td>
<td>Rate, %</td>
</tr>
<tr>
<td>T00_ST-DA01</td>
<td>300</td>
<td>218,477,277</td>
<td>75.9</td>
<td>277,488,186</td>
<td>76.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>195,751,388</td>
<td>86.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T06_ST-DA02</td>
<td>300</td>
<td>213,618,797</td>
<td>74.8</td>
<td>255,413,163</td>
<td>71.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>142,589,424</td>
<td>85.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T12_ST-DA03</td>
<td>300</td>
<td>206,947,451</td>
<td>74.6</td>
<td>275,874,436</td>
<td>75.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>156,344,245</td>
<td>82.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T18_ST-DA04</td>
<td>300</td>
<td>210,270,688</td>
<td>76.5</td>
<td>279,899,765</td>
<td>76.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>155,975,570</td>
<td>84.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T24_ST-DA05</td>
<td>300</td>
<td>190,425,691</td>
<td>78.2</td>
<td>294,733,426</td>
<td>78.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>183,789,389</td>
<td>79.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T30_ST-DA06</td>
<td>300</td>
<td>195,697,940</td>
<td>80.2</td>
<td>271,280,993</td>
<td>77.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>152,421,099</td>
<td>85.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T36_ST-DA07</td>
<td>300</td>
<td>215,565,717</td>
<td>79.9</td>
<td>308,395,082</td>
<td>80.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>169,797,300</td>
<td>84.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T42_ST-DA08</td>
<td>300</td>
<td>153,005,724</td>
<td>85.2</td>
<td>242,223,816</td>
<td>72.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>182,552,562</td>
<td>81.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T48_ST-DA09</td>
<td>300</td>
<td>156,382,336</td>
<td>83.7</td>
<td>240,264,229</td>
<td>73.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>172,108,685</td>
<td>84.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T54_ST-D805</td>
<td>300</td>
<td>222,228,743</td>
<td>76.8</td>
<td>340,039,005</td>
<td>81.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>196,448,375</td>
<td>82.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T60_ST-D806</td>
<td>300</td>
<td>236,437,704</td>
<td>76.3</td>
<td>351,089,370</td>
<td>80.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>199,181,671</td>
<td>82.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T66_ST-D807</td>
<td>300</td>
<td>237,623,498</td>
<td>74.6</td>
<td>339,580,019</td>
<td>75.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>211,538,295</td>
<td>79.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T72_ST-D808</td>
<td>300</td>
<td>208,071,129</td>
<td>80.7</td>
<td>320,303,598</td>
<td>75.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>217,502,039</td>
<td>76.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S2. Numbers of annotated genes and expressed features

<table>
<thead>
<tr>
<th>Gene set</th>
<th>No. of expressed features</th>
<th>CDS</th>
<th>TF</th>
<th>Pseudo</th>
<th>TE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGS A</td>
<td>63,363</td>
<td>2,070</td>
<td>17,344</td>
<td>29,024</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>30,255</td>
<td>1,238</td>
<td>3,040</td>
<td>1,986</td>
<td></td>
</tr>
<tr>
<td>FGS A</td>
<td>39,441</td>
<td>2068</td>
<td>95</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>23,900</td>
<td>1,237</td>
<td>95</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>

A, all; CDS, protein coding gene; E, expressed (RPKM ≥1 in more than one time point); FGS, filtered gene set; pseudo, pseudogene; TE, transposable element; TF, transcription factor; WGS, working gene set.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)