Genome-wide atlas of transcription during maize development

Rajandeep S. Sekhon[1,2,†], **Haining Lin**[3,4,†], **Kevin L. Childs**[3,4], **Candice N. Hansey**[3,4], **C. Robin Buell**[3,4], **Natalia de Leon**[1,2] and **Shawn M. Kaeppler**[1,2,*]

[1]*Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA,*

[2]*Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA,*

[3]*Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824, USA, and*

[4]*Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA*

SUMMARY

Maize is an important model species and a major constituent of human and animal diets. It has also emerged as a potential feedstock and model system for bioenergy research due to recent worldwide interest in developing plant biomass-based, carbon-neutral liquid fuels. To understand how the underlying genome sequence results in specific plant phenotypes, information on the temporal and spatial transcription patterns of genes is crucial. Here we present a comprehensive atlas of global transcription profiles across developmental stages and plant organs. We used a NimbleGen microarray containing 80 301 probe sets to profile transcription patterns in 60 distinct tissues representing 11 major organ systems of inbred line B73. Of the 30 892 probe sets representing the filtered B73 gene models, 91.4% were expressed in at least one tissue. Interestingly, 44.5% of the probe sets were expressed in all tissues, indicating a substantial overlap of gene expression among plant organs. Clustering of maize tissues based on global gene expression profiles resulted in formation of groups of biologically related tissues. We utilized this dataset to examine the expression of genes that encode enzymes in the lignin biosynthetic pathway, and found that expansion of distinct gene families was accompanied by divergent, tissue-specific transcription patterns of the paralogs. This comprehensive expression atlas represents a valuable resource for gene discovery and functional characterization in maize.

Keywords: maize, transcriptome, gene atlas, gene expression, development, microarray.

INTRODUCTION

For over a century, maize (*Zea mays* ssp. *mays* L.) has served as a model system for the understanding of diverse biological phenomena, including heterosis, transposition, paramutation, imprinting and allelic diversity (Bennetzen and Hake, 2009). Maize is also one of the most important food crops, occupying 156 million hectares worldwide and producing 809 million tones of grains in 2009 (http://www.fas.usda.gov/psdonline/). The recent focus on the use of C4 grasses as a sustainable source of lignocellulosic biomass to produce liquid fuels has established a role for maize as a potential bioenergy feedstock (Perlack *et al.*, 2005) and a model system for bioenergy research. In addition, translational research in maize could assist in rapid domestication and development of other closely related C4 grasses such as

switchgrass (*Panicum virgatum*) and miscanthus (*Miscanthus gigantus*) into sustainable biofuel feedstocks (Lawrence and Walbot, 2007).

Developing sustainable biofuel crops will involve improving the productivity and biochemical composition of the biomass, and thus will require further fundamental understanding of maize biology. For instance, a complete understanding of the molecular networks involved in the assembly of grass cell walls will be crucial for efficient breakdown of the lignocellulosic biomass into ethanol (Carpita and McCann, 2008). Similarly, understanding of regulatory mechanisms underlying the biosynthesis, transportation and storage of photosynthates will be crucial for improving the energy content of biofuel feedstocks. The

recent sequencing of the maize genome (Schnable *et al.*, 2009) has provided a framework for the identification and functional characterization of genes and genetic networks for crop improvement and basic research. The availability of global transcriptome profiling technologies, such as DNA microarrays, together with the genome sequence offer the opportunity to understand patterns of transcription in the context of plant growth and development. In plants, atlases of global transcription have been developed for several species, including *Arabidopsis thaliana* (Schmid *et al.*, 2005), *Medicago truncatula* (Benedito *et al.*, 2008), rice (*Oryza sativa*) (Jiao *et al.*, 2009; Wang *et al.*, 2010), soybean (*Glycine max*) (Libault *et al.*, 2010) and barley (*Hordeum vulgare*) (Druka *et al.*, 2006). Although several studies have documented gene expression profiles of individual tissues or biological processes in maize, no single study has documented global transcription patterns of diverse organs using a single platform.

Here we describe an atlas of global gene expression covering major developmental steps during the life cycle of a maize plant. To demonstrate the applicability of our dataset, we present organ- and paralog-specific expression patterns of lignin biosynthetic pathway genes in vegetative organs. The transcriptome data presented here and those publically available at PLEXdb (http://www.plexdb.org), MaizeGDB (http://www.maizegdb.org) and GEO (http://www.ncbi.nlm.nih.gov/geo/) will serve as valuable resources for functional characterization of maize genes.

RESULTS AND DISCUSSION

Generation and quality assessment of the dataset

To document transcription profiles of maize, we designed a NimbleGen microarray. The microarray contained 330 788 probes that were designed to represent in-house gene models identified from B73 BAC sequences (version 1a.49; http://www.maizesequence.org) and PlantGDB-assembled unique transcript (PUT) assemblies from the PlantGDB (http://www.plantgdb.org) (see Experimental procedures). After release of the higher-quality B73 maize genome (Schnable *et al.*, 2009), all probes were further searched against the official maize genome sequence and the mRNA sequences of the maize gene models (AGP_v1, release 4a.53; http://www.maizesequence.org/). Based on the number of times a probe could be mapped to the B73 maize genome, the probes were classified into unique probes, repetitive probes and unmapped probes (see Experimental procedures). The final dataset included 218 980 unique and 100 942 unmapped probes that formed 80 301 probe sets (Figure S1). Of these, 30 892 probe sets corresponded to the official maize cDNA models (including the alternatively spliced isoforms) encoded by 23 740 high-confidence genes. Only those alternately spliced isoforms that had at least one unique probe differentiating them were included.

In cases where the isoforms could not be differentiated, the representative cDNA model (the one with the longest open reading frame among all isoforms of a gene) was chosen. Among the remaining probe sets, 31 290 were represented by probes that mapped to the B73 genome but did not belong to the high-confidence gene models. Finally, a total of 30.5% probes, corresponding to 18 119 in-house gene models, could not be mapped to the B73 genome or the official maize gene models. There are two major reasons for this. First, some BAC sequences that were used for the design of the probes were not included in construction of the official B73 genome sequence. Second, the in-house gene models include PUT assemblies from the PlantGDB, some of which came from genotypes other than B73. Some probes designed for such gene models could not be mapped to the B73 genome, either because of sequence polymorphisms or absence of the sequences from B73.

We profiled transcript levels using RNA samples from 60 diverse tissues representing 11 major organ systems and varying developmental stages of the maize plant (Table 1, Figure S2 and Table S1). The organ systems included germinating seed, root, whole seedling, stem and shoot apical meristem, internodes, cob, tassel and anthers, silk, leaf, husk and seed. Each tissue was represented by three biological replicates.

To assess the quality of the dataset, we focused on the 30 892 probe sets that represent transcripts encoded by the well-annotated set of 23 740 genes in the B73 genome. The representative cDNA model was chosen to represent an alternatively spliced gene for data analyzes performed on a gene basis. Below, the terms 'genes' and 'probe sets' refer to the 23 740 well-annotated genes and the 30 892 probe sets representing transcripts encoded by the well-annotated genes, respectively.

The results for biological replicates for each tissue were highly correlated, with an average Pearson's correlation coefficient value of 0.968 ± 0.001 (Figure S3 and Table S2). The lowest correlation coefficient value was 0.907, and 89% of correlation coefficients were >0.95 ($P < 0.0001$). These correlations indicate that the microarray platform is technically repeatable, and that the results of biological replicates in this study are highly reproducible. As an additional test of data quality, we compared the expression profiles of some of the well-studied genes, and found these to be consistent with earlier reports (Figure S4). For example, *opaque2* and *shrunken2*, two starch biosynthetic pathway genes (Schmidt *et al.*, 1990; Schultz and Juvik, 2004), were almost exclusively expressed in the endosperm. Similarly, expression of *LEAFY COTYLEDON1* (*LEC1*), a transcription factor involved in embryogenesis (Lotan *et al.*, 1998), was specific to embryos, while expression of *ZmTIP2-3*, a root-specific aquaporin gene (Lopez *et al.*, 2004), was only detected in roots. In summary, these observations show that the microarray dataset described here is of high quality.

Table 1 Maize tissues included in the genome-wide gene expression atlas of inbred line B73

Organ group	Tissue description	Growth stage at collection	Tissue designation
Germinating seed	Germinating seed	24 h after imbibition	24H_Germinating seed
Root	Primary root (GH)	6 DAS	6DAS_GH_Primary root
	Primary root	Vegetative emergence (VE); coleoptile barely emerges from the soil surface	VE_Primary root
	Primary root (GH)	Vegetative 1 (V1); first leaf is fully extended	V1_GH_Primary root
Whole seedling	Coleoptile (GH)	6 DAS	6DAS_GH_Coleoptile
	Whole seedling	VE	VE_Whole seedling
Shoot apical meristem (SAM) and young stem	Stem and SAM	V1	V1_Stem and SAM
	Stem and SAM	Vegetative 3 (V3); three fully extended leaves.	V3_Stem and SAM
	Stem and SAM	Vegetative 4 (V4); four fully extended leaves	V4_Stem and SAM
	Shoot tip	Vegetative 5 (V5); five fully extended leaves	V5_Shoot tip
Internodes	First internode	V5	V5_First internode
	First internode	Vegetative 7 (V7); seven extended leaves	V7_First internode
	Fourth internode	Vegetative 9 (V9); nine extended leaves	V9_Fourth internode
Cob	Immature cob	Vegetative 18 (V18); 18 extended leaves	V18_Immature cob
	Pre-pollination cob	Reproductive 1 (R1); silks emerge from the husk	R1_Pre-pollination Cob
Tassel and anthers	Immature tassel	Vegetative 13 (V13)	V13_Immature tassel
	Meiotic tassel	V18	V18_Meiotic tassel
	Anthers	R1	R1_Anthers
Silk	Silk	R1	R1_Silks
Leaves	Pooled leaves	V1	V1_Pooled leaves
	First leaf and sheath	V3	V3_First leaf and sheath
	Topmost leaf	V3	V3_Topmost leaf
	Base of stage 2 leaf	V5	V5_Base of stage-2 leaf
	Tip of stage 2 leaf	V5	V5_Tip of stage-2 leaf
	Base of stage 2 leaf	V7	V7_Base of stage-2 leaf
	Tip of stage 2 leaf	V7	V7_Tip of stage-2 leaf
	Immature leaves	V9	V9_Immature leaves
	Thirteenth leaf	V9	V9_Thirteenth leaf
	Eleventh leaf	V9	V9_Eleventh leaf
	Eighth leaf	V9	V9_Eighth leaf
	Thirteenth leaf	Vegetative tasseling (VT); last branch of the tassel fully emerged	VT_Thirteenth leaf
	Thirteenth leaf	Reproductive 2 (R2); 10–14 days after silk emergence	R2_Thirteenth leaf
Husk	Innermost husk	R1	R1_Innermost husk
	Outer husk	R2	R2_Outer husk
	Innermost husk	R2	R2_Innermost husk
Seed	Whole seed	2–24 DAP (samples collected every other day, 12 tissues)	XDAP_Whole seed (X = days)
	Endosperm	12–24 DAP (samples collected every other day, seven tissues)	XDAP_Endosperm (X = days)
	Embryo	16–24 DAP (samples collected every other day, five tissues)	XDAP_Embryo (X = days)
	Pericarp	18 DAP	18DAP_Pericarp

DAS, days after sowing; DAP, days after pollination.

Overview of the global gene expression trends

Expression of 91.4% of the probe sets was detected in at least one of the 60 tissues. The proportion of probe sets detected in individual tissues ranged between 73 and 81%, indicating that the total number of expressed genes is comparable among tissues. Such high proportion of detected probe sets may be attributed to the diversity of tissues included in the study, and to the fact that the analysis presented here involved only the high-confidence genes in the maize genome. Despite the diversity in plant organs used in our study, 2647 (8.6%) probe sets, corresponding to 2407 genes, were not detected at a level above the arbitrary threshold of 200 on the microarray.

Given that our analysis focused on the high-confidence gene set, it is unlikely that these are pseudogenes or mis-annotations. A plausible explanation is that these genes are expressed under specific environmental conditions or are very specific to organs and/or developmental stages that are not covered in this study. GO Slim enrichment analysis revealed that the non-expressed set was enriched in genes expressed in response to abiotic stimuli ($P < 0.01$), and those involved in transcription factor activity ($P < 0.001$) (Figure S5), consistent with a specialized function of these genes. However, it is also possible that some of the genes in the non-expressed category are expressed at very low levels and thus did not meet our expression cut-off limit.

Identification of constitutively expressed and putative housekeeping genes

In our dataset, using a conservative expression threshold, 44.5% of the probe sets were detected in all tissues, indicating a remarkable overlap of gene expression among biologically distinct plant organs. GO Slim enrichment analysis revealed that this set was significantly ($P < 0.001$) enriched with biological processes that included cellular processes, transport, protein modification, translation and signal transduction (Figure S6). Thus many of the genes in this set are involved in basic biological processes, and are expected to be expressed in all tissues. However, despite detection in all tissues, the levels of expression of constitutively expressed genes were highly variable among tissues, with a coefficient of variation (CV) ranging from 10 to 744%.

Constitutively expressed genes that show stable expression across all tissues are likely to be involved in basal metabolic or 'housekeeping' functions. Stably expressed genes are of practical use as controls in expression experiments. A search of expressed genes with the least variability in expression (CV ≤ 15%) yielded 113 genes (Table S3). The most stable expression (CV 9.6%) was observed for a gene encoding a ubiquitin-conjugating enzyme. GO Slim enrichment analysis revealed over-representation of genes involved in kinase activity ($P < 0.001$), nucleotide binding ($P < 0.001$) and protein modification processes ($P < 0.01$) (Figure S7), which further support a housekeeping function for this set of genes. Notably, none of the genes traditionally used as a control genes for various expression assays, such as *actin*, *ubiquitin*, *glyceraldehyde-3-phosphate dehydrogenase* (GADPH) and *tubulin*, were among the most stably expressed genes (Table S3). Similar to a recently published rice transcriptome analysis (Wang *et al.*, 2010), we found that these traditionally used control genes had variable expression among maize tissues (Figure S8a). In contrast, the 20 most stably expressed genes in our dataset had uniform expression (Figure S8b). Wang *et al.* (2010) also identified a set of the 100 most stably expressed genes in diverse organs of rice. Comparing the rice and maize sets using high stringency (over 65% identity and 60% coverage of peptide sequence), we found 10 orthologs that were stably expressed in both the species (Table S3). These results highlight the value of genome-wide analysis across diverse tissues to select the most appropriate control genes for expression quantification.

Identification and expression dynamics of organ-specific genes

Elucidation and understanding of organ-specific gene expression is a fundamental question of biology with broad applications in basic and applied research. For instance, identification and characterization of such genes can help to unravel the molecular mechanisms underlying differentiation and development of an organ, and offer opportunities for targeted manipulations of gene expression for economic purposes. A search for organ-specific genes in our dataset yielded 863 genes with distinct expression patterns in eight organs (Figure 1a and Table S4). The largest numbers of organ-specific genes (334) were observed for leaves (Figure 1b). This is in contrast with earlier studies in rice, soybean and Arabidopsis, in which relatively few leaf-specific genes were reported (Libault *et al.*, 2010; Schmid *et al.*, 2005; Wang *et al.*, 2010). This result is most likely due to extensive representation of leaf tissues in our study, and indicates that detailed sampling from multiple stages is crucial for better understanding of molecular mechanisms underlying the development of a particular organ. The endosperm, roots and tassels also had a large number of genes with organ-specific expression, similar to that observed in rice (Wang *et al.*, 2010). Very few organ-specific genes were observed for internodes and cobs, two organs that collectively contribute to approximately 60% of total maize stover biomass (Hansey *et al.*, 2010) and are potential targets for biomass improvement. Thus, conventional and transgenic approaches designed to target these organs for improvements in biomass quality will probably also affect other organs. It may be possible to use the organ-specific genes identified in this study as direct targets for manipulation, and also as a source of tissue-specific promoters for transgenic research and development projects. However, it should be noted that expression cut-off alone may not be an ideal method for identification of organ-specific genes, and additional computational approaches and manual inspection may be required to refine this list.

To elucidate the dynamics of gene expression during the development of major maize organs, we calculated the relative gene expression (measured by Z score) (Figure 2). Leaf and endosperm samples had a balanced expression profile, with roughly equal numbers of genes showing higher and lower expression relative to their mean expression across organs. The distribution suppressed and up-regulated genes in these tissues was wide, indicating that a large number of genes deviate from their mean expression. This is consistent with the relatively higher number of organ-specific genes identified in these tissues. Internodes had a narrow distribution, indicating that expression of most genes was close to their mean across all tissues; this observation is in agreement with the observed low number of internode-specific genes. A bimodal distribution was observed in embryo samples, indicating that one group of genes was suppressed relative to their mean expression on the microarray, while another group was up-regulated.

Biologically related tissues have similar global gene expression patterns

To test whether the transcriptome of organs is an indicator of their identity, we clustered the tissues using principal

Figure 1. Organ-specific expression patterns detected in the microarray data.
(a) Heat map of the organ-specific genes, generated by hierarchical clustering based on Pearson's correlation.
(b) Distribution of tissue-specific genes detected in each of the eight selected organs. For organs represented by multiples tissues, information from those tissues was combined.
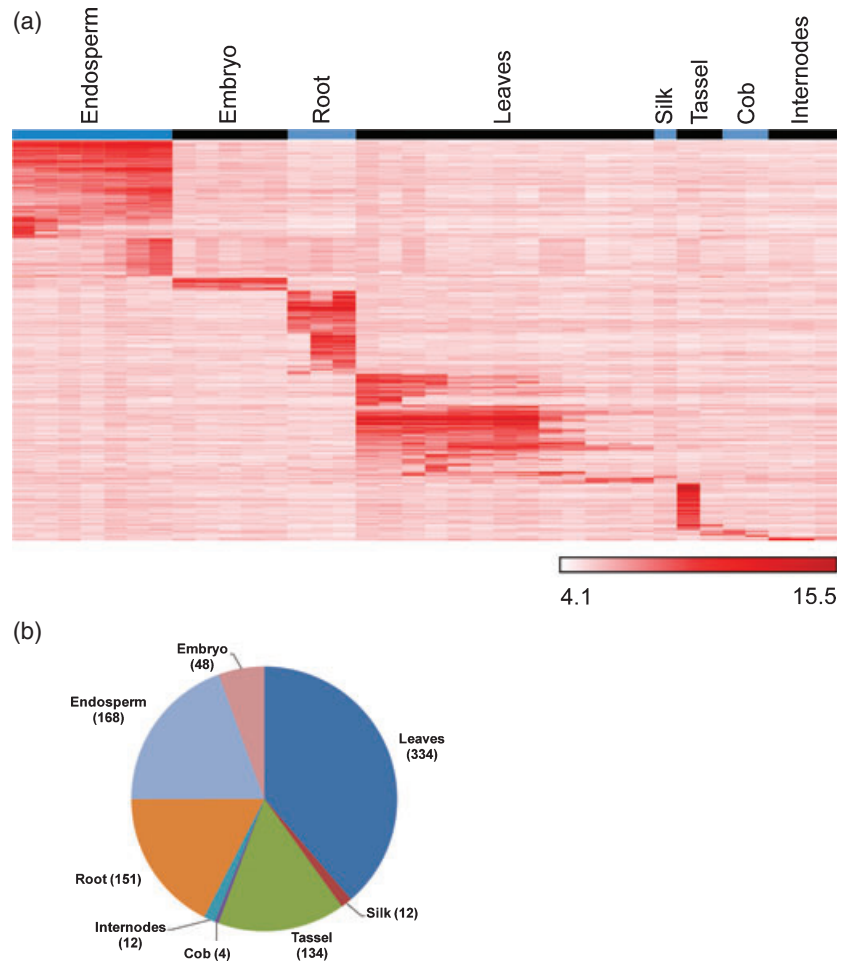
component analysis (PCA). For ease of data presentation, PCA was performed separately for vegetative and seed tissues. PCA for vegetative tissues revealed that tissues were indeed clustered based on their morphological, physiological and developmental similarity (Figure 3a). For instance, organs containing a shoot apical meristem clustered together, while roots formed a separate group, irrespective of growing environment (field or greenhouse). Leaf tissues clustered according to the developmental stage, with immature leaves being distinct from fully mature leaves. An interesting example of contrasting expression profiles within the same organ was evident in developing leaves. These leaves (designated stage 2 leaves by Sylvester *et al.*, 1990) are characterized by a mature, fully developed, green tip and an immature base with undifferentiated plastids and sheaths. The tip of stage 2 leaves clustered with mature leaves, while the base clustered with immature leaves, indicating substantial reprogramming of the transcriptome during leaf development. Indeed, a recent study of expression dynamics during maize leaf development found dramatic differences in the transcriptomes of the base and

tip of developing leaves (Li *et al.*, 2010). They showed that the base of developing leaves was enriched with genes encoding enzymes involved in cell-wall biosynthesis, cell division, cellulose synthesis and auxin signaling, while the tip was enriched in genes involved in photosynthesis and sugar metabolism/transport. We found that the expression profiles of key differentially expressed genes identified by Li *et al.* (2010) followed a similar trend in our dataset (Figure S9). For instance, putative genes encoding a UDP-glucose-6-dehydrogenase, an A-type cyclin, an auxin efflux carrier and several cellulose synthases were over-expressed in the base of stage 2 leaves (Figure S9a). In contrast, putative genes encoding a Rubisco activase, an ATP synthase, a chlorophyll *a/b* binding protein, a hexose carrier protein and a fructose-1,6-bisphosphatase were abundant in the tip of stage 2 leaves (Figure S9b). Thus, differential transcriptomes of developmentally distinct vegetative tissues were apparent from the PCA analysis.

PCA for seed tissues also produced a similar result; all embryos formed a tight cluster, signifying highly similar global expression profiles (Figure 3b). Whole seeds
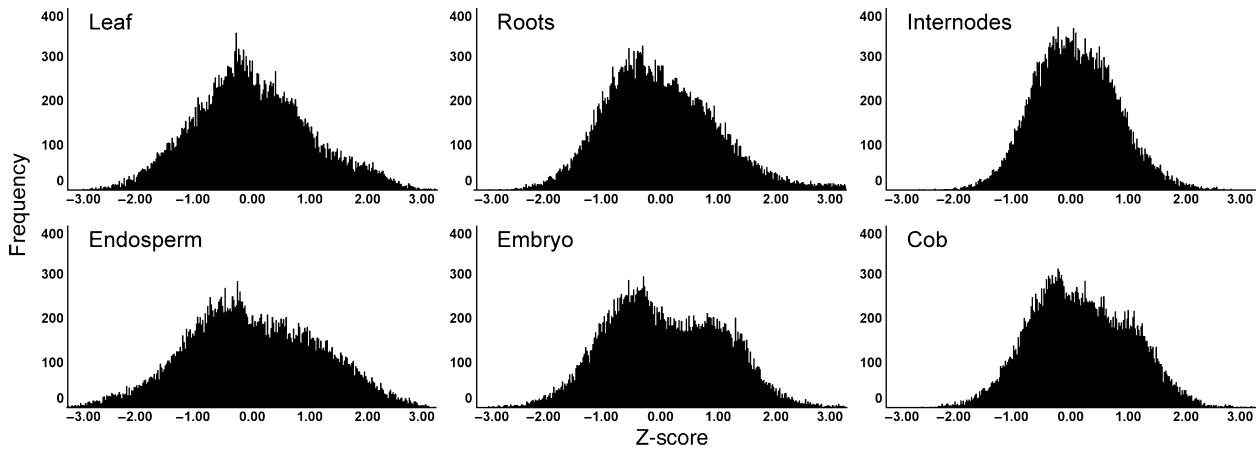
Figure 2. Dynamics of global gene expression in selected maize organs.
Histograms of relative expression levels (measured by Z scores) in six organs. For each of these organs, overall expression was calculated by averaging the RMA-normalized log$_2$-transformed expression values of all the tissues representing that organ. Z scores were calculated using the formula: $Z = (X - X_{mean})/S$, where X is the mean of the log$_2$-transformed expression of a gene in multiple tissues of an organ, and X_{mean} and S are the mean log$_2$-transformed expression and standard deviation of that gene across all selected organs, respectively.

presented a continuously and gradually changing expression profile, especially during early and mid-development. Young seeds at 2, 4, 6 and 8 days after pollination formed a cluster. This pattern is consistent with an important developmental landmark of active mitotic cell proliferation in the endosperm during this period (Sabelli and Larkins, 2009). Likewise, 16, 18, 20, 22 and 24 days after pollination endosperms clustered together, probably signifying active programmed cell death, which starts at approximately 16 days after pollination (Sabelli and Larkins, 2009).

To further corroborate the PCA-based analysis, we used hierarchical clustering, which provided a more detailed view of tissue relatedness (Figure S10). For instance, immature leaves clustered separately from mature green leaves, each forming a distinct group (Figure S10a). Likewise, immature whole seeds of 2–10 days after pollination formed a separate group, while more mature seed tissues at 18–24 days after pollination, when the seed becomes dominated by endosperm, fell into a distinct group (Figure S10b). The similarity of gene expression among related tissues and differences between organ groups are clearly visible from the heat map.

In summary, these observations are consistent with a role for transcription in regulating organ identity in plants. Tissues belonging to the same organ can have very distinct transcriptomes depending on their age and overall developmental stage. Thus, it is important to carefully select the developmental stage of an organ for such studies, and to ensure that the same stage is sampled across treatments within an experiment. This is highly relevant, for instance, in biofuel research, in which understanding and manipulation of the biochemical composition of organs (internodes/leaves) at a precise developmental stage will be required to develop value-added feedstocks.

Dynamic expression patterns of genes involved in lignin biosynthesis

To demonstrate the utility of this data in understanding specific expression pathways, we focused on the expression of genes involved in lignin metabolism. Lignin is an integral component of plant cell walls as it provides structural integrity to the cells and functionality to the vascular system. However, lignin quantity and composition can affect the quality of plants for various agricultural and industrial uses. For instance, lignin content and type can lower cellulosic ethanol yield by physically obstructing the accessibility of cell-wall polysaccharides to hydrolytic enzymes (Li *et al.*, 2008; Moore and Jung, 2001), interfering with the activity of hydrolytic enzymes, and through an inhibitory effect on the microbes used for fermentation of sugars (Keating *et al.*, 2006). Modification of lignin content and/or composition through altered expression of lignin pathway genes offers an attractive approach to improve crop plants for agricultural and bioenergy needs (Li *et al.*, 2008).

Compared with Arabidopsis, the lignin biosynthetic pathway in maize and related grasses is characterized by substantial expansion of various gene families. In a recent study, Penning *et al.* (2009) identified eight major gene families involved in lignin biosynthesis. We examined the expression dynamics of 96 genes belonging to these families in 35 vegetative tissues. Hierarchical clustering indicated that transcription of lignin metabolic genes varies with age as well as the biological relatedness of organs (Figure 4). Roots and the aerial plant parts formed distinct groups, suggesting diversification of the lignin pathway in these organs. Above-ground vegetative organs, especially leaves and internodes, were grouped according to age. In
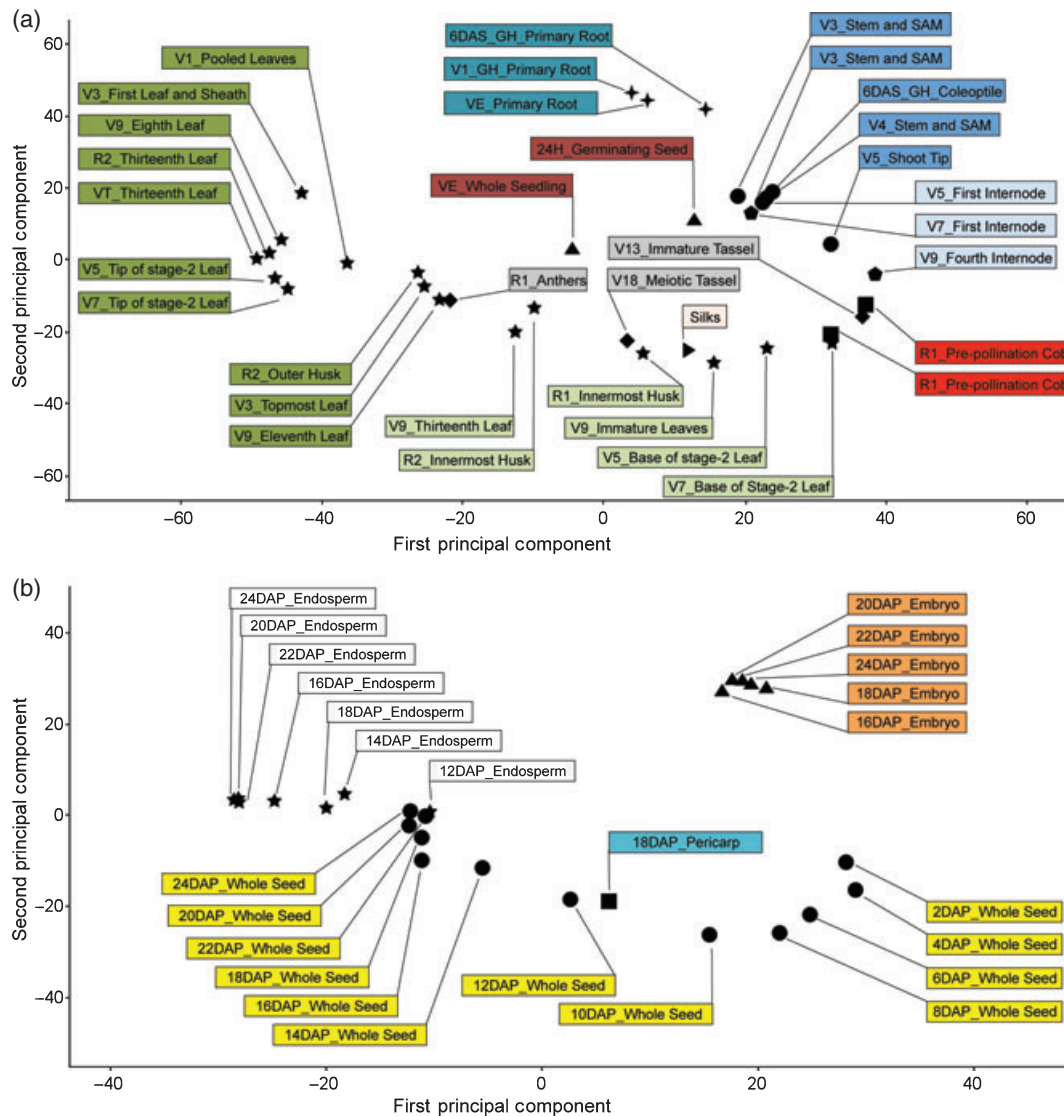
Figure 3. Global gene expression patterns reflect biological relatedness among maize tissues.
Principal component analysis was applied to 35 vegetative tissues (a) and 25 seed tissues (b), based on expression of 30 892 probe sets.

general, most of the lignin genes had relatively higher expression in the immature organs, indicating active secondary cell-wall formation and lignification in these organs early in development.

We observed distinct expression differences among paralogs of genes encoding phenylalanine ammonia lyase (PAL), cinnamoyl CoA reductase (CCR), caffeoyl CoA *O*-methyltransferase (CCoAOMT) and *p*-hydroxycinnamoyl CoA transferase (HCT) (Figure 4). PAL catalyzes the first committed step in the phenylpropanoid pathway, and governs the synthesis of several important secondary metabolites, including lignin, phytoalexins and signal molecules (Dixon *et al.*, 2002, 1983). Six of ten PAL paralogs were constitutively expressed, while four showed differential, organ-specific expression patterns. Organ-specific

expression patterns were more striking in genes later in the pathway. For instance, two of the 18 CCR paralogs (GRMZM2G017285_T01 and GRMZM2G146031_T01), and one of the six CCoAOMT paralogs (GRMZM2G033952_T01), were exclusively expressed in roots. Similarly, of the 38 HCT paralogs reported in B73 (Penning *et al.*, 2009), seven showed no expression in any vegetative or seed (not shown) tissues, suggesting that these may be expressed under specific environmental conditions or at developmental stages not assessed in this study. One of the paralogs (GRMZM2G154216_T01) showed remarkable specificity to endosperm, indicating diversification of the lignin pathway for seed development. Incidentally, HCT is one of only two enzymes (the other being *p*-coumarate 3-hydroxylase, C3H) whose down-regulation has been experimentally shown to
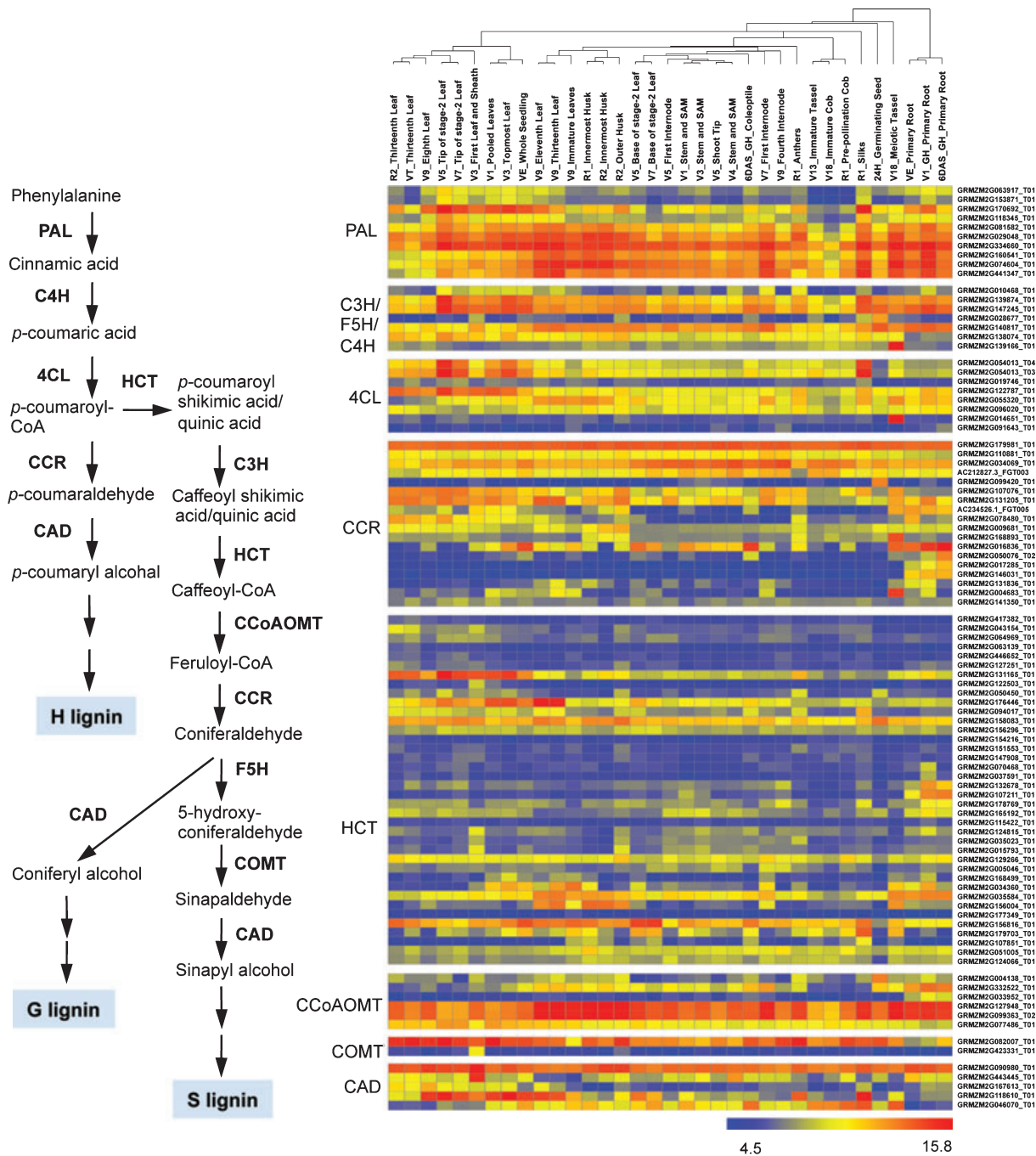
Figure 4. Heat map showing the clustering of vegetative tissues based on expression of lignin pathway genes.
The lignin pathway is based on that published previously (Vanholme *et al.*, 2008). Previously reported lignin pathway genes in the B73 inbred line of maize (Penning *et al.*, 2009) were used for this analysis. The heat map was generated by hierarchical clustering using Pearson's correlation as a measure of similarity. Red, yellow and blue indicate high, medium and low levels of gene expression, respectively.

improve the efficiency of enzymatic saccharification (Chen and Dixon, 2007; Chen *et al.*, 2006). Distinct expression patterns of paralogs suggest that expansion of these gene families in maize is accompanied by diversification in transcriptional regulation and probably sub-functionalization. In summary, these observations demonstrate the complex regulation of plant metabolic pathways, which pose challenges for genetic modifications aimed at improving the economic value of plant products. It is thus not surprising that most of the naturally occurring and artificially induced mutations in lignin pathway genes are associated with reduced plant fitness (Chen and Dixon, 2007; Li *et al.*,

2008; Sattler *et al.*, 2010). Detailed knowledge of tissue- and paralog-specific expression patterns can provide a foundation for targeting specific paralogs to improve maize and related grasses.

CONCLUSIONS

We have generated an extensive expression atlas covering a wide array of tissues and developmental stages of maize using a NimbleGen microarray encompassing 80 301 probe sets, and are providing the data as a community resource. We have demonstrated that the dataset is of high quality and that the results of biological replicates are highly repeatable, based on an analysis of 30 892 probe sets that represented high-confidence maize genes from the filtered maize gene set (Schnable *et al.*, 2009). The quality of the complete set of probes compares very well with this subset (data not shown). In response to improvements in the B73 genome sequence, re-mapping the probe sets will provide transcription profiles of additional genes. The complete dataset is available to the research community through PLEXdb (Wise *et al.*, 2007) under accession number ZM29, through GEO under accession number GSE27004, and through MaizeGDB (Lawrence *et al.*, 2004). The array design used in this study (ID: 090319_Zea_KR_ExpTil) is available from NimbleGen (http://www.nimblegen.com), and information about the probe set can be downloaded from PLEXdb (ZM29) and GEO (GPL12620). This comprehensive maize transcriptome is an excellent resource for functional genomics and gene discovery in maize.

EXPERIMENTAL PROCEDURES

Gene model construction

A multi-step procedure was used to identify maize gene models for synthesis of oligonucleotides for the expression microarray. Figure S1 shows the number of sequences that were used at each stage of the process. Gene models were created by running FGENESH (Salamov and Solovyev, 2000) using a monocot codon usage matrix and repeat masked BAC sequences downloaded from MaizeSequence.org (release 1a.49). Because there is overlap between individual BACs in the tiling path in release 1a.49, an effort was made to remove redundant gene model sequences. The FGENESH gene models were aligned to each other using BLAST (Altschul *et al.*, 1990), and the shorter model was discarded for pairs of models with sequence identity >95% and at least a 300 bp overlap. All FGENESH models <300 bp long were also removed. Release 1a.49 did not provide full coverage of the maize genome, and we supplemented the FGENESH gene predictions with maize PUT transcript assemblies from PlantGDB (release 163a, Duvick *et al.*, 2008). PUT sequences were aligned to FGENESH gene models using BLAST, and the longer of the two sequences was retained for any PUT/FGENESH pair with >95% identity and an overlap >200 bp. PUT assemblies that were longer than 500 bp but did not match any FGENESH gene models were aligned to each other by BLAST to identify highly similar sequences and the shorter sequence was removed. For PUT assemblies between 500 and 700 bp, only those with an ESTScan-predicted ORF (Iseli *et al.*, 1999) were retained. After these processing steps, a total of 67 655 FGENESH/PUT assembly sequences

remained, which included 20 019 FGENESH models that did not have PUT assembly support, 23 061 sequences from the FGENESH/PUT pairs, and 24 575 non-redundant PUTs that did not match any FGENESH models. These 67 655 sequences were used by NimbleGen to design 330 788 60-mer probes to create probe sets for a NimbleGen expression microarray. Details of the probe designing process are available at http://www.nimblegen.com/products/lit/probe_design_2008_06_04.pdf.

Mapping of probes

A total of 330 788 probes were designed for the in-house maize gene models. As the probes were designed based on gene models predicted from BAC sequences of an earlier release and PUT assemblies, they were mapped to the B73 maize genome sequence (AGP_v1) and the cDNA sequences of maize gene models (filtered gene set, release 4a.53) once the pseudomolecules and official genome annotation became available. The cDNA sequences of *Z. mays* gene models (filtered gene set) were downloaded from the Maize Genome Sequencing Project (http://ftp.maizesequence.org/current/filtered-set/ZmB73_4a.53_filtered_cdna.fasta.gz). The genome sequence of *Z. mays* was downloaded from the Maize Genome Sequence Project (http://ftp.maizesequence.org/current/assembly/ZmB73_AGPv1_genome.fasta.gz). Probe sequences were searched against the maize genome sequence and the cDNA sequences of the official maize gene models using the VMATCH program (http://www.vmatch.de/), allowing up to one difference between the probe and the reference sequence. Perl scripts were written to transfer the mapping coordinates of the cDNA sequences back to the genome (either within an exon or across a splice junction) to determine if a probe mapped to various alternative isoforms of one gene. Probes that mapped ≤2 or >2 times across the genome were defined as unique probes and repetitive probes, respectively. This analysis resulted in 218 980 unique probes, 10 866 repetitive probes, and 100 942 unmapped probes (Figure S1).

Plant materials, growing conditions and RNA extraction

Maize inbred B73 was used for constructing the gene atlas. A concise description of the tissues collected to create the gene atlas is presented in Table 1. A more detailed description of sampling is provided in Table S1, and images of selected tissues are shown in Figure S2. Plants were grown in Plano silt loam soil at the West Madison Agricultural Research Station (Verona, WI) during summer 2008. During field preparation, 200 kg per acre of urea (46-0-0) was applied. One day after planting, herbicides including Callisto (142 g per acre; Syngenta, http://www.syngenta-us.com/), Dual II (710 ml per acre; Syngenta) and Simazine (227 g per acre; Agrisolutions, http://www.agrisolutionsinfo.com) were applied. For collection of seed tissues, shoots were covered before silk emergence avoiding any injury to the plants. Plants were self-pollinated on the same day to establish a common initiation point for the harvest timeline. Greenhouse-grown plants were propagated by growing five plants per pot (30 cm top diameter, 28 cm height, 14.5 L volume) containing Metro-Mix 300 (Sun Gro Horticulture, http://www.sungro.com/) with no additional fertilization. The growing conditions were 27°C day and 24°C night temperatures with 16 h light (5:00 am to 9:00 pm) and 8 h dark. Germination was initiated by soaking seeds in distilled water in a Petri dish for 12 h and then placing seeds between layers of moist paper towels for another 12 h to allow germination. The field samples were collected between 8:00 am and 10:00 am, approximately 3 h after sunrise. The greenhouse samples were collected between 8:00 am and 9:00 am, 3 h after turning on the lights.

Three biological replicates were collected for each tissue type. For all the tissues except germinating seeds, a biological replicate was

constituted by collecting and pooling samples from three competitive randomly chosen plants. For germinating seed, ten randomly chosen seeds were pooled to form a biological replicate. The harvested tissues were immediately frozen in liquid nitrogen and stored at −80°C. Total RNA was extracted using TRIZOL reagent (Invitrogen, http://www.invitrogen.com) according to the manufacturer's protocol, and purified using an RNeasy MinElute Cleanup kit (Qiagen, http://www.qiagen.com) according to the manufacturer's instructions.

Microarray hybridization, data extraction and normalization

Isolation of mRNA, cDNA synthesis, probe labeling and hybridization was performed by Roche NimbleGen Inc. (http://www.nimblegen.com/) using the standard protocol for eukaryotic RNA samples and 385K microarrays (http://www.nimblegen.com/products/lit/expression_userguide_v5p0.pdf). Briefly, 10 μg total RNA was used to synthesize cDNA using a SuperScript double-stranded cDNA synthesis kit (Invitrogen) with oligo(dT) primers for amplification. The cDNA samples were labeled with Cy3 using a NimbleGen One-Color DNA labeling kit, and hybridized to slides using a NimbleGen hybridization system (Roche NimbleGen). Scanning and normalization of the data was performed by Roche NimbleGen. The slides were scanned using GenePix 4000B and the microarrays were imported into NimbleScan software. The data were normalized using a robust multi-chip average (RMA) algorithm (Irizarry *et al.*, 2003).

Present calls for expression and identification of constitutively and stably expressed genes

A gene with a RMA-normalized linear expression value of ≥200 in at least one of the 60 tissues was considered to be expressed. The expression cut-off, an arbitrarily chosen conservative value, was five times the mean normalized signal from 165 randomly generated sequences spotted on each slide, which ranged between 27–65 with a mean of 40 across all 180 slides used for the experiment. Genes with a linear expression value of 200 or more in all 60 tissues were considered to be constitutively expressed. Among the constitutively expressed genes, those with a coefficient of variation ($CV = S/X_{mean}$, where S represent the standard deviation and X_{mean} indicates the mean expression of a gene across all the tissues) ≤15% were considered to be stably expressed.

Principal component analysis

To evaluate replicate quality and to study the biological relatedness of tissues, we reduced the data to three dimensions by principal component analysis (PCA) using the Spotfire DecisionSite for Functional Genomics (DSFG) package (http://spotfire.tibco.com/). RMA-normalized $\log_2$-transformed expression values were used for the analysis. This involved performing k-means clustering in order to group genes into 1000 clusters followed by PCA. PCA was performed separately for the 35 vegetative and 25 seed tissues for ease of data presentation.

Hierarchical clustering

Hierarchical clustering was performed using the unweighted pair-group method with arithmetic mean (UPGMA) approach and Pearson's correlation as a similarity measure in the DSFG package. Clustering of vegetative and seed tissues was performed separately.

Organ-specific genes

For identification of organ-specific genes, eight selected organs (leaves, internodes, roots, cob, silk, tassel, embryo, endosperm and pericarp) were considered. Compound organs (those comprising multiple organs, e.g. whole seeds, stem and shoot apical meristem, etc.) were excluded. For tassel, only the immature and meiotic tassels were included. For identification of tissue-specific genes, we used an expression cut-off method: genes with RMA-normalized linear expression values >500 in at least one of the tissues belonging to an organ, and no expression (linear expression value < 200) in all other tissues were considered to be organ-specific.

Z scores

For the organs included in identification of organ-specific genes, we calculated *Z* scores using two steps. First, overall expression for each of the selected organs was calculated by averaging the RMA-normalized $\log_2$-transformed expression values for all the tissues representing that organ. Second, *Z* scores were calculated using the formula: $Z = (X − X_{mean})/S$, where X is the overall expression of a gene in an organ, and X_{mean} and S are the mean expression and standard deviation of that gene across all the selected organs, respectively.

GO Slim enrichment analysis

GO Slim assignments (http://www.geneontology.org/GO.slims.shtml) for the proteins of the filtered gene set were downloaded from MaizeSequence.org (http://ftp.maizesequence.org/current/functional_annotations/ZmB73_4a.53_protein_goslim_plant.txt). A Fisher's exact test with the false discovery rate (FDR) set at 5% as defined by the *Q*-value program (Storey, 2002) was used to identify enriched GO Slim annotations for each subset of the maize genes: non-expressed genes, constitutively expressed genes and stably expressed genes. Only representative gene models were used in this analysis.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Flow chart describing construction of the probe sets used in this study.
Figure S2. Representative tissues used in this study.
Figure S3. Quality of the three biological replicates for each of the 60 tissues used in the study.
Figure S4. Expression of previously studied genes follows the expected expression patterns.
Figure S5. GO Slim enrichment analysis of the non-expressed genes.
Figure S6. GO Slim enrichment analysis of constitutively expressed genes.
Figure S7. GO Slim enrichment analysis of stably expressed genes.
Figure S8. Identification of stably expressed genes.
Figure S9. Up-regulation of selected genes in the base and tip of developing leaves.

Figure S10. Heat map of global gene expression.

Table S1. Maize tissues included in the genome-wide gene expression atlas of inbred line B73.

Table S2. Pearson's correlation coefficients for biological replicates.

Table S3. Most stably expressed genes.

Table S4. Tissue-specific genes.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Benedito, V.A., Torres-Jerez, I., Murray, J.D. *et al.* (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* **55**, 504–513.

Bennetzen, J.L. and Hake, S. (2009) *Handbook of Maize: Genetics and Genomics*. New York: Springer.

Carpita, N.C. and McCann, M.C. (2008) Maize and sorghum: genetic resources for bioenergy grasses. *Trends Plant Sci.* **13**, 415–420.

Chen, F. and Dixon, R.A. (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* **25**, 759–761.

Chen, F., Srinivasa Reddy, M.S., Temple, S., Jackson, L., Shadle, G. and Dixon, R.A. (2006) Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (*Medicago sativa* L.). *Plant J.* **48**, 113–124.

Dixon, R.A., Dey, P.M. and Lamb, C.J. (1983) Phytoalexins: enzymology and molecular biology. *Adv. Enzymol. Relat. Areas Mol. Biol.* **55**, 1–136.

Dixon, R.A., Achnine, L., Kota, P., Liu, C.-J., Srinivasa Reddy, M.S. and Wang, L. (2002) The phenylpropanoid pathway and plant defence – a genomics perspective. *Mol. Plant Pathol.* **3**, 371–390.

Druka, A., Muehlbauer, G., Druka, I. *et al.* (2006) An atlas of gene expression from seed to seed through barley development. *Funct. Integr. Genomics*, **6**, 202–211.

Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M.D., Lawrence, C.J., Lushbough, C. and Brendel, V. (2008) PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.* **36**, D959–D965.

Hansey, C.H., Lorenz, A.J. and de Leon, N. (2010) Cell wall composition and ruminant digestibility of various maize tissues across development. *Bioenergy Res.* **3**, 28–37.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

Iseli, C., Jongeneel, C.V. and Bucher, P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In *Proceedings of the Intelligent Systems for Molecular Biology*. Menlo Park, CA: AAAI Press, pp. 138–148.

Jiao, Y., Tausta, S.L., Gandotra, N. *et al.* (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.* **41**, 258–263.

Keating, J.D., Panganiban, C. and Mansfield, S.D. (2006) Tolerance and adaptation of ethanologenic yeasts to lignocellulosic inhibitory compounds. *Biotechnol. Bioeng.* **93**, 1196–1206.

Lawrence, C.J. and Walbot, V. (2007) Translational genomics for bioenergy production from fuelstock grasses: maize as the model species. *Plant Cell*, **19**, 2091–2094.

Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E. and Brendel, V. (2004) MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* **32**, D393–D397.

Li, X., Weng, J.K. and Chapple, C. (2008) Improvement of biomass through lignin modification. *Plant J.* **54**, 569–581.

Li, P., Ponnala, L., Gandotra, N. *et al.* (2010) The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **42**, 1060–1067.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G. and Stacey, G. (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* **63**, 86–99.

Lopez, F., Bousser, A., Sissoeff, I., Hoarau, J. and Mahé, A. (2004) Characterization in maize of ZmTIP2-3, a root-specific tonoplast intrinsic protein exhibiting aquaporin activity. *J. Exp. Bot.* **55**, 539–541.

Lotan, T., Ohto, M., Yee, K.M., West, M.A., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*, **93**, 1195–1205.

Moore, K.J. and Jung, H.J.G. (2001) Lignin and fiber digestion. *J. Range Manage.* **54**, 420–430.

Penning, B.W., Hunter, C.T. III, Tayengwa, R. *et al.* (2009) Genetic resources for maize cell wall biology. *Plant Physiol.* **151**, 1703–1728.

Perlack, R.D., Wright, L.L., Turhollow, A.F., Graham, R.L., Stokes, B.J. and Erbach, D.C. (2005) *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*. Oak Ridge, TN: Oak Ridge National Laboratory.

Sabelli, P.A. and Larkins, B.A. (2009) The development of endosperm in grasses. *Plant Physiol.* **149**, 14–26.

Salamov, A.A. and Solovyev, V.V. (2000) Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522.

Sattler, S.E., Funnell-Harris, D.L. and Pedersen, J.F. (2010) Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. *Plant Sci.* **178**, 229–238.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. and Lohmann, J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**, 501–506.

Schmidt, R.J., Burr, F.A., Aukerman, M.J. and Burr, B. (1990) Maize regulatory gene opaque-2 encodes a protein with a 'leucine-zipper' motif that binds to zein DNA. *Proc. Natl Acad. Sci. USA*, **87**, 46–50.

Schnable, P.S., Ware, D., Fulton, R.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Schultz, J.A. and Juvik, J.A. (2004) Current models for starch synthesis and the *sugary enhancer1* (*se1*) mutation in *Zea mays*. *Plant Physiol. Biochem.* **42**, 457–464.

Storey, J.D. (2002) A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498.

Sylvester, A.W., Cande, W.Z. and Freeling, M. (1990) Division and differentiation during normal and *liguleless-1* maize leaf development. *Development*, **110**, 985–1000.

Vanholme, R., Morreel, K., Ralph, J. and Boerjan, W. (2008) Lignin engineering. *Curr. Opin. Plant Biol.* **11**, 278–285.

Wang, L., Xie, W., Chen, Y. *et al.* (2010) A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J.* **61**, 752–766.

Wise, R.P., Caldo, R.A., Hong, L., Shen, L., Cannon, E. and Dickerson, J.A. (2007) BarleyBase/PLEXdb: a unified expression profiling database for plants and plant pathogens. *Methods Mol. Biol.* **406**, 347–363.