

# Utility of RNA Sequencing for Analysis of Maize Reproductive Transcriptomes

Rebecca M. Davidson, Candice N. Hansey, Malali Gowda, Kevin L. Childs, Haining Lin, Brienne Vaillancourt, Rajandeep S. Sekhon, Natalia de Leon, Shawn M. Kaeppler, Ning Jiang, and C. Robin Buell\*

## Abstract

Transcriptome sequencing is a powerful method for studying global expression patterns in large, complex genomes. Evaluation of sequence-based expression profiles during reproductive development would provide functional annotation to genes underlying agronomic traits. We generated transcriptome profiles for 12 diverse maize (*Zea mays* L.) reproductive tissues representing male, female, developing seed, and leaf tissues using high throughput transcriptome sequencing. Overall, ~80% of annotated genes were expressed. Comparative analysis between sequence and hybridization-based methods demonstrated the utility of ribonucleic acid sequencing (RNA-seq) for expression determination and differentiation of paralogous genes (~85% of maize genes). Analysis of 4975 gene families across reproductive tissues revealed expression divergence is proportional to family size. In all pairwise comparisons between tissues, 7 (pre- vs. postemergence cobs) to 48% (pollen vs. ovule) of genes were differentially expressed. Genes with expression restricted to a single tissue within this study were identified with the highest numbers observed in leaves, endosperm, and pollen. Coexpression network analysis identified 17 gene modules with complex and shared expression patterns containing many previously described maize genes. The data and analyses in this study provide valuable tools through improved gene annotation, gene family characterization, and a core set of candidate genes to further characterize maize reproductive development and improve grain yield potential.

**M**AIZE (*Zea mays* L.) is an important crop worldwide in terms of both grain and stover production with approximately 36 million ha grown in 2010 within the United States (USDA and NASS, 2011). As demands for corn yield are expected to increase for food, feed, and biofuel production (Godfray et al., 2010), a combination of genetic and genomic resources will be needed to elucidate the underlying genes and deploy preferred alleles into improved varieties. Several large effect mutations affecting reproductive development have been identified in maize including those involved with sexual differentiation, floral branching, and seed maturation (Acosta et al., 2009; Bortiri and Hake, 2007; Gallavotti et al., 2010; Suzuki et al., 2003). In spite of this, little is known about global transcriptional networks across unique floral and seed organs that potentially contribute to complex, multilocus traits. Genomic strategies for maize reproductive studies

R.M. Davidson, C.N. Hansey, M. Gowda, K.L. Childs, H. Lin, B. Vaillancourt, and C.R. Buell, Dep. of Plant Biology, Michigan State Univ., 166 Plant Biology Building, East Lansing, MI 48824; C.N. Hansey, K.L. Childs, H. Lin, B. Vaillancourt, and C.R. Buell, Dep. of Energy Great Lakes Bioenergy Research Center, Michigan State Univ., East Lansing, MI 48824; R.S. Sekhon, N. de Leon, and S.M. Kaeppler, Dep. of Agronomy, Univ. of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706; R.S. Sekhon, N. de Leon, and S.M. Kaeppler, Dep. of Energy Great Lakes Bioenergy Research Center, Univ. of Wisconsin-Madison, Madison, WI 53706; N. Jiang, Dep. of Horticulture, Michigan State Univ., A330 Plant and Soil Sciences Building, East Lansing, MI 48824; M. Gowda, current address: Next-Generation Genomics Lab., Center for Cellular and Molecular Platform, NCBS-GKVK Campus, Bangalore 560065, India. R.M. Davidson and C.N. Hansey contributed equally to this work. Received 19 May 2011. \*Corresponding author (buell@msu.edu).

Published in The Plant Genome 4:191–203. Published 19 Aug. 2011.  
doi: 10.3835/plantgenome2011.05.0015  
© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

**Abbreviations:** cDNA, complementary DNA; DAP, days after pollination; FPKM, fragments per kilobase pair of exon model per million fragments mapped; MeV, Multiple Experiment Viewer Software; mRNA, messenger ribonucleic acid; PCC, Pearson correlation coefficient; RNA-Seq, ribonucleic acid sequencing; SCC, Spearman correlation coefficient; v1, version 1; v2, version 2; WGCNA, weighted gene coexpression network analysis method.

are now feasible with the recent release of the improved version 2 (v2) maize reference genome and gene annotation (Schnable et al., 2009). However, there are significant challenges in analyzing the maize gene set, which is rich in duplicated genes. A deeper understanding of how to evaluate global gene expression patterns in maize reproductive tissues would enhance the utilization of the maize genome and complement our current understanding of the genetic and molecular basis of maize reproduction.

Previous large-scale expression profiling studies in maize have been performed with multiple oligonucleotide microarray platforms (Barr et al., 2010; Gardiner et al., 2005; Hayes et al., 2010; Liu et al., 2008; Sawers et al., 2007; Sekhon et al., 2011; Skibbe et al., 2009; Wang et al., 2010; Zheng et al., 2010; Zhu et al., 2009). These arrays are limited to measuring known sequences and thus cannot fully sample the transcriptome. Recent advances in sequencing technology have led to improved methods for expression profiling through ultrahigh throughput sequencing of messenger ribonucleic acid (mRNA). The RNA-seq method of whole transcriptome profiling involves converting mRNA into complementary DNA (cDNA), which then is sequenced in a massive and parallel manner involving short reads of 36 to 150 bp (Wang et al., 2009b). Short read mapping algorithms (Langmead et al., 2009; Trapnell et al., 2009) and the ability to convert millions of mapped reads to normalized expression values relative to annotated or unannotated gene models (Trapnell et al., 2010) allow for robust expression estimates in organisms with sequenced genomes. Thus, not only does RNA-seq data provide detailed expression phenotypes for genes but it also is informative for improving the quality of gene model annotation.

Though a few RNA-seq studies in maize have been completed (Eveland et al., 2010; Li et al., 2010; Wang et al., 2009a), none have thoroughly examined diverse reproductive tissues. To explore the maize reproductive transcriptome, we generated 10.5 Gbp of whole transcriptome sequence from 12 reproductive tissues from the reference B73 maize inbred line that represent male, female, and developing seed tissue types as well as leaves to provide a contrasting tissue. The transcriptomes of these diverse tissues represent 31,239 genes (79.2%) from the maize v2 annotation (Maize Genome Sequencing Project, 2011; Schnable et al., 2009). We have analyzed these data in the context of gene family expression, differential expression, tissue restricted expression, gene coexpression networks, and network relationships to previously characterized genes and genes of unknown function. This dataset and subsequent analyses provide an invaluable source of information for the community including gene family classification, tissue specific transcriptional networks, and candidate genes for targeted work in improving maize grain yield.

## Materials And Methods

### Plant Materials and RNA Isolation

Plants of the reference maize genotype, B73, were grown under field conditions at an average density of 37,000

plants ha<sup>-1</sup> in East Lansing, MI, during the summer of 2009. The soil type at this location is Marlette fine sandy loam, 2 to 6% slopes. Before planting, fertilizer 19-19-19 N-P-K was applied at a rate of approximately 454 kg ha<sup>-1</sup>. Five weeks after emergence, one tablet of slow release fertilizer (15-11-8, Osmocote plus, Scotts, Marysville, OH) was applied to each plant. Thirteen tissues described in Table 1 were harvested and immediately frozen in liquid N. An independent sample of leaf tissue grown under growth chamber conditions (26/24°C day/night temperatures and 12 h photoperiod) was harvested to assess the repeatability of the data. Total RNA was isolated using a modified hot phenol method. One gram of tissue was ground with liquid N and immediately transferred into a mixture of prewarmed phenol (pH 4.3) and extraction buffer (100 mmol LiCl, 100 mmol Tris pH 8.5, 10 mmol ethylenediaminetetraacetic acid, 1% sodium dodecyl sulfate, and 15 mmol dithiothreitol), vortexed, and incubated at 60°C for 20 min. Chloroform:isoamyl alcohol (24:1) was added and the mixture was vortexed and centrifuged at 11,000 rpm for 10 min at 4°C. The aqueous phase was transferred into phenol:chloroform:isoamyl alcohol (25:24:1), vortexed, and centrifuged at 11,000 rpm for 10 min at 4°C. The previous step was repeated twice with phenol:chloroform:isoamyl alcohol (25:24:1) and twice with chloroform:isoamyl alcohol (24:1). The RNA pellet was precipitated with 4 M LiCl, washed with 70% ethanol, and dissolved in diethylpyrocarbonate water. Genomic DNA was removed by deoxyribonuclease (DNase) treatment using the DNA-Free kit (Ambion, Austin, TX). RNA quantity was determined using a Nanodrop (Thermo Scientific, Wilmington, DE), and RNA quality was checked using an Agilent bioanalyzer (Agilent Technologies, Santa Clara, CA).

### RNA Sequencing Library Construction and Illumina Sequencing

Approximately 5 to 10 µg of total RNA was used for construction of each RNA-seq library. Polyadenylated RNA purification, RNA fragmentation, cDNA synthesis, and polymerase chain reaction (PCR) amplification was performed according to the Illumina RNA-seq protocol (Cat # RS-100-0801, Illumina, Inc., San Diego, CA). Parallel sequencing was performed using an Illumina Genome Analyzer II (Illumina, Inc., San Diego, CA) at the Research Technology Support Facility at Michigan State University (East Lansing, MI) and the Institute for Genome Sciences at the University of Maryland School of Medicine (Baltimore, MD). Single-end sequence reads between 35 and 40 bp were generated, and RNA-seq read quality was evaluated based on the Illumina purity filter, percent low quality reads, and distribution of phred-like scores at each cycle. All data presented passed the quality control filtering based on these metrics. Sequences are available in the Sequence Read Archive at the National Center for Biotechnology Information (Sequence Read Archive study number SRP006463; NCBI, 2011).

## RNA Sequencing Data Analysis

Sequence reads for each tissue sample were mapped to v2 of the B73 reference genome (AGPv2; Maize Genome Sequence Project, 2011; Schnable et al., 2009), hereafter referred to as v2 pseudomolecules, using the quality and splice site aware alignment algorithms, Bowtie version 0.12.7 (Langmead et al., 2009) and TopHat version 1.2.0 (Trapnell et al., 2009) in conjunction with SAMtools version 0.1.7 (Li et al., 2009). The minimum and maximum intron length was set to 5 and 60,000 bp, respectively; all other parameters were set to the default values.

Normalized gene expression levels were calculated using Cufflinks version 0.9.3 (Trapnell et al., 2010) and are reported as fragments per kilobase pair of exon model per million fragments mapped (FPKM). The maximum intron length parameter was set to 60,000 bp. A reference annotation was supplied based on the AGPv2 5b filtered gene set (Maize Genome Sequence Project, 2011), hereafter referred to as v2 annotation, and the v2 pseudomolecules were provided to permit the use of the bias detection and correction algorithm. The quartile normalization option was used to improve differential expression calculations of lowly expressed genes. All other parameters were used at the default settings. Raw FPKM values for all tissues are provided (Supplemental Table S1). Cufflinks provides the upper and lower bound FPKM values for the 95% confidence interval of the abundance of each gene. For summary statistics of the number of genes expressed in each tissue and genes with expression restricted to a single tissue within this study, a gene was considered expressed in a given tissue if the FPKM 95% confidence interval lower boundary was greater than zero. If a gene had a lower boundary value greater than zero in only one tissue, that gene was classified as having expression restricted to that tissue within this study. To evaluate the effect of sampling depth on expression estimates, 5, 10, and 15 million reads were randomly selected from the total pool of reads for all tissues, and for tissues with sufficient reads, 20 and 25 million reads were also selected.

Pearson product-moment correlation analyses of log<sub>2</sub> FPKM values among RNA-seq libraries were performed using PROC CORR with SAS version 9.2 (SAS Institute, 2003), and all log<sub>2</sub> FPKM values less than zero were set to zero. Only tests significant at  $p = 0.05$  are shown. The heat map of correlation values was clustered with hierarchical clustering using a Pearson correlation distance metric and average linkage, and bootstrap support values were calculated from 1000 replicate searches using Multiple Experiment Viewer Software (MeV) version 4.5 (Saeed et al., 2003). Differential expression analysis was conducted using the cuffdiff program within Cufflinks version 0.9.3 (Trapnell et al., 2010) utilizing the mapping results described above. For pairwise gene comparisons between two tissues, cuffdiff first calculates the variance of the transcript abundance in each tissue accounting both for the variation across replicates and

**Table 1. Descriptions of maize tissues used in this study.**

Tissue name <sup>†</sup>	Tissue description	Time of harvest
<b>Leaves</b>		
Leaves (V2)	Pooled leaves	20 d after sowing
<b>Female</b>		
Preemergence cob (V15)	Immature cob	10 d before cob emergence
Postemergence cob (VT)	Whole cob	Silk emergence
Silk (R1)	Mature silk	Anthesis
Ovule (R1)	Mature ovule with stigma and style removed	Anthesis
<b>Male</b>		
Preemergence tassel (V18)	Post meiotic whole tassel including anthers	10 d before tassel emergence
Postemergence tassel (VT)	Post meiotic whole tassel including anthers	Tassel emergence from whorl
Whole Anthers (VT)	Whole anther including pollen dissected from tassel structural tissue	Anthesis
Pollen (VT)	Mature pollen	Anthesis
<b>Seed</b>		
Seed 5 DAP <sup>‡</sup> (R1)	Whole seed	5 DAP
Seed 10 DAP (R2)	Whole seed	10 DAP
Embryo 25 DAP (R4)	Developing embryo	25 DAP
Endosperm 25 DAP (R4)	Developing endosperm	25 DAP

<sup>†</sup>Vegetative (V) and reproductive (R) growth development stages defined in the Iowa State University Extension Corn Field Guide (Abendroth et al., 2011).

<sup>‡</sup>DAP, days after pollination.

the uncertainty in the expression estimate based on mapping uncertainty. When replicates are not available, a Poisson distribution is assumed, and the variation of the count across replicates is equal to the mean of the count. The variance for a gene's abundance is then determined using the variance in abundance for each isoform of that gene. The test statistic  $T$  is calculated as the log of the ratio of FPKM values for the two tissues divided by the variance of the log of the ratio. A two-tailed Student's  $t$  test is then used to determine if genes are differentially expressed. For additional information on how the differential expression test is conducted see Trapnell et al., 2010. The minimum number of alignments at a gene required to test was set to 100. Quartile normalization and a false discovery rate of 0.01 after Benjamini-Hochberg correction for multiple testing were used. The v2 pseudomolecules and v2 annotation were provided as input parameters. All other parameters were used at the default levels.

Wiggle tracks were generated to visualize read coverage across the genome using the wiggles program within TopHat version 1.2.0 (Trapnell et al., 2009). The wiggle tracks were displayed on the Integrated Genome Browser version 6.4.1 (Nicol et al., 2009) using the GFF file for the v2 annotation from maizesequence.org (Maize Genome Sequence Project, 2011).

## RNA Sequencing and Microarray Comparative Analyses

For correlation analyses of RNA-seq to microarray, gene expression data from an extensive maize developmental tissue atlas (Sekhon et al., 2011) was used. Because this microarray data set was analyzed against version 1 (v1) of the reference genome (Schnable et al., 2009), FPKM values for the 13 tissues in this dataset were determined using the v1 pseudomolecules (Maize Genome Sequence Project, 2011) and the 4a.53 filtered gene set (Maize Genome Sequence Project, 2011), hereafter referred to as v1 pseudomolecules and v1 annotation, with the same parameters as described above. For the RNA-seq to microarray comparative analysis only, genes with more than one predicted isoform were removed from both RNA-seq and microarray datasets. The remaining genes present in both datasets were used for downstream analyses. The FPKM values and probe intensities were  $\log_2$  transformed, and Spearman rank correlations were performed using PROC CORR (SAS Institute, 2003). Only tests significant at  $p = 0.05$  are shown. Spearman rank correlations, rather than Pearson correlations, were used due to the discrepancy in scale between the platforms.

### Identification and Analysis of Paralogous Gene Families

Paralogous gene families were identified by OrthoMCL (Chen et al., 2007; Li et al., 2003) with default parameters. To eliminate the situation where alternative isoforms of one gene are grouped in different clusters, only peptide sequences of the representative gene models of the v2 annotation were used. A representative gene model was defined as the gene model that produces the longest peptide sequence among all alternative isoforms of a gene. OrthoMCL gene family assignments for all maize genes are provided (Supplemental Table S1). Pearson product-moment correlations were performed for  $\log_2$  FPKM values of all gene pair combinations within the same gene families using a custom java script, and a random distribution of Pearson correlation coefficients (PCCs) was generated from correlations of 100,000 gene pairs not in the same gene family. The heat map of FPKM values was generated with MeV version 4.5 (Saeed et al., 2003).

### Gene Coexpression Network Analysis

Gene expression values from the 13 tissues in this dataset were used to generate gene networks and to identify modules of highly correlated genes using the weighted gene coexpression network analysis method (WGCNA) (Zhang and Horvath, 2005). All gene FPKM expression values were  $\log_2$  transformed, and any  $\log_2$  FPKM values less than 1 were set to 0. Genes without variation across tissues were filtered out using a coefficient of variation ( $CV = \sigma/\mu$ ) cutoff; genes with a CV less than 1.0 were discarded from further network analysis. An R version 2.10.1 (R Development Core Team, 2011) implementation of the WGCNA was then used to identify gene modules (Langfelder and Horvath, 2008).

Signed networks were constructed using Pearson correlation values for all filtered gene expression values. The  $\beta$  and treecut parameters that are used within WGCNA were chosen as 13 and 0.7, respectively. Eigengenes were calculated using the function module Eigengenes within the WGCNA package (Langfelder and Horvath, 2007). Eigengenes are the first right-singular vector of the singular value decomposition of the module expression data. The heat map of eigengenes for each gene module was constructed using MeV version 4.5 (Saeed et al., 2003). Previously characterized maize genes that have been associated with the maize v2 annotation (Maize Genome Sequence Project, 2011; Schnable and Freeling 2011) were crossreferenced with the genes from each coexpression module.

## Results And Discussion

### Maize Tissue and RNA Sequencing Data Descriptions

The tissues selected for RNA-seq profiling represent different organ types as well as samples of some tissues at multiple developmental time points. The tissues used in this experiment represent four larger groups: male tissues (pre- and postemergence tassels, whole anthers, and pollen), female tissues (pre- and postemergence cobs, mature silks, and ovules), seed tissues (whole seed 5 and 10 d after pollination [DAP], embryo, and endosperm 25 DAP), and a vegetative tissue (leaves) (Table 1). Leaves were included to provide a basis for comparison of reproductive and nonreproductive tissue. This dataset provides an in-depth survey of the maize transcriptome in reproductive tissues.

The potential for biases in comparisons between the libraries was determined based on the total number of purity-filtered reads, the percentage of reads mapped to the v2 pseudomolecules, and the relationship between read depth and tissue complexity as measured by the number of expressed genes. Purity filtered reads ranged from 17.1 to 29.9 million reads per sample and the percentage of reads mapped per library ranged from 84.6 to 90.6% (Table 2) indicating a similar performance of library construction and sequencing across the samples. In addition, no relationship was observed between number of genes detected as expressed and number of reads sampled across a range of tissues (Supplemental Fig. S1). Randomly selected subsets of the total pool of reads for each tissue ranging from 5 to 25 million reads were used to further evaluate the effect of sampling depth on gene expression detection (Supplemental Fig. S2). The simulation demonstrates a clear positive relationship between sampling depth and numbers of expressed genes at lower sequencing depths (5 to 15 million reads). The number of expressed genes, however, begins to plateau at approximately 15 million reads, corresponding to the minimum sampling depth of all libraries in this study. To assess the repeatability of this data in an independent experiment, RNA from an additional leaf sample at approximately the same developmental time point as the field grown leaf

**Table 2. Read mapping and expression summary for 13 maize tissues. Fragments per kilobase pair of exon model per million fragments mapped (FPKM) values were calculated using Cufflinks version 0.9.3 (Trapnell et al., 2010), the version 2 (v2) pseudomolecules, and the v2 annotation.**

Tissue	Number of purity filtered reads	Number of mapped reads	Maximum FPKM	Number of expressed genes <sup>†</sup>	Number of tissue restricted genes <sup>‡</sup>
Leaves	17.1 million	14.5 million	128,735	21,956	492
Preemergence cob	18.3 million	16.0 million	17,256	23,338	70
Postemergence cob	18.3 million	15.8 million	17,784	23,948	44
Silk	18.9 million	16.8 million	257,614	23,015	138
Ovule	29.9 million	26.3 million	33,621	25,003	108
Preemergence tassel	19.8 million	17.0 million	24,576	25,165	88
Postemergence tassel	18.6 million	16.1 million	35,766	24,984	42
Whole anthers	27.0 million	24.0 million	127,465	22,178	96
Pollen	27.9 million	25.3 million	1,004,070	13,418	206
Seed 5 DAP <sup>§</sup>	19.2 million	16.5 million	22,966	24,390	37
Seed 10 DAP	26.4 million	22.9 million	26,927	24,486	105
Embryo 25 DAP	19.9 million	17.0 million	36,127	22,493	188
Endosperm 25 DAP	23.6 million	20.3 million	283,698	22,887	251

<sup>†</sup>Genes with a 95% confidence interval lower boundary FPKM value greater than zero were considered expressed.

<sup>‡</sup>Genes were considered tissue restricted when only one tissue within this study had a 95% confidence interval lower boundary FPKM value greater than zero.

<sup>§</sup>DAP, days after pollination.

sample was also sequenced. The correlation between the two leaf samples was  $R^2 = 0.90$  providing evidence for the repeatability of this dataset.

To further assess the quality of our datasets, we evaluated the expression of eight known genes with characterized expression patterns. As expected based on previous reports, *opaque endosperm2* and *shrunkened-2* were expressed nearly exclusively in the endosperm 25 DAP tissue (Bhave et al., 1990; Schmidt et al., 1990), *leafy cotyledon1* was expressed predominantly in the embryo 25 DAP tissue (Lotan et al., 1998), *NADP malic enzyme3* was highly expressed in the leaves (Ku et al., 1996), *branched silkless1*, although at low levels, was expressed in pre- and postemergence cob tissues (Chuck et al., 2002), *male gametophyte specific1* and *tassel serine threonine kinase1* were both highly expressed in all of the male tissues (Fu et al., 2001; Hamilton et al., 1989), and *waxy1* was expressed in endosperm 25 DAP, pollen, ovule, and floral structures containing these tissues (Nelson and Rines, 1962) (Supplemental Table S2). Thus, our tissues, resulting libraries, and expression sets are consistent with previous studies.

### RNA Sequencing Transcriptome Profiles

Summary statistics revealed transcriptional diversity across reproductive tissues. In all tissues, the minimum FPKM values were zero while maximum FPKM values ranged from 17,256 in preemergence cob to 1,004,070 in pollen (Table 2). Intriguingly, phenotypically unique tissues such as mature silk, pollen, and endosperm 25 DAP had the most extreme range of expression levels. Of the top five most highly expressed genes in pollen, two encode a pollen-specific protein C13 (GRMZM2G022347 and GRMZM2G317406), one encodes an arabinoglucan protein 6 (GRMZM2G175243), and the remaining two encode conserved genes of unknown function (GRMZM2G041448

and GRMZM2G073155). Tissues with extreme maximum FPKM values were also extreme for other summary metrics. For example, pollen had the fewest number of genes expressed among all tissues with only 13,418 genes expressed (Table 2) consistent with a previous microarray study that showed only 10,539 transcripts expressed in mature pollen (Ma et al., 2008). Similarly, within reproductive tissues, endosperm 25 DAP had the most genes with expression restricted to one tissue type, that is, genes with expression in only one of the tissues included in this study. In contrast to reproductive tissues, the leaf sample had the most genes overall with expression restricted to a single tissue within this study (Table 2).

These differences in transcriptome profiles were reflected in correlation and cluster analyses among tissues (Fig. 1). Overall, PCC values ranged from 0.33 (leaves vs. pollen) to 0.97 (preemergence cob vs. postemergence cob). As expected, leaf tissue was poorly correlated with most reproductive tissues (PCC < 0.65) and showed the highest correlation with mature silks (PCC = 0.68). Among female tissues, the two stages of cob and ovule were all highly correlated with each other, and mature silk was more similar to developing seed than to female parts. Correlations greater than 0.73 were observed among tassels and anthers, which is not surprising given that anthers were included in the tassel samples. Pollen had the most unique transcriptome with PCC < 0.5 in all comparisons to nonmale tissues. Finally, among seed tissues, the early development stages (whole seed 5 and 10 DAP) showed high correlation to each other (PCC = 0.95) but lower correlations to endosperm 25 DAP (PCC = 0.78 and 0.80) and to developing embryo (PCC = 0.74 and 0.75). Interestingly, the relatively high correlations observed among cob, seed, ovule, and embryo tissues (PCC > 0.80) indicate a substantial subset of shared gene expression patterns in

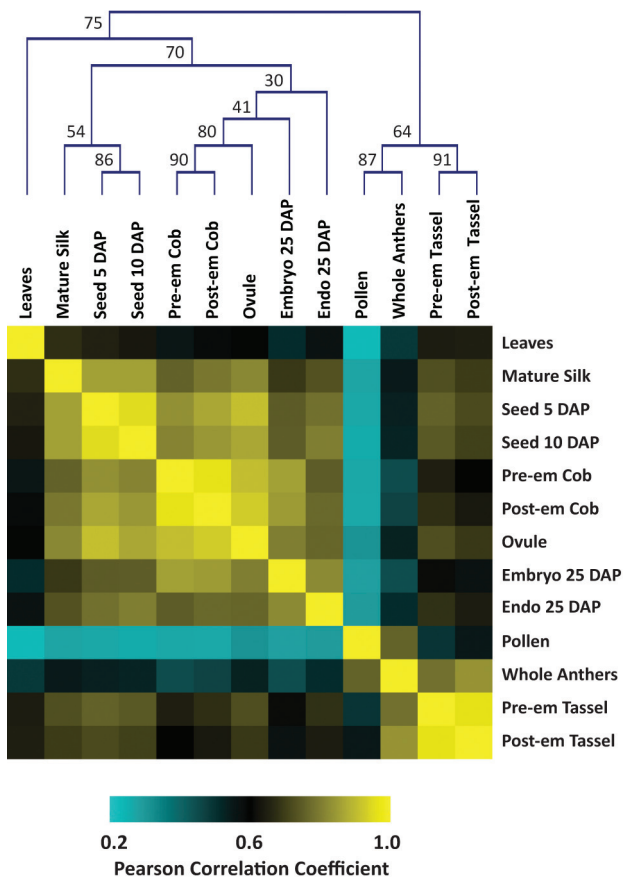


Figure 1. Correlation matrix of 13 maize ribonucleic acid sequencing (RNA-seq) libraries. Normalized transcript abundances for 39,456 genes in the version 2 (v2) annotation were calculated in units of fragments per kilobase pair of exon model per million fragments mapped (FPKM) with Cufflinks version 0.9.3 (Trapnell et al., 2010). Pearson product-moment correlations of  $\log_2$  FPKM values were performed for all versus all tissue samples using SAS (SAS Institute, 2003). The heat map of Pearson correlation coefficients was clustered by hierarchical clustering with a Pearson correlation distance metric and average linkage using Multiple Experiment Viewer Software version 4.5 (Saeed et al., 2003). The bootstrap support values shown on tree nodes denote the percentage of times a given node was supported over 1000 resampling trials. DAP, days after pollination; Pre-em, pre-emergence; Post-em, postemergence.

these tissues despite the cob being exclusively maternally derived and the seed and embryo being both maternally and paternally derived.

### Comparative Analysis between RNA Sequencing and Microarray Data

We compared our RNA-seq data with a similar, microarray-based genome-wide expression dataset from maize (Sekhon et al., 2011). The microarray study assayed gene expression in 60 tissues including 11 that are developmentally similar to our RNA-seq tissues; pollen and post-emergence tassel were the two samples in the RNA-seq dataset without similar tissues in the microarray data. Because the microarray data analysis used v1 annotation, we used FPKM values for the RNA-seq libraries

calculated relative to this gene set as well. Only representative isoforms were used and all genes with more than one predicted isoform were excluded from the analysis. The resulting intersection of genes without predicted isoforms and present in both the microarray and RNA-seq analyses was 17,573 or 54% of the v1 annotation.

A nonparametric correlation statistic was used to compare the RNA-seq and microarray expression datasets. Spearman rank correlation coefficients (SCCs) of  $\log_2$  expression values were calculated between 11 corresponding tissue pairs in the two datasets as well as for three dissimilar tissue pairs as negative controls (Table 3). Among the 11 comparisons between similar tissues (Table 3), the SCC values ranged from 0.68 to 0.82. The highest correlations ( $SCC > 0.8$ ) were observed between the two stages of cob and in tissues where the developmental stages were systematically measured as DAP, including whole seed and embryo. Lower SCC values ( $SCC < 0.8$ ) were observed in leaves, silk, ovule, tassel, anther, and endosperm comparisons where developmental stages and tissue types were slightly different between the studies. Overall, correlation coefficients between analogous tissues (0.68 to 0.82) were greater than comparisons between dissimilar tissues (0.30 to 0.56) indicating that similar expression patterns for this subset of maize genes were detected by the two different platforms.

Scatter plots of  $\log_2$  expression values from various tissue pairs are shown in Fig. 2. These plots illustrate a primary difference between the two expression platforms, which is the broader range of expression detected by RNA-seq compared to microarray (Agarwal et al., 2010) as seen in the low-end expression values between the platforms (Fig. 2). We observed RNA-seq values near zero (FPKM = 1) in the entire range of microarray intensity values similar to previous comparisons of the two technologies (Marioni et al., 2008). This could be caused by cross-hybridization of the oligonucleotide probes to multiple genes and/or background noise in the microarray experiments. This limitation of the microarray platform prohibits accurate detection of expression values for genes belonging to paralogous gene families, which are predominant in plant genomes (Spannagl et al., 2011), underscoring an advantage of RNA-seq for characterization of complete transcriptomes.

### Gene Expression in Paralogous Gene Families

To investigate paralogous gene expression patterns in reproductive tissues, we constructed paralogous gene family groups using the OrthoMCL algorithm (Chen et al., 2007; Li et al., 2003) (Supplemental Table S1). Of the 39,456 predicted genes in the v2 annotation, approximately 85% (33,457) were assigned to 4975 paralogous gene families of two or more members. The distribution of gene family size (Fig. 3A) shows that a large number of gene families (1829) have two members reflective of an ancient whole genome duplication in maize and previously identified nearly identical paralogs (Schnable et al., 2011, 2009). The remaining gene families (3029) contain 3 to 30 members with 117 gene families containing  $>30$  members. The largest gene

family includes 1435 genes, many of which have annotations related to phosphorylation or protein kinase activity. For all genes, we compared qualitative gene expression levels in our RNA-seq dataset with assignment to gene families (Fig. 3B). Of the 6161 single copy genes, 68% had measurable expression in one or more tissues compared to 81% of paralogous genes. Genes classified as not expressed in any of our tissues (8217 total) may be expressed in tissues not represented in our study or expressed below our detection or sampling limits, annotation artifacts, or induced by external stimuli.

It is hypothesized that genes in large gene families exhibit more diversified expression patterns than genes in smaller gene families due to evolutionary mechanisms such as subfunctionalization, neofunctionalization, or pseudogenization (Force et al., 1999; Yim et al., 2009). To assess the similarity of expression patterns among gene family members in the context of flower and seed development, we calculated pairwise Pearson correlations of log<sub>2</sub> FPKM expression values for all gene combinations within each family. Distributions of PCC values (Fig. 3C) show that as many as 65% of within-family gene pairs are correlated at  $\geq 0.5$  in two-member families. This proportion steadily decreases as gene family size increases thus supporting this hypothesis. In the largest gene families tested (gene family size of 21–30 genes), 31% of gene pairs had PCC values greater than or equal to 0.5, which is higher than expected among random correlations (15%; Fig. 3C). This suggests that there are subsets of genes within all sizes of gene families that have retained conserved expression patterns in developing flower and seed tissues.

An example of divergent gene family expression in male reproductive tissues is a 15-member gene family

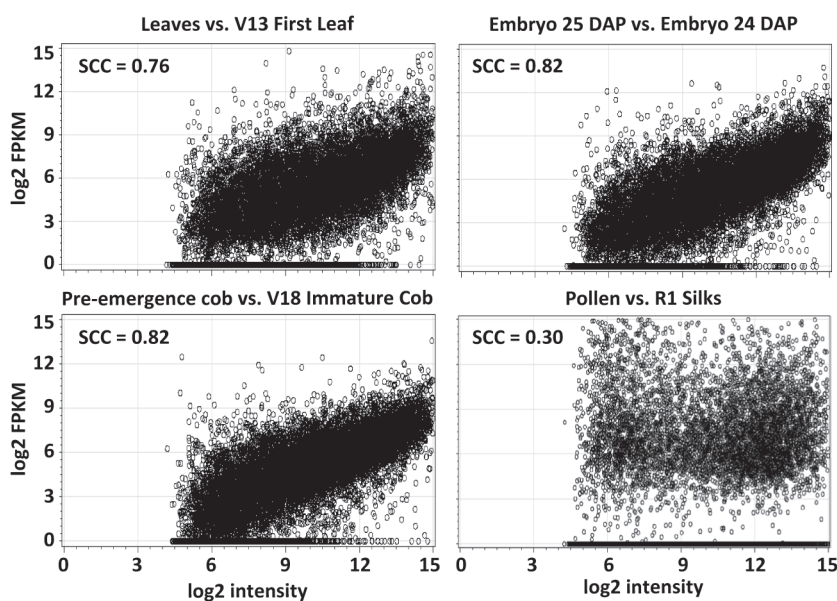
**Table 3. Spearman rank correlations of log<sub>2</sub> expression values between ribonucleic acid sequencing (RNA-seq) and microarray based expression profiles. Eleven RNA-seq libraries generated in this study were compared to the most similar corresponding tissues in a maize microarray study (Sekhon et al., 2011). Three comparisons of dissimilar tissues were included as negative controls. Log<sub>2</sub> expression values for a subset of 17,573 maize genes from the v1 annotation without multiple predicted isoforms and present in both data sets were correlated with the nonparametric Spearman rank correlation statistic (SAS Institute, 2003).**

RNA-seq library <sup>†</sup>	Microarray experiment	SCC <sup>‡</sup>	Tissues types
Leaves (V2)	First leaf (V3)	0.76	Similar
Preemergence cob (V15)	Immature cob (V18)	0.82	Similar
Postemergence cob (VT)	Prepollination cob (R1)	0.81	Similar
Silk (R1)	Silks (R1)	0.78	Similar
Ovule (R1)	Whole seed 2 DAP <sup>§</sup> (R1)	0.79	Similar
Preemergence tassel (V18)	Meiotic tassel (V18)	0.68	Similar
Whole anthers (VT)	Anthers (R1)	0.74	Similar
Whole seed 5 DAP (R1)	Whole seed 4 DAP (R1)	0.80	Similar
Whole seed 10 DAP (R2)	Whole seed 10 DAP (R2)	0.81	Similar
Embryo 25 DAP (R4)	Embryo 24 DAP (R4)	0.82	Similar
Endosperm 25 DAP (R4)	Endosperm 24 DAP (R4)	0.74	Similar
Pollen (VT)	Silks (R1)	0.30	Dissimilar
Leaves (V2)	Endosperm 24 DAP (R4)	0.51	Dissimilar
Embryo 25 DAP (R4)	Anthers (R1)	0.56	Dissimilar

<sup>†</sup>Vegetative (V) and reproductive (R) growth development stages defined in the Iowa State University Extension Corn Field Guide (Abendroth et al., 2011).

<sup>‡</sup>SCC, Spearman correlation coefficient.

<sup>§</sup>DAP, days after pollination.



**Figure 2. Comparison of ribonucleic acid sequencing (RNA-seq) and microarray expression patterns. Microarray expression profiles were obtained from a previous maize atlas experiment (Sekhon et al., 2011). Log<sub>2</sub> transformed expression values of 17,573 maize genes without predicted isoforms and present in both the RNA-seq (log<sub>2</sub> fragments per kilobase pair of exon model per million fragments mapped [FPKM]) and microarray (log<sub>2</sub> intensity) platforms are shown as scatter plots. Spearman correlation coefficients (SCCs) for the comparisons were determined using SAS (SAS Institute, 2003). DAP, days after pollination.**

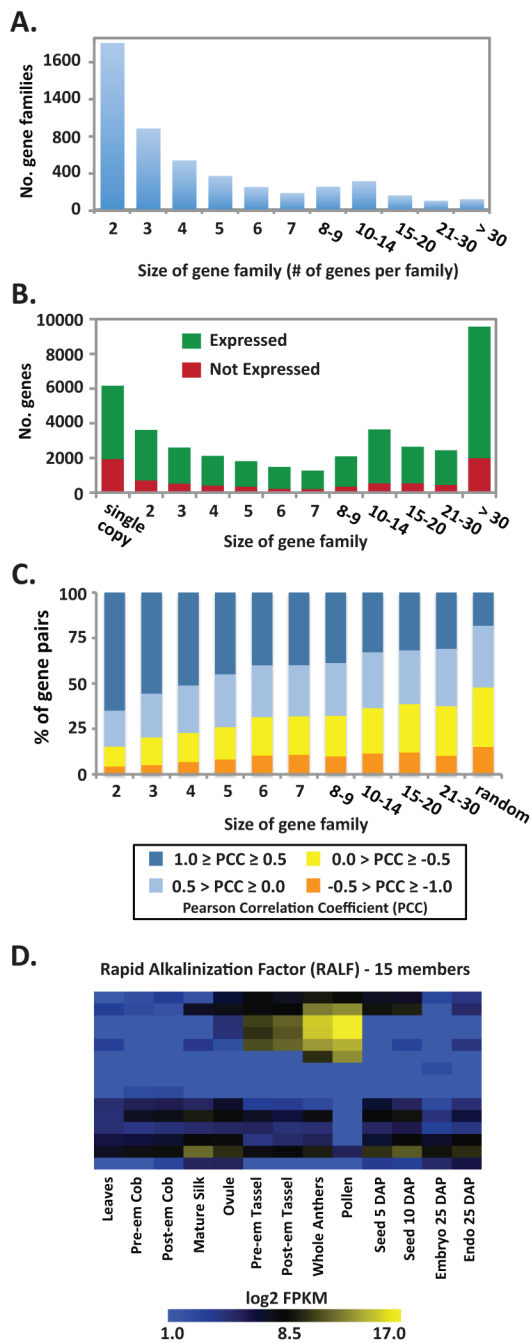


Figure 3. Distribution and expression of maize genes in paralogous families. Using the OrthoMCL algorithm (Li et al., 2003), ~85% (33,457) of genes in the version 2 (v2) annotation were assigned to 4975 paralogous gene families of two or more members. A. Histogram shows the distribution of 4975 gene families. B. All maize genes are shown relative to their gene family assignment and qualitative presence or absence expression level. Genes with low confidence fragments per kilobase pair of exon model per million fragments mapped (FPKM) values equal to zero (defined by Cufflinks [Trapnell et al., 2010]) were defined as not expressed. C. Pairwise correlations of log<sub>2</sub> FPKM expression values for genes within the same gene family were calculated, and Pearson correlation coefficients (PCCs) are summarized by gene family size. Distributions are shown compared to a set of random correlations. D. An example of gene family expression patterns is shown for a 15-member family of genes annotated as rapid alkalinization factors (RALFs). DAP, days after pollination; Pre-em, preemergence; Post-em, postemergence.

annotated as a rapid alkalization factor (RALF; Fig. 3D). Five members show high levels of expression in pollen consistent with previously described roles for these proteins (Covey et al., 2010; Zhang et al., 2010) while the remaining members have low levels or complete lack of expression in pollen. These results demonstrate the utility of RNA-seq for deciphering complex expression patterns of large gene families in plant transcriptomes. More specifically, this dataset is valuable for identifying particular genes and gene family members involved with reproductive organ development in maize.

### Differential Expression and Tissue Restricted Expression Across Tissues

To understand the global transcriptional profile differences of functionally diverse maize tissues, we characterized differential expression of genes expressed across multiple tissues, such as GRMZM2G019404, which encodes a plasma membrane adenosine triphosphatase (ATPase) (Fig. 4A) that is more expressed in leaves, pre-emergence tassel, and whole seed 5 DAP compared to all other tissues. Understanding differences in expression patterns between tissues provides insights into what makes these tissues unique. The number of differentially expressed genes between tissues ranged from 2653 (~6.7% of the genes in the v2 annotation) for preemergence cob versus postemergence cob to 18,962 (~48.1%) for pollen versus ovule (Table 4). Not surprisingly, the three comparisons with the fewest number of differentially expressed genes were those between tissues with repeated measures including cob (pre- and postemergence), tassel (pre- and post emergence), and whole seed (5 and 10 DAP). Comparisons between compound tissues and their component tissues such as postemergence cob versus ovule and ovule versus whole seed 5 DAP were also similar. For example, expression of a putative polyphenol oxidase is restricted to multiple female tissues including silk, ovule, and postemergence cob (Fig. 4B), consistent with its proposed role in oxidation of reproductive tissues after wounding (Sukalovic et al., 2010). Interestingly, 8 of the 10 pairwise tissue comparisons with the highest number of differentially expressed genes included pollen and the remaining two included whole anthers, a compound tissue that contains pollen (Table 4). This set of genes included a putative subtilisin-like protease (GRMZM2G006366; Fig. 4C) that has a rice (*Oryza sativa* L.) homolog known to be involved in anther development (Yoshida and Kuboyama, 2001). Due to limitations in the algorithm used for testing differential expression, genes with an expression value of zero in one of the tissues could not be tested. As pollen has fewer genes expressed than any other tissue, it also had the fewest tests conducted. However, it had the highest (89%) percentage of differentially expressed genes across all pairwise comparisons with the other 12 tissues. Whole anthers was the next highest tissue with 77% significant tests; the tissue with the lowest percentage of significant tests was whole seed 5 DAP with only 57%. These results lend support to the finding that for the few genes that are expressed



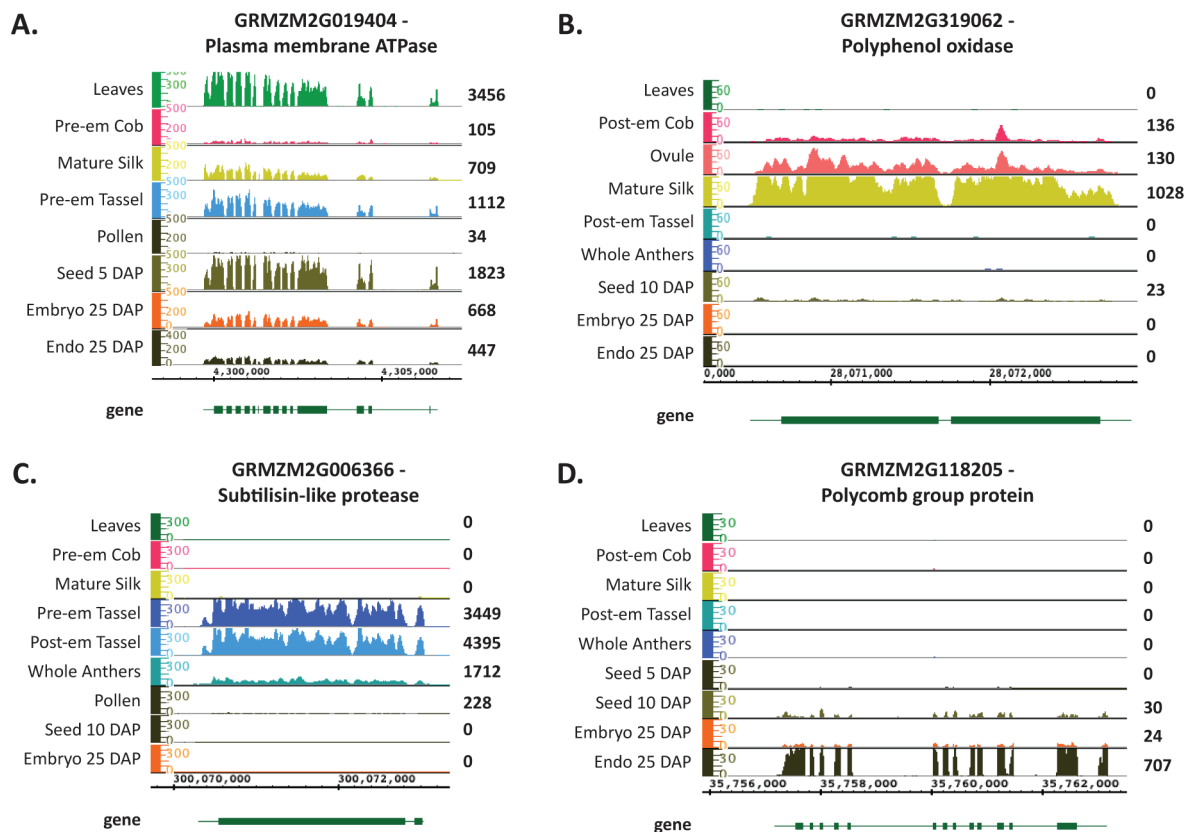


Figure 4. Visualization of ribonucleic acid sequencing (RNA-seq) short read mapping relative to maize gene models. Wiggle tracks indicating chromosomal positions of mapped reads were generated using the wiggles program within TopHat version 1.2.0 (Trapnell et al., 2009) and visualized with the Integrated Genome Browser (Nicol et al., 2009). The maize gene track represents genes in the version 2 (v2) annotation. Figure panels show selected genes with A. differential expression, B. female exclusive expression, C. male exclusive expression, and D. seed exclusive expression. The y axes indicate number of reads mapped. Numbers to the right of the wiggle tracks represent the normalized fragments per kilobase pair of exon model per million fragments mapped (FPKM) expression values for each tissue calculated by Cufflinks (Trapnell et al., 2010). ATPase, adenosine triphosphatase; DAP, days after pollination; Pre-em, preemergence; Post-em, postemergence.

in pollen relative to the other tissues (Table 2), the global expression patterns of those genes are grossly different than that in the other vegetative and reproductive tissues.

In addition to differential expression between tissues, tissue exclusive or restricted gene expression is involved in distinguishing form and function of cells and ultimately whole tissues (Aoyama et al., 1995; Lafos et al., 2011; Lauter et al., 2005). Leaves had the most genes (492) with expression restricted to a single tissue within this study (Table 1; Supplemental Table S3). Excluding leaves, endosperm 25 DAP (251 genes), pollen (206 genes), and embryo 25 DAP (188 genes) had the most genes with expression restricted to a single tissue within this study (Table 1). An example of endosperm restricted expression within the tissues in this study is shown with GRMZM2G118205, which encodes a polycomb protein (Fig. 4D) known to be involved with imprinting processes in maternal tissues (Jullien and Berger, 2009). These differentially expressed genes and genes with expression restricted to a single tissue within this study are strong candidates for further analysis to resolve the molecular basis of maize flower and grain development.

## Gene Coexpression Networks in Maize Reproductive Tissues

Gene network analysis is often performed with gene expression data to identify possible relationships between genes. Typically, gene networks are created based on correlation values derived from gene expression levels across a series of treatments or conditions. Although definitive interactions cannot be concluded from correlation analyses, gene networks do often reflect true biological relationships (Borghini et al., 2010; Ma et al., 2007; Mao et al., 2009; Mutwil et al., 2010; Ren et al., 2010; Xin et al., 2009). Here, we have used WGCNA to generate a correlation-based gene coexpression network and to identify modules of genes where all members of a module are more highly correlated with each other than they are to genes outside the module (Zhang and Horvath, 2005).

The 8751 genes that passed the CV filter, and thus have variation across tissues, were used for network analysis. Of the 8751 genes, a total of 6347 were assigned to 17 gene modules that contained between 49 and 1645 genes each and 2404 genes were not assigned to any module (Fig. 5; Supplemental Table S1). To visualize the expression

**Table 4. Number of differentially expressed genes between each set of tissues. Differential expression analysis was conducted using the cuffdiff program in Cufflinks version 0.9.3 (Trapnell et al., 2010), the version 2 (v2) pseudomolecules, the v2 annotation, and false discovery rate of 0.01.**

	Preemergence cob	Postemergence cob	Silk	Ovule	Preemergence tassel	Postemergence tassel	Whole anthers	Pollen	Seed 5 DAP <sup>†</sup>	Seed 10 DAP	Embryo 25 DAP	Endosperm 25 DAP
Leaves	14,541 (75%) <sup>‡</sup>	14,195 (73%)	12,547 (69%)	14,930 (70%)	13,361 (70%)	12,795 (70%)	13,510 (83%)	13,578 (90%)	13,294 (69%)	14,022 (70%)	14,551 (77%)	13,793 (76%)
Preemergence cob		2,653 (15%)	12,202 (63%)	8,393 (41%)	13,677 (67%)	14,184 (70%)	15,664 (80%)	16,942 (90%)	10,869 (56%)	11,365 (57%)	9,186 (50%)	12,527 (66%)
Postemergence cob			11,543 (60%)	6,493 (32%)	12,869 (63%)	13,460 (67%)	15,494 (79%)	16,972 (90%)	9,496 (50%)	10,519 (53%)	8,689 (47%)	11,953 (63%)
Silk				11,049 (53%)	10,809 (56%)	11,123 (58%)	14,145 (76%)	15,966 (89%)	8,781 (47%)	8,694 (45%)	13,123 (69%)	12,623 (68%)
Ovule					12,263 (57%)	12,961 (60%)	16,316 (76%)	18,962 (90%)	8,296 (40%)	10,472 (50%)	11,189 (54%)	12,649 (61%)
Preemergence tassel						3,522 (19%)	12,514 (68%)	16,117 (88%)	10,160 (51%)	11,425 (56%)	13,691 (68%)	12,444 (64%)
Postemergence tassel							11,301 (66%)	14,942 (87%)	10,743 (55%)	11,972 (58%)	14,059 (71%)	12,891 (68%)
Whole anthers								9,386 (83%)	14,811 (76%)	15,470 (76%)	15,161 (81%)	14,293 (80%)
Pollen									17,029 (90%)	17,886 (90%)	15,841 (90%)	15,275 (90%)
Seed 5 DAP										4,810 (25%)	12,169 (63%)	12,083 (64%)
Seed 10 DAP											12,930 (65%)	12,533 (64%)
Embryo 25 DAP												11,088 (62%)

<sup>†</sup>DAP, days after pollination.

<sup>‡</sup>Numbers in parenthesis indicate the percent of significantly different tests out of the total number of tests that could be performed for each pairwise comparison.

patterns of the genes in each module, eigengenes for each module were calculated and displayed in a heat map (Langfelder and Horvath, 2007) (Fig. 5). The eigengene analysis indicates that each coexpression module represents a set of genes that are strongly expressed in only one or, at most, a few developmentally related tissues. The gene coexpression modules suggest relationships between larger tissue subgroups described in previous sections (Table 1; Fig. 1). For example, module 13 contains a large number of genes that are expressed in pre- and postemergence cobs as well as ovules and 25 DAP embryos. The coexpression analyses also highlight notable differences between similar tissues. Module 9 and module 10 contain relatively small numbers of genes that are almost exclusively expressed in either preemergence or postemergence cobs. Module 8 is the largest module with 1645 genes expressed in all male tissues sampled here while modules 2 through 5 each contain many fewer genes that are specifically expressed in preemergence tassels, postemergence tassels, anthers, and pollen, respectively.

A total of 121 previously characterized maize genes were found within 13 gene modules (Supplemental Table S4), and many are known to have tissue-specific expression similar to the patterns seen in the modules. For example, *pollen extensin-like1*, a pollen-specific cell wall extension gene (Rubinstein et al., 1995), was identified within module 8 that

contains genes that are expressed in pollen containing male tissues. The gene *viviparous1* codes for a transcription factor involved in seed development and maturation (Suzuki et al., 2003) and is a member of the 25 DAP embryo-specific module 16. Module 9 contains genes that are strongly expressed in preemergence cob tissue including the *ramosa1* gene, which is a transcription factor that regulates early inflorescence development (Gallavotti et al., 2010). The identification of a number of known genes in coexpression modules that recapitulate expected expression patterns of those genes provides evidence for the biological relevance of these gene modules. Furthermore, the network analysis performed here provides a powerful tool for identifying candidate genes involved in related developmental pathways.

## Conclusions

Ribonucleic acid sequencing-based expression profiling provides a unique opportunity to explore the transcriptional diversity among tissues on a global level. In this study, we have explored RNA-seq based transcriptional profiles for 12 reproductive tissues and one vegetative tissue in maize. Dramatic differences in whole transcriptome expression patterns were observed between these tissues, in particular for pollen, a tissue that is known to be unique both in terms of form and function.

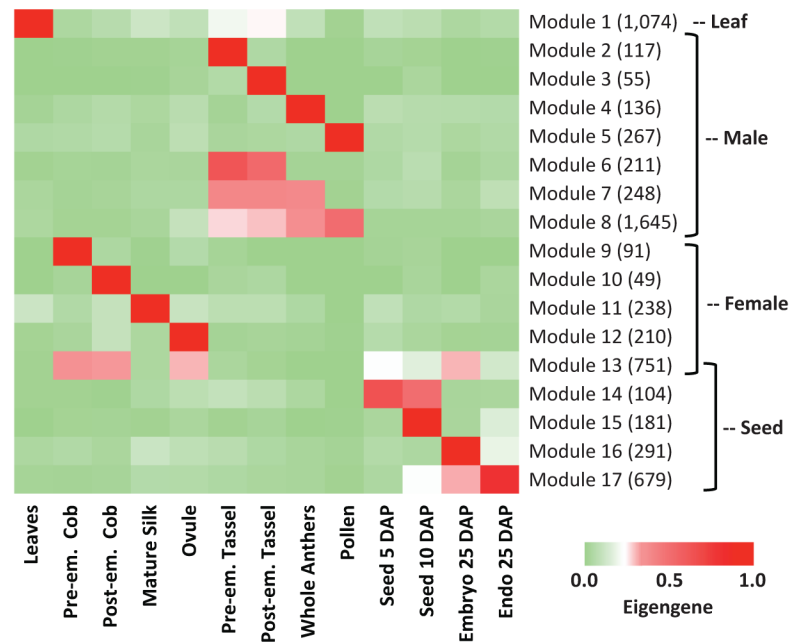


Figure 5. Heat map of the eigengenes representing each gene module. The columns in the heat map represent tissue samples, and the rows represent eigengenes for each of the 17 identified coexpression modules (Langfelder and Horvath, 2007). The numbers of genes in each module are given in parentheses. The cells in the heat map show eigengene values between 0 and 1, which are indicators of relative expression levels of all genes in the module (see Materials and Methods). The labels to the right of the module numbers indicate the larger tissue subgroups where module genes show elevated expression. DAP, days after pollination; Pre-em, preemergence; Post-em, postemergence.

In a comparative analysis with a previous microarray based expression study, we demonstrated the relative advantages of RNA-seq including a broader range of expression values and the ability to differentiate expression of paralogous genes. The capacity to distinguish members of genes families is of critical importance in plant species such as maize where the majority of genes are in paralogous families. Through the use of RNA-seq we were able to identify diverged expression patterns and genes with expression restricted to a single tissue as determined in this study within these families.

Using gene coexpression network analysis, we identified modules of genes that are expressed in very specific stages of floral development as well as genes that are more generally expressed across multiple male, female, or seed tissues. These analyses, as well as the differential gene expression and tissue restricted expression analyses, provide a core set of candidate genes for future analysis to further characterize maize reproductive development and grain yield potential.

### Supplemental Information Available

Supplemental material is available free of charge at <http://www.crops.org/publications/tpg>.

### Acknowledgments

We would like to thank Ann Ferguson and Carolyn Botting for technical assistance. This work was funded in part by a grant to CRB and NJ by the USDA NIFA (2009-65300-05784) and the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). RMD was supported by NSF Minority Postdoctoral Research Fellowship Award #0905696.

### References

- Abendroth, L.J., R.W. Elmore, M.J. Boyer, and S.K. Marlay. 2011. Corn growth and development. PMR 1009. Iowa State University Extension, Ames, IA.
- Acosta, I.F., H. Laparra, S.P. Romero, E. Schmelz, M. Hamberg, J.P. Mottinger, M.A. Moreno, and S.L. Dellaporta. 2009. tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science* 323:262–265. doi:10.1126/science.1164645
- Agarwal, A., D. Koppstein, J. Rozowsky, A. Stoner, L. Habegger, L.W. Hillier, R. Sasidharan, V. Reinke, R.H. Waterston, and M. Gerstein. 2010. Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. *BMC Genomics* 11:383. doi:10.1186/1471-2164-11-383
- Aoyama, T., C.H. Dong, Y. Wu, M. Carabelli, G. Sessa, I. Ruberti, G. Morelli, and N.H. Chua. 1995. Ectopic expression of the Arabidopsis transcriptional activator Athb-1 alters leaf cell fate in tobacco. *Plant Cell* 7:1773–1785. doi:10.1105/tpc.7.11.1773
- Barr, K.L., L.B. Hearne, S. Briesacher, T.L. Clark, and G.E. Davis. 2010. Microbial symbionts in insects influence down-regulation of defense genes in maize. *PLoS ONE* 5:e11339. doi:10.1371/journal.pone.0011339
- Bhave, M.R., S. Lawrence, C. Barton, and L.C. Hannah. 1990. Identification and molecular characterization of shrunken-2 cDNA clones of maize. *Plant Cell* 2:581–588. doi:10.1105/tpc.2.6.581
- Borghini, L., R. Gutzat, J. Futterer, Y. Laizet, L. Hennig, and W. Gruissem. 2010. Arabidopsis RETINOBLASTOMA-RELATED is required for stem cell maintenance, cell differentiation, and lateral organ production. *Plant Cell* 22:1792–1811. doi:10.1105/tpc.110.074591
- Bortiri, E., and S. Hake. 2007. Flowering and determinacy in maize. *J. Exp. Bot.* 58:909–916. doi:10.1093/jxb/erm015
- Chen, F., A.J. Mackey, J.K. Vermunt, and D.S. Roos. 2007. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* 2:e383. doi:10.1371/journal.pone.0000383
- Chuck, G., M. Muszynski, E. Kellogg, S. Hake, and R.J. Schmidt. 2002. The control of spikelet meristem identity by the branched silkless1 gene in maize. *Science* 298:1238–1241. doi:10.1126/science.1076920
- Covey, P.A., C.C. Subbaiah, R.L. Parsons, G. Pearce, F.T. Lay, M.A. Anderson, C.A. Ryan, and P.A. Bedinger. 2010. A pollen-specific RALF

- from tomato that regulates pollen tube elongation. *Plant Physiol.* 153:703–715. doi:10.1104/pp.110.155457
- Eveland, A.L., N. Satoh-Nagasawa, A. Goldshmidt, S. Meyer, M. Beatty, H. Sakai, D. Ware, and D. Jackson. 2010. Digital gene expression signatures for maize development. *Plant Physiol.* 154:1024–1039. doi:10.1104/pp.110.159673
- Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Fu, H., W. Park, X. Yan, Z. Zheng, B. Shen, and H.K. Dooner. 2001. The highly recombinogenic bz locus lies in an unusually gene-rich region of the maize genome. *Proc. Natl. Acad. Sci. USA* 98:8903–8908. doi:10.1073/pnas.141221898
- Gallavotti, A., J.A. Long, S. Stanfield, X. Yang, D. Jackson, E. Vollbrecht, and R.J. Schmidt. 2010. The control of axillary meristem fate in the maize ramosa pathway. *Development* 137:2849–2856. doi:10.1242/dev.051748
- Gardiner, J.M., C.R. Buell, R. Elumalai, D.W. Galbraith, D.A. Henderson, A.L. Iniguez, S.M. Kaepler, J.J. Kim, J. Liu, A. Smith, L. Zheng, and V.L. Chandler. 2005. Design, production, and utilization of long oligonucleotide microarrays for expression analysis in maize. *Maydica* 50:425–435.
- Godfray, H.C., J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, and C. Toulmin. 2010. Food security: The challenge of feeding 9 billion people. *Science* 327:812–818. doi:10.1126/science.1185383
- Hamilton, D.A., D.M. Bashe, J.R. Stinson, and J.P. Mascarenhas. 1989. Characterization of a pollen-specific genomic clone from maize. *Sex. Plant Reprod.* 2:208–212. doi:10.1007/BF00195580
- Hayes, K.R., M. Beatty, X. Meng, C.R. Simmons, J.E. Habben, and O.N. Danilevskaya. 2010. Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator. *PLoS ONE* 5:e12887. doi:10.1371/journal.pone.0012887
- Jullien, P.E., and F. Berger. 2009. Gamete-specific epigenetic mechanisms shape genomic imprinting. *Curr. Opin. Plant Biol.* 12:637–642. doi:10.1016/j.pbi.2009.07.004
- Ku, M.S., T. Kano-Murakami, and M. Matsuoka. 1996. Evolution and expression of C4 photosynthesis genes. *Plant Physiol.* 111:949–957. doi:10.1104/pp.111.4.949
- Lafos, M., P. Kroll, M.L. Hohenstatt, F.L. Thorpe, O. Clarenz, and D. Schubert. 2011. Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. *PLoS Genet.* 7:e1002040. doi:10.1371/journal.pgen.1002040
- Langfelder, P., and S. Horvath. 2007. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* 1:54. doi:10.1186/1752-0509-1-54
- Langfelder, P., and S. Horvath. 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi:10.1186/1471-2105-9-559
- Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi:10.1186/gb-2009-10-3-r25
- Lauter, N., A. Kampani, S. Carlson, M. Goebel, and S.P. Moose. 2005. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *Proc. Natl. Acad. Sci. USA* 102:9412–9417. doi:10.1073/pnas.0503927102
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. doi:10.1093/bioinformatics/btp352
- Li, L., C.J. Stoeckert, and D.S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189. doi:10.1101/gr.1224503
- Li, P., L. Ponnala, N. Gandotra, L. Wang, Y. Si, S.L. Tausta, T.H. Kebrom, N. Provart, R. Patel, C.R. Myers, E.J. Reidel, R. Turgeon, P. Liu, Q. Sun, T. Nelson, and T.P. Brutnell. 2010. The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* 42:1060–1067. doi:10.1038/ng.703
- Liu, X., J. Fu, D. Gu, W. Liu, T. Liu, Y. Peng, J. Wang, and G. Wang. 2008. Genome-wide analysis of gene expression profiles during the kernel development of maize (*Zea mays* L.). *Genomics* 91:378–387. doi:10.1016/j.ygeno.2007.12.002
- Lotan, T., M. Ohto, K.M. Yee, M.A. West, R. Lo, R.W. Kwong, K. Yamagishi, R.L. Fischer, R.B. Goldberg, and J.J. Harada. 1998. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryonic development in vegetative cells. *Cell* 93:1195–1205. doi:10.1016/S0092-8674(00)81463-4
- Ma, J., D.S. Skibbe, J. Fernandes, and V. Walbot. 2008. Male reproductive development: Gene expression profiling of maize anther and pollen ontogeny. *Genome Biol.* 9:R181. doi:10.1186/gb-2008-9-12-r181
- Ma, S., Q. Gong, and H.J. Bohnert. 2007. An Arabidopsis gene network based on the graphical Gaussian model. *Genome Res.* 17:1614–1625. doi:10.1101/gr.6911207
- Maize Genome Sequencing Project. 2011. Maize genome sequence project file transfer protocol (FTP) site. Available at <http://ftp.maizesequence.org> (verified 15 Aug. 2011). Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Mao, L., J.L. Van Hemert, S. Dash, and J.A. Dickerson. 2009. Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics* 10:346. doi:10.1186/1471-2105-10-346
- Marioni, J.C., C.E. Mason, S.M. Mane, M. Stephens, and Y. Gilad. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18:1509–1517. doi:10.1101/gr.079558.108
- Mutwil, M., B. Usadel, M. Schütte, A. Loraine, O. Ebenhoh, and S. Persson. 2010. Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol.* 152:29–43. doi:10.1104/pp.109.145318
- National Center for Biotechnology Information (NCBI). 2011. NCBI sequence read archive. Available at <http://www.ncbi.nlm.nih.gov/sra> (verified 15 Aug. 2011). National Institutes of Health, National Center for Biotechnology Information, Bethesda, MD.
- Nelson, O.E., and H.W. Rines. 1962. The enzymatic deficiency in the waxy mutant of maize. *Biochem. Biophys. Res. Commun.* 9:297–300. doi:10.1016/0006-291X(62)90043-8
- Nicol, J.W., G.A. Helt, S.G. Blanchard, A. Raja, and A.E. Loraine. 2009. The integrated genome browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25:2730–2731. doi:10.1093/bioinformatics/btp472
- R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ren, G., Q. Zhou, S. Wu, Y. Zhang, L. Zhang, J. Huang, Z. Sun, and B. Kuai. 2010. Reverse genetic identification of CRN1 and its distinctive role in chlorophyll degradation in Arabidopsis. *J. Integr. Plant Biol.* 52:496–504.
- Rubinstein, A.L., A.H. Broadwater, K.B. Lowrey, and P.A. Bedinger. 1995. Pex1, a pollen-specific gene with an extensin-like domain. *Proc. Natl. Acad. Sci. USA* 92:3086–3090. doi:10.1073/pnas.92.8.3086
- Saeed, A.I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush. 2003. TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
- SAS Institute. 2003. The SAS system for windows version 9.1. SAS Inst., Cary, NC.
- Sawers, R.J., P. Liu, K. Anufrikova, J.T. Hwang, and T.P. Brutnell. 2007. A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf. *BMC Genomics* 8:12. doi:10.1186/1471-2164-8-12
- Schmidt, R.J., F.A. Burr, M.J. Aukerman, and B. Burr. 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. doi:10.1073/pnas.87.1.46
- Schnable, J.C., and M. Freeling. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS ONE* 6:e17855. doi:10.1371/journal.pone.0017855

- Schnable, J.C., N.M. Springer, and M. Freeling. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. USA* 108:4069–4074. doi:10.1073/pnas.1101368108
- Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F.S. Wei, S. Pasternak, C.Z. Liang, J.W. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Krukowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F.Y. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W.Z. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R.F. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J.K. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambrose, S. Muller, W. Spooner, A. Narechania, L.Y. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, P. Van Buren, M.W. Vaughn, et al. 2009. The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115. doi:10.1126/science.1178534
- Sekhon, R.S., H. Lin, K.L. Childs, C.N. Hansey, C. Robin Buell, N. de Leon, and S.M. Kaeppeler. 2011. Genome-wide atlas of transcription during maize development. *Plant J.* 66:553–563. doi:10.1111/j.1365-313X.2011.04527.x
- Skibbe, D.S., J.F. Fernandes, K.F. Medzihradsky, A.L. Burlingame, and V. Walbot. 2009. Mutator transposon activity reprograms the transcriptomes and proteomes of developing maize anthers. *Plant J.* 59:622–633. doi:10.1111/j.1365-313X.2009.03901.x
- Spannagl, M., K. Mayer, J. Durner, G. Haberer, and J.A. Frohlich. 2011. Exploring the genomes: From Arabidopsis to crops. *J. Plant Physiol.* 168:3–8. doi:10.1016/j.jplph.2010.07.008
- Sukalovic, V.H.T., S. Veljovic-Jovanovic, J.D. Maksimovic, V. Maksimovic, and Z. Pajic. 2010. Characterisation of phenol oxidase and peroxidase from maize silk. *Plant Biol.* 12:406–413. doi:10.1111/j.1438-8677.2009.00237.x
- Suzuki, M., M.G. Ketterling, Q.B. Li, and D.R. McCarty. 2003. *Viviparous1* alters global gene expression patterns through regulation of abscisic acid signaling. *Plant Physiol.* 132:1664–1677. doi:10.1104/pp.103.022475
- Trapnell, C., L. Pachter, and S.L. Salzberg. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111. doi:10.1093/bioinformatics/btp120
- Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.U. Wold, and L. Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28:511–515. doi:10.1038/nbt.1621
- USDA and National Agricultural Statistics Service (NASS). 2011. Crop production 2010 summary. Available at [http://usda.mannlib.cornell.edu/usda/current/CropProdSu/CropProdSu-01-12-2011\\_new\\_format.pdf](http://usda.mannlib.cornell.edu/usda/current/CropProdSu/CropProdSu-01-12-2011_new_format.pdf) (verified 9 Aug. 2011). USDA and National Agricultural Statistics Services, Washington, DC.
- Wang, D., J.A. Osés-Prieto, K.H. Li, J.F. Fernandes, A.L. Burlingame, and V. Walbot. 2010. The male sterile 8 mutation of maize disrupts the temporal progression of the transcriptome and results in the misregulation of metabolic functions. *Plant J.* 63:939–951. doi:10.1111/j.1365-313X.2010.04294.x
- Wang, X.F., A.A. Elling, X.Y. Li, N. Li, Z.Y. Peng, G.M. He, H. Sun, Y.J. Qi, X.S. Liu, and X.W. Deng. 2009a. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* 21:1053–1069. doi:10.1105/tpc.109.065714
- Wang, Z., M. Gerstein, and M. Snyder. 2009b. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10:57–63. doi:10.1038/nrg2484
- Xin, Z., A. Wang, G. Yang, P. Gao, and Z.L. Zheng. 2009. The Arabidopsis A4 subfamily of lectin receptor kinases negatively regulates abscisic acid response in seed germination. *Plant Physiol.* 149:434–444. doi:10.1104/pp.108.130583
- Yim, W.C., B.M. Lee, and C.S. Jang. 2009. Expression diversity and evolutionary dynamics of rice duplicate genes. *Mol. Genet. Genomics* 281:483–493. doi:10.1007/s00438-009-0425-y
- Yoshida, K.T., and T. Kuboyama. 2001. A subtilisin-like serine protease specifically expressed in reproductive organs in rice. *Sex. Plant Reprod.* 13:193–199. doi:10.1007/s004970000059
- Zhang, B., and S. Horvath. 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:Article17.
- Zhang, G.Y., J.A. Wu, and X.W. Wang. 2010. Cloning and expression analysis of a pollen preferential rapid alkalinization factor gene, BoRALF1, from broccoli flowers. *Mol. Biol. Rep.* 37:3273–3281. doi:10.1007/s11033-009-9912-9
- Zheng, J., J. Fu, M. Gou, J. Huai, Y. Liu, M. Jian, Q. Huang, X. Guo, Z. Dong, H. Wang, and G. Wang. 2010. Genome-wide transcriptome analysis of two maize inbred lines under drought stress. *Plant Mol. Biol.* 72:407–421. doi:10.1007/s11103-009-9579-6
- Zhu, Y., J. Fu, J. Zhang, T. Liu, Z. Jia, J. Wang, Y. Jin, Y. Lian, M. Wang, J. Zheng, W. Hou, and G. Wang. 2009. Genome-wide analysis of gene expression profiles during ear development of maize. *Plant Mol. Biol.* 70:63–77. doi:10.1007/s11103-009-9457-2