Combined Large-Scale Phenotyping and Transcriptomics in Maize Reveals a Robust Growth Regulatory Network[OPEN]

Joke Baute, Dorota Herman, Frederik Coppens, Jolien De Block, Bram Slabbinck, Matteo Dell’Acqua, Mario Enrico Pè, Steven Maere, Hilde Nelissen, and Dirk Inzé*

Department of Plant Systems Biology, Vlaams Instituut voor Biotechnologie, 9052 Ghent, Belgium (J.B., D.H., F.C., J.D.B., B.S., S.M., H.N., D.I.); Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium (J.B., D.H., F.C., J.D.B., B.S., S.M., H.N., D.I.); and Institute of Life Sciences, Scuola Superiore Sant’Anna, 56127 Pisa, Italy (M.D., M.E.P.)

ORCID IDs: 0000-0003-0507-6904 (J.B.); 0000-0001-6565-5145 (F.C.); 0000-0003-0047-4380 (J.D.B.); 0000-0002-4961-6231 (M.D.); 0000-0002-5341-136X (S.M.); 0000-0001-7494-1290 (H.N.); 0000-0002-3217-8407 (D.I.).

Leaves are vital organs for biomass and seed production because of their role in the generation of metabolic energy and organic compounds. A better understanding of the molecular networks underlying leaf development is crucial to sustain global requirements for food and renewable energy. Here, we combined transcriptome profiling of proliferative leaf tissue with in-depth phenotyping of the fourth leaf at later stages of development in 197 recombinant inbred lines of two different maize (Zea mays) populations. Previously, correlation analysis in a classical biparental mapping population identified 1,740 genes correlated with at least one of 14 traits. Here, we extended these results with data from a multiparent advanced generation intercross population. As expected, the phenotypic variability was found to be larger in the latter population than in the biparental population, although general conclusions on the correlations among the traits are comparable. Data integration from the two diverse populations allowed us to identify a set of 226 genes that are robustly associated with diverse leaf traits. This set of genes is enriched for transcriptional regulators and genes involved in protein synthesis and cell wall metabolism. In order to investigate the molecular network context of the candidate gene set, we integrated our data with publicly available functional genomics data and identified a growth regulatory network of 185 genes. Our results illustrate the power of combining in-depth phenotyping with transcriptomics in mapping populations to dissect the genetic control of complex traits and present a set of candidate genes for use in biomass improvement.

In plants, leaves are the main organs for fundamental biological functions, such as photosynthesis and respiration. The size, shape, and number of leaves strongly determine a plant’s photosynthetic capacity and the efficiency with which water and nutrients are used. Consequently, the size of the leaves influences plant biomass and yield (Linkies et al., 2010; Pérez-Pérez et al., 2010). Therefore, understanding the mechanisms controlling various aspects of leaf development, such as growth rate and final size, is of great interest, in particular since the demands in food, feed, and renewable bioenergy are envisaged to increase strongly in the next decades (Godfray et al., 2010). Leaf size is a complex trait determined by the interplay of several factors. The dynamics of leaf development have been studied in detail in various plant species at the organ and cellular levels, but insight into the underlying molecular mechanisms remains limited. Typically, leaf development starts with the recruitment of founder cells at the peripheral zone of the shoot apical meristem, followed by a phase of cell proliferation and successive cell expansion. During this postprimordial development, leaf differentiation progresses basipetally in monocots: the region containing proliferating cells becomes gradually restricted to the base of the organ; as a consequence, a growing leaf shows a developmental gradient with dividing cells at the base, fully differentiated mature cells at the tip, and in between a region of expanding cells (Poethig, 1984; Sylvester et al., 1990). At the molecular level, some genes that affect leaf size have been identified, primarily in the model species Arabidopsis (Arabidopsis thaliana; for review, see Gonzalez et al., 2012; Hepworth and Lenhard, 2014). Although regulators of both cell division and cell expansion have been identified, mutants or transgenic
lines with larger leaves tend to be composed of more cells rather than larger cells (Niklas, 1994; Gonzalez et al., 2010). For instance, leaves of maize (Zea mays) plants with altered levels of GA are affected in their growth rates, and the size of the division zone (DZ) is changed correspondingly (Nelissen et al., 2012). Examples in Arabidopsis of genes that are regulators of final leaf size by influencing cell proliferation are AVP1, JAW, and BRI1 (Gonzalez et al., 2010), GROWTH-REGULATING FACTOR1 (GRF1) and GRF2 (Kim and Kende, 2004), GRF5 (Horiguchi et al., 2005), DA1 and ENHANCER OF DA1 (Li et al., 2008), ANGUSTIFOLIA3 (AN3)/GRF-INTERACTING FACTOR (Lee et al., 2009), and KLUH (Kazama et al., 2010). These examples illustrate that cell proliferation seems to be a key contributing factor to final leaf size.

Exploiting natural variation has been proposed as a complementary approach to the traditional, gene-centric reverse and forward genetics approaches to identify new genes (Weigel, 2012). Knockdown or overexpression of single genes does not capture the extensive genetic variation present in natural populations, which results from a combination of single-nucleotide substitutions, insertions, deletions, copy number variations, epigenetic changes, and expression differences. In the monocotyledonous species maize, the intraspecific variation is large and offers great potential to relate genotype to phenotype. Also, several mapping populations are available, but until now, only a few were used to determine the genetic control of leaf-related traits via quantitative trait locus (QTL) or genome-wide association studies (Pellesschi et al., 2006; Ku et al., 2010, 2012; Tian et al., 2011; Dignat et al., 2013). Although several small-effect QTLs were identified in these studies, further fine-mapping using complementary approaches or a priori knowledge is required to find the genes underlying the quantitative trait.

Previous studies found evidence that the phenotypic diversity in maize is mainly under transcriptional control (Li et al., 2012; Wallace et al., 2014). The recent availability of cost-efficient and high-throughput sequencing technologies to analyze transcriptomes provides new opportunities to gain further insights into the molecular basis of leaf size. In maize, several recent studies applied next-generation sequencing technologies to assess transcriptomic differences between leaf developmental stages and leaf cell types. Transcriptomes of different regions of a growing leaf, representing different developmental stages, were compared by Li et al. (2010) and Pick et al. (2011). Transcriptional dynamics during early development of embryonic leaves were surveyed by Liu et al. (2013b) and Yu et al. (2015). Additionally, the regulatory and functional differentiation of various leaf cell types was examined by transcriptome analysis (Li et al., 2010; Wang et al., 2013, 2014). The transcriptional variation between tissue types and cell types during development (Li et al., 2010; Wang et al., 2013, 2014) illustrates the importance of focusing the analysis of a given process on those tissues where the process takes place. Since leaf growth is driven by proliferation and expansion, zooming in on proliferative and/or expanding tissue is required to identify the regulatory networks underlying leaf development. The developmental gradient in growing maize leaves and the large size of the leaf makes it possible to dissect these specific growth zones for further analysis (Nelissen et al., 2012). Since it was recently suggested that it is the final number of cells that primarily determines final leaf size (Gonzalez et al., 2010; Nelissen et al., 2012), we focus our transcriptome analysis specifically on proliferative tissue of the growing leaf.

Although analyses of transcriptional variation during maize leaf development have provided us with new insights, all these studies were restricted to one genetic background (Li et al., 2010; Pick et al., 2011; Liu et al., 2013b; Wang et al., 2013, 2014; Yu et al., 2015). Adding an additional layer of information, phenotypic variation in mapping populations, and combining this with transcriptome variation in these populations offers new opportunities to identify genes and regulatory mechanisms that are at the basis of phenotypic differences (Andorf et al., 2012).

Recently, we associated phenotypic variation with transcriptome variation in 103 lines of a biparental recombinant inbred line (RIL) population (Baute et al., 2015). We described the relationship of leaf size traits, such as final leaf area and leaf weight, and transcriptional variation in fully proliferative tissue sampled during early leaf development. Although the genetic and phenotypic variation in a classical biparental RIL mapping population provides a valuable source of information, the possibility to detect variation in expression that is associated with phenotypic variation remains limited, since it depends on the polymorphisms between only two parents. In multiparent advanced generation intercross (MAGIC) populations, RILs are generated from multiple parents by mixing the genomes of the founder lines through several rounds of mating, followed by inbreeding to obtain a set of stable homozygous lines (Churchill et al., 2004; Cavanagh et al., 2008; Kover et al., 2009). Such a MAGIC population was recently established for maize, and a panel of 529 lines was genotyped (Dell’Acqua et al., 2015). The population has a larger genetic diversity than biparental populations, and as such, the number of components in the regulatory network that can be identified is expected to be higher. Moreover, integrating results from different populations may allow for the identification of the most robust players in the growth-related molecular network across different lines. In this study, we performed detailed phenotyping and transcriptome analysis of 94 lines of the MAGIC population and combined this with the previously described phenotyping and transcriptome analysis of 103 lines of the biparental B73 × H99 population. We identified a set of 226 genes with expression levels in the DZ of the growing leaf (anti)correlating with leaf phenotype measurements in both populations. Some of these
genes have homologs in Arabidopsis with a known function in leaf development. However, the majority of the genes had no known function or were not annotated before as having a role in this process, implying that these might be interesting candidates to further decipher the molecular network underlying leaf development. Additionally, integration of publicly available functional genomics data (De Bodt et al., 2012) led to the identification of a subset of 185 genes that are interconnected through expression correlations or protein-protein interactions.

RESULTS AND DISCUSSION

Correlation between Leaf Size and Shoot Traits Is Similar in the Two Different Populations

Previously, in-depth phenotyping of 103 lines of a biparental RIL population derived from the inbred parents B73 and H99 (Marino et al., 2009) was combined with transcriptome profiling to dissect leaf size, growth, and shoot-related traits into phenotypic and molecular components (Baute et al., 2015). Here, a similar analysis was conducted on the recently established multiparent MAGIC population (Dell’Acqua et al., 2015), and results of both analyses were integrated. Concerning phenotyping, final leaf size-related traits, such as leaf length (LL), leaf width (Lwi), leaf area (LA), and leaf weight (Lwe), were complemented with measurements that capture growth kinetics, such as growth rate (leaf elongation rate [LER]) and duration (emergence, time point of maximal LER [Tm], time point when leaf 4 reaches its final length [Te], and leaf elongation duration [LED5-e]; Voorend et al., 2014), and cellular measurements, such as the size of the cell DZ. In addition, whole-shoot variables were measured at the seedling stage: fresh weight, dry weight, leaf number (LN), and vegetative (V)-stage (maize leaf stage; Baute et al., 2015; Fig. 1; Supplemental Table S1).

Pearson correlation coefficients (PCCs) between the traits were determined based on the data obtained for the MAGIC population (Table I), for the combined data of both populations (Supplemental Table S2), and compared with our previous analysis for the biparental RIL population (Table I; Baute et al., 2015). All three analyses gave comparable results and supported the separation of the traits into three groups: leaf size traits (LL, Lwe, LA, Lwi, LER, and DZ size), shoot-related

Figure 1. Schematic representation of maize leaf development as a function of time and phenotyping strategy. Shown is the leaf growth of B73. LL was measured daily from emergence from the surrounding leaves until 27 d after sowing, LER, Tm, Te, and LED5-e were deduced from these daily measurements. DZ size was determined 2 d after leaf emergence, at the same time that leaf material was sampled for RNA sequencing. End measurements, LL, LA, Lwe, Lwi, fresh weight (FW), dry weight (DW), V-stage, and LN, were determined 27 d after sowing, when leaf 4 had reached its mature size for all RILs.
traits (fresh weight, dry weight, LN, and V-stage), and timing-related traits (emergence, $T_m$, $T_e$, and LED$_{50}$; Table I; Supplemental Table S2; Supplemental Fig. S1). Also between groups of traits, the correlations were largely the same in the two populations.

In general, the PCCs were higher for the MAGIC population than for the biparental RIL population, especially among leaf, shoot, and timing traits and between leaf-shoot and leaf-timing traits, which may be due to the fact that the phenotype variation in the MAGIC population is generally larger than that of the biparental RIL population (i.e. the phenotype distributions are broader; Supplemental Table S1; Supplemental Fig. S2), which may suppress the negative influence of stochastic and measurement noise on PCC values. A larger phenotypic variation in the eight-way RIL population was expected, given the increased variation in the phenotypes of the parents compared with the biparental mapping population (Supplemental Fig. S2). As an exception, the PCC between LER and DZ size was lower in the MAGIC population than in the biparental population, 0.347 and 0.507, respectively. Also, the anticorrelations between the timing traits and shoot traits fresh weight and dry weight were slightly stronger in the biparental population than in the MAGIC population (Table I). Possibly, the biparental population already covers a large part of the variability for the timing and shoot traits, while additional power is available in the MAGIC population for the leaf size traits. This is confirmed if we determine the positions of the parental lines in the distributions of the two populations (Supplemental Fig. S2): for the leaf size traits LL, Lwe, LA, and Lwi, values for the parental lines B73 and H99 are very similar, while they are more diverse for the other traits.

Correlation between Traits Is Fully Supported at the Transcriptome Level

In both populations, DZ size correlated positively with the leaf size traits and, to some extent, with shoot fresh weight and dry weight (Table I; Supplemental Table S2; Supplemental Fig. S2). This supports the hypothesis that the number of dividing cells is one of the key factors in the determination of final organ size and that the transcriptional differences between genotypes in proliferative tissue may inform us on important players that determine final size traits. We previously performed RNA sequencing of proliferative leaf tissue of 103 lines of the biparental RIL population (Baute et al., 2015) and extended this here with RNA sequencing of comparable tissue of 94 lines of the MAGIC population.

Table I. PCC for the analyzed traits, grouped as leaf size, shoot-related, and timing-related traits for biparental and MAGIC RIL populations

<table>
<thead>
<tr>
<th>Trait</th>
<th>Biparental RIL Population (Bottom Diagonal)</th>
<th>MAGIC RIL Population (Top Diagonal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LER</td>
<td>0.772** 0.517** 0.466** 0.151 0.347** 0.607** 0.659**</td>
<td>0.187 0.081 -0.069 -0.082 -0.054 -0.045</td>
</tr>
<tr>
<td>LL</td>
<td>0.738** 0.775** 0.782** 0.487** 0.615** 0.355** 0.427**</td>
<td>-0.259 -0.435** 0.239** 0.421** 0.387** 0.502**</td>
</tr>
<tr>
<td>Lwe</td>
<td>0.493** 0.750** 0.904** 0.787** 0.562** 0.418** 0.419**</td>
<td>-0.286 -0.500** 0.338** 0.496** 0.461** 0.556**</td>
</tr>
<tr>
<td>LA</td>
<td>0.479** 0.781** 0.915** 0.880** 0.603** 0.289** 0.356**</td>
<td>-0.354** -0.556** 0.399** 0.585** 0.528** 0.660**</td>
</tr>
<tr>
<td>Lwi</td>
<td>0.063 0.316** 0.688** 0.794** 0.478** 0.102 0.14</td>
<td>-0.435** -0.623** 0.475** 0.616** 0.655** 0.659**</td>
</tr>
<tr>
<td>DZ size</td>
<td>0.507** 0.594** 0.516** 0.526** 0.308** 0.279** 0.315**</td>
<td>-0.137 -0.301** 0.142 0.320** 0.248** 0.423**</td>
</tr>
<tr>
<td>Fresh weight</td>
<td>0.502** 0.215** 0.413** 0.297** 0.17 0.211**</td>
<td>0.179 0.939** 0.614** 0.456** -0.473** -0.455** -0.481** -0.342**</td>
</tr>
<tr>
<td>Dry weight</td>
<td>0.502** 0.206** 0.311** 0.245** 0.098 0.219**</td>
<td>0.083 0.435** 0.326** 0.834** -0.725** -0.887** -0.866** -0.860**</td>
</tr>
<tr>
<td>LN</td>
<td>0.176 -0.255 -0.258 -0.273 -0.267 -0.251 0.521** 0.377**</td>
<td>0.834** -0.735** -0.816** -0.818** -0.752**</td>
</tr>
<tr>
<td>V-stage</td>
<td>0.083 -0.423** -0.420 -0.472** -0.375** -0.266 0.435** 0.326**</td>
<td>0.834** -0.725** -0.887** -0.866** -0.860**</td>
</tr>
<tr>
<td>Emergence</td>
<td>0.007 0.245** 0.258** 0.291** 0.307** 0.097</td>
<td>-0.455** -0.462** -0.649** -0.599** -0.865** 0.911** 0.742**</td>
</tr>
<tr>
<td>$T_0$</td>
<td>-0.105 0.379** 0.328** 0.414** 0.365** 0.113</td>
<td>-0.582** -0.557** -0.621** -0.701** 0.620** 0.976** 0.960**</td>
</tr>
<tr>
<td>$T_n$</td>
<td>-0.048 0.374** 0.334** 0.413** 0.379** 0.118</td>
<td>-0.568** -0.557** -0.601** -0.675** 0.650** 0.979** 0.880**</td>
</tr>
<tr>
<td>LED$_{50}$</td>
<td>-0.114 0.424** 0.332** 0.418** 0.298** 0.147</td>
<td>-0.470** -0.420** -0.566** -0.656** 0.461** 0.894** 0.787**</td>
</tr>
</tbody>
</table>

Significant correlations are indicated by asterisks (**, $P < 0.01$ and *, $P < 0.05$); highly significant positive correlations are indicated in italics, and highly significant negative correlations are indicated in bold.
Linear correlation between phenotypes and transcript levels was determined by calculating PCCs between the expression level of each transcript and each trait in both populations separately. The \( q_{0.99} \) and \( q_{0.01} \) PCC (i.e. the correlation coefficient of the 1% best [anti] correlating transcripts) were determined before and after permutation of the trait data (for details, see "Materials and Methods"). For the majority of the traits, the \( q_{0.99} \) and \( q_{0.01} \) PCC values were significantly higher than those expected at random (Fig. 2), indicating that the gene sets identified by this arbitrarily chosen limit of 1% contain genes whose expression levels in proliferative tissue of a growing leaf correlate significantly with final size measurements. For the leaf size traits in particular, the \( q_{0.99} \) and \( q_{0.01} \) PCC were higher in the MAGIC population than in the biparental RIL population. As the cutoffs defined by permuting the data were very comparable for both populations (Fig. 2), this resulted in a higher number of genes with a PCC greater than the \( q_{0.99,\text{random}} \) PCC in the MAGIC population (Table II), indicating that the higher variability in the MAGIC population facilitates the identification of significant transcript-phenotype correlations for leaf size traits. On the other hand, \( q_{0.99} \) and \( q_{0.01} \) PCC for fresh weight and dry weight were considerably lower in the MAGIC population than in the biparental RIL population, suggesting that the relationship between gene expression and fresh weight and dry weight traits in the eight-way population may be more complex and nonlinear in nature than in the two-way population, due to an increased number of different alleles that play a role. Accordingly, the number of genes with a PCC higher than the \( q_{0.99,\text{random}} \) PCC or lower than the \( q_{0.01,\text{random}} \) PCC for dry weight and fresh weight in the MAGIC population was lower than in the biparental population, and the number of genes in the intersection of both populations was limited and not higher than expected by chance (Table II). Also for LER, the percentage of genes found in common for the two populations was small and not higher than expected by chance \((P > 0.05)\), while for all other traits, the overlap was significant (Table II). In both populations, the \( q_{0.99} \) and \( q_{0.01} \) PCC for LER were lower than for the other traits (Fig. 2). The lack in significant overlap between the two populations of transcripts whose expression correlated to LER, fresh weight, and dry weight indicates that another range of growth mechanisms and networks may be active and/or captured for these three traits in the different populations.

Further analyses were restricted to the 1% best correlating and anticorrelating genes for each trait, or 286 genes for each trait, referred to below as the correlated and anticorrelated gene sets. Importantly, although some traits had \( q_{0.99} \) and \( q_{0.01} \) PCC close to the corresponding \( q_{\text{random}} \) PCC, in all cases the selected 286 genes had a PCC higher or lower than the \( q_{0.99} \) and \( q_{0.01} \) thresholds, respectively, and thus higher or lower than the corresponding \( q_{\text{random}} \) PCC. The total number of genes thus selected for at least one trait was 22% lower in the MAGIC population (1,367) than in the biparental RIL population (1,740), since the gene selection for the MAGIC population contains a higher proportion of genes that correlate with several traits compared with the biparental RIL population (Fig. 3A). This might be due to the higher correlation between the leaf size traits in the MAGIC population. In the MAGIC population, we found 21 genes (Supplemental Table S3) associated with nine traits, the maximum in the biparental RIL population being eight traits. One of these 21 genes, \( \text{GRMZM2G389768} \), with homology to cold shock

![Figure 2](https://www.plantphysiol.org/content/170/2/1852/F2.large.jpg) The 0.99 and 0.01 quantiles of distribution of Pearson correlation between transcript expression levels and traits. Real data are shown in dark gray (biparental RIL population) and light gray (MAGIC RIL population) bars; permuted data are shown in dark gray (biparental RIL population) and light gray (MAGIC RIL population) lines. Error bars indicate the SD of the permuted data \((n = 1,000)\) in dark gray for the biparental population and in light gray for the MAGIC population. DW, Dry weight; FW, fresh weight.
domain proteins, correlated with eight traits in the biparental RIL population. In both populations, the numbers of (anti)correlating genes shared between traits was higher within the three trait groups (leaf size, timing, and shoot) than between these groups (Fig. 3, B and C), in accordance with the correlations found at the phenotype level. As for the biparental population (Baute et al., 2015), no opposite gene-trait correlations with traits that were categorized in the same group were found for the MAGIC population (numbers in blue in Fig. 3, B and C). Opposite correlation of a transcript with multiple traits was very limited in both populations, 84 and 28 genes in the two-way and eight-way populations, respectively, and was observed only for timing-related traits versus shoot-related traits, traits that are also anticorrelated at the phenotype level.

Combining the Two Populations Identifies Genes That Are Robustly Associated with the Traits

In the separate populations, expression levels of 1,740 and 1,367 genes in the biparental and MAGIC populations, respectively, correlated with at least one of the traits. Of these genes, 226 were (anti)correlating with at least one of the traits in both populations, a strong reduction compared with the numbers found in the populations separately (Supplemental Table S3). The expression levels of the 226 (anti)correlating genes in all RILs and their association with the traits were visualized in a clustered heat map (Fig. 4), revealing a clear gradient in expression levels that were coinciding with (Fig. 4A) or opposite to (Fig. 4B) the phenotypic variation observed in the RILs. The 226 genes in common for both populations were unevenly distributed over the different traits. The percentage of (anti) correlating genes in the quantiles that was shared between the two populations ranged from a few percent for LER, fresh weight, and dry weight up to 29% for DZ size (Table II). For all traits except LER, fresh weight, and dry weight, the number of genes in the intersection was larger than expected by chance (hypergeometric test; \( P < 0.05 \)), and overall, the number of genes in the intersection was highest for leaf size traits. It is worth noting that for none of the traits were there correlating genes in one population and anticorrelating genes in the other population. The strong reduction of the number of correlated genes by combining the two populations, combined with the fact that the overlap remains significantly higher than expected by chance for the majority of the traits, indicates that this approach is efficient in identifying the genes that are more robustly associated with a specific trait, regardless of the population context, and thus might be more relevant to characterize functionally.

### Table II. Number of genes correlating with traits in biparental and MAGIC RIL populations, and number and percentage of genes in the intersections

<table>
<thead>
<tr>
<th>Trait</th>
<th>Correlation (Random)</th>
<th>Correlation (Quantiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Biparental</td>
<td>MAGIC</td>
</tr>
<tr>
<td>LL</td>
<td>2,206</td>
<td>3,111</td>
</tr>
<tr>
<td>Lwe</td>
<td>2,396</td>
<td>3,141</td>
</tr>
<tr>
<td>LA</td>
<td>2,230</td>
<td>3,165</td>
</tr>
<tr>
<td>Lwi</td>
<td>2,477</td>
<td>2,504</td>
</tr>
<tr>
<td>LER</td>
<td>1,073</td>
<td>1,174</td>
</tr>
<tr>
<td>DZ size</td>
<td>1,927</td>
<td>2,228</td>
</tr>
<tr>
<td>Emergence</td>
<td>2,490</td>
<td>1,463</td>
</tr>
<tr>
<td>T(_f)</td>
<td>3,171</td>
<td>2,368</td>
</tr>
<tr>
<td>T(_e)</td>
<td>3,260</td>
<td>2,892</td>
</tr>
<tr>
<td>LED(_{50})</td>
<td>3,003</td>
<td>3,419</td>
</tr>
<tr>
<td>Fresh weight</td>
<td>2,259</td>
<td>530</td>
</tr>
<tr>
<td>Dry weight</td>
<td>1,707</td>
<td>401</td>
</tr>
</tbody>
</table>

*aNumbers of genes with PCC greater than \( q_{\text{random}} \) PCC. \( b \)Genes in \( q_{0.01} \) and \( q_{0.99} \) in the intersection of two data sets.

Enriched Functional Categories Are Partially Overlapping for the Two Populations

* All genes were assigned to MapMan functional categories (Thimm et al., 2004), and tests for the enrichment of functional categories in the correlating gene sets for the different traits were performed for the two populations separately, to verify if gene sets were enriched for comparable categories in the two populations (Supplemental Fig. S3). For positively correlating genes, the major enriched functional category in both populations was regulation of transcription. The functional categories hormone metabolism, protein modifications, and protein degradation were enriched for several traits in the biparental RIL population but not or for only one trait in the MAGIC population. Regulation of transcription, protein synthesis, and cell wall synthesis and degradation were significantly enriched categories for the negatively correlated gene sets for multiple traits in both populations. Thus, the major enriched functional categories are regulation of transcription, protein synthesis, and cell wall synthesis.
and degradation, with 56, 28, and 15 genes in common between the two populations.

Next, we compared the frequencies of all functional categories in the intersection of both populations (i.e. the 226 genes) with their frequencies in the populations independently (i.e. the 1,740 and 1,367 genes in the biparental and MAGIC populations, respectively) to try to understand the nature of processes conserved...
Figure 4. Expression levels of the 226 correlating and anticorrelating genes in all RILs. A, Correlated genes. B, Anticorrelated genes. Columns represent RILs of the two populations that were hierarchically clustered, and rows represent gene expression profiles with, on top, above the white separation, a heat map of the phenotypic traits. Green indicates low values and red indicates high values.
across the two populations. The functional categories regulation of transcription, protein synthesis, and cell wall synthesis and degradation were significantly ($P < 0.05$) enriched 1.8-, 3-, and 2.5-fold, respectively, in the intersection compared with the complete gene set correlated with the biparental population, while compared with the MAGIC population, the two functional categories regulation of transcription and cell wall synthesis and degradation were enriched significantly ($P < 0.05$) 1.4- and 1.8-fold, respectively. This underscores the usefulness of focusing on the intersection for identifying the processes conserved in different populations, since largely the same processes are enriched in both analyses.

The three major functional categories we identified, regulation of transcription, protein synthesis, and cell wall synthesis and degradation, are described in more detail in the following sections.

**Transcription Factors Involved in Hormone Regulation, Leaf Architecture Traits, and Chromatin Structure Correlate Robustly with Leaf Size**

To obtain a better understanding of the nature of the regulatory processes during leaf growth, we focused on the 56 transcription factor (TF) genes whose expression levels were correlating with traits in both populations. Breakdown of the functional category regulation of transcription into the different TF families revealed that the major classes of TFs in the gene sets that are positively correlated with the traits are ARGONAUTE, MADS box, SQUAMOSA PROMOTER-BINDING PROTEIN (SBP), SET, GRF, and bZIP, while the major classes of TFs that are negatively correlated with the traits are bHLH, GATA, and TRIHELIX (Fig. 4). As determined with the Bio-Analytic Resource for maize gene expression (http://bar.utoronto.ca/epf_maize), about one-third of these TFs are expressed specifically in the DZ (Fig. 5). This is less than expected, since, according to Li et al. (2010), approximately 70% of all TFs show DZ-specific expression. Among the 56 TFs, we could identify 22 reported before to be linked to growth (Fig. 5). Many of these are related to hormone signaling, leaf architecture, and chromatin structure, three classes that are discussed in more detail below.

Our data set contains three GATA-type TFs that show a negative correlation with the leaf size-related traits (Fig. 5). In Arabidopsis, two redundant GATA-type TFs, GATA, NITRATE-INDUCIBLE, CARBON-METABOLISM INVOLVED (GNC) and GNC-LIKE/CYTOKININ-RESPONSIVE GATA FACTOR1 (GNL/CGA1), were identified to regulate multiple aspects of plant development by repressing GA signaling. In agreement with the anticorrelation between the transcript level and leaf size, double mutants displayed increased rosette sizes, while overexpression resulted in smaller plants (Richter et al., 2010). Moreover, it was shown that these GATA factors act downstream of AUXIN RESPONSE FACTOR2 (ARF2), which was also identified in the subset of 56 TFs. GNC and GNL overexpressors show phenotypic similarities with arf2 mutants (Richter et al., 2013). In agreement, ARF2 expression levels were positively correlated with several traits (i.e. LL, Lwe, and LA), while expression levels of the GATA-type TFs were negatively correlated with these traits (Fig. 5). In addition, other TFs that are likely implicated in hormone regulation, such as Arabidopsis Response Regulator (ARR), cytokinin signaling; [GRMZM2G129954; Ren et al., 2009] and GRAS (GA signaling; GRMZM2G097456) family members (Hirsch and Oldroyd, 2009), also correlated with the traits in both populations. A bHLH TF (GRMZM2G159456) with homology to the Arabidopsis and rice (Oryza sativa) paclobutrazol-resistant family of TFs, which mediate growth responses to multiple environmental and hormonal signals (Zhang et al., 2009; Bai et al., 2012), was negatively correlated with leaf size-related traits (Fig. 5).

Also correlated with leaf size traits were some TFs that are known to define leaf architecture traits. For example, the expression profile of LIGULELESS2 (LG2), a bZIP TF involved in establishing the position of the ligule (Walsh et al., 1998), was found to be positively correlated with leaf size- and timing-related traits (Fig. 5). Mutations in LG2 affect leaf architecture due to a change in the leaf angle. However, no clear effect on leaf size has been reported (Tian et al., 2011), although LG2 was mapped within a meta-QTL for leaf length and leaf width (Ku et al., 2012). Moreover, LG2 functions in the same pathway as LIGULELESS NARROW (LGN) and lg2 transcripts are reduced in lgn mutants, which display a severe reduction in leaf size and total plant height (Moon et al., 2013).

As another example, the expression levels of two SBP/SBP-LIKE (SPL) genes, which regulate a wide variety of processes associated with shoot development, correlated with leaf size traits in our data set (Fig. 4): one maize-specific gene (GRMZM2G414805) and one gene (GRMZM2G067624) with homology to Arabidopsis SPL4, which is involved in vegetative phase change (Wu and Poethig, 2006). Another class of genes known to affect leaf architecture traits are the TCP TFs that contain a bHLH domain and are involved in the coordination of cell proliferation, cell differentiation, and growth; as such, they play a role in leaf development (Ori et al., 2007). One of the genes identified as correlated with leaf size traits, GRMZM2G465091, shows homology to Arabidopsis class I TCP proteins. Quadraple and pentuple loss-of-function mutants of class I TCP genes in Arabidopsis have larger but fewer rosette leaves (Aguilar-Martínez and Sinha, 2013), in agreement with the negative correlation between leaf size traits and expression levels of this maize homologous gene.

A third set of TFs that correlated with leaf size traits were related to chromatin structure. Three positively correlating SET domain TF family proteins were identified with homology to Arabidopsis SU(VAR)3-9 RELATED4 (SUVR4) and ASHI-RELATED3 (ASHR3),
involved in histone modification (Veiseth et al., 2011; Kumpf et al., 2014). ASHR3 is required for coordinated DNA replication and cell division, and the \textit{ashr3} mutant has a reduced root apical meristem size and primary root size (Kumpf et al., 2014). Moreover, it was shown that expression levels of \textit{ASHR3} are controlled by the E2Fa/E2Fb-DPa TF complex (Kumpf et al., 2014); in agreement, expression levels of E2F/DP TF (GRMZM2G462623) correlated positively with final leaf size traits and DZ size in both RIL populations (Fig. 5).

Maintenance of epigenetic signatures by setting up the appropriate epigenetic marks is essential to regulate gene expression and establish euchromatin and heterochromatin. Furthermore, histone modification

---

**Figure 5.** TFs with expression profiles correlating or anticorrelating with at least one of the traits in both populations. Bright red indicates positive correlation in the biparental population, dark red indicates positive correlation in the MAGIC population, dark green indicates negative correlation in the biparental population, and light green indicates negative correlation in the MAGIC population. DZ-specific TFs are indicated in boldface (Li et al., 2010). DW, Dry weight; FW, fresh weight.
pathways are intertwined with DNA methylation, for instance by interaction between chromatin modifiers such as SET domain proteins and DNA methyltransferases (Cedar and Bergman, 2009) and/or methyl-CpG-binding domain proteins (MBDs; Zemach and Grafi, 2007), which read out the DNA methylation pattern. The subset of 56 TFs also includes a DNA methyltransferase, of which the expression levels in the DZ were positively correlated with Lwe, LA, DZ size, and the timing-related traits, and a MBD also positively correlated with the traits (Fig. 5). In agreement, the down-regulation of some Arabidopsis MBD genes results in developmental defects comparable to the down-regulation of genes with a role in chromatin remodeling and RNA-mediated silencing (Berg et al., 2003; Peng et al., 2006). Also, two ARGONAUTE (AGO) genes with homology to Arabidopsis AGO4 involved in RNA-directed DNA methylation (Zilberman et al., 2004) showed a positive correlation with leaf size traits and timing-related traits (Fig. 5). In addition to the co-variant histone modifications, chromatin remodeling also depends on ATP-dependent chromatin remodeling complexes that move, eject, or restructure nucleosomes. For instance, the transcription of GRFs is regulated by recruitment of the SWITCH/SUCCROSE NONFERMENTING (SWI/SNF) chromatin-remodeling complexes to their promoters by the transcriptional coactivator AN3 (Vercruyssen et al., 2014; Nelissen et al., 2015). The phenotypes upon differential expression of AN3, BOMBAH, or SWI/SNF-ASSOCIATED PROTEIN73B, subunits of the SWI/SNF complex, reflect the importance of this chromatin-remodeling complex in the regulation of leaf growth (Farrona et al., 2004; Horiguchi et al., 2005; Vercruyssen et al., 2014), and constitutive overexpression of GRF1 and GRF2 also increased leaf and cotyledon size (Kim et al., 2003). In agreement, the subset of 56 TFs enclosed two GRFs that displayed a positive correlation with leaf size- and timing-related traits (Fig. 5).

More than one-third of the 56 TFs that were correlated with leaf-related traits in the two populations had homologs with a known role in growth, while most of the other TFs had an unknown function. We primarily identified genes required for hormone signaling, leaf architecture, and chromatin remodeling, next to TFs involved in other processes, such as Suc signaling (homeobox TF family; Hanson et al., 2001) and cell cycle regulation (Trihelix TF; Breuer et al., 2012).

Protein Synthesis Is Negatively Correlated to Leaf Growth

In the functional category protein synthesis, all 26 genes shared between the two populations are ribosomal proteins. Ribosomes provide the basis for protein production, which is essential to sustain cell growth. Therefore, the majority of the genes that encode ribosomal proteins are highly expressed at the base of the leaf in the DZ (Li et al., 2010), and all selected ribosomal proteins in our data set, except for the chloroplast ribosomal proteins that showed the highest expression toward the elongation zone, were specifically expressed in the DZ, as determined by the Bio-Analytic Resource expression viewer (Fig. 6). The energy status of the cell tunes the transcription of ribosomal RNA and the production and maintenance of ribosomal proteins, which are substantial indirect costs of protein synthesis. Since protein synthesis is one of the most energy-consuming processes in the cell, one can expect selective pressure to achieve a frugal use of the translational machinery (Perry, 2007). Thus, a more efficient translational machinery that minimizes the indirect costs of protein synthesis is likely to allow a higher rate of cell proliferation and growth (Lempiaenen and Shore, 2009; Piques et al., 2009), a possible explanation for the anticorrelation we observe between transcript levels of genes encoding for ribosomal proteins and leaf size and timing traits.

The importance of the translational machinery is reflected in the effect that many mutations in ribosomal proteins have on leaf development (for review, see Byrne, 2009; Horiguchi et al., 2012). For instance, Arabidopsis RPS13A encodes a 40S ribosomal protein S13 involved in early leaf development. An insertion mutant of RPS13A gives rise to a whole range of phenotypic abnormalities, including an altered leaf shape of the first leaves and an increased number of leaves compared with the wild type (Ito et al., 2000). In our data set, transcript levels of the homologous maize gene GRMZM2G158034 were anticorrelating with the leaf size and timing traits in both populations (Fig. 6). Also, the expression levels of several chloroplast ribosomal proteins were anticorrelated with the leaf size traits (Fig. 6). Many of the nucleus-encoded components of the plastid ribosomes are essential for plant growth and development, since their absence results in diverse phenotypic effects, such as embryo lethality, paleness, and reduced overall sizes (Magnard et al., 2004; Asakura et al., 2012; Romani et al., 2012). The Arabidopsis TF CHLOROPLAST IMPORT APPARATUS2 (CIA2) was shown to up-regulate the expression of genes encoding chloroplast ribosomal proteins to accomplish the high protein demands of chloroplasts (Sun et al., 2009). Three of the eight chloroplast ribosomal proteins in our data set were homologs of the Arabidopsis genes up-regulated by CIA2. Also, the maize homolog of CIA2, GRMZM2G062885, was negatively correlated with leaf size traits (Fig. 4).

Leaf Growth Is Strongly Correlated with Cell Wall Characteristics

To allow cells to grow, cell wall expansion is indispensable, and this was also reflected in the overrepresentation of the functional category cell wall synthesis and degradation for genes with expression levels correlated with leaf size traits. The BioArray tool revealed that the 15 correlating genes in this functional category
are all highly expressed in the DZ (Li et al., 2010). Five of these genes are involved in cellulose synthesis, of which three were positively correlated with the traits while two showed anticorrelation (Fig. 7). Expression levels of the catalytic subunits of the cellulose synthase complex \textit{CELLULOSE SYNTHASE3} (ZmCESA3; GRMZM2G039454), ZmCESA5 (GRMZM2G111642), and GRMZM2G099101 were positively correlating with leaf size traits (Fig. 7). Mutations in the respective Arabidopsis orthologs \textit{CESA1}, \textit{CESA3}, and \textit{KORRIGAN} (KOR), which interact with specific CESA complexes (Mansoori et al., 2014), cause an abnormal plant morphology and severe dwarfism due to defects in cell wall formation (Sato et al., 2001; Beeckman et al., 2002; Persson et al., 2007). On the other hand, expression levels of GRMZM2G027794 and GRMZM2G074792, which show homology to the CESA-like C family in Arabidopsis, were negatively correlated with leaf size traits (Fig. 7). This family is most likely involved in the biosynthesis of the glucan backbone of the hemicellulosic polysaccharide xyloglucan rather than cellulose (Liepman and Cavalier, 2012). Lower amounts of xyloglucans in growing tissue possibly increase the accessibility of the primary cell wall for cell wall-loosening enzymes to promote cell wall extension and cell expansion (Pauly et al., 2001). In addition, four genes, GRMZM2G165357, GRMZM2G007404, GRMZM2G328500, and GRMZM5G862540, involved in the biosynthesis of UDP-Xyl (Harper and Bar-Peled, 2002; Kärkönen et al., 2005; Reboul et al., 2011), a nucleotide sugar required for the synthesis of diverse plant cell wall polysaccharides including xyloglucan, had expression levels in the DZ that also negatively correlated with the leaf size traits (Fig. 7).

In addition to cell wall biosynthesis genes, genes involved in cell wall degradation were also identified. Expression levels of the cell wall-loosening enzyme endo-1,4-β-glucanase/cellulase (GRMZM2G331566) were positively correlated with LL and timing-related traits in our data set (Fig. 7). In accordance, it was demonstrated in transgenic Arabidopsis plants that the expression of poplar (\textit{Populus} spp.) \textit{CELLULASE1} resulted in enhanced growth rates (Park et al., 2003). Expression levels of β-\textit{EXPANSIN4}, another cell wall-loosening enzyme, were negatively correlated with leaf size traits (Fig. 7), although up- and down-regulation of expansin expression in Arabidopsis and rice resulted in larger and smaller leaves, respectively (Cho and Cosgrove, 2000; Choi et al., 2003; Goh et al., 2012). However, more pleiotropic phenotypes were observed after constitutive modification of expansin gene expression, including a reduction in overall plant growth (Rochange and McQueen-Mason, 2000), demonstrating that the effect of expansin expression on leaf growth is

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure6}
\caption{Ribosomal proteins with expression profiles correlating or anticorrelating with at least one of the traits in both populations. Bright red indicates positive correlation in the biparental population, dark red indicates positive correlation in the MAGIC population, dark green indicates negative correlation in the biparental population, and light green indicates negative correlation in the MAGIC population. Asterisks indicate genes that are homologs of the Arabidopsis chloroplast ribosomal genes up-regulated by \textit{CIA2} (Sun et al., 2009). DZ-specific genes are indicated in boldface (Li et al., 2010). DW, Dry weight; FW, fresh weight.}
\end{figure}
context dependent and influenced by the growth phase of the leaf (Sloan et al., 2009).

Intriguingly, for many of the cell wall-related genes that were negatively correlated with leaf and shoot size traits in our data set, down-regulation of these genes or their orthologs has been shown previously to result in a growth reduction (Sato et al., 2001; Roberts et al., 2004; Rautengarten et al., 2011) or even embryo lethality (Goubet et al., 2003). We hypothesize that the difference between the direction of the correlation in our data set and the phenotypes of Arabidopsis homologs might be due to the comparison of more subtle variations in expression because of allelic effects in natural variants with more drastic effects of knockout or constitutive overexpression often in one genetic background in Arabidopsis. Most likely, to obtain growth-promoting effects, more subtle changes in expression levels should be applied during specific phases of development and/or by combinatorial perturbation of several genes.

Toward a Robust Growth Regulatory Network

Of the 226 genes we identified with expression levels in the DZ (anti)correlating with at least one of the traits in both populations, some were described previously to be associated with leaf size and growth. However, the majority of these genes was not identified until now as linked to growth, and 48 of these genes even had no assigned function. To obtain a better insight into the putative coregulation of these 226 genes, we used CORNET Corn (De Bodt et al., 2012) to identify networks of coexpressed genes based on the expression of these query genes in two publicly available expression compendia. Additionally, protein-protein interactions based on experimental and computational data (primarily inferred from Arabidopsis) were added to this network.

The resulting network incorporates 185 genes and 943 edges (Fig. 8). Of the other 41 genes, for seven no reliable probe sets were identified on the arrays used to generate the CORNET data sets, while for 34 genes no coexpression or interaction links were found.

Functional information and correlation to the three trait groups were visualized in the network for each gene, and four subnetworks were determined using CytoCluster (Fig. 8). Two of these subnetworks were enriched in the functional category protein synthesis. ribosomal proteins; subnetwork 1 contained genes anticorrelating primarily with the timing traits, and subnetwork 4 was primarily anticorrelated with leaf size traits (Fig. 8). The genes within the two subnetworks were highly connected to each other and to a limited set of other genes. For subnetwork 1, one of these other genes, GRMZM2G038032, shows homology to Arabidopsis RECEPTOR FOR ACTIVATED C PROTEIN KINASE1 (RACK1) genes involved in ribosome biogenesis (Guo et al., 2011); it was shown that loss-of-function mutations in RACK1 genes in Arabidopsis severely affect rosette leaf production and root growth (Guo and Chen, 2008). The ribosomal genes in this first subnetwork were also highly connected to GRMZM2G067877, a homolog of an Arabidopsis mitochondrial adenine nucleotide transporter (ADNT1) that plays a role in energy supply in heterotrophic tissues. This protein catalyzes the exchange between cytosolic AMP and intramitochondrial ATP, and plants with decreased ADNT1 expression show an inhibition of root growth (Palmieri et al., 2008). A third gene that was coexpressed with the ribosomal genes in subnetwork 1 is GRMZM2G389768, a putative COLD SHOCK DOMAIN (CSP) gene. In Arabidopsis it was shown that CSP1 associates with polysomes by binding to a specific set of mRNAs and acts as a chaperone under low temperatures. CSP1 primarily binds mRNAs encoding for energy-consuming processes, such as ribosome...
biogenesis, to maintain the translation of these specific mRNAs under stress conditions (Juntawong et al., 2013). Thus, this first subnetwork of coexpressed genes seems to be linked to energy consumption and supply, suggesting that the duration of leaf growth is associated with energy availability, as the genes in this subnetwork are anticorrelated with timing traits.

The network contained a second subset of strongly coexpressed genes encoding ribosomal proteins and a limited number of other negatively correlated genes, subnetwork 4 (Fig. 8). In contrast to subnetwork 1, these genes showed anticorrelating primarily to the leaf size traits (Fig. 8). Of the 31 nodes in this subnetwork, nine were ribosomal proteins all targeted to the chloroplasts. For most of these genes, the effect of perturbed expression was not analyzed before now (Tiller and Bock, 2014). However, two of the Arabidopsis orthologs of these ribosomal proteins, RPL21C and RPL3, were shown to be essential for plastid development and embryogenesis (Yin et al., 2012; Tiller and Bock, 2014), while reduced gene expression of the plastid-specific ribosomal protein PSRP5 resulted in severely delayed plant growth due to reduced plastid translation (Tiller et al., 2012). Additionally, this subnetwork contained one GATA TF (GRMZM2G039586) homologous to Arabidopsis GCN, described to function in chloroplast development, likely by regulating chloroplast division (Chiang et al., 2012). Interestingly, the majority of the other genes in this subnetwork were chloroplast targeted and involved in chloroplast development and/or function. We identified genes functioning in chloroplast biogenesis, for instance tetrapyrrole biosynthesis genes (Huang et al., 2009; Quesada et al., 2013) and the FZO-like protein FZL coding for a dynamin-related membrane-remodeling protein (Gao et al., 2006); genes with metabolic functions in chloroplast coding, for instance, for components of the NAD(P)H dehydrogenase complex (Ishida et al., 2009) essential for the supply
of ATP for photosynthesis (Endo et al., 2008); and genes involved in plastid gene expression, such as the ribosomal proteins, components of the plastid-encoded RNA polymerase complex, the pkB-type carbohydrate kinase FRUCTOKINASE-LIKE PROTEIN (Gilkerson et al., 2012), MurE (Garcia et al., 2008), and PLASTID TRANSCRIPTIONALLY ACTIVE CHROMOSOME6 (Pfalz et al., 2006), CIA2 (Sun et al., 2009), genes involved in chloroplast protein import (Chiu and Li, 2008), and pentatricopeptide repeat proteins (Liu et al., 2013a). As a consequence, many mutants in these genes are chlorotic, white, or pale green due to abnormal plastid development and reduced photosynthetic competence. However, evidence for the impact of plastid development and plastid protein synthesis on leaf development and morphology independent of photosynthetic capacity has been suggested for a long time (for review, see Tiller and Bock, 2014).

Since the genes with a role in leaf development that are currently described are nucleus encoded, it seems likely that this effect of plastid translation on leaf development is the consequence of plastid-nucleus communication known as chloroplast retrograde signaling (Larkin, 2014). Recently, it was shown that chloroplast retrograde signaling regulates the spatial expression levels of genes involved in the expansion of leaf primordia, probably to avoid the expansion of leaves with reduced photosynthesis due to impaired plastid activity (Tameshige et al., 2013). Furthermore, in Arabidopsis, it was postulated that chloroplast retrograde signaling is an important regulator of the onset of cell expansion and photosynthesis (Andriankaja et al., 2012). Although the chloroplasts are still undifferentiated in the most basal part of the DZ we sampled for RNA sequencing (Majeran et al., 2010), it can be assumed that signaling between plastids and the nucleus starts early on, as plastid biogenesis is highly coordinated with organ development to ensure the presence of the appropriate number of plastids of the correct type in each stage of development (Terry and Smith, 2013).

Subnetwork 4 with nucleus-encoded genes targeted to the chloroplasts and functioning in chloroplast development, and some possibly also in chloroplast retrograde signaling (Strand et al., 2003; Garcia et al., 2008; Gilkerson et al., 2012), showed a variation in expression levels in both RIL populations. This expression variation in the DZ between RILs and the anticorrelation with leaf size traits suggests an early coordination of extrachloroplast processes with chloroplast function via chloroplast retrograde signaling.

Besides the two highly interconnected subnetworks of genes encoding ribosomal proteins, two other subnetworks were identified. Subnetwork 2 consisted of 22 nodes, primarily anticorrelated with leaf size traits (Fig. 8). This subnetwork was enriched in cell wall biogenesis and degradation: six of the 22 genes in this subnetwork were involved in this process, and all negatively correlated with leaf size or timing traits. These cell wall-related genes as well as other genes of subnetwork 2 were coexpressed with GRMZM2G045534, an ATP-citrate lyase whose down-regulation of Arabidopsis orthologs results in pleiotropic phenotypes, including reduced size (Fatland et al., 2005). Also, three TFs were highly coexpressed with the genes in subnetwork 2: GATA TF (GRMZM2G101058) described above, a bZIP TF (GRMZM2G052102) of which the Arabidopsis orthologous mutant shows cell wall defects in pollen (Gibalová et al., 2009), and a TRIHELIX TF (GRMZM2G314660). The latter shows homology to Arabidopsis GTLI, a master regulator that negatively regulates cell growth (Breuer et al., 2012). Other examples of genes in this subnetwork that were already functionally annotated were ZmPIN1b, an auxin efflux carrier (Carraro et al., 2006), and Zmβ-GLUCOSIDASE1, involved in the (in)activation of cytokinins (Brzobohaty et al., 1993). Subnetwork 3 consisted of 45 genes that were primarily positively correlated to leaf size and timing traits (Fig. 8). This subnetwork was enriched for two functional categories, regulation of transcription (12 genes) and cell wall synthesis and degradation (three genes). The node with the most edges in this subnetwork was a cell-wall-related gene, KOR (GRMZM2G099101), and was coexpressed with eight TFs: three bHLH TFs, GRMZM2G074438, GRMZM2G128807, and GRMZM2G158281, the latter showing orthology to Arabidopsis LONESOME HIGHWAY, which is required for establishing and maintaining the normal vascular cell number and pattern in roots (Ohashi-Ito and Bergmann, 2007); two SET domain TFs, GRMZM2G409224 and GRMZM2G318803, the latter being a homolog of Arabidopsis SUVR4; E2F/DP (GRMZM2G462623) and GRF (GRMZM2G099862), both having homologs in Arabidopsis with known roles in leaf development (De Veylder et al., 2002; Kim et al., 2003); and an AGO gene with homology to Arabidopsis AGO4 (GRMZM2G589579). Other genes coexpressed in this subnetwork with know annotations include the cell cycle gene CYCLIN A2 (GRMZM2G017081), involved in vein development in Arabidopsis (Donner and Scarella, 2013), and BRASSINOSTEROID SIGNALING KINASE2 (GRMZM2G054634), affecting growth through its role in the initial steps of brassinosteroid signal transduction in Arabidopsis (Sreeramulu et al., 2013). Still other genes in this subnetwork were primarily linked to metabolic processes and/or are not well described to date.

Within the set of 226 genes identified in the two-way and eight-way RIL populations, we found 185 genes to be coexpressed and four subnetworks could be distinguished, of which two were enriched in the functional category protein synthesis. Ribosomal proteins. Of the 185 genes, 32 were TFs, and for some a function in leaf development and growth had already been described (see above). Coexpression relationships between these TFs and other genes may suggest potential regulatory influences; however, clear causal links between these TFs and potential targets cannot be established from the coexpression network alone, and further analyses are necessary to confirm these links.
CONCLUSION

In this study, we report an integrated analysis of transcript measurements in proliferative tissue and in-depth phenotyping of leaf size parameters of 197 RILs belonging to two different maize segregating populations to establish how variation in transcript levels relates to phenotypic changes. A set of 226 genes was identified for which the expression levels in the DZ correlated significantly with final leaf size and timing or shoot traits in both populations, a strong reduction compared with the 1,740 and 1,364 (anti)correlating genes identified separately in the biparental and MAGIC populations, respectively. Thus, combining data from different populations led to the identification of a limited set of genes, including 56 TFs, that may play a more general role in the molecular networks underlying the establishment of the various phenotypic traits across maize accessions. Of these 226 correlated genes, 185 were present in a highly connected network of coexpressed genes constructed from public expression compendia. The occurrence of several genes known to affect leaf size in this correlated gene list confirms that, already in the dividing cells of growing leaves, the molecular networks that will establish final leaf phenotypes are present. Besides these known genes, genes without a clear function or with no known role in leaf growth were also identified. In summary, our comparative analysis of transcriptome and phenotype variation in two RIL populations supports the conclusions reported earlier for one of the populations (Baute et al., 2015), extends our understanding of system-level processes that are active in leaf growth across different populations, and resulted in the identification of a set of interesting candidate genes for future follow-up studies.

MATERIALS AND METHODS

Plant Growth and Sampling

In this study, a subset of lines of two maize (Zea mays) RIL populations was used. A subset of 103 lines of a biparental RIL population derived from a cross between parental lines B73 and H99 (Marino et al., 2009) and a subset of 94 lines of a MAGIC maize population (Dell’Acqua et al., 2015) was phenotyped and sampled for RNA sequencing. The subset of 94 lines was chosen randomly from the set of 529 lines that was genotyped and phenotyped in the biparental and MAGIC populations separately. As described before (Baute et al., 2015), for each trait in two populations, the q0.01 and q0.99 quantiles of the set of transcript-trait PCC values, respectively, referred to as the q0.01,random and q0.99,random, were taken as the reference correlation coefficients expected by chance across all RIL samples of one population. The permutation of each trait was conducted in a randomized design each time experiments in a single growth chamber. For each population, the q0.01 and q0.99 quantiles of the set of transcript-trait PCC values were calculated and compared with a reference correlation coefficient. These reference correlation coefficients were determined by permuting the trait values 1,000 times and calculating the q0.01 and q0.99 quantiles for each permutation. The mean q0.01 and q0.99 quantiles after permutation, q0.01,random and q0.99,random, were taken as the reference correlation coefficients expected by chance across all RIL samples of one population. The permutation of each trait was conducted in R software using the coin function for the calculation of PCC and the sample function for trait permutation. We focused on the genes with correlation coefficients lower and higher than the (real) q0.01 and q0.99 quantiles, respectively, referred to as the anticorrelating and correlating gene sets, respectively.

Transcriptome Analysis

RNA sequencing analysis was performed as described before, resulting in a set of 15,051 retained transcripts (Baute et al., 2015). In brief, after library preparation with the TruSeq RNA Sample Preparation Kit version 2 (Illumina), the quality of the raw data was verified with FastQC version 0.9.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, quality filtering was performed using FASTX-Toolkit version 0.10.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were subsequently mapped to the maize B73 reference genome (5b) using GSNAP (Wu and Nacu, 2010), allowing maximally five mismatches. The concordantly paired reads that uniquely map to the genome were used for quantification on the gene level with htseq-count from the HTSeq.py python package (Anders et al., 2015). To avoid artifacts in the mapping and consequently in transcript quantification because of diversity in the maize inbred lines (Morgante et al., 2007), we selected genes conserved in the eight parental lines of the MAGIC population (Dell’Acqua et al., 2015). Therefore, RNA sequencing data of proliferative tissue for these eight inbred lines were mapped to the B73 reference genome. A coverage cutoff was applied, using the R/Bioconductor package with default HTSFilt parameter settings (Rau et al., 2013), retaining 19,948 genes that were expressed in at least one of the parents. Next, single-nucleotide polymorphism calling was performed, and we selected genes with no more than 1.75% single-nucleotide polymorphisms, resulting in a set of 15,051 genes.

RNA sequencing count data were subsequently normalized for library size with the default normalization method in the DESeq2 package version 1.2.10 (Love et al., 2014). Next, transcripts expressed (nonzero count) in less than 5% of samples were removed. On the remaining transcripts, the inverse hyperbolic sine transformation was applied with the asinh function in R software. Additionally, another filter was applied that removed the 5% least varying transcripts, based on the coefficient of variation, resulting in the selection of 14,255 transcripts in the biparental population and 14,297 transcripts in the MAGIC population.

Correlation Analysis

PCCs among the traits were calculated on the means of the lines in SPSS (SPSS), separately for the two populations. For correlation analysis between transcript levels and trait variation, PCCs were calculated between each transcript and all traits over all RILs for the two populations separately. As described before (Baute et al., 2015), for each trait in each population, the mean q0.01 function of the set of transcript-trait PCC values were calculated and compared with a reference correlation coefficient. These reference correlation coefficients were determined by permuting the trait values 1,000 times and calculating the q0.01 and q0.99 quantiles for each permutation. The mean q0.01 and q0.99 quantiles after permutation, q0.01,random and q0.99,random, were taken as the reference correlation coefficients expected by chance across all RIL samples of one population. The permutation of each trait was conducted in R software using the coin function for the calculation of PCC and the sample function for trait permutation. We focused on the genes with correlation coefficients lower and higher than the (real) q0.01 and q0.99 quantiles of the set of transcript-trait correlation coefficient values, respectively, referred to as the anticorrelating and correlating gene sets, respectively.
Principal component analysis was performed as a dimensionality reduction technique on the centered and scaled phenotype data, using the prcomp function in R.

To determine whether the number of (anti)correlating genes in the intersection of both populations was higher than expected by chance, we used hypergeometric tests.

Visualization of Expression Patterns

The expression patterns of the set of 226 genes (anti)correlating with at least one trait in both populations and the phenotypic measurements for all RILs were visualized in MeV (Saeed et al., 2003). Data were adjusted by normalizing the genes/row, and color scale limits were set at −3 and +3 as lower limit and upper limit, respectively, since these numbers approached the minimal and maximal data values after normalization. Next, we performed hierarchical clustering of the RILs, based on Euclidean distance and average linkage clustering, to visualize the gradient in expression levels and phenotypic measurements over all RILs.

Functional Enrichment Analysis

Functional enrichment analyses were based on pathways defined in MapMan (Thimm et al., 2004) and performed as described before (Baute et al., 2015).

Network Analysis

Coexpression analysis between the 226 genes for which expression levels in the DZ (anti)correlated with at least one of the traits in both populations was performed with the online tool CORNET (De Bodt et al., 2012), but with its in-house version, where the probe sets of publicly available Affymetrix and Nimblegen expression data sets for maize are mapped to gene models B73_RefGen_v2. For our analysis, we selected both expression data sets, Affymetrix and Nimblegen. The generated coexpression network was based on PCCs of 0.7 and higher, and protein-protein interactions between query genes based on both experimental and predicted data from CORNET were included. Functional enrichment analyses were based on pathways described in Supplemental Table S3. Functional Enrichment Analysis

Visualization of Expression Patterns

The expression patterns of the set of 226 genes (anti)correlating with at least one trait in both populations and the phenotypic measurements for all RILs were visualized in MeV (Saeed et al., 2003). Data were adjusted by normalizing the genes/row, and color scale limits were set at −3 and +3 as lower limit and upper limit, respectively, since these numbers approached the minimal and maximal data values after normalization. Next, we performed hierarchical clustering of the RILs, based on Euclidean distance and average linkage clustering, to visualize the gradient in expression levels and phenotypic measurements over all RILs.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. PCA analysis of the phenotype data.

Supplemental Figure S2. Histograms of the leaf size and seedling related traits in the bi-parental RIL and MAGIC RIL population.

Supplemental Figure S3. Enrichment of functional categories for correlating and anti-correlating genes in the bi-parental and MAGIC RIL populations.

Supplemental Table S1. Mean, maximum, minimum and percentage difference of the traits within the two populations.

Supplemental Table S2. Pearson CC for the analyzed traits, grouped as leaf size, shoot related and timing related traits combined for bi-parental and MAGIC RIL populations.

Supplemental Table S3. Genes with expression levels (anti-)correlating with at least one of the traits in one of the populations.

Supplemental Table S4. Phenotypic measurements for the MAGIC population.

ACKNOWLEDGMENTS

We thank Dr. Annick Bleys for help with article submission and Dr. Michel Van Bel for help in updating CORNET Corn.

Received December 2, 2015; accepted January 7, 2016; published January 11, 2016.

LITERATURE CITED


Chiu CC, Li HM (2008) Ti40 is important for reinsertion of proteins from the chloroplast stroma into the inner membrane. Plant J 56: 793–801


Copyright © 2016 American Society of Plant Biologists. All rights reserved.
dynamics of maize embryonic leaves during seed germination. Proc Natl Acad Sci USA 110: 3979–3984

Copyright © 2016 American Society of Plant Biologists. All rights reserved.


