The transcription regulatory code of a maize leaf

Xiaoyu Tu[1,2,6], María Katherine Mejía-Guerra[3,6,*], Jose A Valdes Franco[4], David Tzeng[2], Po-Yu Chu[2], Xiuru Dai[1], Pinghua Li[1,*], Edward S Buckler[3,4,5], Silin Zhong[2,*]

[1]College of Agronomy, Shandong Agricultural University, China.
[2]State Key Laboratory of Agrobiotechnology, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
[3]Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA
[4]School of Integrative Plant Sciences, Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA
[5]Agricultural Research Service, United States Department of Agriculture, Ithaca, NY, USA
[6]These authors contributed equally to this work.
[*]Correspondence: mm2842@cornell.edu (M.K.M.); pinghuali@sdau.edu.cn (P.L.); silin.zhong@cuhk.edu.hk (S.Z.)

SUMMARY

The transcription regulatory network underlying essential and complex functionality inside the cell is defined by the combinatorial actions of transcription factors (TFs). Although maize is an important crop and has a long history as a model for genetic research, its regulatory circuit is largely unknown. Here, we developed a scalable ChIP-seq approach that enabled us to determine the binding profiles of 104 maize leaf TFs, and identified 2,147,346 TF binding sites genome-wide. TF binding occurs in clusters covering ~2% of the genome, and shows enrichment for sequence variations associated with complex agronomic traits and eQTLs. We have constructed a regulatory network for the maize leaf that covers over 70% of the expressed genes, which has scale-free topology and functional modularity. Machine-learning analyses were used to identify TF sequence preferences, and showed that co-binding is key for TF specificity. Furthermore, the trained models were used to predict and compare the regulatory networks in other species and showed that the core network is evolutionarily conserved. Together, our results provide an extensive description of the architecture, organizing principle and evolution of the leaf transcription regulatory network.

INTRODUCTION

For all living cells, a basic control task is to determine gene expression throughout time and space during developmental processes and responses to stimuli (Sorrells and Johnson, 2015). The code that governs gene expression is stored in the non-coding region of the genome. Sequence-specific DNA binding proteins known as transcription factors (TFs) can read the code by binding to cis-regulatory elements to activate or repress gene expression. Each gene can receive instructions from multiple TFs, and each TF can target thousands of genes, forming a transcription regulatory network, which underlies essential and complex functionalities inside the cell (Neph et al., 2012). Hence, it is critical to study the mechanism, architecture and behaviour, as well as conservation and diversification of the regulatory network in different species.

The budding yeast has probably the best mapped transcription regulatory network and is a unicellular model for eukaryote, with around 200 TFs (Harbison et al., 2004; Lee et al., 2002). For multicellular organisms with large genomes and thousands of TFs, complete network reconstruction has proven to be a formidable task. For example, the ENCODE consortium has mapped the binding of 88 TFs in five cell lines to study its architecture and dynamics (Gerstein et al., 2012), while a more complete regulatory network was constructed in human colorectal cancer cells with data from 112 TFs (Yan et al., 2013). But similar efforts are seldom feasible for individual laboratories, and have yet to be attempted in plants.

Maize is one of the most economically important crops worldwide. It is also the best-studied and most tractable genetic system among the cereal crops, making it an ideal model for studying this group of important crop plants (Schnable et al., 2009). It has been shown that the maize non-coding region harboring the cis-regulatory elements is a major contributor to the observable phenotypic differences and adaptation. For example, 70% of maize GWAS hits are located in the non-coding regions that has yet been functionally annotated (Wallace et al., 2014). Recently, it was shown that open chromatin sequence variations within these non-coding regions could explain up to 40% of the phenotypic variations of key agronomic traits (Rodgers-Melnick et al., 2016). However, it is difficult to assess their true functions without knowing the precise locations of the cis-regulatory elements and the TFs that recognize them. Despite the importance of TF binding data, ChIP-seq experiments in plants are limited by antibody availability and difficulties in transforming crops to express the epitope fusion protein. As a result, TF binding information is only available for a handful of maize TFs, and even in the model species like Arabidopsis and rice, such studies are too few in number to produce a general picture of the plant transcription regulatory network.

In this study we have developed a high-throughput ChIP-seq approach to map the binding of 104 TFs expressed in the maize leaf. We found that similar to the mammalian ones, plant TF binding are highly clustered, overlap with open chromatin regions, show enrichment of GWAS hits and eQTL, and are evolutionarily conserved. We constructed a high-confidence regulatory network that display scale-free topological features, and trained machine-learning models to predict TF binding from sequence, co-localization information, and annotate non-coding regions in other

grasses. Our findings provide a first glimpse into the plant transcription regulatory network and reveals its architecture, organizing principle and evolutionary trajectory.

RESULTS

TF ChIP-seq using maize leaf protoplast

To transform maize with TF constructs for ChIP-seq, we developed an efficient protoplast isolation and transformation method (Figure 1A; STAR Methods). We found that the high mortality rate and low transformation efficiency associated with the conventional methods are due to excessive wounding of the cells during the isolation. Instead of cutting the leaf into thin slices, we peeled the lower epidermis to allow the cell wall digesting enzymes to gently penetrate the leaf and release the protoplasts. Using this method, we can obtain ~10^7 intact mesophyll cells from the leaves of a 9-day-old seedling, and achieved over 90% transformation efficiency estimated by the GFP spike-in control.

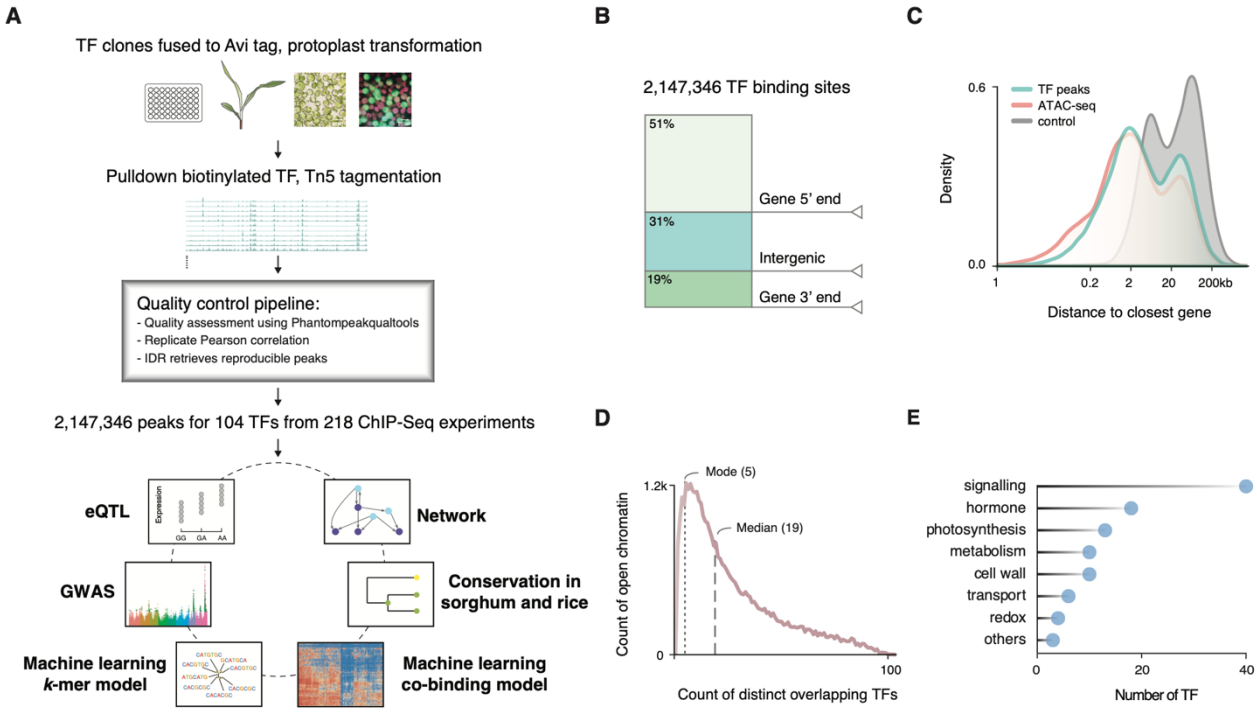

Figure 1. The regulatory landscape of the maize leaf

(A) Overview of the experiment and analytical approaches.
(B) Genome-wide distribution of TF binding loci.
(C) Density plots corresponding to distances between annotated genes (grey), distances of TF binding loci (green) and open chromatin regions (pink) to closest annotated gene after excluding regions within or overlapping with gene bodies.
(D) Distribution of the count of distinct TFs that overlap with open chromatin regions.
(E) Functional classification of the 104 TF based on their target genes.

In addition, we fused the TF coding sequence to the Avi epitope tag, which can be biotinylated by the co-expressed bacterial biotin ligase BirA. The biotinylated TF can then be pull-down directly using streptavidin magnetic beads, without the need for antibodies. It also has high signal to noise ratio due to the strong affinity between streptavidin and biotin that permits the use of stringent washes under denaturing conditions. To obtain large quantities of ultrapure TF-Avi fusion DNA for transformation, we also developed a plasmid preparation method using low cost sand powder as binding matrix and detergent to remove endotoxin without expensive cation exchange chromatography.

Transposase Tn5 tagmentation is one of the most convenient library preparation methods available (Picelli et al., 2014). However, homebrew Tn5 tends to lose activity rapidly due to protein aggregation. To improve the reproducibility, we fused the Tn5 to the C-terminal of the *E. coli* elongation factor Ts, which is often used as N-terminal tag in the purification of aggregation prone proteins (Han et al., 2007). This hyper-stable Tn5 transposase enabled us to significantly improve the consistency and reduce the cost of ChIP-seq library construction. Using this method, we have successfully generated 218 ChIP-seq profiles for 121 TFs that are expressed in the developing section of the maize leaf based on previous RNA-seq data (Dong et al., 2017) (Figure 1A; Data S1).

TFs bind to genome in dense clusters

From the published mammalian TF ChIP-seq, we anticipated that plant TF binding might also form dense clusters and frequently locate within open chromatin regions (Yan et al., 2013). First, we used the ENCODE2 uniform pipeline to process the ChIP-seq data (Landt et al., 2012) (Figure 1A; Figure S1A). We identified a total of 2,147,346 reproducible TF binding peaks, and the number of peaks varies between TFs, with a median value of ~16K (Interquartile range, between 7,664-32,566 peaks, Figure S1B). These TF binding peaks clustered in 144,890 non-overlapping loci covering ~2% of the genome (Data S2). Similar saturation of TF binding clusters was observed when sufficiently large number of TFs expressed in colorectal cancer cells have been sampled (Yan et al., 2013). Next, we measured chromatin accessibility in the same tissue using ATAC-seq, and identified 38,713 biologically reproducible open chromatin regions. We found that the TF binding loci and open chromatin regions show similar genome-wide distributions, frequently proximal to gene bodies (+/- 2.5kb), with preferences for the 5' end (Figures 1B and 1C).

We also found that the distance between TF binding site to gene after excluding regions that overlap with gene body is bimodal (Figure 1C). Despite the larger space available for distal regulation, we observed that regions between 10-100kb constitute ~15% of the TF binding loci, and ~17% of the open chromatin regions. Layering ATAC-seq and ChIP-seq data showed that TF binding loci and open chromatins overlap (Figure S1C, P-value $< 10^{-5}$). On average, ~74% of the peaks for a given TF (IQR_{25-75} 64%-87%) intersect with open chromatin regions (Figure S1D) confirming the relevance of the identified TF binding sites within the chromatin context. Collectively, 98% of the open chromatin regions overlap with TF peaks (Figure 1D), with a mode of 5, and a median of 19 distinct TFs for each region, suggesting there are a large number of possible TF combinations that co-regulate transcription.

As none of the 104 TFs selected have been previously examined by ChIP-seq, and most of them have not even been functionally characterized, we used GO-term and MAPMAN functional category enrichment analysis to classify them based on their targets (Figure 1E; Data S3). Majorities of the TFs are grouped into signaling, hormone and photosynthesis categories. For example, targets of the maize ZIM TFs show enrichment in the GO terms "response to wounding" and "jasmonic acid metabolism", consistent with the role of their homologs in other plant species (Pauwels and Goossens, 2011). The PIF and GLK TFs targets are enriched for "circadian rhythm", “light harvesting” and "apocarotenoid metabolic process", and were assigned to the photosynthesis category. The *ZmMyb38*/*ZmCOMT1 targets* were associated with terms such as "regulation of flavonoid biosynthetic process", and "phenylpropanoid biosynthetic process" and were assigned to the metabolic group (Yang et al., 2017).

Previous studies showed that the high TF occupancy regions in the genome are often associated with important functions (Boyer et al., 2005; Heyndrickx et al., 2014). We identified 2,037 open chromatin regions at the top 5% of the TF occupancy distribution (Data S4), and their surrounding genes are indeed enriched for regulatory GO-terms (Data S5). Notably, we found that *Vgt1*, the most important QTL for flowering time (Salvi et al., 2007), is associated with a high TF occupancy region bound by 76 TFs, and locates at a distance of ~72 kb from the TSS of *ZmRAP2.7*, a TF-coding gene whose expression is known to be regulated by *Vgt1*. Six of the TFs bound to this region (*i.e.,* PRR5, ELF3, COL3, COL7, COL18 and DOF3/PBF1) have previously been linked to flowering time variations through genetic studies (Alter et al., 2016; Li et al., 2016).

Taken together, our data provide a detailed insight into the biological functions of maize TFs. Similar to the animal genome, plant TF bindings formed dense clusters, and high TF occupancy regions also play an important role in shaping complex phenotypic variation such as flowering time.

TF binding sites show sequence conservation and enrichment for eQTL and GWAS hits

TF binding is a key determinant of transcriptional regulation and, if purifying selection is effective in these regions, they should exhibit low sequence diversity. We examined the conservation of the TF binding sites by assessing the overall nucleotide diversity represented in the maize HapMap (Bukowski et al., 2018) while controlling for overall SNP density in function to the distance of TF’s peak summit (Figure 2A). The result confirmed that sequence variation is, in fact, reduced.

While binding of most TFs is constrained, TF and their binding sites are key to local adaptation or domestication. For example, TF *ZmTB1* is known to play an important role in maize domestication (Doebley et al., 1997). Hence, we predict that TF binding loci could be enriched for common SNP variations controlling gene expression and downstream traits. This was first tested in a panel of 282 inbred breeding lines for their effect on mRNA expression using common and likely adaptive variants (Kremling et al., 2018). We found twofold enrichment of TF binding loci (95% credible interval 2.26-2.46), similar to the enrichment around 5'UTRs (95% credible interval 1.80-2.11) and 3'UTRs (95% credible interval 2.37-2.73) (Figure. 2B). Distal and proximal TF binding loci are

both enriched when examined separately (Figure S2A). The enrichment pattern is ubiquitous across the 104 TFs that we examined (Figure S2B; Data S6). Overall, these support our hypothesis that common variation in mRNA expression is controlled by TF binding site variation.

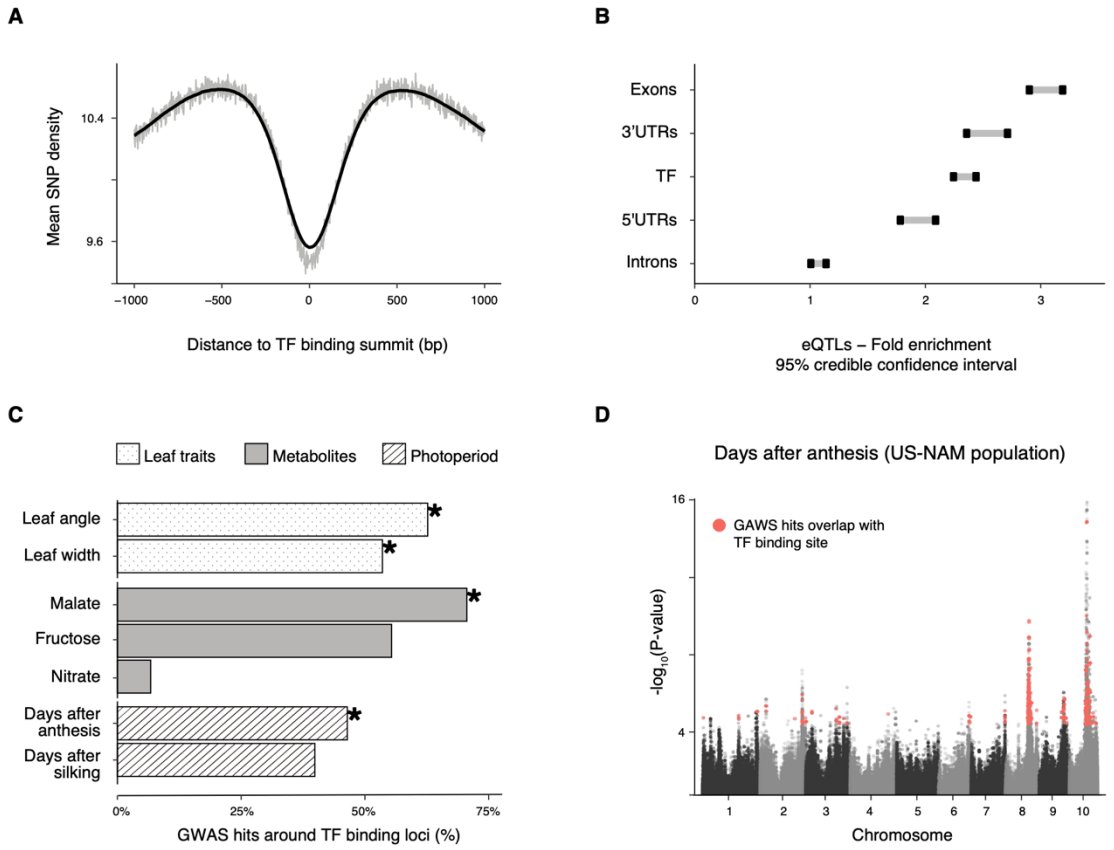

Figure 2. TF binding sites show low sequence diversity and enrichment in functional variation

(A) Mean SNPs density calculated in sliding windows of 100 bp bins flanking TF binding summits, and normalized by mappability rate.
(B) 95% confidence interval for the enrichment of SNPs associated to variation in mRNA levels (eQTLs) vs. non eQTLs SNPs, and relative to control regions for different sets of genomic regions.
(C) Proportion of phenotype-associated GWAS hits for an assortment of traits overlapping to TF binding loci. Traits in which the enrichment was statistically significant are labelled with an asterisk.
(D) Manhattan plot of GWAS for days after anthesis. Highlighted GWAS hits overlap with binding regions for a group of TFs (PRR5, ELF3, COL3, COL7, COL18 and DOF3/PBF1) associated to photoperiod variation.

Next, we tested the overlap with functional variations associated with traits others than gene expression. To this end, we calculated the enrichment in GWAS hits for seven traits related to metabolites (Zhang et al., 2015), leaf architecture (Tian et al., 2011), and photoperiodicity (Buckler et al., 2009) measured in the US NAM population. Overall, TF binding loci are enriched for four of the traits (Figure 2C). We noticed that simple traits, such as metabolites, showed few TFs enriched for GWAS hits (*e.g.* malate and nitrate, Data S7). In contrast, many TF bindings overlap with GWAS hits for complex traits, which are known to be polygenic and influenced by a large number of genetic variants (*e.g.,* days after silking and days after anthesis, Figure 2D; Figure S2C). A further look at TFs enriched in GWAS hits for photoperiodicity revealed that 51% of the TFs enriched in days after anthesis, and 35% in days after silking, also bound to the *Vgt1/ZmRAP2.7* high TF occupancy region, and they could be new regulators of maize flowering time.

Taken together, our observation that TF binding regions are conserved and frequently overlap with sequence variations associated with phenotypic changes meet our initial expectations. The general trend for GWAS enrichment also supports our hypothesis that non-coding region variations related to traits are mediated by TFs. Furthermore, our finding highlighted the potential of using large-scale TF ChIP-Seq data to connect sequence variation in *cis* to *trans*-regulators to highlight the regulatory mechanism implicated in complex phenotypes.

A scale-free transcription regulatory network in plant

Real-world networks such as the social network, mammalian regulatory network and protein-protein interaction network, frequently exhibit a scale-free topology (Barabási and Albert, 1999). To assess the feasibility of our data to pinpoint true regulatory relationships, we tested if the derived network could fit the topology. We reshaped the regulatory data into a graph (Figure 3A), adopting the ENCODE probabilistic framework to identify high confidence proximal interactions (P-value < 0.05) (Cheng et al., 2011; Gerstein et al., 2012; Landt et al., 2012). This approach renders a network with 20,179 nodes (~45% of the annotated genes and ~77% of the leaf expressed genes, Data S8). We evaluated the in-degree distribution (Figures 3B and 3C), or number of edges towards each node, and found it to follow a linear trend in the log-scale (R^2 = 0.882, P-value $< 2.2e-16$), as expected for power-law distribution (goodness of fit P-value = 0.67), which is a landmark of scale-free networks.

In a real-world network, nodes that appear more connected than others are called “hubs”, and they are critical for information flow (Barabási and Albert, 1999). We defined these hub genes as target nodes in the top percentile of the in-degree distribution (99th percentile, for a total of 206 genes, Data S5). Interestingly, we found half of the hub genes were located nearby the TF highly occupied regions. Similar to the result from the analysis of highly occupied regions, hub genes showed enrichment for regulatory functions (*i.e.,* GO term: "*transcription factor activity*", enrichment 4-fold), consistent to the expected role for a hub node in the transcriptional regulatory network.

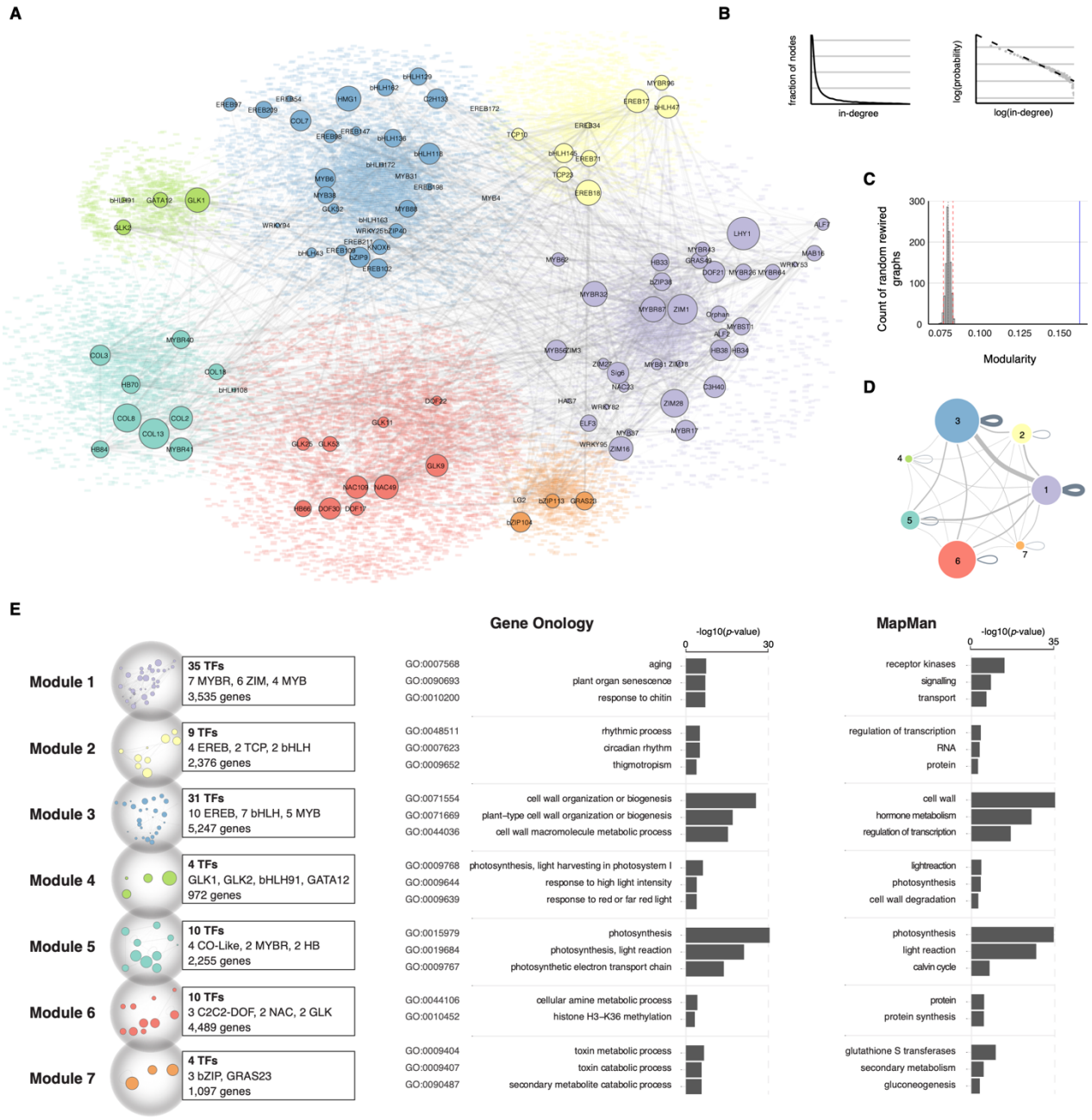

Figure 3. The maize leaf transcription regulatory network exhibit properties of real-world networks

(A) Graph diagram showing the 104 TF and their connections. The seven modules are labelled in different color. TF nodes are shown as circles with size representing gene expression level, and their target genes are shown as color dots in the background.

(B) Distribution of node in-degree values. The in-degree distribution fits the power-law (dashed line).

(C) Modularity of the network (blue line) is higher than the random control (histogram of maximum modularity values of 1000 rewired graphs).

(D) Graph diagram showing TF to gene connections within and between modules. Each module is represented as circle with size proportional to the number of nodes within.

(E) GO-term and MapMan category enrichment analysis of genes in each module.

Structural and functional modularity of the network

Biological networks often exhibit topological and/or functional modularity (Dittrich et al., 2008; Olesen et al., 2007). We tested for this predicted property by contrasting the maximum modularity in our network to a null distribution from an ensemble of random rewired graphs (H_0: 1000 rewired graphs) (Clauset et al., 2004). The result confirmed that our network exhibits a significant increase in modularity (P-value < 0.05, Figure 3C). We identified seven modules, each containing from ~27% to ~5% of the total nodes (Gephi resolution 1, Data S9). It should be noted that these modules are not isolated. Only ~40% of the total edges occurring within each module, suggesting that TFs often regulate targets outside their own modules, and there are large information flows between modules (Figure 3D).

We predicted that topological modularity could be related to function in known biological processes. Hence, we performed GO-term and MapMan functional category enrichment analyses for genes in each module, and found that they were indeed enriched for specific functions (Figure 3E). For example, we found two photosynthesis-related modules. Module 4 is enriched in targets of *GLK1/GLK2*, which are known regulators of photosynthesis (Waters et al., 2009), while model 5 is enriched in targets of the CONSTANT(CO)-like TFs, which are associated with roles in flowering time, circadian clock and light signaling (Wang et al., 2013) (Data S10).

These modules contain thousands of genes from different pathways, and are too large to be assessed as a whole. We hypothesize that as the network topology can already provide clues to biological function at this scale, potential regulators of a smaller pathway might be identified based on connectivity at a local scale. We first tested this in the conserved chlorophyll biosynthesis pathway, which is known to be regulated by two GLK TFs, as their mutations caused disruption of the photosynthesis apparatus gene expression (Rossini et al., 2001; Waters et al., 2009). We calculated the sum of the log transformed p-values that the ENCODE probabilistic model generated for each TF-target interaction, based on binding intensity and proximity, and used it to infer the contribution of each TF to the pathway (Figure 4A). We found that the top regulators are indeed the two GLKs and an unknown MYBR26. Despite the function of the MYBR26 has yet been studied, its Arabidopsis homologs are involved in circadian regulation, confirming our hypothesis (Nguyen and Lee, 2016).

Next, we used this strategy to examine the maize C_4 photosynthesis pathway without well-defined regulator. It turns out that the top five TFs in connectivity ranking are CO-Like (COL) TFs (Figures 4B and 4C). Previous studies of COL TFs in other plant species have showed that they play an important role in the regulation of flowering and photoperiod (Wang et al., 2013). Without maize COL mutants, we searched different maize CRISPR/Cas9 populations and found one line with a frame-shift deletion in the first exon of COL8 (Figure S3). The homozygous mutant has a pale-green phenotype and seedling lethality, supporting our hypothesis that the COL TFs are important for photosynthesis (Figure 4D). Interestingly, we also found that the key C_4 photosynthesis genes with cell-specific expression pattern are associated with cell-specific H3K27me3 marks, suggesting that they are regulated not just by a complex TF network, but also at the epigenome level (Figure 4E).

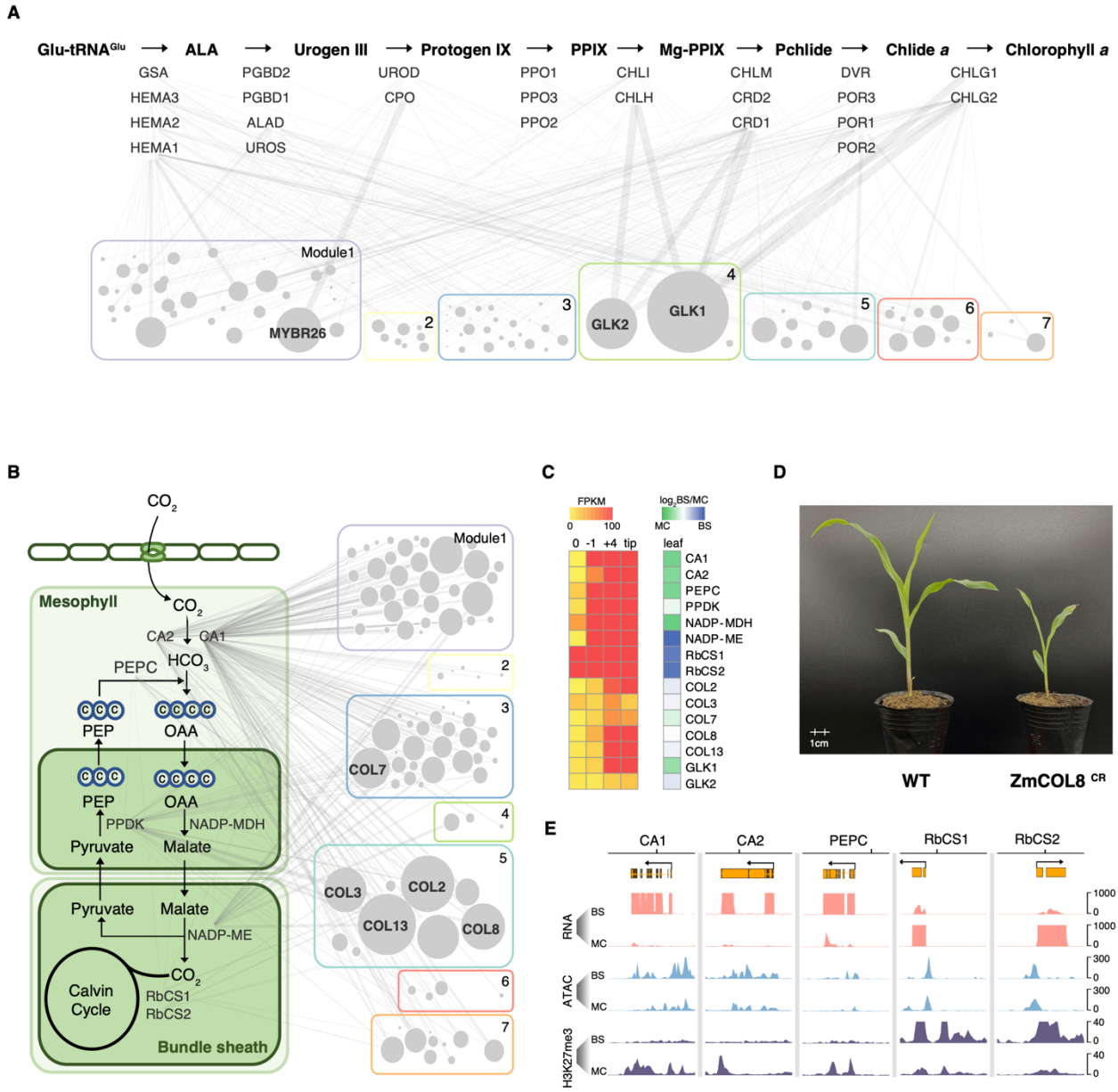

Figure 4. Identify key regulators based on local TF connectivity

(A) The chlorophyll biosynthesis pathway. The genes encoding key enzymes for each step are shown below the arrow. The width of the edge between TF and gene represent regulatory potential based on the TIP model. The 104 TF nodes are represented as circles and grouped into 7 modules, with size proportional to the sum of the regulatory potential.

(B) The maize photosynthesis pathway. Key genes encoding enzymes responsible for CO_2 fixation and transport between mesophyll and bundle sheath cells are shown.

(C) Heatmap shows the expression gradient of core C_4 photosynthesis genes in four leaf sections from base to tip, as well as their mesophyll and bundle sheath cell-specific expression pattern.

(D) Pale-green leaf and seedling lethal phenotypes of the COL8 mutant.

(E) Cell-type specific H3K27me3 associated with core C_4 genes. MC: mesophyll cell; BS: bundle sheath cells.

TF sequence binding preferences are similar among family members and conserved through angiosperm evolution

To model TF binding from sequence we applied a "bag-of-*k*-mers" model (Mejía-Guerra and Buckler, 2019) to discriminate TF binding regions from other regions in the genome, which resulted in reliable models for all the TFs (5-fold cross-validation, average accuracy for each TF > 70%) (Figures 5 and S4; Data S10). Using average *k*-mer weights from the models, we derived a distance matrix among TFs, and summarized TFs relationships (Data S11). After removal of singleton families, we observed that for 85% of the TFs families, most of their members (>= 50%) cluster into the same group in a dendrogram (Figure 5A and Figure S4A).

This observation prompted us to evaluate whether TF sequence preferences have persisted across angiosperm evolution, as TF protein families are frequently well conserved. Using the top 1% of the predictive *k*-mers for each TF, we examined their similarity to a large collection of Arabidopsis TF binding position weight matrices (PWMs) derived from DAP-seq experiments (O'Malley et al., 2016). After removal of families that did not have a counterpart (or were poorly represented), 50 out of 81 (61%) of the evaluated TFs preferentially match PWMs to their corresponding family in Arabidopsis (P-value < 0.001, Figure 5C).

Our estimations of conservation through Arabidopsis are likely understated, as some TF families can recognize more than one type of motif, which might not be properly represented in the current databases. For instance, DAP-seq data is not available for the Arabidopsis homologous to *ZmGLK1*, which together with *ZmGLK2* forms an independent cluster from the others GLKs. Yet, our *k*-mer model for *ZmGLK1* (Data S11) agrees with the core of a motif (*i.e.,* "RGAT", R = A/G) enriched in the promoters of genes up-regulated by Arabidopsis GLK1 (Waters et al., 2009; Zubo et al., 2018). Overall, the sequence preference conservation suggests a strong constraint over more than 150 million years, which also agrees with the reduced SNP variation at the TF peaks (Figure 2A).

TF binding co-localization is context specific

With a genome larger than two billion bp, any given TF (recognition sites avg. 6.8 bp) could have affinity for roughly a third of a million locations across the genome (Mejía-Guerra and Buckler, 2019; O'Malley et al., 2016). Under this scenario, achieving the observed TF binding specificity would likely require extra cues. We predict that co-binding and combinatorial recognition of *cis*-elements is key for specificity. To test this, we created machine-learning models based on co-localization information to learn non-linear dependencies among TFs using the ENCODE pipeline (Gerstein et al., 2012). To fit a model for each TF (*i.e.,* focus TF or context), we built a co-localization matrix, by overlapping peaks for the focus-TF with peaks of all remaining TFs (*i.e.,* partner TF). The co-localization model was aimed to discriminate between the true co-localization matrix and a randomized version of the same (Friedman and Popescu, 2008). The output of each model is a set of combinatorial rules that can predict TF binding. For each TF, the average of 10 models with independent randomized matrices have area under the receiver operating curve >

0.9 (Data S12). The high performance of the model agrees with the hypothesis that co-localization has vast information content to determine binding specificity.

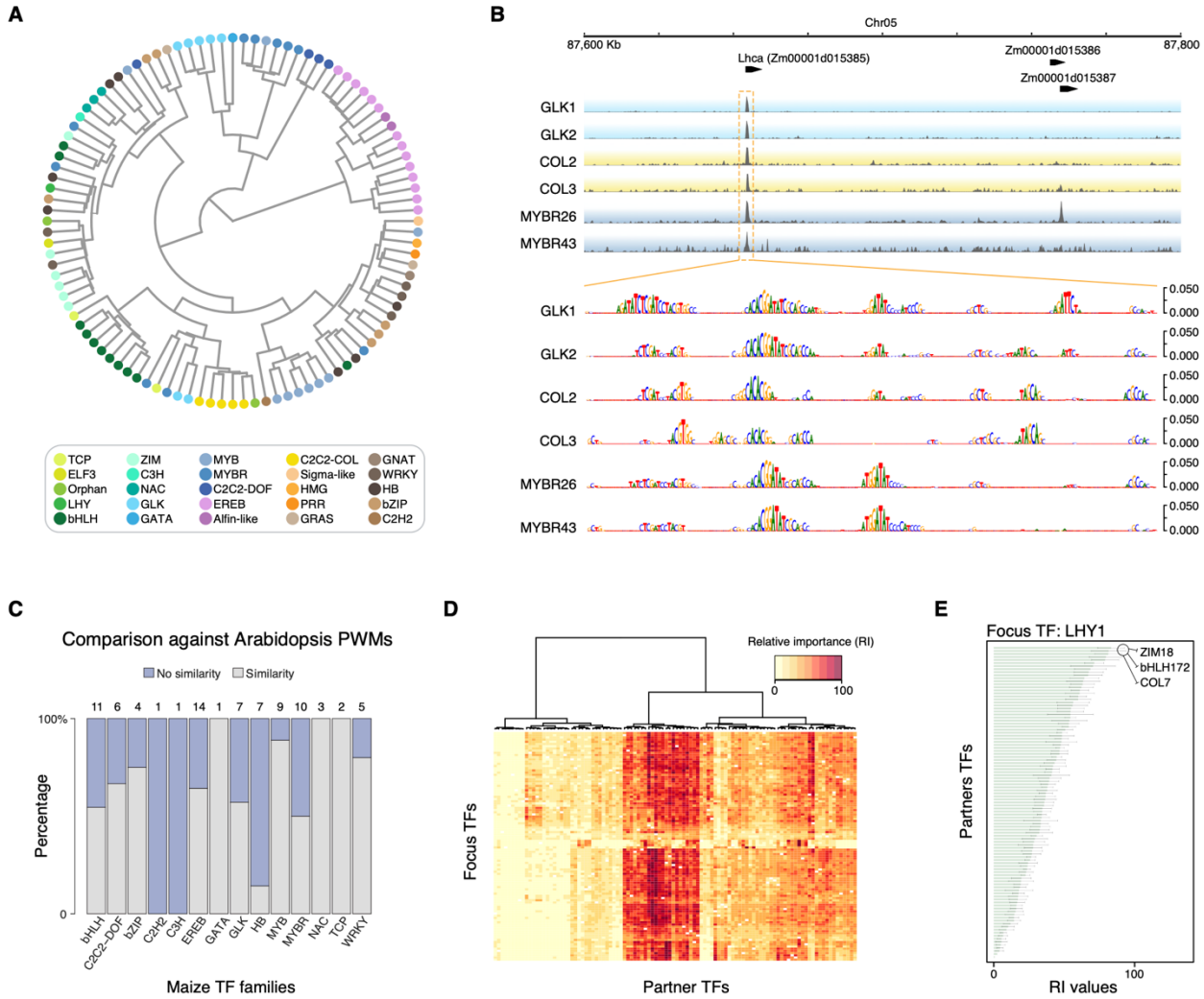

Figure 5. Machine learning models of TF recognition sequence and co-binding

(A) 104 TF clustered by sequence binding similarity (predicted *k*-mer weights) derived from the sequence models.

(B) Example of the promoter region of the gene encoding the light-harvesting complex of PSI (*Lhca*, Zm00001d015385). Occlusion analysis for TFs targeting *Lhca* showing scores of putative regulatory positions at base pair resolution derived from the sequence models.

(C) Conservation of recognition sequence between Arabidopsis and maize TF homologues. Numbers of TFs examined are shown on top, and the TF family names are shown at the bottom.

(D) Cluster of 104 TF based on their relative importance value with co-binding TFs.

(E) Top three co-binding partner TFs of LHY1 based on RI ranking.

Using the rules derived from the co-localization models, we scored the relative importance (RI) of each partner TF for the joint distribution of the set of peaks for a given context (Figure 5D; Data S13). To obtain a global view, we calculated the average RI of a TF across all focus TFs. We observed that the whole set shows a trend towards medium to low average RI values (*i.e.,* ≤ 60 RI, more context specific), with fewer TFs predictive for a large number of focus TFs (*i.e.,* > 60 RI, high-combinatorial potential). Clustering according to the RI values revealed a group of TFs that have large RI across many focus TFs. The large RI score means that a TF is highly predictive of the binding in a specific context. For example, among the 104 TFs we examined, LATE ELONGATED HYPOCOTYL (LHY, Zm00001d024546) is the most highly expressed one in the differentiating leaf (Li et al., 2010). LHY encodes a MYB TF that is a central oscillator in the plant circadian clock (Nagel and Kay, 2012), and the its top three predicted partner TFs are ZIM18, bHLH172 and COL7 (Figure 5E). Although their functions have yet been characterized in maize, their Arabidopsis homologs are involved in jasmonic acid signaling, iron homeostasis and flowering time regulation, all of which are tightly coupled to the circadian clock (Sanchez and Kay, 2016).

Taken together, TFs co-binding could be the key to explain how TFs with similar sequence preferences could target different genes and control different biological functions. The co-localization model revealed a large combinatorial space for TF binding sites that likely favors the occurrence of specific combinations, which could facilitate rapid diversification of the regulatory network during speciation.

Conservation of the TF regulatory interactions among grasses

Finally, we used the machine learning models to investigate the conservation of the regulatory network in other grasses. To do so we performed ATAC-seq in sorghum and rice, and obtained open chromatin sequences of their maize synteny genes. Next, we inferred network edge conservation based on whether the sequence model of maize TFs could predict binding in the open chromatin of the synteny target gene in sorghum and rice (Figure 6A). For example, we could predict TF binding for 68% of the open chromatin regions of maize synteny genes in sorghum. Looking at the edges to syntenic genes, we found that ~30% of the observed edges in the maize network were conserved to sorghum, and ~18% were conserved to rice (Figure 6B).

Comparison between the human and mouse regulatory networks has shown that the core network, which consists of TF-to-TF connections, are frequently more conserved that the individual DNA bases among regulatory regions (Stergachis et al., 2014). If the same is true in plants, we would expect TF genes to be enriched in the set of conserved targets. Hence, we performed GO enrichment analysis for the conserved edges and the results agrees with our hypothesis (Figure 6C). In addition, we also examined the constitution of conserved edges, and found that the TF-to-TF edges were over-represented than TF-to-nonTF edges in both sorghum and rice (Figure 6B).

ENCODE and modENCODE also found that the prevalence of human and mouse TF recognition sites in the open chromatin regions are under selection (Stergachis et al., 2014). To test this for

plant TF, we calculated the number of TF model matches in the maize, sorghum and rice ATAC-seq peaks, and found that they are indeed correlated (Figure 6D). In addition, we found that the TF target conservation rates in rice and sorghum are also similar, suggesting that syntenic TFs are under similar selection pressure (Figure 6E). Together with previous observation that the core network is more frequently conserved, our findings suggest that plants and animals might have evolved similar transcription networks in terms of structure and mechanism, and the rewiring of the network occurs in a hierarchical fashion, with nodes at the top of the transmission of information retaining more conserved edges than those at the bottom.

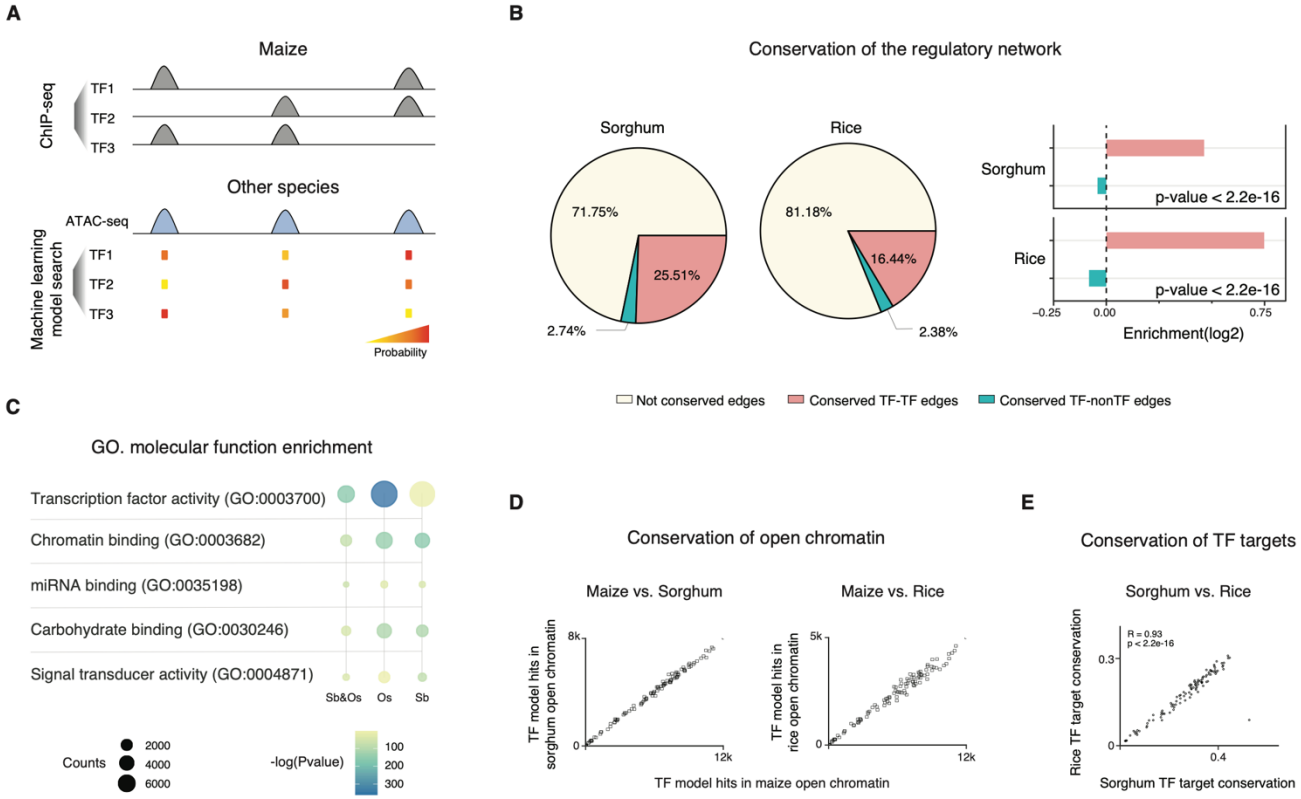

Figure 6. Conservation of the TF regulatory interactions

(A) Maize TF binding models are used to search the open chromatin regions of maize synteny gene in rice and sorghum, to infer conservation of the regulatory interaction.
(B) TF to TF edges are more conserved than TF to non-TF gene edges.
(C) GO-term enrichment analysis of the network edges conserved in sorghum, rice and both.
(D) Number of recognition sequences of each TF found in maize, sorghum and rice open chromatin based on the TF binding models.
(E) The axis indicates the percentage of targets conservation of each maize TF that has a synteny partner in rice and sorghum.

DISCUSSION

For a long time, plant researches have been limited to studying either the function of a single gene or perform genome-wide association analysis for everything. The former often resulted in over-simplification, while the latter failed to determine causality. Neither is able to resolve complex and robust biological processes such as transcription regulation. A bottom-up study towards a systems biology approach is required to understand complex system, and we need to study the system as a whole, not the individual components.

In this study, we attempted to systematically identify the emergent properties of the gene transcriptional regulation in a plant tissue. To do so, we resolved binding profiles for 104 TFs expressed in the maize leaf. We constructed a high-confidence regulatory network that covers ~45% of the genes in the maize genome, and ~77% of the leaf expressed genes. We found the network is scale-free and shows a modular organization at the topological and functional levels. A key advantage of the scale-free topology is the increased tolerance to random failures. However, the high connectivity hubs are the vulnerabilities in a scale-free network. Using chlorophyll biosynthesis and C_4 photosynthesis as examples, we showed that mutation of a highly connected TF could disrupt plant growth and development (Figure 4). This also demonstrated how network connectivity data could be used to predict biological functions.

The observed structural and functional split of the network could indicate how gene expression is fine tuned. The presence of multiple TFs across modules to regulate genes in the same pathway suggests intricate modes of actions to coordinate transcription, in contrast to the classic view of a singular or few master regulators. Altogether, the topological and functional features of the regulatory graph not only provide validation of known TF to target gene interactions, but also provide additional clues to understand a vast number of uncharacterized genes and TFs.

The finding that TF co-binding is widespread in plant is consistent with the results of mammalian studies (Gerstein et al., 2012). Co-binding could be a key evolutionary advantage, because from an information standpoint, it enables a large variety of transcriptional outputs using synergistic and combinatorial input of a small number of TFs. For instance, taking sets of five distinct TFs (*i.e.,* the mode of distinct TF binding sites in maize open chromatin regions) from the 104 TF repertoire would generate a ~91 million possible combinations without considering TF repetitions. In mammalian cells, high-combinatorial *trans* factors, such as MAX and P300, have key roles in enhancer function (Ram et al., 2011; Shlyueva et al., 2014). In our dataset, we also found high-combinatorial TFs, most of which has not being characterized, and deserve further attention. Some of them are homologues to Arabidopsis TFs with known roles in hormone signaling pathways (*e.g.* ZIM TFs in JA signalling). Some of them belong to families that are known for protein-protein interactions, such as the MYB-bHLHs and the ZIM-MYB-bHLHs TFs (Qi et al., 2011; Ramsay and Glover, 2005; Tian et al., 2016). These could suggest that plant cells could use these factors to integrate multiple inputs to coordinate the transcriptional response.

It has long been proposed that the changes in the mechanisms controlling gene expression are the main driving force of evolution (King and Wilson, 1975). Around 20% of the human and mouse TF bindings are conserved despite over 90 Mya of divergence (Stergachis et al., 2014). Rice diverged from the common ancestor of maize and sorghum ~50 Mya, and we found over 18% of the rice network edges are conserved in maize. However, stringent cut off used for our TF binding model search might underestimate the conservation rate, and future cross-species comparison using a uniformed methodology would be greatly beneficial. On a practical level, we expect the presented TF binding and co-binding models to be integrated into pipelines to predict effects of non-coding variants, both common and rare, on TF binding, to pinpoint causal sites. As the possibility of being able to predict and generate novel variation not seen in nature could fundamentally change future plant breeding.

In the future, we need to study and compare transcription regulatory networks in a broad variety of species and in different tissues or different environmental conditions, to gain insights into the architectures and circuits that enable different outputs. The finding that core C_4 photosynthesis genes are regulated by histone modification (Figure 4E) showed that it will is also important to integrate regulations at additional levels such as epigenome, proteome and metabolome.

STAR★METHODS

Detailed methods are provided in the supplementary material including the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABLE
- EXPERIMENTAL METHOD AND SUBJECT DETAILS
- METHOD DETAILS
    - TF ChIP-seq
        - Preparation of hyper-stable Tn5 transposase
        - Preparation of transfection-grade plasmids
        - Transient expression of biotinylated TF in leaf protoplasts
        - Bio-ChIP-Seq library preparation
    - RNA-seq
        - Isolation of bundle sheath strands from maize leaf
        - Isolation of mesophyll and bundle sheath cell RNA
        - Non-strand specific RNA-seq library preparation
    - Histone ChIP-seq
        - Crosslinking and nuclei lysis
        - Chromatin shearing and immunoprecipitation
        - On-bead histone ChIP-Seq library preparation
    - ATAC-seq
    - Data Processing
        - Alignment of ChIP-seq and ATAC-seq Reads
        - Post-processing and Quality Control Steps - TF ChIP-seq data
        - Post-processing and Quality Control Steps – ATAC-seq data
    - Data Analyzing
        - Clustering of TF peaks
        - Overlap ChIP-seq and ATAC-seq data
        - Sequence variation in TF binding regions
        - Sequence variation analyses - Enrichment of cis-eQTLs
        - Sequence variation analyses - Enrichment of GWAS hits
        - Inference of a transcriptional regulatory graph
        - Network analyses
        - Sequence model – description
        - Occlusion and saturation mutagenesis maps of TF binding regions
        - ZmTF sequence preferences – comparison with Arabidopsis thaliana
        - Co-localization model – description
        - Co-localization rule learning
        - Calculation of relative importance (RI) score, and differential importance (DI) score
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABLE
- SUPPLEMENTARY FIGURES

ACKNOWLEDGMENTS

We thank B. Du, X. Wang, R. Foo, J. Li and M. Huang for technical support, G. Ramstein for advice in the analysis of GWAS data, and D. Grierson and S. Miller for proofreading the manuscript.

AUTHOR CONTRIBUTIONS

S. Zhong, E.S. Buckler, and P. Li designed and supervised the research; X. Tu. developed techniques, performed experiments and processed the raw data with assistance from D. Tzeng, P. Chu and X. Dai. M.K. Mejia-Guerra designed and performed computational analysis, J.A.Valdes Franco performed population genetics analysis; M.K. Mejia-Guerra wrote the paper. All the authors read the paper and agree with the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Alter, P., Bircheneder, S., Zhou, L.-Z., Schlüter, U., Gahrtz, M., Sonnewald, U., Dresselhaus, T., 2016. Flowering Time-Regulated Genes in Maize Include the Transcription Factor ZmMADS11[OPEN]. Plant Physiol 172, 389–404. https://doi.org/10.1104/pp.16.00285

Barabási, A.-L., Albert, R., 1999. Emergence of Scaling in Random Networks. Science 286, 509–512. https://doi.org/10.1126/science.286.5439.509

Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. ICWSM 8, 361–362.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., Young, R.A., 2005. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell 122, 947–956. https://doi.org/10.1016/j.cell.2005.08.020

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J.C., Goodman, M.M., Harjes, C., Guill, K., Kroon, D.E., Larsson, S., Lepak, N.K., Li, H., Mitchell, S.E., Pressoir, G., Peiffer, J.A., Rosas, M.O., Rocheford, T.R., Romay, M.C., Romero, S., Salvo, S., Villeda, H.S., Silva, H.S. da, Sun, Q., Tian, F., Upadyayula, N., Ware, D., Yates, H., Yu, J., Zhang, Z., Kresovich, S., McMullen, M.D., 2009. The Genetic Architecture of Maize Flowering Time. Science 325, 714–718. https://doi.org/10.1126/science.1174276

Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., Wang, B., Xu, D., Yang, B., Xie, C., Fan, L., Gao, S., Xu, X., Zhang, G., Li, Y., Jiao, Y., Doebley, J.F., Ross-Ibarra, J., Lorant, A., Buffalo, V., Romay, M.C., Buckler, E.S., Ware, D., Lai, J., Sun, Q., Xu, Y., 2018. Construction of the third-generation Zea mays haplotype map. Gigascience 7. https://doi.org/10.1093/gigascience/gix134

Burdo, B., Gray, J., Goetting-Minesky, M.P., Wittler, B., Hunt, M., Li, T., Velliquette, D., Thomas, J., Gentzel, I., Brito, M. dos S., Mejía-Guerra, M.K., Connolly, L.N., Qaisi, D., Li, W., Casas, M.I., Doseff, A.I., Grotewold, E., 2014. The Maize TFome – development of a transcription factor open reading frame collection for functional genomics. Plant J 80, 356–366. https://doi.org/10.1111/tpj.12623

Cheng, C., Min, R., Gerstein, M., 2011. TIP: A probabilistic method for identifying transcription factor target genes from ChIP-seq binding profiles. Bioinformatics 27, 3221–3227. https://doi.org/10.1093/bioinformatics/btr552

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E 70, 066111. https://doi.org/10.1103/PhysRevE.70.066111

Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Rev. 51, 661–703. https://doi.org/10.1137/070710111

Cuellar-Partida, G., Buske, F.A., McLeay, R.C., Whitington, T., Noble, W.S., Bailey, T.L., 2012. Epigenetic priors for identifying active transcription factor binding sites. Bioinformatics 28, 56–62. https://doi.org/10.1093/bioinformatics/btr614

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dittrich, M.T., Klau, G.W., Rosenwald, A., Dandekar, T., Müller, T., 2008. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. Bioinformatics 24, i223–i231. https://doi.org/10.1093/bioinformatics/btn161

Doebley, J., Stec, A., Hubbard, L., 1997. The evolution of apical dominance in maize. Nature 386, 485–488. https://doi.org/10.1038/386485a0

Dong, P., Tu, X., Chu, P.-Y., Lü, P., Zhu, N., Grierson, D., Du, B., Li, P., Zhong, S., 2017. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. Molecular Plant 10, 1497–1509. https://doi.org/10.1016/j.molp.2017.11.005

Feng, J., Liu, T., Qin, B., Zhang, Y., Liu, X.S., 2012. Identifying ChIP-seq enrichment using MACS. Nat Protoc 7, 1728–1740. https://doi.org/10.1038/nprot.2012.101

Friedman, J.H., Popescu, B.E., 2008. Predictive learning via rule ensembles. Ann. Appl. Stat. 2, 916–954. https://doi.org/10.1214/07-AOAS148

Gaudinier, A., Rodriguez-Medina, J., Zhang, L., Olson, A., Liseron-Monfils, C., Bågman, A.-M., Foret, J., Abbitt, S., Tang, M., Li, B., Runcie, D.E., Kliebenstein, D.J., Shen, B., Frank, M.J., Ware, D., Brady, S.M., 2018. Transcriptional regulation of nitrogen-associated metabolism and growth. Nature 563, 259–264. https://doi.org/10.1038/s41586-018-0656-3

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., Malinverni, R., 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289–291. https://doi.org/10.1093/bioinformatics/btv562

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman, N., Bhardwaj, N., Boyle, A.P., Cayting, P., Charos, A., Chen, D.Z., Cheng, Y., Clarke, D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A., Jain, P., Kasowski, M., Lacroute, P., Leng, J., Lian, J., Monahan, H., O'Geen, H., Ouyang, Z., Partridge, E.C., Patacsil, D., Pauli, F., Raha, D., Ramirez, L., Reddy, T.E., Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K.Y., Zilberman-Schapira, G., Batzoglou, S., Sidow, A., Farnham, P.J., Myers, R.M., Weissman, S.M., Snyder, M., 2012. Architecture of the human regulatory network derived from ENCODE data. Nature 489, 91–100. https://doi.org/10.1038/nature11245

Gillespie, C.S., 2015. Fitting Heavy Tailed Distributions: The poweRlaw Package. Journal of Statistical Software 64, 1–16. https://doi.org/10.18637/jss.v064.i02

Han, K.-Y., Song, J.-A., Ahn, K.-Y., Park, J.-S., Seo, H.-S., Lee, J., 2007. Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive Escherichia coli protein, SlyD. Protein Eng Des Sel 20, 543–549. https://doi.org/10.1093/protein/gzm055

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J., Jennings, E.G., Zeitlinger, J., Pokholok, D.K., Kellis, M., Rolfe, P.A., Takusagawa, K.T., Lander, E.S., Gifford, D.K., Fraenkel, E., Young, R.A., 2004. Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99–104. https://doi.org/10.1038/nature02800

Heyndrickx, K.S., Velde, J.V. de, Wang, C., Weigel, D., Vandepoele, K., 2014. A Functional and Evolutionary Perspective on Transcription Factor Binding in Arabidopsis thaliana. The Plant Cell 26, 3894–3910. https://doi.org/10.1105/tpc.114.130591

Kharchenko, P.V., Tolstorukov, M.Y., Park, P.J., 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol 26, 1351–1359. https://doi.org/10.1038/nbt.1508

King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. Science 188, 107–116. https://doi.org/10.1126/science.1090005

Kremling, K.A.G., Chen, S.-Y., Su, M.-H., Lepak, N.K., Romay, M.C., Swarts, K.L., Lu, F., Lorant, A., Bradbury, P.J., Buckler, E.S., 2018. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature 555, 520–523. https://doi.org/10.1038/nature25966

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin,

M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., Snyder, M., 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831. https://doi.org/10.1101/gr.136184.111

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.-B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A., 2002. Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298, 799–804. https://doi.org/10.1126/science.1075090

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers, C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., Brutnell, T.P., 2010. The developmental dynamics of the maize leaf transcriptome. Nature Genetics 42, 1060–1067. https://doi.org/10.1038/ng.703

Li, Q., Brown, J.B., Huang, H., Bickel, P.J., 2011. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1779. https://doi.org/10.1214/11-AOAS466

Li, Yong-xiang, Li, C., Bradbury, P.J., Liu, X., Lu, F., Romay, C.M., Glaubitz, J.C., Wu, X., Peng, B., Shi, Y., Song, Y., Zhang, D., Buckler, E.S., Zhang, Z., Li, Yu, Wang, T., 2016. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. The Plant Journal 86, 391–402. https://doi.org/10.1111/tpj.13174

Mejía-Guerra, M.K., Buckler, E.S., 2019. A k-mer grammar analysis to uncover maize regulatory architecture. BMC Plant Biology 19, 103. https://doi.org/10.1186/s12870-019-1693-2

Müllner, D., 2013. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. Journal of Statistical Software 53, 1–18. https://doi.org/10.18637/jss.v053.i09

Nagel, D.H., Kay, S.A., 2012. Complexity in the Wiring and Regulation of Plant Circadian Networks. Current Biology 22, R648–R657. https://doi.org/10.1016/j.cub.2012.07.025

Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., Stamatoyannopoulos, J.A., 2012. Circuitry and Dynamics of Human Transcription Factor Regulatory Networks. Cell 150, 1274–1286. https://doi.org/10.1016/j.cell.2012.04.040

Nguyen, N.H., Lee, H., 2016. MYB-related transcription factors function as regulators of the circadian clock and anthocyanin biosynthesis in Arabidopsis. Plant Signal Behav 11. https://doi.org/10.1080/15592324.2016.1139278

Olesen, J.M., Bascompte, J., Dupont, Y.L., Jordano, P., 2007. The modularity of pollination networks. PNAS 104, 19891–19896. https://doi.org/10.1073/pnas.0706375104

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., Ecker, J.R., 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165, 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Pauwels, L., Goossens, A., 2011. The JAZ Proteins: A Crucial Interface in the Jasmonate Signaling Cascade. Plant Cell 23, 3089–3100. https://doi.org/10.1105/tpc.111.089300

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M.,

Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Picelli, S., Björklund, Å.K., Reinius, B., Sagasser, S., Winberg, G., Sandberg, R., 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24, 2033–2040. https://doi.org/10.1101/gr.177881.114

Qi, T., Song, S., Ren, Q., Wu, D., Huang, H., Chen, Y., Fan, M., Peng, W., Ren, C., Xie, D., 2011. The Jasmonate-ZIM-Domain Proteins Interact with the WD-Repeat/bHLH/MYB Complexes to Regulate Jasmonate-Mediated Anthocyanin Accumulation and Trichome Initiation in Arabidopsis thaliana. The Plant Cell 23, 1795–1814. https://doi.org/10.1105/tpc.111.083261

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., Durham, T., Zhang, X., Donaghey, J., Epstein, C.B., Regev, A., Bernstein, B.E., 2011. Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-wide Location Analysis in Human Cells. Cell 147, 1628–1639. https://doi.org/10.1016/j.cell.2011.09.057

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160–W165. https://doi.org/10.1093/nar/gkw257

Ramsay, N.A., Glover, B.J., 2005. MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends in Plant Science 10, 63–70. https://doi.org/10.1016/j.tplants.2004.12.011

Rodgers-Melnick, E., Vera, D.L., Bass, H.W., Buckler, E.S., 2016. Open chromatin reveals the functional maize genome. PNAS 113, E3177–E3184. https://doi.org/10.1073/pnas.1525244113

Rossini, L., Cribb, L., Martin, D.J., Langdale, J.A., 2001. The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. The Plant Cell 13, 1231–1244. https://doi.org/10.1105/tpc.13.5.1231

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K.A., Meeley, R., Ananiev, E.V., Svitashev, S., Bruggemann, E., Li, B., Hainey, C.F., Radovic, S., Zaina, G., Rafalski, J.-A., Tingey, S.V., Miao, G.-H., Phillips, R.L., Tuberosa, R., 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. PNAS 104, 11376–11381. https://doi.org/10.1073/pnas.0704145104

Sanchez, S.E., Kay, S.A., 2016. The Plant Circadian Clock: From a Simple Timekeeper to a Complex Developmental Manager. Cold Spring Harb Perspect Biol 8. https://doi.org/10.1101/cshperspect.a027748

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Buren, P.V., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu,

Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., Wilson, R.K., 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326, 1112–1115. https://doi.org/10.1126/science.1178534

Sheen, J., 2001. Signal Transduction in Maize and Arabidopsis Mesophyll Protoplasts. Plant Physiology 127, 1466–1475. https://doi.org/10.1104/pp.010820

Shlyueva, D., Stampfel, G., Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nature Reviews Genetics 15, 272–286. https://doi.org/10.1038/nrg3682

Sorrells, T.R., Johnson, A.D., 2015. Making Sense of Transcription Networks. Cell 161, 714–723. https://doi.org/10.1016/j.cell.2015.04.014

Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., Thurman, R.E., Vong, S., Bates, D., Neri, F., Diegel, M., Giste, E., Dunn, D., Vierstra, J., Hansen, R.S., Johnson, A.K., Sabo, P.J., Wilken, M.S., Reh, T.A., Treuting, P.M., Kaul, R., Groudine, M., Bender, M.A., Borenstein, E., Stamatoyannopoulos, J.A., 2014. Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature 515, 365–370. https://doi.org/10.1038/nature13972

Tanaka, E., Bailey, T., Grant, C.E., Noble, W.S., Keich, U., 2011. Improved similarity scores for comparing motifs. Bioinformatics 27, 1603–1609. https://doi.org/10.1093/bioinformatics/btr257

Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B., Buckler, E.S., 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics 43, 159–162. https://doi.org/10.1038/ng.746

Tian, H., Qi, T., Li, Y., Wang, C., Ren, C., Song, S., Huang, H., 2016. Regulation of the WD-repeat/bHLH/MYB complex by gibberellin and jasmonate. Plant Signaling & Behavior 11, e1204061. https://doi.org/10.1080/15592324.2016.1204061

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., Vandepoele, K., 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res 46, D1190–D1196. https://doi.org/10.1093/nar/gkx1002

Wallace, J.G., Bradbury, P.J., Zhang, N., Gibon, Y., Stitt, M., Buckler, E.S., 2014. Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize. PLOS Genetics 10, e1004845. https://doi.org/10.1371/journal.pgen.1004845

Wang, P., Kelly, S., Fouracre, J.P., Langdale, J.A., 2013. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. The Plant Journal 75, 656–670. https://doi.org/10.1111/tpj.12229

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., Langdale, J.A., 2009. GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. The Plant Cell 21, 1109–1128. https://doi.org/10.1105/tpc.108.065250

Yan, J., Enge, M., Whitington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M., Taipale, J., 2013. Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites. Cell 154, 801–813. https://doi.org/10.1016/j.cell.2013.07.034

Yang, F., Li, W., Jiang, N., Yu, H., Morohashi, K., Ouma, W.Z., Morales-Mantilla, D.E., Gomez-Cano, F.A., Mukundi, E., Prada-Salcedo, L.D., Velazquez, R.A., Valentin, J., Mejía-Guerra, M.K., Gray, J.,

Doseff, A.I., Grotewold, E., 2017. A Maize Gene Regulatory Network for Phenolic Metabolism. Molecular Plant 10, 498–515. https://doi.org/10.1016/j.molp.2016.10.020

Zhang, N., Gibon, Y., Wallace, J.G., Lepak, N., Li, P., Dedow, L., Chen, C., So, Y.-S., Kremling, K., Bradbury, P.J., Brutnell, T., Stitt, M., Buckler, E.S., 2015. Genome-Wide Association of Carbon and Nitrogen Metabolism in the Maize Nested Association Mapping Population. Plant Physiology 168, 575–583. https://doi.org/10.1104/pp.15.00025

Zhong, S., Joung, J.-G., Zheng, Y., Chen, Y., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., Giovannoni, J.J., 2011. High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation. Cold Spring Harb Protoc 2011, pdb.prot5652. https://doi.org/10.1101/pdb.prot5652

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44, 821–824. https://doi.org/10.1038/ng.2310

Zubo, Y.O., Blakley, I.C., Franco-Zorrilla, J.M., Yamburenko, M.V., Solano, R., Kieber, J.J., Loraine, A.E., Schaller, G.E., 2018. Coordination of Chloroplast Development through the Action of the GNC and GLK Transcription Factor Families. Plant Physiology 178, 130–147. https://doi.org/10.1104/pp.18.00414