LARGE-SCALE BIOLOGY ARTICLE

Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules

Kelsey A. Maher,a,b,1 Marko Bajic,a,c,1 Kaisa Kajala,d,2 Mauricio Reynoso,e Germain Pauluzzi,e Donnelly A. West,f Kristina Zumstein,f Margaret Woodhouse,f Kerry Bubb,g Michael W. Dorrity,g Christine Queitsch,g Julia Bailey-Serres,e Neelima Sinha,f Siobhan M. Brady,d and Roger B. Deala,3

a Department of Biology, Emory University, Atlanta, Georgia 30322
b Graduate Program in Biochemistry, Cell, and Developmental Biology, Emory University, Atlanta, Georgia 30322
c Graduate Program in Genetics and Molecular Biology, Emory University, Atlanta, Georgia 30322
d Department of Plant Biology and Genome Center, University of California, Davis, California 95616
e Center for Plant Cell Biology, Botany and Plant Sciences Department, University of California, Riverside, California 92521
f Department of Plant Biology, University of California, Davis, California 95616
g University of Washington, School of Medicine, Department of Genome Sciences, Seattle, Washington 98195

ORCID IDs: 0000-0001-6483-7473 (K.K.); 0000-0003-0624-8034 (D.A.W.); 0000-0001-9516-8081 (K.Z.); 0000-0002-8568-7125 (J.B.-S.); 0000-0002-1494-7065 (N.S.); 0000-0001-9424-8055 (S.M.B.); 0000-0003-1289-688X (R.B.D.)

The transcriptional regulatory structure of plant genomes remains poorly defined relative to animals. It is unclear how many cis-regulatory elements exist, where these elements lie relative to promoters, and how these features are conserved across plant species. We employed the assay for transposase-accessible chromatin (ATAC-seq) in four plant species (Arabidopsis thaliana, Medicago truncatula, Solanum lycopersicum, and Oryza sativa) to delineate open chromatin regions and transcription factor (TF) binding sites across each genome. Despite 10-fold variation in intergenic space among species, the majority of open chromatin regions lie within 3 kb upstream of a transcription start site in all species. We find a common set of four TFs that appear to regulate conserved gene sets in the root tips of all four species, suggesting that TF-gene networks are generally conserved. Comparative ATAC-seq profiling of Arabidopsis root hair and non-hair cell types revealed extensive similarity as well as many cell-type-specific differences. Analyzing TF binding sites in differentially accessible regions identified a MYB-driven regulatory module unique to the hair cell, which appears to control both cell fate regulators and abiotic stress responses. Our analyses revealed common regulatory principles among species and shed light on the mechanisms producing cell-type-specific transcriptomes during development.

INTRODUCTION

The transcription of protein coding genes is controlled by regulatory DNA elements, including both the core promoter and more distal enhancer elements (Lee and Young, 2000). The core promoter is a short DNA region surrounding the transcription start site (TSS), at which RNA polymerase II and general transcription factors are recruited. Enhancer elements act as platforms for recruiting both positive- and negative-acting transcription factors (TFs) and serve to integrate multiple signaling inputs in order to dictate the spatial and temporal control of transcription from the core promoter. As such, enhancer functions are critical for directing transcriptional output during cell differentiation and development, as well as coordinating transcriptional responses to environmental change (Ong and Corces, 2011). Despite their importance, only a small number of bona fide enhancers have been characterized in plants, and we lack a global view of their general distribution and action in plant genomes (Weber et al., 2016).

In large part, our limited knowledge of plant cis-regulatory elements arises from the unique difficulties in identifying these elements. While some enhancers exist near their target core promoter, others can be thousands of base pairs upstream or downstream, or even within the transcribed region of a gene body (Ong and Corces, 2011; Spitz and Furlong, 2012). Furthermore, enhancers generally do not display universal sequence conservation, aside from sharing of individual TF binding sites, which makes them very challenging to locate. By contrast, core promoters can be readily identified through mapping the 5' ends of transcripts (Morton et al., 2014; Mejia-Guerra et al., 2015). It was recently discovered that many enhancer elements in animal genomes could be identified with relatively high confidence based on a unique combination of flanking histone posttranslational
modifications (PTMs), such as an enrichment for H3K27ac and H3K4me1. This characteristic histone PTM signature has led to the annotation of such elements in several animal models and specialized cell types (Heintzman et al., 2009; Bonn et al., 2012). However, the only currently known association between plant cis-regulatory elements and histone PTMs appears to be a modest correlation with H3K27me3 (Zhang et al., 2012b; Zhu et al., 2015). Though encouraging, this mark is not unique to these elements and cannot be used to identify enhancers on its own.

A long-known and general feature of sequence-specific DNA binding proteins is their ability to displace nucleosomes upon DNA binding, leading to an increase in nucleosome accessibility around the binding region (Gross and Garrard, 1988; Henikoff, 2008). In particular, DNaseI treatment of nuclei coupled with high-throughput sequencing (DNase-seq) has been used to probe chromatin accessibility. This technology has served as an important tool in identifying regulatory elements throughout animal genomes (Thurman et al., 2012) and more recently in certain plant genomes (Zhang et al., 2012a, 2012b; Pajoro et al., 2014; Sullivan et al., 2014). In addition, a differential micrococcal nuclease sensitivity assay has also been used to probe functional regions of the maize genome, demonstrating the versatility of this approach (Vera et al., 2014; Rodgers-Melnick et al., 2016).

DNase-seq has been used successfully to identify open chromatin regions in different tissues of both rice (Oryza sativa) and Arabidopsis thaliana (Zhang et al., 2012a; Pajoro et al., 2014; Zhu et al., 2015). Over a dozen of the intergenic DNase-hypersensitive sites in Arabidopsis were tested and shown to act as enhancer elements by activating a minimal promoter-reporter cassette, demonstrating that chromatin accessibility is an important factor in enhancer identification (Zhu et al., 2015). Collectively, these DNase-seq studies show that the majority of open chromatin sites exist outside of genes in rice and Arabidopsis, that differences in open chromatin sites can be identified between tissues, and that a large proportion of intergenic open chromatin sites are in fact regulatory, at least in Arabidopsis. Another recent significant advance came from using DNase-seq to examine the changes in Arabidopsis chromatin accessibility and TF occupancy that occur during development and in response to abiotic stress (Sullivan et al., 2014). This work showed that TF-to-TF regulatory network connectivity appears to be similar between Arabidopsis, human, and Caenorhabditis elegans and that such networks were extensively “rewired” in response to stress. This study also showed that many genetic variants linked to complex traits were preferentially located in accessible chromatin regions, portending the potential for harnessing natural variation in regulatory DNA for plant breeding.

We are still left with many open questions regarding the general conservation of transcriptional regulatory landscapes across plant genomes. For example, it remains unclear how many cis-regulatory elements generally exist in plant genomes, where they reside in relation to their target genes, and to what extent these features are conserved across plant genomes. Furthermore, it is not clear how the cis-regulatory elements within a single genome confer cell-type-specific transcriptional activity—and thus cell type identity—during development. In this study, we seek to build on previous work and to address some of these outstanding questions by analyzing chromatin accessibility across multiple, diverse plant species, and between two distinct cell types.

From a methodological perspective, the DNase-seq procedure is relatively labor-intensive and requires a large number of starting nuclei for DNasel treatment, which can be a major drawback for conducting cell-type-specific profiling investigations. More recently, the assay for transposase-accessible chromatin with sequencing (ATAC-seq) was developed as an alternative approach (Buenrostro et al., 2013). ATAC-seq employs treatment of isolated nuclei with an engineered transposase that simultaneously cleaves DNA and inserts sequencing adapters, such that cleaved fragments originating from open chromatin can be converted into a high-throughput sequencing library by PCR. Sequencing of the resulting library provides readout highly similar to that of DNase-seq, but ATAC-seq requires far fewer nuclei (Buenrostro et al., 2015). The relatively simple procedure for ATAC-seq and its low nuclei input, combined with its recent application in Arabidopsis and rice (Wilkins et al., 2016; Bajic et al., 2018; Lu et al., 2017), has made it widely useful for assaying plant DNA regulatory regions. In this study, we first optimized ATAC-seq for use with crude nuclei and nuclei isolated by INTACT (isolation of nuclei tagged in vivo and cut with sites for ligation using aptamers), demonstrating the versatility of this approach (Vera et al., 2014; Rodgers-Melnick et al., 2016). When comparing the two Arabidopsis root epidermal cell types, we found that their open chromatin profiles are qualitatively very similar. However, many quantitative differences between cell types were identified, and these regions often contained binding motifs for TFs that were more highly expressed in one cell type than the other. Further analysis of several such cell-type-enriched TFs led to the discovery of a hair cell transcriptional regulatory module driven by ABA INSENSITIVE5 (ABI5) and MYB DOMAIN PROTEIN77 (MYB77), which were previously shown to impact root development in Arabidopsis (Oyama et al., 1997; Shin et al., 2007). When comparing the two Arabidopsis root epidermal cell types, we found that their open chromatin profiles are qualitatively very similar. However, many quantitative differences between cell types were identified, and these regions often contained binding motifs for TFs that were more highly expressed in one cell type than the other. Further analysis of several such cell-type-enriched TFs led to the discovery of a hair cell transcriptional regulatory module driven by ABA INSENSITIVE5 (ABI5) and MYB DOMAIN PROTEIN77 (MYB77), which were previously shown to impact root development in Arabidopsis (Oyama et al., 1997; Shin et al., 2007).
Overall, our work suggests that the cis-regulatory structure of these four plant genomes is strikingly similar and that TF-target gene modules are also generally conserved across species. Furthermore, early differential expression of high-level TFs between the Arabidopsis hair and non-hair cells appears to drive a TF cascade that at least partially explains distinctions between hair and non-hair cell transcriptomes. Our data also highlight the utility of comparative chromatin profiling approaches and will be widely useful for hypothesis generation and testing.

RESULTS AND DISCUSSION

Application of ATAC-Seq in Arabidopsis Root Tips

The ATAC-seq method was introduced in 2013 and has since been widely adopted in many systems (Buenrostro et al., 2013; Mo et al., 2015; Scharer et al., 2016; Lu et al., 2017). This technique utilizes a hyperactive Tn5 transposase that is preloaded with sequencing adapters as a probe for chromatin accessibility. When purified nuclei are treated with the transposase complex, the enzyme freely enters nuclei and cleaves accessible DNA, both around nucleosomes and at nucleosome-depleted regions arising from the binding of TFs to DNA. Upon cleavage of DNA, the transposon integrates sequencing adapters, fragmenting the DNA sample in the process. Regions of higher accessibility will be cleaved by the transposase more frequently and generate more fragments—and ultimately more reads—once the sample is sequenced. Conversely, less accessible regions will have fewer fragments and reads. After PCR amplification of the raw DNA fragments, paired-end sequencing of the ATAC-seq library can reveal nucleosome-depleted regions where TFs are bound.

In this study, we set out to apply ATAC-seq to multiple plant species as well as different cell types from a single species. As such, we first established procedures for using the method with Arabidopsis, starting with root tip nuclei affinity purified by INTACT. We also established a protocol to use nuclei purified by detergent lysis of organelles followed by sucrose sedimentation, with the goal of broadening the application of ATAC-seq to nontransgenic starting tissue. We began with an Arabidopsis INTACT transgenic line constitutively expressing both the nuclear envelope targeting fusion protein (NTF) and biotin ligase (BirA) transgenes. Coexpression of these transgenes results in all the nuclei in the plant becoming biotinylated and, thus, amenable to purification with streptavidin beads (Deal and Henikoff, 2010; Sullivan et al., 2014). Transgenic INTACT plants were grown on vertically oriented nutrient agar plates to facilitate root growth, and total nuclei were isolated from the 1-cm root tip region. These nuclei were further purified either by treatment with 1% (v/v) Triton X-100 and sedimentation through a sucrose cushion (“Crude” purification) or affinity purified using streptavidin-coated magnetic beads (INTACT purification). In both cases 50,000 nuclei from each purification strategy were used as the input for ATAC-seq (Figure 1A). Overall, both Crude and INTACT-purified nuclei yielded very similar results (Figures 1B and 1C; Supplemental Figure 1). One clear difference that emerged was the number of reads that map to organellar DNA between the nuclear preparation methods. While the total reads of Crude nuclei preparations mapped ~50% to organellar genomes and 50% to the nuclear genome, the total reads of INTACT-purified nuclei consistently mapped over 90% to the nuclear genome (Table 1). The issue of organellar genomes contaminating ATAC-seq reactions is a common one, resulting in a large percentage of organelle-derived reads that must be discarded before further analysis. This issue was also recently shown to be remedied by increasing the purity of nuclei prior to ATAC-seq by use of fluorescence-activated nuclei sorting (Lu et al., 2017). To compare between data sets for the Crude and INTACT preparation strategies, we analyzed the enrichment of ATAC-seq reads using Hotspot peak mapping software (John et al., 2011). Though designed for use with DNase-seq data, Hotspot can also be readily used with ATAC-seq data. The number of enriched regions found with this algorithm did not differ greatly between nuclei preparation types, nor did the SPOT score (a signal specificity measurement representing the proportion of sequenced reads that fall into enriched regions) (Table 1). These results suggest that the data sets are generally comparable regardless of the nuclei purification method.

Visualization of the Crude- and INTACT-ATAC-seq data sets in a genome browser revealed that they were highly similar to one another and to DNase-seq data from whole root tissue (Figure 1B). Further evidence of similarity among these data sets was found by examining the normalized read count signal in all data sets (both ATAC-seq and DNase-seq) within the regions called “enriched” in the INTACT-ATAC-seq data set. For this and all subsequent peak calling in this study, we used the findpeaks algorithm in the HOMER package (Heinz et al., 2010), which we found to be more versatile and user-friendly than Hotspot. Using this approach, we identified 23,288 enriched regions in our INTACT-ATAC-seq data. We refer to these peaks, or enriched regions, in the ATAC-seq data as THSs. We examined the signal at these regions in the whole root DNase-seq data set and both Crude- and INTACT-ATAC-seq data sets using heat maps and average plots. These analyses showed that THSs detected in INTACT-ATAC-seq tended to be enriched in both Crude-ATAC-seq and DNase-seq signal (Figure 1C). In addition, the majority of enriched regions (19,516 of 23,288) were found to overlap between the root tip INTACT-ATAC-seq and the whole-root DNase-seq data (Figure 1D), and the signal intensity over DNase-seq or ATAC-seq enriched regions was highly correlated between the data sets (Supplemental Figure 1).

To examine the distribution of hypersensitive sites among data sets, we identified enriched regions in both types of ATAC-seq data sets and the DNase-seq data set and then mapped these regions to genomic features. We found that the distribution of open chromatin regions relative to gene features was nearly indistinguishable among the data sets (Figure 1E). In all cases, the majority of THSs (~75%) were outside of transcribed regions, with most falling within 2 kb upstream of a TSS and within 1 kb downstream of a transcript termination site (TTS).

Overall, these results show that ATAC-seq can be performed effectively using either Crude or INTACT-purified nuclei and that the data in either case are highly comparable to that of DNase-seq. While the use of crudely purified nuclei should be widely useful for assaying any tissue of choice without a need for transgenics, it comes with the drawback that ~50% of the obtained reads will be from organellar DNA. The use of INTACT-purified nuclei greatly increases the cost efficiency of the procedure and can also
provide access to specific cell types, but requires preestablished transgenic lines.

Comparison of Root Tip Open Chromatin Profiles among Four Species

Having established an efficient procedure for using ATAC-seq on INTACT affinity-purified nuclei, we used this tool to compare the open chromatin landscapes among four different plant species. In addition to the Arabidopsis INTACT line described above, we also generated constitutive INTACT transgenic plants of *M. truncatula*, rice, and tomato. Seedlings of each species were grown on vertically oriented nutrient plates for 1 week after radicle emergence, and nuclei from the 1-cm root tip regions of each seedling were isolated and purified with streptavidin beads. ATAC-seq was performed in at least two biological replicates for each species, starting with 50,000 purified nuclei in each case. Visualization of the mapped reads across each genome showed notable consistencies in the data for all four species. In all cases, the reads localize to discrete peaks that are distributed across the genome, as expected (Figure 2A). Examination of a syntenic region found in all four genomes suggested at least some degree of consistency in the patterns of transposase accessibility around orthologous genes (Figure 2A).

To specifically identify regions of each genome that were enriched in ATAC-seq signal (THSs), we used the HOMER find-peaks function on each biological replicate experiment. For further analysis, we retained only THS regions that were found in at least two biological replicates of ATAC-seq in each species. These reproducible THSs were then mapped to genomic features in each species in order to examine their distributions. As seen previously for Arabidopsis, the majority of THSs (~70–80%) were found...
outside of transcribed regions in all four species (Figure 2B). For this analysis, we classified these extragenic THSs (THSs found anywhere outside of transcribed regions) as proximal upstream (<2 kb upstream of the TSS), proximal downstream (<1 kb downstream of the TTS), or intergenic (>2 kb upstream from a TSS or >1 kb downstream from a TTS). The proportion of THSs in the proximal upstream and intergenic regions varied greatly with genome size and, thus, the amount of intergenic space in the genome. For example, a full 52% of THSs in Arabidopsis, 15% in rice, 23% in tomato, and 7% in M. truncatula (820 Mb). The next analysis focused on the distance of THSs from the nearest gene. It is worth noting that the distribution of THSs, as proximal as expected, with 12% of the THSs in intergenic regions for Arabidopsis, 30% for rice and 50% for tomato. The proportion of THSs that are more than 10 kb away from the TSS, whereas tomato genes tend to have a larger number of elements per gene based on clade-specific alterations in gene regulation, DNA copy number changes, or simply the greater abundance of transposons and other repeat elements is not entirely clear. Compared with the other species, tomato THSs are much more abundant and tend to be smaller in size than those of the other species, and the tomato ATAC-seq data generally appear to have a lower signal-to-noise ratio (Figure 2A). While it is

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Plastid Mapped Reads (%)</th>
<th>Mitochondrial Mapped Reads (%)</th>
<th>Nuclear Mapped Reads (%)</th>
<th>Total Nuclear Mapped Reads (×10^6)</th>
<th>Total Hotspot Enriched Regions Called</th>
<th>SPOT Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude 1</td>
<td>25.33</td>
<td>22.15</td>
<td>52.52</td>
<td>40.6</td>
<td>43,599</td>
<td>0.4339</td>
</tr>
<tr>
<td>Crude 2</td>
<td>24.40</td>
<td>21.03</td>
<td>54.58</td>
<td>31.0</td>
<td>43,043</td>
<td>0.4086</td>
</tr>
<tr>
<td>Crude 3</td>
<td>25.13</td>
<td>23.17</td>
<td>51.70</td>
<td>35.8</td>
<td>42,469</td>
<td>0.4471</td>
</tr>
<tr>
<td>INTACT 1</td>
<td>4.82</td>
<td>2.44</td>
<td>92.94</td>
<td>34.6</td>
<td>36,463</td>
<td>0.4167</td>
</tr>
<tr>
<td>INTACT 2</td>
<td>3.51</td>
<td>2.03</td>
<td>94.46</td>
<td>34.0</td>
<td>41,305</td>
<td>0.4004</td>
</tr>
<tr>
<td>INTACT 3</td>
<td>2.81</td>
<td>1.61</td>
<td>95.57</td>
<td>89.7</td>
<td>55,857</td>
<td>0.4896</td>
</tr>
</tbody>
</table>

ATAC-seq was performed in biological triplicate for both Crude and INTACT-purified nuclei. For each replicate, the table shows the percentage of reads mapping to organelle and nuclear genomes, the total number of enriched regions identified by the peak calling program Hotspot, as well as the SPOT score for each data set. The SPOT score is a measure of specificity describing the proportion of reads that fall in enriched regions, with higher scores indicating higher specificity.
unclear why the data from tomato are distinct in these ways, it is clear that tomato THSs occupy mostly genic regions of the genome, as expected, and are highly reproducible between biological replicate experiments (Supplemental Figure 2).

Collectively, these results suggest that there is a relatively small number of regulatory elements per gene in plants. These elements tend to be focused near the promoter rather than at more distal sites, as has been observed in animal, particularly mammalian, genomes (Stadhouders et al., 2012). The assumptions implicit in this argument are that open chromatin sites near a TSS reflect regulatory elements that regulate that TSS and not a more distant one and that upstream elements contribute the majority of regulatory effects. These assumptions appear to be generally validated by many reporter assays showing that an upstream fragment of several kilobases is frequently sufficient to recapitulate native transcription patterns (Medford et al., 1991; Masucci et al., 1996; Ruzicka et al., 2007; Tittarelli et al., 2009; Li et al., 2012), as well as our observation that upstream THSs are the most abundant class of open chromatin sites.

Open Chromatin Features Are Not Directly Conserved among Orthologous Genes

Given that many of the properties of open chromatin regions were shared among Arabidopsis, M. truncatula, rice, and tomato, we next asked whether the numbers and locations of THSs—and thus putative regulatory elements—were conserved among orthologous genes across species. For these analyses, we identified 373 syntenic orthologs (Supplemental Data Set 2) that were found in all four genomes and asked whether members of each ortholog set harbored a similar number of open chromatin regions across the species. Again, using root tip THSs present in at least two
biological replicates for each species, we counted the number of THSs within 5 kb upstream of the TSS for each ortholog in each species. We then examined these data for similarities and differences in upstream THS number (Figure 3A). While no clear trend of strong conservation in the number of upstream THSs emerged from this analysis, there was a small subset of orthologs that did have upstream THSs in similar numbers across species. However, this was a very small proportion of the total. As seen in earlier analyses, tomato genes tended to have a larger number of upstream THSs compared with the other species, and most of the 373 orthologs in tomato did have at least one upstream THS. This was not the case in the other three species, where many of the orthologs had no detectable upstream THSs within 5 kb of the TSS. Among the four species, Arabidopsis and M. truncatula showed the greatest similarity in upstream THS number, but even in this case the similarity was minimal despite the relatively closer phylogenetic relationship between these two organisms.

We next examined the distribution of open chromatin regions across the upstream regions of these 373 orthologous genes relative to their expression level in Arabidopsis, reasoning that there could be patterns of open chromatin similarity based on THS positions, rather than numbers. For this analysis, we examined the normalized ATAC-seq signal across the upstream region of all 373 orthologous genes, from −5000 bp to +100 bp relative to the TSS of each gene (Figure 3B). Orthologs were then ranked within the heat map based on the transcript level of each Arabidopsis ortholog in the root tip (Li et al., 2016a), from highest to lowest expression. For each Arabidopsis ortholog, we also included the upstream THS number to ascertain how this feature might correlate with transcript level for Arabidopsis. While there was some consistency among species in that open chromatin often overlapped with the TSS, we did not observe any clear pattern in transposase hypersensitivity within the upstream regions of these orthologs. K-means clustering of the heat maps similarly did not reveal evidence for conservation of open chromatin patterns among orthologs (Supplemental Figure 3A). An important caveat to this analysis is that many of these syntenic orthologs may not be functional homologs, or “expressologs” (Patel et al., 2012), due to subfunctionalization within gene families. As such, we identified a smaller group (52) of expressologs on which to perform a similar test (Supplemental Data Set 3). While these expressolog genes have both maximally high protein level similarity and expression pattern similarity, including expression in the root, there was also no clear correspondence in upstream THS number among them (Supplemental Figure 3B).

There does not appear to be strong conservation in the number and location of open chromatin sites at orthologous genes across species. Assuming that these genes are still under control of common TFs, this suggests that regulatory elements could be free to migrate, and perhaps split or fuse, while retaining the regulatory parameters of the target gene in question.

One interesting finding from these analyses was that the pattern of upstream THS number does not correlate with expression level, at least for Arabidopsis (Figure 3B). Thus, THSs must not simply represent activating events upstream of the TSS but may also represent binding of repressive factors. Furthermore, we found no correlation between upstream THS number and expression entropy among all genes in the Arabidopsis genome, suggesting a more complex relationship between regulatory element distribution and target gene transcription (Supplemental Figure 3C).

**Evidence for Co-regulation of Common Gene Sets by Multiple TFs across Species**

While there does not appear to be a consistent pattern in the number or placement of open chromatin regions around orthologs or expressologs, we wanted to examine whether it would be possible to find common regulators of specific gene sets among species using a deeper level of analysis. To do this, we first searched for common TF motifs in root tip THSs across the four species. Using the THSs that were found in at least two replicates for each species, we employed the MEME-ChIP motif analysis package (Machanick and Bailey, 2011; Ma et al., 2014) to identify overrepresented motifs of known TFs. We discovered 30 motifs that were both overrepresented and common among all species (Supplemental Data Set 4). We narrowed our list of candidate TFs by considering a variety of factors, including the expression of each TF in the root tip, any known mutant root phenotypes involving those TFs, and whether genome-wide binding information was available for each candidate in Arabidopsis. Ultimately, we selected four TFs for further analysis: HY5, ABSCISIC ACID RESPONSIVE ELEMENTS BINDING FACTORS (ABF3), C-REPEAT/DRE BINDING FACTOR2 (CBF2), and MYB77. It is worth noting that among these factors, both HY5 and MYB77 had been previously implicated in root development (Oyama et al., 1997; Zhao et al., 2014). Like HY5 and MYB77, CBF2 and ABF3 have been implicated in stress responses as well as abscisic acid (ABA) signaling (Kang et al., 2002; Knight et al., 2004). Furthermore, overexpression of ABF3 leads to increased tolerance to multiple abiotic stresses in Arabidopsis, rice, cotton (Gossypium hirsutum), and alfalfa (Oh et al., 2005; Abdeen et al., 2010; Wang et al., 2016; Kerr et al., 2017). Given this evidence, we decided to focus on these factors for further study.

We first sought to define the target genes for each of these four TFs in Arabidopsis by combining our chromatin accessibility data with published genome-wide binding data for each factor in Arabidopsis (Table 2). Because an accessible chromatin region (a THS) represents the displacement of nucleosomes by a DNA binding protein, we reasoned that our THS profiles for a given tissue would represent virtually all possible protein binding sites in the epigenomes of root tip cells. Similarly, by using in vitro genomic binding data (DAP-seq) (O’Malley et al., 2016) or ChIP-seq data from a highly heterogeneous tissue, we could identify the spectrum of possible binding sites for that TF, such that the intersection of these data sets would represent the binding sites for that TF in the sample of interest. While there are caveats to this approach, we reasoned that it was more likely to generate false negatives than false positives and would give us a set of high confidence target genes to analyze for each TF. In this regard, ChIP-seq data may be more robust because they represent in vivo binding, while DAP-seq is an in vitro assay and may not capture binding sites that depend on chromatin properties or interactions with other TFs. On the other hand, ChIP-seq data are inherently limited by the cell types present in the sample used.

We first tested this approach in Arabidopsis with each of the four TFs of interest. Using THSs from the Arabidopsis root tip that were
found in at least two biological replicates, we used the motif identification tool FIMO (Grant et al., 2011) to identify THSs that contained a significant occurrence of the TF motif of interest. The THSs that contained a significant motif match were considered predicted binding sites. We then identified predicted binding sites that also overlapped with a known binding site for that TF (a DAP-seq or ChiP-seq peak), and these were considered high confidence binding sites for that TF in the root tip (Supplemental Figure 4). The predicted binding sites (motif-containing THSs) were themselves very good predictors of the true binding sites for these four TFs (Table 2). For example, of the 1316 Arabidopsis root tip THSs with an occurrence of the ABF3 motif (Mathelier et al.,

Figure 3. Characterization of Open Chromatin Regions and Regulatory Elements in Arabidopsis, M. truncatula, Tomato, and Rice.

(A) Heat map showing the number of upstream THSs at each of 373 syntenic orthologs in each species. Each row of the heat map represents a syntenic ortholog, and the number of THSs within 5 kb upstream of the TSS is indicated with a black-to-red color scale for each ortholog in each species. Hierarchical clustering was performed on orthologs using uncentered correlation and average linkage.

(B) Normalized ATAC-seq signals upstream of orthologous genes. Each row of the heat maps represents the upstream region of one of the 373 syntenic orthologs in each species. ATAC-seq signal is shown across each ortholog from +100 to −5000 bp relative to the TSS, where blue is high signal and white is no signal. Heat maps are ordered by transcript level of each Arabidopsis ortholog in the root tip, from highest (top) to lowest (bottom). The leftmost heat map in black-to-red scale indicates the number of upstream THSs from −2100 to −25000 bp associated with each of the Arabidopsis orthologs, on the same scale as in (A).

(C) Overlap of predicted target genes for HY5, ABF3, CBF2, and MYB77 in the Arabidopsis root tip. Predicted binding sites for each factor are those THSs that also contain a significant motif occurrence for that factor. Venn diagram shows the numbers of genes with predicted binding sites for each factor alone and in combination with other factors. Significance of target gene set overlap between each TF pair was calculated using a hypergeometric test with a population including all Arabidopsis genes reproducibly associated with an ATAC-seq peak in the root tip (13,714 total genes). For each overlap, we considered all genes cotargeted by the two factors.

(D) Conveying data similar to that in (C), the clustered bar graph shows the percentage of total target genes that fall into a given regulatory category (targeted by a single TF or combination of TFs) in each species.
2.1. Target genes, 297 (23%) are also targeted by HY5 (hypergeometric expected by chance (Figure 3C). For example, of the 1271 ABF3 TFs, with gene sets being targeted by one, two, three, or all four of these TFs to a degree that was far higher than what would be expected by chance (Figure 3C). For example, of the 1271 ABF3 target genes, 297 (23%) are also targeted by HY5 (hypergeometric P = 2.1 × 10^−56). Among these 297 genes, 46 are targeted by ABF3, HY5, and CBF2, and seven are targeted by all four TFs. We also asked where the binding sites driving this pattern were located relative to the target genes. To do this, we considered only binding sites within the 5-kb upstream region of a TSS and repeated the same THS (Supplemental Figure 5A). These results collectively suggest that these four TFs have important roles in root tip gene regulation both individually and in combination and that the majority of their binding sites (~80%) fall within the 5-kb region upstream of the TSS for target genes. In addition, we find that the binding sites for multiple TFs often occur in the same THS (Supplemental Figure 5B).

We next sought to examine the target genes and proportions of target gene overlaps between the four species to address the conservation of coregulatory relationships among these four TFs. Given that no TF binding data are available for the other three species and knowing that the majority of our predicted binding sites in Arabidopsis corresponded to known binding sites (Table 2; 61–97%), we opted to also use the predicted binding sites for each of the four TFs in M. truncatula, tomato, and rice, with the knowledge that these sets may contain some false positives. For these analyses, we used the Arabidopsis TF motifs, since these have not been directly defined for the other species, with the caveat that the DNA binding specificity of these factors may not be identical among species.

We again used FIMO to identify significant occurrences of each TF motif within the root tip THSs found in at least two biological replicates for each of our four species. We then mapped the predicted binding sites of each TF to the nearest TSS to define target genes for each TF in each species (Supplemental Data Set 5). We then analyzed the overlap of TFs at target genes in each species using four-way Venn diagrams, similar to Figure 3C. To compare regulatory associations across species, we considered each of the 15 categories in every species-specific four-way Venn diagram as a regulatory category. For example, one regulatory category consists of the genes targeted only by ABF3 alone, another would be those targeted only by HY5 and ABF3 at the exclusion of the other two TFs, and so on. For each regulatory category in each species, we calculated the percentage of the total target genes in that category (number of genes in the regulatory category/total number of genes targeted by any of the four TFs) and then compared these percentages between species (Figure 3D). We found remarkably consistent proportions of the target genes in nearly all regulatory categories across all four species. However, notable deviations from this consistency among species were seen in the proportion of rice genes targeted by MYB77 alone and rice genes targeted by CBF2 and HY5 together. In most cases, the proportions of target genes in different regulatory categories were most similar between Arabidopsis and M. truncatula, and these were generally more similar to tomato than to rice, consistent with the evolutionary distances between the species (Vanneste et al., 2014). Commonly over-represented Gene Ontology (GO) terms among gene sets in particular regulatory categories across species further support the notion of regulatory conservation (Supplemental Figure 5C), although these analyses are limited by the depth of GO annotation in some of these species.

These findings suggest that while neither synteny orthologous gene sets nor expressolog gene sets tend to share open chromatin patterns, the genes under control of specific TFs or specific

<table>
<thead>
<tr>
<th>TF Family</th>
<th>Average Expression in Arabidopsis Root Tip (RPKM)</th>
<th>Percentage of Motif Occurrences in the Arabidopsis Genome That Overlap with a Known Binding Site</th>
<th>Percentage of Motif-Containing THSs in Arabidopsis That Overlap with a Known Binding Site (High-Confidence Binding Sites)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT5G11260 (HY5)</td>
<td>bZIP</td>
<td>37</td>
<td>30.3% (7,156/23,541)</td>
</tr>
<tr>
<td>AT4G34000 (ABF3)</td>
<td>bZIP</td>
<td>1</td>
<td>53.3% (3,821/7,164)</td>
</tr>
<tr>
<td>AT4G25470 (CBF2)</td>
<td>AP2</td>
<td>16</td>
<td>64.2% (13,144/20,457)</td>
</tr>
<tr>
<td>AT3G50060 (MYB77)</td>
<td>MYB/SANT</td>
<td>6</td>
<td>52.5% (9,147/17,402)</td>
</tr>
</tbody>
</table>

THSs found in at least two replicates for each species were analyzed for overrepresented TF motifs. Four of the 30 TFs that were significantly enriched in THSs of all four species are shown in the table. Significant occurrences of each TF motif were identified across the Arabidopsis genome, and the percentage of these motif occurrences that fall within known binding sites for that factor (based on published ChIP-seq or DAP-seq data sets) are indicated in column 4. The final column indicates the percentage of Arabidopsis root tip THSs that contain a motif for each factor and also overlap with a known binding site for the factor. These are considered high-confidence binding sites (Supplemental Figure 4).
Landscapes of Arabidopsis Root Epidermal Cell Types

Having examined questions of regulatory conservation between species, we then explored regulatory element and TF relationships between cell types within a single species. In this case, we chose to focus on the root epidermal hair and non-hair cell types in Arabidopsis. Since these two cell types are derived from a common progenitor, they are prime candidates to offer insight into the epigenetic alterations that occur during—and likely drive—cell differentiation. Specifically, we investigated to what extent the open chromatin landscapes would differ between cell types and whether differences in THSs could pinpoint the sites of differential transcriptional regulation. Furthermore, we wanted to understand whether we could use this information to examine the TF-to-TF regulatory connections that underlie the transcriptomic and physiological differences between these cell types.

We used two previously described INTACT transgenic lines as starting material for these experiments: one having biotin-labeled nuclei exclusively in the root hair (H) cells and another with labeled nuclei only in the root epidermal non-hair (NH) cells (Deal and Henikoff, 2010). Nuclei were purified from each fully differentiated cell type by INTACT, and 50,000 nuclei of each type were subjected to ATAC-seq. Visualization of these cell-type-specific data sets in a genome browser, along with the Arabidopsis whole 1-cm root tip ATAC-seq data, showed a high overall degree of similarity among the three data sets (Figure 4A). Comparison of the ATAC-seq signal intensity at common THS regions genome-wide revealed that these two cell types have open chromatin patterns that are highly similar to one another, but distinct from that of the whole root tip (Supplemental Figure 7).

To identify regions of differential accessibility between the cell types and the whole root tip, we considered THS regions that were found in at least two biological replicates of each cell type or tissue. The total number of these reproducible THSs was 32,942 in the whole root tip, 35,552 for the H cells, and 28,912 for the NH cells. The majority of these sites (18,742) were common (overlapping) in all three sample types (Figure 4B) and thus likely represent regulatory sites that are utilized in multiple Arabidopsis root cell types. We also found 6582 THSs that were common to both root epidermal cell types but were not found in the whole root tip, suggesting that these may represent epithelial-specific regulatory elements. In a search for unique THSs in each of the three sample types (those not overlapping with a THS in any other sample), we found 10,455 THSs that were unique to the whole root tip, 7537 unique to the H cells, and 2574 that were unique to the NH cells. We refer to these regions as differential THSs (dTHSs). The dTHSs identified only in the H or NH cell type were of further interest because they may represent regulatory elements that drive the transcriptomic differences between these two epidermal cell types.

To examine the extent of chromatin accessibility differences at these dTHSs, we visualized the accessibility signals from each cell type at both H cell dTHSs and NH cell dTHSs. First, using the 7537 regions identified as H cell dTHSs, we used heat maps and average plots to examine the normalized ATAC-seq read count across these regions in each cell type (Figure 4C, left panel). We then repeated this analysis using the 2574 NH cell dTHSs (Figure 4C, right panel). In each case, it was clear that the regions we identified as dTHSs showed significant differences in chromatin accessibility between the two cell types. However, the differences in chromatin accessibility between cell types were quantitative (varying intensity) rather than qualitative (all-or-nothing). This indicates that, at large, the dTHSs represent sites that are highly accessible in one cell type and less so in the other, rather than being strictly present in one and absent in the other. Therefore, we refer to these sites from this point on as cell-type-enriched dTHSs to convey the notion of quantitative differences between cell types.

To identify the genes that might be impacted by cell-type-enriched dTHSs, we mapped each dTHS to its nearest TSS and considered that to be the target gene. We found that the 7537 H-enriched dTHSs mapped to 6008 genes, while the 2574 NH-enriched dTHSs mapped to 2295 genes. Thus, the majority of genes that are associated with a dTHS are only associated with one such site. This is consistent with our previous finding that most Arabidopsis genes are associated with a single upstream THS (Figure 2D).

We then asked how the set of genes associated with dTHSs overlapped with those whose transcripts show differential abundance between the two cell types. Using data from a recent comprehensive RNA-seq analysis of flow sorted Arabidopsis root cell types (Li et al., 2016a), we identified sets of transcripts that were more highly expressed in H versus NH cell types. To be considered a cell-type-enriched gene, we required a gene to have a transcript level with twofold or greater difference in abundance between H and NH cell types, as well as at least five reads per...
kilobase per million mapped reads (RPKM) in the cell type with a higher transcript level. Using this relatively conservative approach, we derived a list of 3282 H cell-enriched genes and 2731 NH cell-enriched genes. We then asked whether the genes associated with cell-type-enriched dTHSs were also cell-type-enriched genes (Figure 4D). Of the 3282 H cell-enriched genes, 743 were associated with an H cell-enriched dTHS, 258 were associated with a NH cell-enriched dTHS, and 108 genes were associated with a dTHS in both cell types. Among the 2731 NH cell-enriched genes, 156 were associated with a NH cell-enriched dTHS, 516 were associated with a H cell-enriched dTHS, and 52 genes showed dTHSs in both cell types. These results suggest that cell-type-enriched expression of a gene is frequently associated with a dTHS in the cell type where the gene is highly expressed but is also often associated with a dTHS in the cell type where that gene is repressed. This highlights the importance of transcriptional activating events in the former case and repressive events in the latter. Interestingly, for a smaller set of cell-type-enriched genes, we observed dTHSs at a given gene in both cell types, indicating regulatory activity at the gene in both cell types.

We next asked what proportion of the transcriptome differences between H and NH cells might be explained based on differential chromatin accessibility. Of the 3282 H cell-enriched genes, 1109 have a dTHS in one or both of the cell types, and among the 2731 NH cell-specific genes, 724 have a dTHS in one or both cell types. Assuming that each dTHS represents a regulatory event contributing to the differential expression of its identified target gene, we could explain differential expression of 33% of the H cell-enriched genes and 27% of the NH cell-enriched genes. The remaining ~70% of the identified cell-type-enriched genes without clear chromatin accessibility differences may be explained in numerous ways. These genes may not require a change in chromatin accessibility, changes in chromatin accessibility may fall below our limit of detection, or these transcripts may be primarily regulated at the posttranscriptional level rather than at the chromatin accessibility level that we measured.
Another key question relates to the significance of the cell-type-enriched dTHSs that do not map to differentially expressed genes. These could be explained by an inability to detect all differentially expressed genes, perhaps simply due to the stringency of our definition of cell-type-enriched genes. An important biological possibility to consider is that many of these regulatory regions do not in fact regulate the closest gene, but rather act over a distance such that they are orphaned from their true target genes in our analysis. Another possibility is that many of the different protein binding events represented by these dTHSs are unrelated to transcriptional regulation.

Overall, the accessible chromatin landscapes of the root epidermal H and NH cells appear to be nearly identical in a qualitative sense, but differ significantly at several thousand sites in each cell type. The reasons for the quantitative, rather than all-or-nothing, nature of this phenomenon are not entirely clear. Are the accessibility differences between cell types reflective of unique protein assemblages at the same element in different cell types, or do they instead reflect differences in abundance of the same proteins at an element in different cell types? While these questions certainly warrant further investigation and experimentation, we can gain further insight into the regulatory differences between cell types through deeper examination of the differentially accessible chromatin regions in each.

**TF Motifs in Cell-Type-Specific THSs Identify Regulators and Their Target Genes**

As a means of identifying specific TFs that might be important in specifying the H and NH cell fates, we sought to identify overrepresented motifs in the differentially accessible regions of each cell type. We used each set of cell-type-enriched dTHSs as input for MEME-ChIP analyses (Machanick and Bailey, 2011) and examined the resulting lists of overrepresented motifs. We initially found 219 motifs that were significantly overrepresented relative to genomic background only in H-cell-enriched dTHSs and 12 that were significantly overrepresented only in NH-cell-enriched dTHSs (Supplemental Data Set 6). To narrow our list of candidate TFs to pursue, we vetted these lists of potential cell-type-enriched TFs by considering their transcript levels in each cell type as well as the availability of genome-wide binding data. Based on the available data, we narrowed our search to five transcription factors of interest: four H-cell-enriched TF genes (MYB33, ABI5, NAC083, and At5g04390) and one NH-enriched TF gene (WRKY27) (Table 3).

We next attempted to directly identify the binding sites for each TF by differential ATAC-seq footprinting between the cell types. The logic behind this approach is the same as that for DNase-seq footprinting: that the regions around a TF binding site are hypersensitive to the nuclease or transposase due to nucleosome displacement, but the sites of physical contact between the TF and DNA will be protected from transposon insertion/cutting and thus leave behind a characteristic “footprint” of reduced accessibility on a background of high accessibility (Hesselberth et al., 2009; Vierstra and Stamatoyannopoulos, 2016). We reasoned that we could identify binding sites for each of these cell-type-enriched TFs by comparing the footprint signal at each predicted binding site (a motif occurrence within a THS) between H and NH cells.

For this analysis, we examined the transposase integration patterns around the motifs of each TF in both cell types as well as in purified genomic DNA subjected to ATAC-seq, to control for transposase sequence bias. It was recently reported in Arabidopsis that many TF motifs exhibit conspicuous transposase integration bias on naked DNA (Lu et al., 2017), and our results were in line with these findings for all five TFs of interest here (Supplemental Figure 8). While we observed footprint-like patterns in the motif-containing THSs in our ATAC-seq data, these patterns in each case were also evident on purified genomic DNA. As such, it was not possible to distinguish true binding sites from these data, as any footprint signal arising from TF binding was already obscured by the transposase integration bias. For unknown reasons, many TF motif DNA sequences seem to inherently evoke hyper- and/or hypo-integration by the transposase, and this automatically obscures any potentially informative footprint signal that could be obtained by integration during ATAC-seq on nuclei. Similar technical concerns have also been raised for DNase footprinting (Sung et al., 2016). These results suggest that the ATAC-seq footprinting approach may be useful for certain TFs, but these will likely need to be examined on a case-by-case basis.

**Table 3. TF Motifs Overrepresented in Cell Type-Enriched dTHSs**

<table>
<thead>
<tr>
<th>TF</th>
<th>Family</th>
<th>Cell Specificity</th>
<th>Hair/Non-Hair FPKM Ratio</th>
<th>Percentage of Genomic Motif Occurrences in Known Binding Sites</th>
<th>Percentage of Motif-Containing THSs That Overlap a Known Binding Site (High-Confidence Binding Sites)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT5g06100 (MYB33)</td>
<td>MYB</td>
<td>Hair</td>
<td>2000</td>
<td>42.2% (7,038/16,655)</td>
<td>69.9% (1,473/2,106)</td>
</tr>
<tr>
<td>AT2g36270 (ABI5)</td>
<td>bZIP</td>
<td>Hair</td>
<td>3.3</td>
<td>26.2% (7,261/27,656)</td>
<td>57.5% (2,814/4,891)</td>
</tr>
<tr>
<td>At5g04390</td>
<td>C2H2</td>
<td>Hair</td>
<td>17</td>
<td>9.5% (3,850/40,305)</td>
<td>8.5% (282/3,290)</td>
</tr>
<tr>
<td>At5g13180 (NAC083)</td>
<td>NAC</td>
<td>Hair</td>
<td>2</td>
<td>51.3% (13,762/26,815)</td>
<td>48.3% (1,169/2,419)</td>
</tr>
<tr>
<td>ATG5G2830 (WRKY27)</td>
<td>WRKY</td>
<td>Non-hair</td>
<td>0.35</td>
<td>15.3% (4,458/29,126)</td>
<td>23.6% (1,169/2,419)</td>
</tr>
</tbody>
</table>

Cell-type-enriched dTHSs were analyzed for overrepresented TF motifs using MEME-ChIP software, and several significantly matching factors are shown in the table. Cell specificity indicates the cell-type-enriched dTHS set from which each factor was exclusively enriched, and hair/non-hair FPKM ratio indicates expression specificity of each factor using RNA-seq data from Li et al. (2016a). Significant occurrences of each TF motif were identified across the Arabidopsis genome, and the percentage of these motif occurrences that fall within known binding sites for that factor (based on published ChIP-seq or DAP-seq data sets) are indicated in column 5. Percentages are calculated by the number of motif occurrences in known binding sites/total number of motif occurrences in the genome. Column 6 indicates the percentage of THSs from the relevant cell type that contain a motif for a factor and also overlap with a known binding site for the factor (high-confidence binding sites).
Given this issue and the resulting lack of evidence for footprints of our TFs of interest, we decided to take the approach of defining TF target sites as we did for our studies of root tip TFs.

As described earlier, we defined high confidence binding sites for the five TFs of interest as TF motif-containing THSs in the cell type of interest (predicted binding sites) that also overlapped with an enriched region for the TF in publicly available DAP-seq data (O’Malley et al., 2016) or ChIP-seq data (Supplemental Figure 4). Assigning these high confidence binding sites to their nearest TSS allowed us to define thousands of target genes for these factors in the root epidermal cell types (Table 3; Supplemental Data Set 7). Compared with our analysis of root tip TFs, our capability to predict target sites based on motif occurrences in THSs was much reduced for the four H-cell-enriched and one NH-cell-enriched TFs examined here. For further analyses, we decided to focus on three of the TFs that were more highly expressed in the H cell type and had the largest number of high confidence target genes: ABI5, MYB33, and NAC083.

We first asked how many of the high confidence target genes for these TFs were also preferentially expressed in one cell type or the other. We found that for all three TFs, a large percentage of the total target genes are H cell enriched in their expression (17–21%), while many others are NH cell enriched (6–9%) (Figure 5A). These results are intriguing as they suggest that the activities of these TFs may be generally context dependent. At the same time, however, the majority of the target genes for each TF were not more highly expressed in one cell type compared with the other.

Each of these H-cell-enriched TFs could activate other H-cell-enriched genes, but what are their functions at regulatory elements near genes that are expressed at low levels in the H cell and

![Figure 5](image-url)
high levels in the NH cell? One possibility is that these factors are activators of transcription in the context of H-cell-enriched genes but act as repressors or are neutral toward the target genes that are NH-cell-enriched in their expression. This may reflect context dependency in the sense that the effect on transcription of a target gene may depend on the local milieu of other factors.

We next examined whether ABI5, MYB33, and NAC083 target any of the same genes. Similar to the root tip TFs examined previously, we found that these three TFs also appear to have extensive coregulatory relationships (Figure 5B). For example, 207 target genes were shared between ABI5 and NAC083, 238 were shared between ABI5 and MYB33, and 50 target genes were shared by all three factors. We further analyzed the genes that were cotargeted by ABI5 and MYB33, finding that 57 of the cotargeted genes were H-cell enriched. As such, we performed GO analysis on the H-cell-enriched targets as well as the full set of target genes to gain insight into the functions of this coregulatory relationship (Figure 5C). Many of the ABI5/MYB33 target genes were annotated as being involved in responses to ABA as well as water, salt, and cold stress. This is consistent with the known roles of these proteins in ABA signaling (Finkelstein and Lynch, 2000; Reyes and Chua, 2007). Interestingly, seven of the 57 ABI5/ MYB33 target genes that were H cell-enriched were also annotated with the term regulation of transcription, suggesting that ABI5 and MYB33 may be at the apex of a transcriptional regulatory cascade in the H cell type.

Identification of a New Regulatory Module in the Root Hair Cell Type

Based on our findings that ABI5 and MYB33 cotarget seven H-cell-enriched TFs, we decided to investigate this potential pathway further. Among the seven TFs putatively coregulated by ABI5 and MYB33 and having H-cell-enriched transcript expression were DEAR5, ERF11, At3g49930, SCL8, NAC087, and two additional MYB factors: MYB44 and MYB77. Aside from MYB77, none of these TFs had been previously reported to produce root-specific phenotypes when mutated. MYB77 was previously shown to interact with auxin response factors (Shin et al., 2007) and to be involved in lateral root development through promotion of auxin-responsive gene expression (Shin et al., 2007). Interestingly, the ABA receptor, PYL9, was shown to physically interact with both MYB77 and MYB44 and to promote auxin-responsive transcription by MYB77 (Zhao et al., 2014). MYB44 has also been implicated in ABA signaling through direct interaction with an additional ABA receptor, PYL9 (Li et al., 2014), as well as repression of jasmonic acid (JA)-responsive transcription (Jung et al., 2010). These factors have additionally been implicated in salicylic acid (SA) and ethylene signaling (Yanhui et al., 2006; Shim et al., 2013). Given that MYB44 and MYB77 are paralogs (Dubos et al., 2010) that appear to integrate multiple hormone response pathways in a partly redundant manner (Jaradat et al., 2013), we decided to identify high confidence target genes (Supplemental Figure 4) for each of them for further study.

We again defined high confidence binding sites as THSs in H cells that contain a significant motif occurrence for the factor and also overlap with a DAP-seq or ChIP-seq enriched region for that factor. Using this approach, we found that MYB44 and MYB77 each target over 1000 genes individually and cotarget 483 genes (Figure 6A). In addition, MYB44 and MYB77 appear to regulate one another, while MYB77 also appears to target itself. This feature of self-reinforcing coregulation could serve as an amplifying and sustaining mechanism to maintain the activity of this module once activated by ABI5, MYB33, and potentially other upstream factors.

To gain a deeper understanding of the impact of MYB44 and MYB77 on downstream processes, we performed GO analysis of the target genes for each factor. First considering all target genes, regardless of their expression in the H cell type, we found a variety of overrepresented GO terms for each that were consistent with the known roles of these factors in hormone signaling (Figure 6B). For example, both factors targeted a large number of genes annotated with the terms response to ABA stimulus, response to ethylene stimulus, and response to SA stimulus. Additionally, MYB44 alone targeted many genes with the annotation response to JA stimulus, consistent with its previously reported role as a negative regulator of JA signaling (Jung et al., 2010). Interestingly, the largest overrepresented gene functional category for both factors was transcription factor activity (102 genes for MYB77 and 183 genes for MYB44). This indeed further suggests that these factors initiate a cascade of transcriptional effects. The next-largest overrepresented term was plasmodesma, indicating that production and/or regulation of cell-cell connecting structures are likely controlled by these factors. Plasmodesmata are important for numerous epidermal functions including cell-to-cell movement of TFs such as CPC and TRY (Schellmann et al., 2002; Wada et al., 2002) and transport of other macromolecules and metabolites (Lucas and Lee, 2004).

We also analyzed overrepresented ontology terms in the MYB77 and MYB44 targets that were classified as H-cell-enriched genes. Among the MYB77 target genes in this category were known regulators of H cell fate, while numerous H-cell-enriched MYB44 target genes were annotated as being involved in response to water and phosphate starvation (Figure 6C). The ontology category that was overrepresented in both target lists was negative regulation of transcription (six MYB77 targets and seven MYB44 targets), suggesting that these factors exert additional specific effects on the H cell transcriptome by regulating a subset of potentially repressive TFs.

The fact that MYB77 and MYB44 target a large number of genes that show H-cell-enriched expression suggests that these factors serve as activators of transcription at these targets, and this is supported by published accounts of transcriptional control by these factors (Persak and Pitzschke, 2014). However, both factors also target NH-cell-enriched genes as well as genes without preferential expression between the cell types. This phenomenon was also observed for the H-enriched TFs ABI5, MYB33, and NAC083 (Figure 5), suggesting that certain TFs may generally serve as activators but may also have context-dependent repressive functions. Such a functional switch could occur through direct mechanisms such as structural alteration by alternative splicing or posttranslational modification, functional alteration by partnering with a specific TF or chromatin-modifying complex, or perhaps indirectly by binding to a target site to occlude the binding of other factors necessary for transcriptional activation. The numerous reports of dual function transcription factors in animals and plants support the notion that this may be a general
phenomenon (Ikeda et al., 2009; Boyle and Després, 2010; Li et al., 2016b).

Collectively these results suggest that the MYB44/MYB77 module in the H cell specifies a cascade of downstream transcriptional regulation, some of which is positive and some of which is negative. This module likely represents an important hub in controlling H cell fate as well as a variety of physiological functions and environmental responses in this cell type. The fact that MYB77 was also discovered in our analyses of root tip TFs suggests that this factor likely has a broader role in other cell types during early development.

Figure 6. A Transcriptional Regulatory Module in the Root Hair Cell Type.

(A) Diagram of the proposed regulatory module under control of ABI5 and MYB33. As referenced in Figure 5C, ABI5 and MYB33 cotarget seven TFs that are preferentially expressed in the hair cell relative to the non-hair cell type. The family classification of each of the seven TFs is denoted in the figure key. Among the seven hair-cell-specific target TFs are two MYB family members, MYB77 and MYB44. High-confidence binding sites for these two MYB factors were again defined as open chromatin regions in the hair cell that contain a significant motif occurrence for the factor and also overlap with a known enriched region for that factor from DAP-seq or ChIP-seq data. Each high-confidence binding site was then assigned to the nearest TSS to define the target gene for that site. This analysis revealed that MYB44 and MYB77 target each other, and MYB77 targets itself. Both factors target thousands of additional genes, 483 of which are in common (Venn diagram on the lower right of the schematic). Arrows pointing down from MYB77 and MYB44 indicate GO analyses of that factor’s target genes.

(B) and (C) The upper tables (B) represent enriched annotation terms for all target genes of the factor, regardless of differential expression between H and NH cells, while the lower tables (C) represent enrichment of terms within target genes that are preferentially expressed in the hair cell relative to the non-hair cell. Annotation term levels are indicated as cellular component (CC), biological process (BP), molecular function (MF), or KEGG pathway (KEGG). For each annotation, the number of target genes associated with that term is shown to the right of the term, followed by the FDR-corrected P value for the term enrichment in the rightmost column. Groups of terms boxed in gray are those that differ between MYB44 and MYB77. The structure of the module suggests that ABI5 and MYB33 drive a cascade of TFs including MYB77 and MYB44, which act to amplify this signal and also further regulate many additional TFs. Additional target genes of MYB77 and MYB44 include hair cell differentiation factors, hormone response genes, secondary metabolic genes, and genes encoding components of important cellular structures such as plasmodesmata.
root development, in addition to a role in specification of the H cell versus the NH cell fate. An important next step will be to perform genetic manipulations of these factors (knockout and inducible overexpression, for example), in order to test and elaborate on the specific predictions made by our model.

Summary and Conclusions

In this study, we used ATAC-seq profiling of accessible chromatin to investigate questions regarding the transcriptional regulatory landscape of plant genomes and its conservation across species. We also investigated the similarities and differences in open chromatin landscapes in two root cell types that arise from a common progenitor, allowing us to identify and analyze TFs that act specifically in one cell type versus the other. Overall, we are able to gain several new insights from this work.

In optimization of our ATAC-seq procedures, we found that the assay can be performed effectively on crudely purified nuclei but that this approach is limited by the large proportion of reads arising from organelle genomes (Table 1). This issue is ameliorated by the use of the INTACT system to affinity-purify nuclei for ATAC-seq, which also provides access to individual cell types. Consistent with previous reports, we found that the data derived from ATAC-seq are highly similar to those from DNase-seq (Figure 1). In comparing our root tip ATAC-seq data to DNase-seq data from whole roots, we found that some hypersensitive regions were detected in one assay but not the other. This discrepancy is most likely attributable to differences in starting tissue and laboratory conditions, rather than biological differences in the chromatin regions sensitive to DNaseI versus the hyperactive Tn5 transposase. This interpretation would fit with the large number of differences also observed in THS overlap between Arabidopsis root tip and epidermal cell types.

In a comparison of open chromatin among the root tip epi-genomes of Arabidopsis, M. truncatula, tomato, and rice, we found the genomic distribution of THSs in each were highly similar. About 75% of THSs lie outside of transcribed regions, and the majority of these THSs are found within 3 kb upstream of the TSS in all species (Figure 2). Thus, the distance of upstream THSs from the TSS is relatively consistent among species and is not directly proportional to genome size or intergenic space for these representative plant species. Among genes with an upstream THS, 70% of these genes in Arabidopsis, M. truncatula, and rice have a single such feature, 20% have two upstream THSs, and <10% have three or more. By contrast, only 27% of tomato genes with an upstream THS have a single THS, 20% have two, and the proportion with 4 to 10 THSs is 2 to 7 times higher than that for any other species examined. This increase in THS number in tomato could be reflective of an increase in the number of regulatory elements per gene, but it is perhaps more likely a result of the greater number of long-terminal repeat retrotransposons near genes in this species (Xu and Du, 2014). In either case, our investigation revealed that open chromatin sites—and by extension transcriptional regulatory elements—in all four species are focused in the TSS-proximal upstream regions and are relatively few in number per gene. This suggests that transcriptional regulatory elements in plants are generally fewer in number and are closer to the genes they regulate than those of animal genomes. For example, the median distance from an enhancer to its target TSSs in Drosophila was found to be 10 kb, and it was estimated that each gene had an average of four enhancers (Kvon et al., 2014). It was also recently reported that in human T cells, the median distance between enhancers and promoters was 130 kb, far greater than the distances we have observed here across plant species (Mumbach et al., 2017).

Analysis of overrepresented TF motifs in THSs across species suggested that many of the same TFs are at play in early root development in all species. Perhaps more surprisingly, correlation of specific gene sets by multiple TFs seems to be frequently maintained across species (Figure 3). Taken together with the lack of shared open chromatin profiles among orthologous genes and expressologs, these findings suggest that transcriptional regulatory elements may relocate over evolutionary time within a window of several kilobases upstream of the TSS, but regulatory control by specific TFs is relatively stable.

Our comparison of the two Arabidopsis root epidermal cell types, the H and NH cells, revealed that open chromatin profiles were highly similar between cell types. By examining THSs that were exclusive to one cell type, we were able to find several thousand THSs that were quantitatively more accessible in each cell type compared with the other (Figure 4). Mapping of these dTHSs to their nearest genes revealed that in each cell type there were many dTHSs that were near genes expressed more abundantly in that cell type, as well as many near genes with the opposite expression pattern. This suggests that some dTHSs represented transcriptional activating events whereas others were repressive in nature.

Analysis of TF motifs at these dTHSs between cell types identified a suite of TFs that were more highly expressed in H cells and whose motifs were significantly overrepresented in H-cell-enriched dTHSs. Analysis of three of these TFs—AB15, MYB33, and NAC083—revealed that each factor targets a large number of H-cell-enriched genes as well as a smaller number of NH-cell-enriched genes (Figure 5). These factors also have many overlapping target genes among them, and AB15 and MYB33 both target seven additional H-cell-enriched TFs. Among these seven H-enriched TFs are two additional MYB factors: MYB77 and MYB44 (Figure 6). Examination of the high confidence target genes of MYB77 and MYB44 revealed that these paralogous factors appeared to regulate each other as well as many other common target genes, including large numbers of other TF genes. Hundreds of the MYB77 and MYB44 target genes were also more highly expressed in the H cell relative to the NH cell, suggesting that these factors set off a broad transcriptional cascade in the H cell type. In addition, they appear to directly regulate many H-cell-enriched genes involved in cell fate specification and water and phosphate acquisition. This type of cooperative action by pairs of MYB paralogs has also been documented recently in Arabidopsis and other species (Miliar and Gubler, 2005; Matus et al., 2017; Wang et al., 2017), and the fact that many target genes for each MYB factor are not regulated by the other may reflect a degree of subfunctionalization between the paralogs.

An important question arising from our results is whether classifying a TF as strictly an activator or repressor is generally accurate in most cases. For example, the H-cell-enriched TFs that we examined all have apparent target genes that are highly expressed in
the H cell type as well as targets that are expressed at very low levels, if at all, in the H cell type. In fact, these latter genes are often much more highly expressed in the NH cell type. Given that a number of these TFs have been shown to activate transcription in specific cases, this suggests that they promote the transcription of H-cell-enriched targets and either repress or have no effect on NH-cell-enriched target genes. One explanation for this phenomenon is that these TFs have “dual functionality” as activators and repressors, depending on the context (Bauer et al., 2010). However, it is equally possible that these factors do not play a direct role in gene repression. For example, the binding of an activator near a repressed gene may be functionally irrelevant to the regulation of that gene, or it may be that other gene-specific repressors are also bound nearby and override the activity of the activator. This phenomenon will be worth exploring as it may deepen our understanding of the intricacies of transcriptional control.

In this study, we outline a widely applicable approach for combining chromatin accessibility profiling with available genome-wide binding data to construct models of TF regulatory networks. The putative TF regulatory pathways we have illuminated through our comparison across species and cell types provide important hypotheses regarding the evolution of gene regulatory mechanisms in plants and the mechanisms of cell fate specification that are now open to experimental analysis.

**METHODS**

**Plant Materials and Growth Conditions**

Plants used in this study were of the Arabidopsis thaliana Col-0 ecotype, the A17 ecotype of Medicago truncatula, the M82 LA3475 cultivar of tomato (Solanum lycopersicum), and the Nipponbare cultivar of rice (Oryza sativa). Transgenic plants of each species for INTACT were produced by transformation with a binary vector carrying both a constitutively expressed bsd ligase and constitutively expressed NTF protein containing a nuclear outer membrane association domain (Ron et al., 2014). The binary vector used for M. truncatula was identical to the tomato vector (Ron et al., 2014) but was constructed in a pB7WG vector containing the phosphinothricin resistance gene for plant selection and it retains the original AtACT2 promoter. The binary vector used for rice is described elsewhere (Reynoso et al., 2017). Transformation of rice was performed at UC Riverside and then resuspended in 10 mL of NPB and split into two 0.5-mL fractions in new tubes. Nuclei from one fraction were purified by INTACT using streptavidin-coated magnetic beads as previously described (Bajic et al., 2018) and kept on ice prior to counting and subsequent transposase integration reaction. Nuclei from the other fraction were purified by nonionic detergent lysis of organelles and sucrose sedimentation, as previously described (Bajic et al., 2018). Briefly, these nuclei in 0.5 mL of NPB were pelleted at 1200g for 10 min at 4°C, decanted, and resuspended thoroughly in 1 mL of cold EB2 (0.25 M sucrose, 10 mM Tris, pH 8, 10 mM MgCl2, 0.1% Triton X-100, and 1× Roche Complete protease inhibitors). Nuclei were then pelleted at 1200g for 10 min at 4°C, decanted, and resuspended in 300 μL of EB3 (1.7 M sucrose, 10 mM Tris, pH 8, 2 mM MgCl2, 0.15% Triton X-100, and 1× Roche Complete protease inhibitors). This suspension was then layered gently on top of 300 μL of fresh EB3 in a 1.5-mL tube and centrifuged at 16,000g for 10 min at 4°C. Pelleted nuclei were then resuspended in 1 mL of cold NPB and kept on ice prior to counting and transposase integration.

For INTACT purification of total nuclei from root tip of M. truncatula, tomato, and rice, as well as purification of Arabidopsis root hair and non-hair cell nuclei, 1 to 3 g of starting tissue was used. In all cases, nuclei were purified by INTACT and nuclei yields were quantified as described previously (Bajic et al., 2018).

**ATAC-Seq**

Freshly purified nuclei to be used for ATAC-seq were kept on ice prior to the transposase integration reaction and never frozen. Transposase integration reactions and sequencing library preparations were then performed as previously described (Bajic et al., 2018). In brief, 50,000 purified nuclei or 50 ng of Arabidopsis leaf genomic DNA was used in each 50 μL transposase integration reaction for 30 min at 37°C using Nextera reagents (Illumina; FC-121-1030). DNA fragments were purified using the Minelute PCR purification kit (Qiagen), eluted in 11 μL of elution buffer, and the entirety of each sample was then amplified using High Fidelity PCR Mix (NEB) and custom bar-coded primers for 9 to 12 total PCR cycles. These amplified ATAC-seq libraries were purified using AMPure XP beads (Beckman Coulter), quantified by qPCR with the NEBioNext Library Quantification Kit (NEB), and analyzed on a Bioanalyzer High Sensitivity DNA Chip (Agilent) prior to pooling and sequencing.

**High-Throughput Sequencing**

Sequencing was performed using the Illumina NextSeq 500 or HiSeq 2000 instrument at the Georgia Genomics Facility at the University of Georgia. Sequencing reads were either single-end 90-nucleotide or paired-end 36-nucleotide and all libraries that were to be directly compared were pooled and sequenced on the same flow cell.
Sequence Read Mapping, Processing, and Visualization

Sequencing reads were mapped to their corresponding genome of origin using Bowtie2 software (Langmead and Salzberg, 2012) with default parameters. Genome builds used in this study were Arabidopsis version TAIR10, M. truncatula version Mt4.0, Tomato version SL2.4, and Rice version IRGSP 1.0.30. Mapped reads in .sam format were converted to .bam format and sorted using Samtools version Mt4.0, Tomato version SL2.4, and Rice version IRGSP 1.0.30. Mapped reads in .sam format were converted to .bam format and sorted using Samtools 0.1.19 (Li et al., 2009). Mapped reads were then filtered using Samtools to retain only those reads with a mapping quality score of 2 or higher (Samtools “view” command with option “–q 2” to set mapping quality cutoff). Arabidopsis ATAC-seq reads were further filtered with Samtools to remove those mapping to either the chloroplast or mitochondrial genomes, and root hair and non-hair cell data sets were also subsampled such that the experiments within a biological replicate had the same number of mapped reads prior to further analysis. For normalization and visualization, the filtered, sorted .bam files were converted to bigwig format using the “barcoverag” script in deepTools 2.0 (Ramirez et al., 2016) with a bin size of 1 bp and RPKM normalization. Use of the term normalization in this article refers to this process. Heat maps and average plots displaying ATAC-seq data were also generated using the “computeMatrix” and “plotHeatmap” functions in the deepTools package. Genome browser images were made using the Integrative Genomics Viewer (IGV) 2.3.68 (Thorvaldsdóttir et al., 2013) with bigwig files processed as described above.

Identification of Orthologous Genes among Species

Orthologous genes among species were selected exclusively from syntenic regions of the four genomes. Syntenic orthologs were identified using a combination of CoGe SynFind (https://genomevolution.org/CoGe/SynFind.pl) with default parameters, and CoGe SynMap (https://genomevolution.org/coge/SynMap.pl) with the QuotaAlign feature selected and a minimum of six aligned pairs required (Lyons and Freeling, 2008; Lyons et al., 2008).

Peak Calling to Detect THSs

Peak calling on ATAC-seq data was performed using the “FindPeaks” function of the HOMER package (Heinz et al., 2010). The parameters “–region” and “–minDist 150” were used to allow identification of variable length peaks and to set a minimum distance of 150 bp between peaks before they are merged into a single peak, respectively. We refer to the peaks called in this way as transposase hypersensitive sites or THSs.

Genomic Distribution of THSs

For each genome, the distribution of THSs relative to genomic features was assessed using the PAVIS web tool (Huang et al., 2013) with “upstream” regions set as the 2000 bp upstream of the annotated transcription start site and “downstream” regions set as 1000 bp downstream of the transcription termination site.

TF Motif Analyses

ATAC-seq THSs that were found in two replicates of each sample were used for motif analysis. The regions were adjusted to the same size (600 bp for root tip THSs or 300 bp for cell-type-specific dTHSs). The MEME-ChIP pipeline (Machanick and Bailey, 2011) was run on the repeat-masked fasta files representing each THS set to identify overrepresented motifs, using default parameters. For further analysis, we used the motifs derived from the DREME, MEME, and CentriMo programs that were significant matches (E value < 0.05) to known motifs. Known motifs from both cis-BP (Weirauch et al., 2014) and the DAP-seq database (O’Malley et al., 2016) were used in all motif searches.

Assignment of THSs to Genes

For each ATAC-seq data set, the THSs were assigned to genes using the “TSS” function of the PeakAnnotator 1.4 program (Salmon-Divon et al., 2010). This program assigns each peak/THS to the closest TSS, whether upstream or downstream, and reports the distance from the peak center to the TSS based on the genome annotations described above.

ATAC-Seq Footprinting

To examine motif-centered footprints for TFs of interest, we used the “dnase_average_profile.py” script in the pyDNavase package (Piper et al., 2013). The script was used in ATAC-seq mode [“-A” parameter] with otherwise default parameters.

Defining High-Confidence Target Sites for Transcription Factors

We used FIMO (Grant et al., 2011) to identify motif occurrences for TFs of interest, and significant motif occurrences were considered to be those with a P value < 0.0001. Genome-wide high confidence binding sites for a given transcription factor were defined as transposase hypersensitive sites in a given cell type or tissue that also contain a significant motif occurrence for the factor and also overlap with a known enriched region for that factor from DAP-seq or ChiP-seq data (see also Supplemental Figure 2 for a schematic diagram of this process).

GO Analysis

GO analyses using only Arabidopsis genes were performed using the GeneCodis 3.0 program (Nogales-Cadenas et al., 2009; Tabas-Madrid et al., 2012). Hypergeometric tests were used with P value correction using the false discovery rate (FDR) method. AgriGO was used for comparative GO analysis of gene lists among species, using default parameters (Du et al., 2010; Tian et al., 2017).

Accession Numbers

The raw and processed ATAC-seq data described here have been deposited in the NCBI Gene Expression Omnibus database under record number GSE101482. The characteristics of each data set (individual accession number, read numbers, mapping characteristics, and THS statistics) are included in Supplemental Data Set 8. For comparison to our ATAC-seq data from root tips, we used a published DNase-seq data set from 7-d-old whole Arabidopsis roots (SRX391990), which was generated from the same INTACT transgenic line used in our experiments (Sullivan et al., 2014). Publicly available CHIP-seq and DAP-seq data sets were also used to identify genomic binding sites for transcription factors of interest. These include ABF3 (AT4G34000; SRX1720080) and MYB44 (AT5G67300; SRX1720040) (Song et al., 2016), HYS (AT5G11260; SRX1412757), CBF2 (AT4G25470; SRX1412036), MYB77 (AT3G50600; SRX1412453), ABI5 (AT2G36270; SRX670505), MYB33 (AT5G06100; SRX1412418), NAC083 (AT5G13180; SRX1412546), MYB77 (AT3G50060; SRX1412453), WRKY27 (AT5G52830; SRX1412681), and At5g04390 (SRX1412214) (O’Malley et al., 2016). Raw reads from these files were mapped and processed as described above for ATAC-seq data, including peak calling with the HOMER package. Published RNA-seq data from Arabidopsis root hair and non-hair cells (Li et al., 2016a) were used to define transcripts that were specifically enriched in the root hair cell relative to the non-hair cell (hair-cell-enriched genes), and vice versa (non-hair-enriched genes). We defined cell-type-enriched genes as those whose transcripts were at least 2-fold more abundant in one cell type than the other and had an abundance of at least five RPKM in the cell type with higher expression.

Supplemental Data

Supplemental Figure 1. Comparison of read counts at enriched regions in DNase-seq versus ATAC-seq and Crude-ATAC-seq versus INTACT-ATAC-seq.
Supplemental Figure 2. Analysis of reproducibility in tomato ATAC-seq data
Supplemental Figure 3. Analysis of ATAC-seq signals at orthologous genes.
Supplemental Figure 4. Defining high-confidence binding sites and target genes for each TF.
Supplemental Figure 5. Overlaps of root tip transcription factor target genes.
Supplemental Figure 6. Wild-type and hy5-1 root tip morphology and gravitropism phenotypes.
Supplemental Figure 7. Comparison of ATAC-seq read counts between data sets.
Supplemental Figure 8. Footprinting at motifs of cell-type-enriched TFs in genomic DNA and cell-type-specific ATAC-seq data sets.
Supplemental Data Set 2. Syntenic orthologous genes in all four species.
Supplemental Data Set 3. Expressolog gene sets in four species.
Supplemental Data Set 4. Motifs common to THSs in all species.
Supplemental Data Set 5. Predicted target genes for ABF3, CBF2, HY5, and MYB77 in all four species.
Supplemental Data Set 7. Binding sites and target genes for cell-type-enriched TFs.
Supplemental Data Set 8. ATAC-seq data set characteristics.

ACKNOWLEDGMENTS

We thank Paja Sijacic and Shannon Torres for constructive criticism of the manuscript. This work was supported by funding from the National Science Foundation (Plant Genome Research Program Grant IOS-123843) to J.B.-S., N.S., S.M.B., and R.B.D. D.A.W. was supported in part by funding from the Elise Taylor Stocking Memorial Fellowship, and K.K. was supported in part by the Finnish Cultural Foundation.

AUTHOR CONTRIBUTIONS

R.B.D., N.S., J.B.-S., K.A.M., M.B., K.K., M.R., G.P., and D.A.W. performed all experiments on Arabidopsis root tips as well as hair and non-hair cells. M.B. performed all experiments on *M. truncatula* root tips. K.K., D.A.W., and K.Z. performed all experiments on tomato root tips, and M.R. and G.P. performed all experiments on rice root tips. M.W. performed all analyses of syntenic regions and identification of orthologous genes among species. K.B., M.D., and C.Q. analyzed ATAC-seq data sets with Hotspot software and also contributed expertise in other analyses. R.B.D., K.A.M., and M.B. analyzed the data. R.B.D. drafted the manuscript with subsequent input and editing from all authors.

Received July 31, 2017; revised October 30, 2017; accepted December 6, 2017; published December 11, 2017.

REFERENCES


