Diversification and conservation of DNA binding specificities of SPL family of transcription factors

Miaomiao Li[1,2,*], Tao Yao[2], Mary Galli[3], Wanru Lin[1], Yilin Zhou[1], Jin-Gui Chen[2], Andrea Gallavotti[3] and Shao-shan Carol Huang[1,*]

[1] Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA

[2] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[3] Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ 08854-8020, USA.

*Correspondence: s.c.huang@nyu.edu, lim7@ornl.gov

Abstract

SQUAMOSA Promoter-Binding Protein-Like (SPL) transcription factors play vital roles in plant development and stress responses. In this study, we report a comprehensive DNA Affinity Purification sequencing (DAP-seq) analysis for 14 of the 16 SPL transcription factors in *Arabidopsis thaliana*, providing valuable insights into their DNA-binding specificities. We performed Gene Ontology (GO) analysis of the target genes to reveal their convergent and diverse biological functions among SPL family proteins. Comparative analysis between the paralogs AtSPL9 and AtSPL15 revealed differences in their binding motifs, suggesting divergent regulatory functions. Additionally, we expanded our investigation to homologs of AtSPL9/15 in *Zea mays* (ZmSBP8/30) and *Triticum aestivum* (TaSPL7/13), identifying conserved and unique DNA-binding patterns across species. These findings provide key resources for understanding the molecular mechanisms of SPL transcription factors in regulating plant development and evolution across different species.

Introduction

Transcription factors (TFs) are crucial regulators of gene expression, playing a central role in controlling various biological processes by binding to specific DNA sequences and modulating the transcription of target genes. The expansion of TF families through gene duplication events and subsequent functional diversification is a key evolutionary mechanism that enhances the complexity of gene regulation in plants (Panchy et al. 2016). This expansion allows for the development of novel regulatory networks and the fine-tuning of gene expression in response to environmental and developmental cues. This process, known as functional diversification, can result in the emergence of TFs with altered DNA-binding specificities, thereby contributing to the evolution of complex regulatory circuits (Panchy et al. 2016, McKeown et.al 2014, Rogers et.al 2018). A comprehensive analysis of DNA-binding specificities across an entire TF family is a powerful way to understand the full range of their regulatory potential. This can reveal how different TFs within a family contribute to the regulation of distinct biological pathways, offering insights into the mechanisms underlying the evolution of gene regulatory networks.

As a plant-specific TF family, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKEs* (SPLs) have highly conserved SQUAMOSA Promoter-Binding Protein (SBP) domains with approximately 76 amino acids and containing two zinc finger-like structures and a nuclear localization signal motif (Birkenbihl et al. 2005). The first two SBP members were identified in *Antirrhinum majus* floral meristem and found to act during early flower development (Klein 1996). In *Arabidopsis*, the SPL genes are central to the age-dependent pathway of flowering (Wang et al. 2009, Wang 2014, Xu et al. 2016). Many SPL members in *Arabidopsis* are targeted by miR156/157, and the miR156/SPL module, plays an important role in diverse development stages including shoot meristem growth, flower development and the phase change from vegetative growth to reproductive growth, as well as the response to abiotic environmental stress. SPLs directly activate the expression of critical flowering genes, such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)*, *FLOWERING LOCUS T (FT)*, *FRUITFULL (FUL)*, and *APETALA1 (AP1)* (Wang et al. 2009, Yamaguchi et al. 2009, Kim et al. 2012), as well as auxin related genes, while also repressing the expression of *AUXIN RESPONSE FACTOR (ARF)* genes by binding certain promoters such as those of *ARF6* and *ARF8*, thereby balancing auxin signaling (Nagpal et al. 2005).

In *Arabidopsis*, SPL9 is one of the most extensively studied members of the SPL family (Wang et al. 2009, Wu et al. 2009, Xu et al. 2016). rSPL9 is a miR156-insensitive mutant bypassing the

juvenile period and showing less response to aging. Within the *Arabidopsis* SPL family, SPL9 and SPL15 are the closest paralogs derived from genome duplication and have redundant biological functions (Yang et al. 2008, Preston and Hileman, 2013). The *spl9/spl15* double mutants exhibit alternations in branching and flowering patterns, highlighting the critical role of SPL9 and SPL15 in regulating shoot branching (Schwarz et al. 2008). However, SPL15 and SPL9 showed distinct flowering phenotypes under short-day conditions (Hyun et al. 2016). SPL homologs in crops play critical roles in controlling plant architecture, flowering time, and reproductive development. In rice, one of the most studied SPL genes, *OsSPL14* (also known as *Ideal Plant Architecture1, IPA1*), closely related to *Arabidopsis SPL9* and *SPL15*, is a key regulator of tiller number and panicle size (Wang et al. 2018, Song et al. 2017). In maize (*Zea mays*), *ZmSBP8* and *ZmSBP30* are homologous to *Arabidopsis SPL9/15*. The *zmsbp8/30* mutants exhibit notable changes in kernel row number, ear size, and cob architecture, ultimately affecting grain yield (Chuck et al. 2014, Wei et al. 2018). In hexaploid bread wheat (*Triticum aestivum*), the *SPL9/15* homologs across the three subgenomes are *TaSPL7A/B/D* and *TaSPL13A/B/D.* CRISPR/Cas9 knockout of *TaSPL7A/B/D* and *TaSPL15A/B/D* resulted in increased tiller number, decreased spike length, and reduced spikelet number (Pei et al. 2023). These proteins are thus essential regulators of agronomic traits that directly impact crop yield, and identifying and editing the targets of the TFs could provide powerful tools for crop improvement.

However, the differences in DNA binding specificities among paralogs within the SPL family and their functional homologies across plant species, particularly in crops, remain less understood. In this study, we use DNA affinity purification-sequencing (DAP-seq, Bartlett et al. 2017) to map the genome-wide binding sites of the entire SPL family in *Arabidopsis* and systematically analyze their DNA binding specificities. Additionally, we examined the conservation of potential SPL target genes across *Arabidopsis*, maize, and wheat, along with the DNA sequence preferences of the homologous SPLs in these species. By focusing on the expansion of the SBP/SPL TF family, our study provides insights into how TF duplication and subsequent functional diversification have contributed to the increasing complexity of gene regulation and the evolution of TF-DNA binding specificities in plants.

Results

Genome-wide binding site analysis of *Arabidopsis* SPL family by DAP-seq

We applied DAP-seq to all 16 members of the SPL family in *Arabidopsis*, successfully identifying between 854 to 34,301 peaks (read enrichment q-value ≤ 0.01) for 14 members. By using ampDAP-seq, which assays protein binding on a PCR-amplified DNA library, we obtained 2,189 peaks for SPL16 (Figure 1A). With the exception of SPL11 and SPL13, all other SPLs had thousands of peaks, indicating strong DNA binding activities. In contrast, SPL11 (854 peaks) or SPL13 (no peaks) may have potential DNA binding activities that are less detectable under the same experimental conditions in DAP-seq, or they might also require co-factors to enhance their DNA binding activity. Around half of the binding events were located in proximal promoter regions of -1000 bp to 500 bp from the transcription start sites (TSS) (Figure 1D). The distribution of binding location relative to genomic features was similar for all SPL members (Figure 1D). Accordingly, the genome-wide binding site map showed both conserved and variable binding profiles across all SPL members at the promoter regions of the well-known downstream targets, for example, the flowering development gene *SOC1* and the an auxin response gene *ARF8* (Figure 1B and 1C).

To gain insight into the potential biological functions of SPL family members based on their DNA binding profiles, we performed Gene Ontology (GO) enrichment analysis of genes associated with the top 3000 binding events, ranked by DAP-seq binding site scores. The top ten enriched biological processes across the SPL family members revealed their shared and distinct functions (Figure 1E). Notably, hormonal responses, such as response to auxin and abscisic acid, were highly enriched for SPL15 and SPL9. Hypoxia and oxygen-related responses were more enriched for SPL15, SPL1, SPL2, SPL4 and SPL12. In contrast, the GO term "photosynthesis, light harvesting in photosystem I" was enriched for SPL8, suggesting a role in capturing light energy. Interestingly, metal ion homeostasis processes, such as cellular copper ion homeostasis and transition metal ion transport, were predominantly enriched for SPL7 DAP-seq targets, consistent with its previously characterized function in regulating copper uptake and distribution, which is essential for plants' adaptation to low copper conditions (Yamasaki et al. 2019, Bernal et al. 2012, Araki et al. 2018, Ramamurthy et al. 2018).

We next evaluated the potential association between the DAP-seq predicted target genes and gene expression using a Gene Set Enrichment Analysis (GSEA) test, comparing the DAP-seq predicted target genes to differential RNA expression in the root of *spl7* mutant (Ramamurthy et al. 2018). Among the SPL members that had strong GO term enrichment (Figure 1F), DAP-seq targets of SPL7 were the most enriched for down-regulated DEGs between *spl7* mutant and

wild-type Col-0 (Figure 1F), suggesting that DAP-seq can predict transcriptionally regulated genes by the relevant TFs for further functional studies.

Comparison of genome-wide binding and DNA sequence motifs of the SPL family

To investigate the relationships between the SPL family members in terms of their DNA binding specificities, we first clustered the SPLs based on their patterns of binding in the whole genome reported by DAP-seq. To do this, we determined the regions of hypervariable DAP-seq read signals among all the SPL binding profiles, calculated the pairwise distances between the individual SPLs (Tu et al. 2021), and performed hierarchical clustering of the distance matrix (Figure 2A). Two major subgroups emerged: Class A containing four SPLs (SPL7, 2, 10 and 11) and Class B containing the rest of the SPLs (SPL1, 12, 8, 6, 3, 5, 4, 14, 15 and 9). We then performed *de novo* motif discovery using sequences under the 1000 peaks with the highest significance of DAP-seq read enrichment for each SPL. We found the most significant motifs from the Class A members shared the core GTAC motif, while Class B members shared the core GTACGG motif (Fig. 2B). These results suggest Class A and Class B members could bind to distinct sets of genes. For example, the well-known gene HSL2 (also known as VAL2), which promotes vegetative phase change (Fouracre et al., 2012), showed strong binding signal at its promoter region from all Class B members that preferred the GTACG motif. In contrast, relatively low binding signal was observed for Class A members that preferred the GTAC motif (Figure 2C). Additionally, in Class B, motifs for SPL9 and SPL15 (close relatives derived from a duplication event) contained weaker 'GG' consensus sequence flanking the core 'GTAC' sequence, suggesting these two paralogs bind DNA with more flexibility outside of the core motif .

Differential DNA binding specificities of SPL9 and SPL15 are revealed by DAP-seq

To assess the functional coherence or divergence of the evolutionarily conserved SPL9 and SPL15, we conducted differential binding analysis between the SPL9 and SPL15 DAP-seq binding sites (Ross-Innes et al. 2012). We identified 5,066 SPL15-prefered and 3,122 SPL9-prefered binding sites at FDR threshold of 0.05 (Figure 3A), about 18% and 15% of the binding sites for these two SPLs, indicating potential functional distinctions between them (Figure 3A). Strikingly, distinct biological functions were enriched for genes associated with the SPL15- and SPL9-preferred binding events (Figure 3B). SPL15-preferred gene targets were strongly enriched in processes related to ion homeostasis, positive regulation of transcription, response to oxygen and hypoxia, and auxin response. In contrast, SPL9-preferred gene targets were

mainly enriched in secondary cell wall biogenesis, responses to photooxidative stress and regulation of the ABA-activated signaling pathway. These results suggest that beyond the conserved functions, SPL15 and SPL9 could have distinct roles in regulating different biological processes, and DNA binding preferences could mediate their specialized functions in plant development and stress responses.

To understand the mechanism underlying the differential binding activities between SPL9 and SPL15, we applied *de novo* motif discovery using the sequences in the 1000 differentially bound peaks with the highest increase in binding by SPL15 and SPL9, respectively (Figure 3C). We found that the most significant motifs in the SPL15-preferred binding sites contained the consensus sequences 'GTACgg' and 'TGTACT', while SPL9 preferred motifs with the consensus sequences 'CGG' and 'GTAC'. For example, the differentially bound peak at the promoter region of the gene *DI21* showed strong binding by SPL9 but not by SPL15, and it contained a binding site sequence GTAC (Figure 3D). In contrast, at the promoter region of the *ERD14* gene that contained a SPL15-preferred binding site without enriched signals for SPL9, the binding sequence GTACGG was found (Figure 3E). To validate that the differential binding events could influence gene expression, we conducted reporter assays using these two regions (Yoo et al.2007). We cloned a 758 bp sequence from the *DI21* promoter region covering the 200 bp SPL9-preferred DAP-seq peak into the reporter plasmid *pUC-GUS* (*DI21pro::GUS*) to drive the GUS reporter gene expression. When we transiently expressed SPL9 and SPL15 by transfecting SPL15 and SPL9 separately, we found that SPL9 induced significantly higher reporter activation compared to SPL15 (Figure 3D). For the SPL5-preferred DAP-seq peak at the *ERD14* promoter, we cloned a 1193 bp *ERD14* promoter sequence containing this peak into a reporter plasmid as *ERD14pro::GUS* and performed a similar reporter assay by transient expression of SPL9 and SPL15. Consistent with the SPL15-preferred binding, we observed transient expression of SPL15 induced significantly higher reporter gene activation than SPL9 (Figure 3E). These results support our hypothesis that SPL15 and SPL9 have distinct binding preferences to their target motifs, resulting in differences in target gene regulation.

Cross-species comparison across *Arabidopsis*, maize and wheat

To understand the differences in DNA binding specificities among SPL homologs across plant species, we compared the conserved genes associated with DAP-seq peaks of AtSPL9/15 in *Arabidopsis*, ZmSBP8/30 in maize (Ricci et al., 2019) and TaSPL13 in wheat (including three paralogs from the A/B/D subgenome) (Pei et al. 2023) on their respective genomes. By

comparing the *Arabidopsis* homologs in maize that were targets of ZmSBP8/30 to the target genes of AtSPL9/15 in *Arabidopsis*, we identified 4,557 unique target genes for AtSPL9/15, 1,125 unique target genes for ZmSBP8/30, and 492 conserved target genes shared between the two species (Figure 4A). Similarly, when comparing AtSPL9/15 in *Arabidopsis* and TaSPL7/13 in wheat, we found 4,017 unique target genes for *Arabidopsis* AtSPL9/15, 2,709 unique target genes for wheat TaSPL7/13, and 939 conserved target genes (Figure 4B).

To gain insight into the function of these conserved target genes, we performed GO enrichment analysis (Figure 4C). We found four major clusters of enriched biological processes. Cluster 2 and Cluster 4 highlighted functions conserved across the three species, primarily related to cell wall biogenesis and transcriptional regionalization. Cluster 1 revealed functions that were partially conserved, such as xylem and cell wall development, while also suggesting wheat-specific conserved processes such as cell wall modification, secondary metabolite biosynthesis, and lignin metabolic processes. These GO terms suggest the potential role of TaSPL7/13 in specific aspects of wheat architecture, possibly related to its structural integrity and stress responses. Cluster 3 GO terms were significantly enriched for SPL targets conserved between *Arabidopsis* and maize, including processes associated with different developmental stages, such as floral organ development, cotyledon development, and post-embryonic development. Additionally, it included processes such as plant organ specification and gibberellin response, which are closely related to flowering time and overall plant development. Overall, these results suggest that different biological functions and developmental processes are targeted by the SPL shared target genes in different species, offering valuable insights into the evolutionary trajectories of the SPL homologs in the three species (Figure 4C).

Since conservation of binding motifs are often used as the basis for cross-species prediction of *cis*-regulatory elements, we analyzed the enriched DNA sequence motifs in the DAP-seq peaks associated with the conserved target genes in the three species. Unexpectedly, we identified three different types of SBP protein binding motifs, although they all shared or partially shared the core motif ‘GTAC’ (top panels in Figure 4D). In the *Arabidopsis* genome (genome size approximately 135Mb), AtSPL9/15 recognized motifs with a ‘GTAC’ core and flanking ‘GG’ sequence. In the maize genome (genome size approximately 2.3 Gb), ZmSBP8/30 recognized a tandem repeat of the ‘GTAC’ sequence, ‘GTACTGAC’. In the wheat genome (genome size

approximately 17 Gb), TaSPL7/13 bound to a motif containing the consensus sequence GTACTAC/GTAGTAC. These striking motif variations suggest differences in genome-wide DNA recognition mechanisms among the SPL TFs in these three species.

The two overlapping or back-to-back motif types were prominently observed in the cereal plants, maize and wheat, suggesting an enhancement in binding specificity of the SPLs in these much larger and more complex genomes. This may be due to the formation of TF dimers, which can increase recognition specificity by the proteins. To explore the potential for homodimer formation and DNA binding across the three species, we utilized AlphaFold3 (Abramson et al., 2024) to predict homodimers based on their preferred motif types. The middle panels of Figure 4D show the DNA sequence fragments in the presence of the predicted SBP homodimers, while the bottom panels show the Predicted Alignment Error (PAE) plots. Notably, the homodimer prediction confidence was much higher for wheat SPLs, particularly for TaSPL7/13 interacting with the overlapping 'GTAGTAC' motif (Figure 4D). These results suggest that motif variation and dimer formation may play crucial roles in the DNA-binding specificities and regulatory capabilities of SPL transcription factors in complex plant genomes.

Discussion

TFs regulate gene expression by binding to specific DNA sequences. Variations in their binding sites can lead to differential regulation of genes within a TF family and in different species. In our analysis of the DAP-seq binding profiles of the *Arabidopsis* SPL family, we observed two major classes of binding site motifs, while the enriched GO terms of the predicted target genes were much more diverse among the family members. Furthermore, we found that even slight changes in DNA binding motifs between the duplicated homologs SPL9/15 could lead to diverse functional outcomes. These variations help differentiate the roles of closely related TFs in regulating downstream targets and ensuring functional specificity.

The evolution of TF-DNA interactions between species is a complex and dynamic process that involves multiple factors, including changes in the TF proteins themselves, their DNA-binding motifs, and the genomic context in which they operate. Gene duplication is common in plant genomes, especially in TF families, which can lead to new functions or specialization. After duplication, some TF genes may retain the original function by preserving the DNA binding

preferences of the ancestral genes, while other TF genes may acquire mutations that allow them to bind different DNA sequences and thus regulate different sets of genes. Duplicated TFs might evolve to bind novel DNA-binding motifs, enabling the species to respond to new environmental challenges, such as stress or disease, or to regulate different developmental pathways in plant adaptation. In our study, we observed subfunctionalization among the SPL/SBP homologs in *Arabidopsis*, maize, and wheat, where the conserved functions were partitioned, and each species evolved unique target sites, leading to regulatory specialization. Specifically, we used comparative genomics approaches to identify conserved and divergent TF binding sites through DAP-seq, discover the associated motifs, and analyze changes in *cis*-regulatory elements across species to elucidate the regulatory networks unique to each species. These findings support the use of DAP-seq to reveal evolutionary trajectories of TF networks.

Another interesting finding emerged when we explored the DNA binding motifs among the TF homologs between species, uncovering longer motif types bound by the SPL/SBP TF homologs in maize and wheat. It is well known that many TF families, such as AP2/ERF, bZIP, homeodomain and MADS-box, form dimers and higher-order complexes to regulate gene expression (Strader et al. 2022, Li et al. 2023). These dimers often recognize longer, more complex motifs than individual monomers, increasing the DNA-binding specificity of TFs, specifically in complex genomes. Our findings suggest that the SPL family may become capable of forming dimers to enhance the specificity of their DNA binding targets in large and complex plant genomes.

Materials and Methods

Plant materials

Arabidopsis reference accession Col-0 (CS70000) was grown in soil at 22 °C under long-day (16 h light/8 h dark) conditions for three weeks. Rosette leaves were collected, and flash frozen with liquid nitrogen prior to genomic DNA (gDNA) isolation for DAP-seq DNA library preparation.

DAP-seq experiments

The *Arabidopsis* SPL DAP-seq methods experiments were performed according to published protocols of (Bartlett et al. 2017; and Li et al. 2022). Briefly, the DAP-seq gDNA library was

prepared as a standard high throughput gDNA sequencing library for the Illumina platform. gDNA was fragmented to an average of 200 bp using Covaris S220 Sonicator. Fragmented gDNA was end-repaired using the End-It DNA End-Repair Kit (Lucigen, ER81050, ordered in 2021) and incubated at room temperature for 45 min followed by an A-tailing reaction using Klenow (3′→5′ exo-) (NEB, M0212), incubated at 37 °C for 30 min. A-tailed DNA fragments and annealed adapters were ligated in a ligation reaction using T4 DNA Ligase (Promega, M1804) incubated at room temperature for 3 h. The DNA was precipitated and suspended in Elution Buffer resulting in the gDNA libraries used in DAP-seq. A SPL protein expression reaction was prepared using TNT SP6 Coupled Reticulocyte Lysate System (Promega, L4600) incubated with 1000 ng *pIX-HALO-SPL* plasmids for 3 h at 30 °C. The reaction was then mixed with 10 µl Magne HaloTag Beads (Promega, G7282) and 50 µl wash buffer (PBS + 0.05% NP40) on a rotator for 1 h at room temperature. The beads with protein were then washed five times on a magnet with 100 µl wash buffer to purify HaloTag-fused protein. The protein-bound beads were incubated with 100 ng adapter-ligated gDNA library in 100 µl wash buffer for 2 h. The beads were then washed 5 times with wash buffer to remove unbound ligated DNA fragments. The beads were suspended in 30 µl elution buffer, heated at 98 °C for 10 min, and put on ice immediately for 5 min to denature the protein and release the bound DNA fragments. 25 µl of the supernatant was used for the PCR enrichment step.

Reporter assay in *Arabidopsis* protoplasts

To generate the effector plasmids used in the reporter assay, *TF AtSPL9* and *AtSPL15* were cloned in *pUC19-35S-DC* using LR reactions. A 758 bp sequence from the AtDI21 promoter and a 1,193 bp sequence from the *AtERD14* promoter was synthesized from by Twist BioScience and subcloned into *pUC19-DC-GUS* as reporter plasmids. The combinations of effector plasmids, reporter plasmids, and reference plasmids (*pUC19-35S-LUC*) were transformed into *Arabidopsis* leaf protoplasts. Middle sections of four-week-old fully expanded leaves were cut out, sliced into strips and immersed into enzyme solution containing 0.4 M mannitol, 20 mM KCl, 20 mM MES, 10 mM $CaCl_2$, 5 mM β-mercaptoethanol, 0.1% BSA, 0.4% macerozyme R10, and 1.5% cellulase R10. The mixture was incubated at room temperature for 2 h before filtering the protoplasts through a 75 µm nylon mesh and washing them with W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES). After centrifuging at 1000 rpm for 3 min at 4 °C, the protoplasts were resuspended with MMg solution (0.8 M mannitol, 1 M $MgCl_2$ and 0.2 M MES) to obtain a concentration of 200,000 cells per ml. Next, 6 µg effector, and 3 µg reporter and 100 ng reference plasmids were co-transfected into 100 µl of protoplasts using the PEG-calcium

mediated transfection method, followed by incubation in darkness for 18 to 20 h at room temperature. The GUS activity assay was conducted as described in Tiwari et al. 2003 and measured using a Fluoroskan microplate reader. The MUG (4-Methylumbelliferyl β-d-glucuronide) (Sigma-Aldrich, M9130) and luciferase assay system (Promega, E1500) were used to perform GUS and LUC activity assays, respectively. Relative GUS activity was calculated via normalization to LUC activity, and the data are presented as three independent biological replicates.

DAP-seq data processing and analysis

Read processing, peak calling and motif discovery

The *Arabidopsis* DAP-seq libraries were sequenced on an Illumina NextSeq 500 instrument. Adapter sequences were trimmed from the single read FASTQ files by Trim Galore (version 0.6.6) and Cutadapt version 3.1(Martin et al. 2011) with quality cutoff of 20. The trimmed reads were mapped to the *Arabidopsis* reference genome sequence TAIR10 using Bowtie2 (Langmead et al. 2012) version 2.2.9 with default parameters. Aligned reads were filtered by mapping quality score of at least 30 using SAMtools (Danecek et al. 2012) version 1.11. Peak calling was done by the GEM peak caller (Gao et al. 2012) (version 3.3) on the filtered mapped reads with the default read distribution, TAIR10 nuclear chromosome sequences, q-value threshold of 0.01 (--q 2), and parameters "--f SAM --t 1 --min 200 --k_min 5 --k_max 14 --k_seqs 2000 --k_neg_dinu_shuffle --outNP --outBED --outMEME --outJASPAR --outHOMER --print_bound_seqs --print_aligned_seqs". Peaks were called for each replicate individually or by merging the two replicates using GEM's multi-replicate mode, with samples from experiments of empty vector pIXHALO as control. To create a blacklist of regions that contained highly enriched but artifact signals, peak calling was done for a set of six pIXHALO empty vector controls using MACS3 (Zhang et al. 2008), which could find broader regions of read enrichment compared to the point-source binding events reported by GEM. The peaks called by MACS3 (version 3.0.0a5) for these control samples with parameters "--keep-dup auto --nomodel --extsize 200 -q 0.1" were merged and the peak regions shared by at least two of these control samples were designated to be blacklist regions. DAP-seq peaks overlapping these regions were removed prior to all downstream analysis.

BigWig files of normalized read signals were created using the MAPQ 30 filtered alignment BAM files by the bamCoverage program in the deepTools package (Ramirez et al. 2016) (version 3.5.0) with the following parameters: "--binSize 1 --normalizeUsing RPKM --

ignoreForNormalization Mt Pt". The read normalized bigwig files were visualized in JBrowse 2 (Diesh et al. 2023).

For motif discovery, the top 1000 binding events for each TF were obtained by sorting the GEM output narrowPeak files first by increasing q-value then by decreasing signal value. Sequences of 100 bp centered at the peak summits were extracted from the TAIR10 genome sequence and used as input for MEME motif discovery (Bailey et al. 1994) with a Markov background model of order 2 computed from the peak sequences and parameters "-mod zoops -nmotifs 5 -minw 5 -maxw 15 -dna -searchsize 0 -revcomp -csites 1000". The resulting PWM motifs were imported into R by the universalmotif package (Tremblay et al. 2024), aligned by the DiffLogo package (Nettling et al. 2024), and plotted by the ggseqlogo package (Nettling et al. 2015).

The maize DAP-seq data (Ricci et al. 2019) were obtained from NCBI GEO accession GSE120304. Reads were quality trimmed using trimmomatic (Bolger et al.2014) and mapped to the *Z. mays* genome sequence assembly Zm-B73-REFERENCE-NAM-5.0 with Bowtie 2 using default parameters (Langmead et al. 2012). Mapped reads were filtered to retain only reads with MAPQ greater than 30 using "samtools view -q 30" (Danecek et al.2012). Stringent criteria were established to exclude artifactual binding regions by the generation of a blacklist that captured the majority of non-specific peaks, consisting of sites bound in nearly all TF datasets and the negative control HALO-GST sample (Galli et al. 2018). Peaks were called with GEM3 (Guo et al. 2012) using a standard threshold method of adjusted p-value of 0.00001 (--q 5).

The wheat DAP-seq (Pei et al. 2023) data were obtained from NCBI GEO accession GSE188699. Adapter sequences were trimmed from the pair-end FASTQ files by Trim Galore (version 0.6.6) and Cutadapt version 3.1 (Martin et al. 2011). The trimmed pair-end reads were mapped to the *T. aestivum cv. Chinese Spring v2.1* genome sequence assembly using Bowtie2 (Langmead et al. 2012) version 2.2.9 with default parameters. Aligned reads were filtered by mapping quality score of at least 20 using SAMtools (Danecek et al. 2021) version 1.11. Peak calling was done by the MACS3 peak caller (Zhang et al. 2008), version 3.0.0a5 with input library as the control and parameters "-f BAMPE -g 11453938932.0 --keep-dup auto --nomodel --call-summits -q 0.05". Blacklist regions of artifactual read enrichment were created by calling peaks on two replicates of the input libraries individually, merging the peaks and computing the regions called in both replicates of the input. Peaks overlapping with blacklist regions were removed prior to all downstream analysis.

Arabidopsis DAP-seq sample clustering

MANorm2_utils (version 1.0.0) were first used to find the number of DAP-seq reads of all *Arabidopsis* SPL DAP-seq samples contained in the GEM peak regions that occurred in at least one replicate. MAnorm2 R package (Tu et al. 2021) (version 1.2.2) was used on this count matrix to perform normalization by a hierarchical normalization approach: the replicates for each TF were first normalized, then all the TFs were normalized to a pseudo-reference, a mean-variant-curve was then fitted using local regression and occupied regions only. Regions with significantly hypervariable read signals with adjusted p-value less than 0.01 were used to compute distance between each TF pair. The pairwise distance matrix was then used in hierarchical clustering of TFs with complete linkage method.

Peak annotation, Gene Ontology Enrichment Analysis

To associate *Arabidopsis* DAP-seq peaks to genes, the top 3000 merged replicate GEM peaks for each TF were annotated with a gene that had the closest transcription start site (TSS) using the annotatePeak function in the ChIPseeker package with the protein coding genes in Araport11 (Wang et al. 2022). Genes that have a DAP-seq peak within -1000 bp upstream and 500 bp downstream from the TSS were designated potential target genes. The R package clusterProfiler (Xu et al. 2014) was used to identify the top 10 most enriched GO categories in the Biological Process ontology annotated in the org.At.tair.db database and calculate the enrichment P-values of the GO terms in all the TFs. The enrichment P-values were corrected by the Benjamini & Hochberg method and a matrix of −log10 adjusted P-values were created with enriched GO terms on the rows and DAP-seq TFs on the columns and plotted as a heatmap by the ComplexHeatmap package. Clustering dendrogram was obtained by hierarchical clustering of the −log10 adjusted P-value matrix with Euclidean distance between rows and columns as distances and the complete linkage method.

spl7 RNA-seq data processing and Gene Set Enrichment Analysis

RNA-seq gene counts matrix for *spl7* mutant and Col-0 wild-type roots in low copper treatment and control conditions were downloaded from NCBI GEO accession GSE104916 (Ramamurthy et al. 2018). Using the R package DESeq2 (Love et al. 2014), the count matrix was imported into R, pre-filtered by keeping only genes that had read counts more than 10 in at least 3 replicates, normalized and analyzed for differential expression with the design formula genotype+treatment+genotype:treatment. Differentially expressed genes between *spl7* and Col-0 WT in control condition were extracted from the contrast for the genotype factor between *spl7* and WT, with adjusted p-value threshold of 0.1. The genes were then ranked by the log2 fold

change between WT and *spl7* and compared to the DAP-seq predicted target genes by Gene Set Enrichment Analysis test (GSEA) function in clusterProfiler.

Analysis of differential binding between AtSPL9 and AtSPL15

Differential binding analysis between AtSPL9 and AtSPL15 DAP-seq samples was performed by the R package DiffBind (Ross-Innes et al. 2012) version 3.6.5. The samples were normalized using the native DESeq2 normalization method RLE using reads in peaks, and differentially bound peaks were calculated by DESeq2 with the AtSPL9 as the reference and q-value threshold of 0.05. The significantly differentially bound peaks that have had a positive fold change were designated as AtSPL15-preferred and a negative fold change were designated as AtSPL9-preferred. Sequences were extracted from the TAIR10 reference genome for the top 1000 AtSPL15-preferred or AtSPL9-preferred peaks sorted by adjusted p-values for motif discovery using MEME-ChIP (Machanick et al. 2011) version 5.3.0 with the parameters "-meme-mod anr -meme-searchsize 0 -minw 5". Genes were assigned to the top 3000 AtSPL15-preferred or AtSPL9-preferred peaks and GO enrichment was performed as described above.

Cross-species comparison of AtSPL9/15 and homologs in maize and wheat

Maize DAP-seq peaks were annotated by the R package ChIPseeker using the Zm00001eb.1 annotation. Wheat DAP-seq peaks were annotated by the R package ChIPseeker using the IWGSC RefSeq v2.1 annotation. Genes that had a DAP-seq peak within -10000 bp upstream and 500 bp downstream from the TSS were designated potential target genes. Homologs between maize and *Arabidopsis* and between wheat and *Arabidopsis* were taken from the Best.hit.arabi.name column in the PhytozomeV13 annotation_info file for maize and wheat, respectively. Conserved target genes between maize and *Arabidopsis* were defined to be genes associated with shared peaks between ZmSBP8 and ZmSBP30 that are homologs to genes associated with shared peaks between AtSPL9 and AtSPL15. Conserved target genes between wheat and *Arabidopsis* were defined to be genes associated with shared peaks between at least two of TaSPL7A/B/D and at least two of TaSPL13A/B/D that were homologs to genes associated with shared peaks between AtSPL9/15. Sequences within the peaks that were assigned to the conserved target genes were extracted from their respective genomes, and motif discovery was performed by MEMEChIP with parameters "-ccut 0 -meme-mod anr -minw 4 -meme-nmotifs 5". Structures of homodimer interacting with DNA were predicted by AlphaFold3 (Abramson et al., 2024) and visualized by PAE viewer (Elfmann and Stülke, 2023).

Acknowledgements

We thank Joseph Ecker for the gift of TF expression plasmids. We also thank Aurelia Li for technical assistance. Sequencing was performed by the NYU Genomics Core Facility with generous subsidies from the Zegar Family Foundation. This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise.

Funding

This work was supported by the NIH award R35GM138143 to S.C.H. and NSF Plant Genome Research Project grant IOS-1916804 to S.C.H. and A.G. This material is also in part based upon work supported by the Center for Bioenergy Innovation (CBI), U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number ERKP886. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the Office of Science of the U.S. Department of Energy under Contract Number DE-AC05-00OR22725.

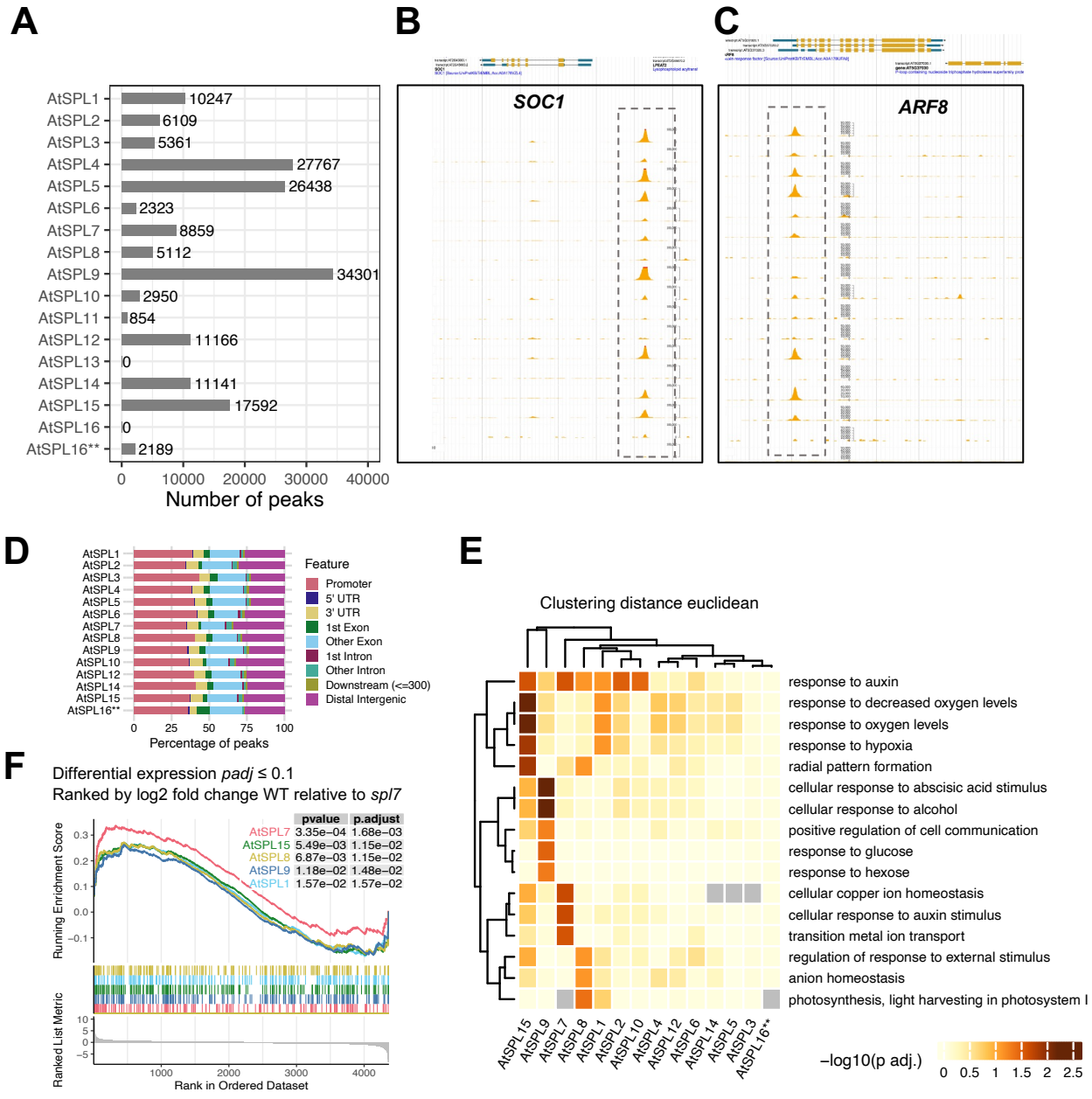

Figure 1. DAP-seq identifies genome-wide binding events for SPL TFs that are biologically relevant. (**A**) Number of DAP-seq peaks for all SPL members in *Arabidopsis*. (**B**) *Arabidopsis* SPL DAP-seq binding signal at the promoter region of a known SPL target gene *SOC1*. (**C**) *Arabidopsis* SPL DAP-seq binding signal at the promoter of *ARF8*. (**D**) Distribution of AtSPL DAP-seq peaks at genome annotation features. Promoter is defined as -1000 bp upstream to 500 bp downstream of TSS. (**E**) Enriched Gene Ontology biological process terms DAP-seq predicted targets of the AtSPL family. (**F**) GSEA of SPL DAP-seq targets and the differentially expressed genes in *spl7* mutant vs. Col-0 wild type.

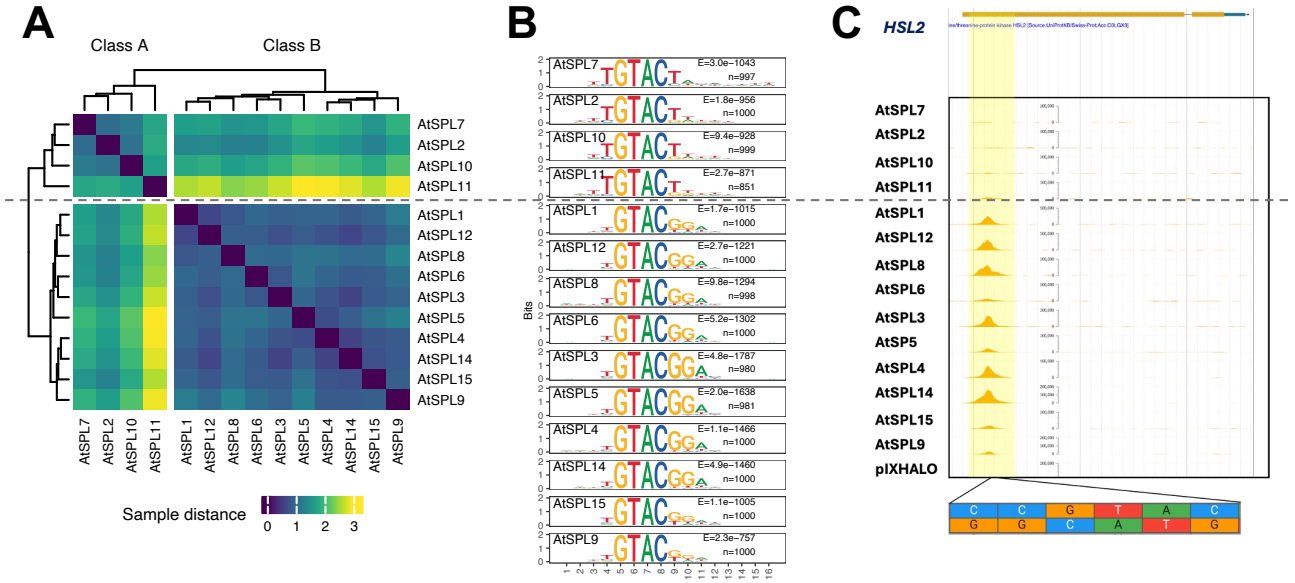

Figure 2. Genome-wide DNA binding comparison and DNA motifs of the AtSPL family. (**A**) Hierarchical clustering of AtSPL DAP-seq binding profiles splits the family members into Class A and Class B. (**B**) PWM models of the enriched motifs from the top 1000 DAP-seq peaks for each AtSPL. (**C**) Class A and Class B AtSPLs show different DAP-seq binding signals at the promoter of phase change related gene *AtHSL2* (*AtVAL2*).

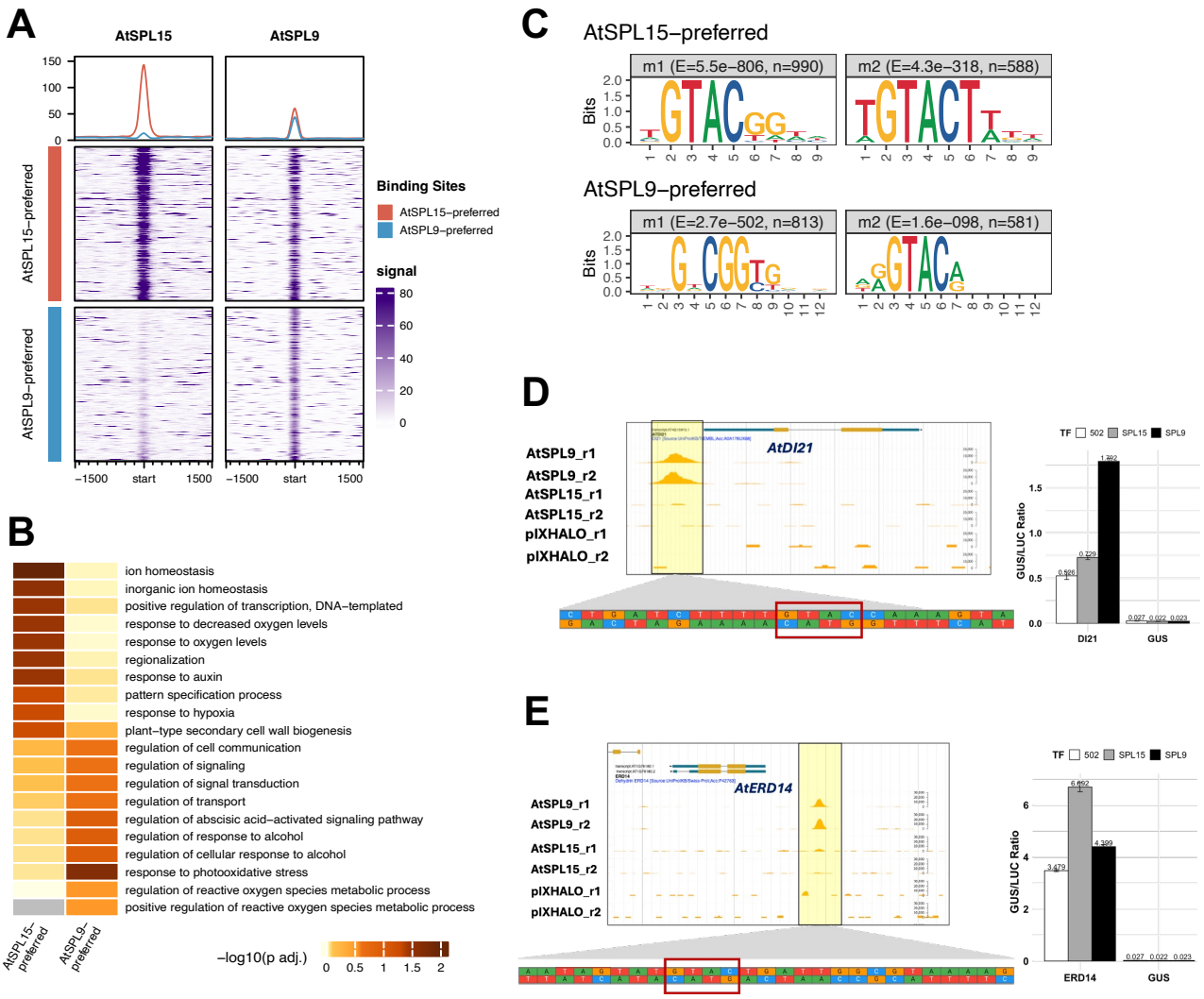

Figure 3. AtSPL9 vs AtSPL15 differential binding events, genes and motifs. (**A**) DAP-seq read signals at differentially bound peaks of AtSPL15 and AtSPL9. (**B**) Top 10 enriched GO biological process terms for genes associated with AtSPL15-preferred and AtSPL9-preferred peaks. (**C**) PWM motif models enriched in AtSPL15- and AtSPL9-preferred binding sites. (**D**) AtSPL9-preferred binding and validation in reporter assay. Left: AtSPL9-preferred binding site upstream of gene *AtDI21*. Right: transient expression of AtSPL9 and AtSPL15 induced significantly different expression of the reporter *AtDI21::GUS*. (**E**) AtSPL15-preferred binding and validation in reporter assay. Left: AtSPL15-preferred binding site upstream of gene *AtERD14* Right: transient expression of AtSPL9 and AtSPL15 induced significantly different expression of the reporter *AtERD14::GUS*.

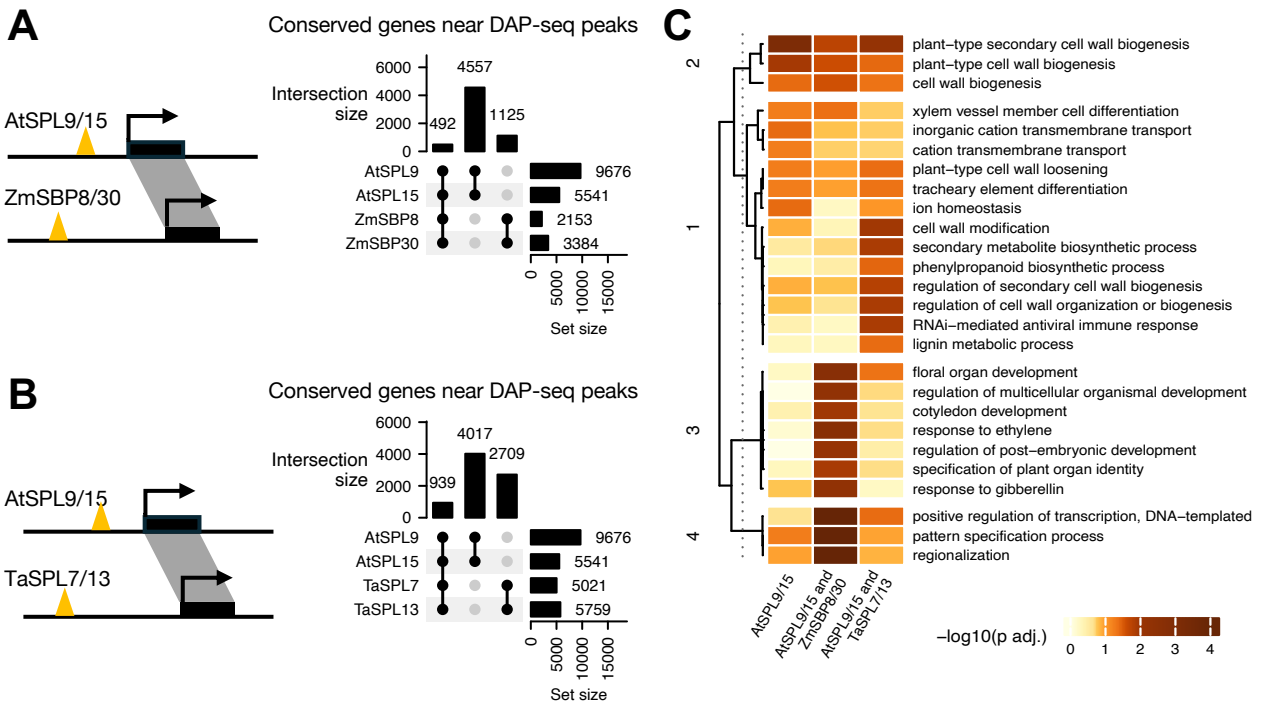

Figure 4. Comparative analysis of DAP-seq target genes and binding sites revealed variation of DNA binding among SPL homologs. (**A**) Upset plot comparing the target gene overlap between *Arabidopsis* SPL9/15 and maize ZmSBP8/30. (**B**) UpSet plot comparing the target gene overlap between *Arabidopsis* SPL9/15 and wheat TaSPL7/13. (**C**) GO enrichment analysis shows biological processes enriched for targets conserved between SPL homologs between *Arabidopsis* and maize and between *Arabidopsis* and wheat. (**D**) DNA binding motif discovered by MEME for DAP-seq peaks associated with conserved SPL targets in *Arabidopsis*, maize and

wheat and the corresponding AlphaFold3 predicted structures of SPL homodimer-DNA motif interaction.

References

Abramson, J., et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*.

Araki, R., Mermod, M., Yamasaki, H., Kamiya, T., Fujiwara, T., and Shikanai, T. (2018) SPL7 locally regulates copper-homeostasis-related genes in *Arabidopsis*. *Journal of Plant Physiology*, 224–225, 137–143.

Bailey, T.L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, 2, 28–36.

Bartlett, A., O'Malley, R., Huang, S.S. et al. (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols*, 12, 1659–1672.

Bernal, M., Casero, D., Singh, V., Wilson, G.T., Grande, A., Yang, H., Dodani, S.C., Pellegrini, M., Huijser, P., Connolly, E.L., and Merchant, S.S. (2012) Transcriptome sequencing identifies SPL7-regulated copper acquisition genes FRO4/FRO5 and the copper dependence of iron homeostasis in *Arabidopsis*. *Plant Cell*, 24, 738–761.

Birkenbihl, R.P., Jach, G., Saedler, H., and Huijser, P. (2005) Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *Journal of Molecular Biology*, 352, 585–596.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Chuck, G.S., Brown, P.J., Meeley, R., and Hake, S. (2014) Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. *Proceedings of the National Academy of Sciences USA*, 111, 18775–18780.

Crow, K.D., and Wagner, G.P. (2006) What is the role of genome duplication in the evolution of complexity and diversity? *Molecular Biology and Evolution*, 23, 887–892.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, 10, 1–4.

Diesh, C., Le, B., Edwards, B., Hays, M., Chan, B., Rao, A., Paten, B., and Nussbaum, H. (2023) JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology*, 24, 74.

Elfmann, C., and Stülke, J. (2023) PAE viewer: a webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks. *Nucleic Acids Research*, 51, W404–W410.

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T., Nemhauser, J.L., Schmitz, R.J., and Gallavotti, A. (2018) The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nature Communications*, 9, 4526.

Guo, Y., Shibin, M., and Gordon, D.K. (2012) High-resolution genome-wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Computational Biology*, 8, e1002638.

He, J., Xu, M., Willmann, M.R., McCormick, K., Hu, T., Yang, L., Starker, C.G., Voytas, D.F., Meyers, B.C., and Poethig, R.S. (2018) Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in *Arabidopsis thaliana*. *PLoS Genetics*, 14, e1007337.

Hyun, Y., Richter, R., Vincent, C., Martinez-Gallegos, R., Porri, A., and Coupland, G. (2016) Multi-layered regulation of SPL15 and cooperation with SOC1 integrate endogenous flowering pathways at the *Arabidopsis* shoot meristem. *Developmental Cell*, 37, 254–266.

Kim, J.J., Lee, J.H., Kim, W., Jung, H.S., Huijser, P., and Ahn, J.H. (2012) The microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3 module regulates ambient temperature-responsive flowering via FLOWERING LOCUS T in *Arabidopsis*. *Plant Physiology*, 159, 461–478.

Klein, J., Saedler, H., and Huijser, P. (1996) A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. *Molecular and General Genetics*, 250, 7–16.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.

Li, M., and Huang, S.S.C. (2022) DNA Affinity Purification Sequencing (DAP-Seq) for mapping genome-wide transcription factor binding sites in plants. In: Bilichak, A., Laurie, J.D. (eds) *Accelerated Breeding of Cereal Crops*. Springer Protocols Handbooks. Humana, New York, NY.

Li, M., Yao, T., Lin, W., Hinckley, W.E., Galli, M., Muchero, W., Gallavotti, A., Chen, J.G., and Huang, S.C. (2023) Double DAP-seq uncovered synergistic DNA binding of interacting bZIP transcription factors. *Nature Communications*, 14, 2600.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.

Machanick, P., and Bailey, T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, 27, 1696–1697.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 3.

McKeown, A.N., Bridgham, J.T., Anderson, D.W., Murphy, M.N., Ortlund, E.A., and Thornton, J.W. (2014) Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell*, 159, 58–68.

Nettling, M., Treutler, H., Grau, J., Keilwagen, J., Posch, S., and Grosse, I. (2015) DiffLogo: a comparative visualization of sequence motifs. *BMC Bioinformatics*, 16, 387.

Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016) Evolution of gene duplication in plants. *Plant Physiology*, 171, 2294–2316.

Pei, H., Yu, Z., Sun, Z., Guo, X., Zhou, Q., and Wang, X. (2023) Low-affinity SPL binding sites contribute to subgenome expression divergence in allohexaploid wheat. *Science China Life Sciences*, 66, 819–834.

Preston, J.C., and Hileman, L.C. (2013) Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Frontiers in Plant Science*, 4, 80.

Ramamurthy, R.K., Xiang, Q., Hsieh, E.J., Liu, K., Zhang, C., and Waters, B.M. (2018) New aspects of iron-copper crosstalk uncovered by transcriptomic characterization of *Col-0* and the copper uptake mutant *spl7* in *Arabidopsis thaliana*. *Metallomics*, 10, 1824–1840.

Ramirez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44, W160–W165.

Ricci, W.A., Lu, Z., Ji, L., Marand, A.P., Ethridge, C.L., Murphy, N.G., Noshay, J., Galli, M., Mejia-Guerra, M.K., Colomé-Tatché, M., et al. (2019) Widespread long-range *cis*-regulatory elements in the maize genome. *Nature Plants*, 5, 1237–1249.

Rogers, J.M., and Bulyk, M.L. (2018) Diversification of transcription factor-DNA interactions and the evolution of gene regulatory networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 10, e1423.

Ross-Innes, C.S., Brown, G.D., Carroll, J.S., and Lin, C.Y. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, 481, 389–393.

Schwarz, S., Grande, A.V., Bujdoso, N., Saedler, H., and Huijser, P. (2008) The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in *Arabidopsis*. *Plant Molecular Biology*, 67, 183–195.

Song, X., Lu, Z., Yu, H., et al. (2017) IPA1 functions as a downstream transcription factor repressed by D53 in strigolactone signaling in rice. *Cell Research*, 27, 1128–1141.

Strader, L., Weijers, D., and Wagner, D. (2022) Plant transcription factors—being in the right place with the right company. *Current Opinion in Plant Biology*, 65, 102136.

Tiwari, S.B., Hagen, G., and Guilfoyle, T.J. (2003) The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell*, 15, 533–543.

Tremblay, B.J.-M. (2024) universalmotif: An R package for biological motif analysis. *Journal of Open Source Software*, 9, 7012.

Tu, S., Kong, L., Wang, J., Sun, Y., Zhou, X., Xiao, Y., Zhang, S., Zhou, Y., Li, L., He, H., et al. (2021) MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Research*, 31, 131–145.

Wagih, O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, 33, 3645–3647.

Wang, J.W., Czech, B., and Weigel, D. (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell*, 138, 738–749.

Wang, J., Zhou, L., Shi, H., Chern, M., Yu, H., Yi, H., He, M., Yin, J., Zhu, X., Li, Y., Li, W., Liu, J., Wang, J., Chen, X., Qing, H., Wang, Y., Liu, G., Wang, W., Li, P., Wu, X., Zhu, L., Zhou, J.M., Ronald, P.C., Li, S., Li, J., and Chen, X. (2018) A single transcription factor promotes both yield and immunity in rice. *Science*, 361, 1026–1028.

Wei, H., Zhao, Y., Xie, Y., and Wang, H. (2018) Exploiting SPL genes to improve maize plant architecture tailored for high-density planting. *Journal of Experimental Botany*, 69, 4675–4688.

Xu, M., Hu, T., Smith, M., and Poethig, R.S. (2016) Epigenetic regulation of vegetative phase change in *Arabidopsis*. *Plant Cell*, 28, 28–41.

Xu, S., Zhang, Z., Liu, X., Zhao, Z., Li, F., Wang, J., Liu, Y., and Guo, Y. (2024) Using clusterProfiler to characterize multiomics data. *Nature Protocols*.

Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y., and Shikanai, T. (2009) SQUAMOSA Promoter Binding Protein-Like7 is a central regulator for copper homeostasis in *Arabidopsis*. *Plant Cell*, 21, 347–361.

Yang, Z., Wang, X., Gu, S., Hu, Z., Xu, H., and Xu, C. (2008) Comparative study of SBP-box gene family in *Arabidopsis* and rice. *Gene*, 407, 1–11.

Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., and Ahn, J.H. (2005) CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in *Arabidopsis*. *Plant Physiology*, 139, 770–778.