MANUSCRIPT TITLE

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

AUTHORS

Miaomiao Li[1,2,3]*, Tao Yao[2,3], Mary Galli[4], Luigi F. Di Costanzo[5], Wanru Lin[1], Yilin Zhou[1], Jin-Gui Chen[2,3], Andrea Gallavotti[4] 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] Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA

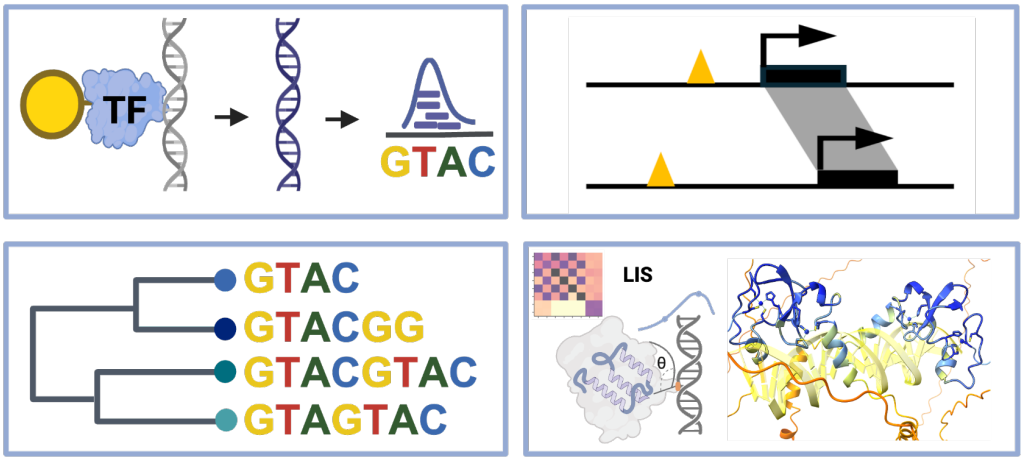
[4] Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ 08854-8020, USA.

[5] Department of Agricultural Sciences, University of Naples Federico II, Via Università 100, 80055 Portici, NA, Italy

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

Notice: This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

GRAPHICAL ABSTRACT



ABSTRACT

SQUAMOSA Promoter-Binding Protein-Like (SPL) transcription factors are key regulators of plant development and stress responses. Here, we present a comprehensive DNA affinity purification sequencing (DAP-seq) analysis of 14 of the 16 SPL members in *Arabidopsis thaliana*, revealing two major classes based on distinct DNA-binding motifs. Gene Ontology enrichment of target genes indicated both shared and specialized functions among AtSPLs, including roles in hormone signaling, metal ion homeostasis, and developmental transitions. Comparative analysis of closely related paralogs, AtSPL9 and AtSPL15, uncovered divergence in their genomic binding locations and target gene regulation, particularly in auxin and abscisic acid pathways. These differences were supported by motif analysis and protoplast reporter assays. Extending our study across species, we analyzed DAP-seq data from maize (*Zea mays*) and wheat (*Triticum aestivum)* SPL homologs, identifying conserved target genes but species-specific DNA motifs. Notably, wheat SPLs preferentially bind to longer, palindromic motifs distinct from the GTAC core motif in *Arabidopsis*. Using AlphaFold3 structure modeling, we show that these motif differences may arise from enhanced protein-protein interactions that stabilize dimeric binding. Our results highlight how SPL family members diversify through DNA-binding evolution and suggest that transcription factor dimerization contributes to motif complexity and regulatory specialization across plant genomes.

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 are a key evolutionary mechanism that enhances the complexity of gene regulation in plants (1). 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 (1–3). 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 shared and 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, which are approximately 76 amino acids long and contain two zinc finger-like structures and a nuclear localization signal motif (4). The first two SBP proteins discovered, *AmSBP1 a*nd *AmSBP2*, were identified in *Antirrhinum majus* floral meristem and found to act during early flower development (5). In *Arabidopsis*, the SPL genes are central to the age-dependent pathway of flowering (6–8). Many SPL members in *Arabidopsis* are targeted by miR156/157, and this 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 (9–12). SPLs in *Arabidopsis* 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)* (6, 13, 14), as well as auxin related genes. They also repress the expression of *AUXIN RESPONSE FACTOR (ARF)* genes by binding to their promoters such as those of *ARF6* and *ARF8*, thereby balancing auxin signaling (15).

In *Arabidopsis*, *AtSPL9* is one of the most extensively studied members of the SPL family (6, 7, 16). *rSPL9* is a miR156-insensitive mutant, showing reduced juvenile period and less response to aging (6, 17, 18). Within the *Arabidopsis* SPL family, *AtSPL9* and *AtSPL15* are the closest paralogs resulting from genome duplication and have redundant biological functions (19, 20). The *spl9/spl15* double mutants exhibit alternations in branching and flowering patterns,

80 highlighting the critical role of AtSPL9 and AtSPL15 in regulating shoot branching (21).
81 However, AtSPL15 and AtSPL9 coordinate distinct flowering phenotypes under short-day
82 conditions (22). SPL homologs in crops play critical roles in controlling plant architecture,
83 flowering time, and reproductive development. In rice (*Oryza sativa*), one of the most studied
84 SPL genes, *OsSPL14*, also known as *Ideal Plant Architecture1 (IPA1*) and closely related to
85 *AtSPL9* and *AtSPL15*, is a key regulator of tiller number and panicle size (8, 23). In maize (*Zea*
86 *mays*), *ZmSBP8* and *ZmSBP30* are homologous to *AtSPL9/15*. The *zmsbp8/30* mutants exhibit
87 notable changes in kernel row number, ear size, and cob architecture, ultimately affecting grain
88 yield (24, 25). In hexaploid bread wheat (*Triticum aestivum*), the *AtSPL9/15* homologs across
89 the three subgenomes are *TaSPL7A/B/D* and *TaSPL13A/B/D.* CRISPR/Cas9 knockout of
90 *TaSPL7A/B/D* and *TaSPL15A/B/D* resulted in increased tiller number, decreased spike length,
91 and reduced spikelet number (23, 26). These proteins are thus essential regulators of
92 agronomic traits that directly impact crop yield, so identifying and editing the targets of the TFs
93 could provide powerful tools for crop improvement.

94 Despite extensive research on the SPL family, key questions remain about the differences in
95 DNA binding specificities among paralogs and the functional similarities and differences of these
96 genes across plant species, especially in crops. In this study, we used DNA affinity purification-
97 sequencing (DAP-seq) (27, 28), to map the genome-wide binding sites of the entire SPL family
98 in *Arabidopsis* and systematically analyze their DNA binding specificities. Additionally, we
99 utilized additional DAP-seq datasets for homologous SPLs in maize and wheat to examine the
100 conservation of potential target genes SPLs across the three species, uncovering changes of
101 their DNA binding specificities that could be mediated by a dimerization mechanism. By
102 focusing on the expansion of the SBP/SPL TF family, our study provides insights into how TF
103 duplication and subsequent functional diversification, mediated by evolution of DNA binding
104 specificities, have contributed to the complexity of gene regulation in plants.

105 **MATERIALS AND METHODS**

106 **Plant materials**

107 *Arabidopsis* reference accession Col-0 (CS70000) was grown in soil at 21 °C under long-day
108 (16 h light and 8 h dark) conditions for four weeks. Rosette leaves were collected and flash
109 frozen in liquid nitrogen prior to genomic DNA (gDNA) isolation for DAP-seq DNA library
110 preparation.

111 **DAP-seq experiments**

112 The *Arabidopsis* SPL DAP and ampDAP-seq experiments were performed according to
113 published protocols of (27, 29). First, frozen leaf tissues were ground into a fine powder using a
114 mortar and pestle, dissolved in extraction buffer (0.1 M TRIS base, 0.5 M sodium chloride, 0.05
115 M EDTA, pH 8.0), and subject to phenol-chloroform extraction to remove proteins and other
116 contaminants, resulting in purified gDNA. DAP-seq gDNA library was then prepared as a
117 standard DNA sequencing library for the Illumina platform. gDNA was fragmented to an average
118 of 200 bp using Covaris S220 Sonicator. Fragmented gDNA was end-repaired using the End-It
119 DNA End-Repair Kit (Lucigen, ER81050, ordered in 2021) and incubated at room temperature
120 for 45 min, followed by an A-tailing reaction using Klenow (3′→5′ exo-) (NEB, M0212) incubated
121 at 37 °C for 30 min. A-tailed DNA fragments and annealed adapters were ligated in a ligation
122 reaction using T4 DNA Ligase (Promega, M1804) incubated at room temperature for 3 h. The
123 DNA was precipitated and suspended in elution buffer (10 mM Tris-Cl, pH 8.5), resulting in the
124 gDNA libraries used in DAP-seq. A SPL protein expression reaction was prepared using TNT
125 SP6 Coupled Reticulocyte Lysate System (Promega, L4600) incubated with 1000 ng pIX-HALO-
126 SPL plasmids for 3 h at 30 °C. 10 µl of Magne HaloTag Beads (Promega, G7282) and 50 µl of
127 wash buffer (PBS with 0.05% NP40) were added to the reaction, bringing the final volume to
128 100 µl, and the mixture was incubated on a rotator for 1 hour at room temperature. The proteins
129 immobilized on beads were then washed five times on a magnet with 100 µl wash buffer to
130 purify HaloTag-fused SPL protein. After washing, the protein-bound beads were incubated with
131 100 ng adapter-ligated gDNA library in 100 µl wash buffer for 2 h on a rotator. The mixture was
132 then washed five times with wash buffer to remove unbound ligated DNA fragments. The beads
133 were suspended in 30 µl elution buffer, heated at 98 °C for 10 min, and put on ice immediately
134 for more than 5 min to denature the protein and release the bound DNA fragments. 25 µl of the
135 supernatant was used for the PCR enrichment step. The ampDAP library is generated by
136 amplifying DNA from the gDNA library by PCR, followed by incubation with the purified TF
137 protein as for the DAP-seq protocol. The PCR reactions for making ampDAP-seq libraries and
138 sequencing libraries following protein binding were prepared as 1 µl of Phusion DNA
139 Polymerase (New England Biolabs, M0530), 10 µl of 5x Phusion HF Buffer, 2.5 µl of 10 mM
140 dNTPs, 1 µl each 25 µM index primers, 25 µl of eluted DNA, add water to 50 µl. The PCR
141 condition is as follows: 98 °C for 2 min, 15-19 cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C
142 for 1 min, final extension at 72 °C for 10 min. PCR-amplified samples were run in 1% agarose
143 gel and cut to purify fragments from 200 bp to 600 bp with Zymoclean Gel DNA Recovery Kit
144 (Zymo Research, D4007). The purified DNA fragments were measured by Qubit HS dsDNA
145 Assay Kit (ThermoFisher, Q32854) and sequenced on the Illumina platform.

146 **Reporter assay in *Arabidopsis* protoplasts**

147 To generate the effector plasmids used in the reporter assay, *AtSPL9* and *AtSPL15* CDS
148 fragments were cloned in *pUC19-35S-DC* using LR reactions. A 758 bp sequence from the
149 AtDI21 promoter (Supplementary Table S1) was synthesized by Twist BioScience and
150 subcloned into *pUC19-DC-GUS* as reporter plasmid. The combinations of effector plasmid,
151 reporter plasmid, and reference plasmid (*pUC19-35S-LUC*) were transformed into *Arabidopsis*
152 leaf protoplasts. Middle sections of four-week-old fully expanded leaves were cut out, sliced into
153 strips and immersed into enzyme solution containing 0.4 M mannitol, 20 mM KCl, 20 mM MES,
154 10 mM $CaCl_2$, 5 mM β-mercaptoethanol, 0.1% BSA, 0.4% macerozyme R10, and 1.5% cellulase
155 R10. The mixture was incubated at room temperature for 2 h before filtering the protoplasts
156 through a 75 µm nylon mesh and washing them with W5 solution (154 mM NaCl, 125 mM $CaCl_2$,
157 5 mM KCl, and 2 mM MES). After centrifuging at 1000 rpm for 3 min at 4 °C, the protoplasts
158 were resuspended with MMg solution (0.8 M mannitol, 1 M $MgCl_2$ and 0.2 M MES) to obtain a
159 concentration of 200,000 cells per ml. Next, 6 µg effector, 3 µg reporter and 100 ng reference
160 plasmids were co-transfected into 100 µl of protoplasts using the PEG-calcium mediated
161 transfection method, followed by incubation in darkness for 18 to 20 h at room temperature. The
162 GUS activity assay was conducted (30) and measured using a Fluoroskan microplate reader
163 (30). The MUG (4-Methylumbelliferyl β-d-glucuronide) (Sigma-Aldrich, M9130) and luciferase
164 assay system (Promega, E1500) were used to perform GUS and LUC activity assays,
165 respectively. Relative GUS activity was calculated via normalization to LUC activity, and the
166 data are presented as three independent biological replicates.

167 **DAP-seq data processing and analysis**

168 <u>*Read processing, peak calling and motif discovery*</u>

169 The *Arabidopsis* DAP-seq libraries were sequenced on an Illumina NextSeq 500 instrument.
170 The reads were basecalled using Picard IlluminaBasecallsToFastq (31) version 2.23.8, with
171 APPLY_EAMSS_FILTER set to false. Following basecalling, the reads were demultiplexed
172 using Pheniqs (32) version 2.1.0. The entire process was executed using a custom nextflow
173 pipeline, GENEFLOW (33). Adapter sequences were trimmed from the single read FASTQ files
174 by Trim Galore (version 0.6.6) and Cutadapt version 3.1 (34) with quality cutoff of 20. The
175 trimmed reads were mapped to the *Arabidopsis* reference genome sequence TAIR10 using
176 Bowtie2 (35)version 2.2.9 with default parameters. Aligned reads were filtered by mapping
177 quality (MAPQ) score of at least 30 using SAMtools (36) version 1.11. Peak calling was done by
178 the GEM peak caller (37) version 3.3 on the filtered mapped reads with the default read

179 distribution, TAIR10 nuclear chromosome sequences, q-value threshold of 0.01 (--q 2), and
180 parameters "--f SAM --t 1 --min 200 --k_min 5 --k_max 14 --k_seqs 2000 --k_neg_dinu_shuffle --
181 outNP --outBED --outMEME --outJASPAR --outHOMER --print_bound_seqs --
182 print_aligned_seqs". Peaks were called for each replicate individually or by merging the two
183 replicates using GEM's multi-replicate mode, with samples from experiments of empty vector
184 pIXHALO as control. To create a blacklist of regions that contained highly enriched but artifact
185 signals, peak calling was done for a set of six pIXHALO empty vector controls using MACS3
186 (38), which could find broader regions of read enrichment compared to the point-source binding
187 events reported by GEM. The peaks called by MACS3 (version 3.0.0a5) for these control
188 samples with parameters "--keep-dup auto --nomodel --extsize 200 -q 0.1" were merged and the
189 peak regions shared by at least two of these control samples were designated to be blacklist
190 regions. DAP-seq peaks overlapping these regions were removed prior to all downstream
191 analysis. BigWig files of normalized read signals were created using the MAPQ 30 filtered
192 alignment BAM files by the bamCoverage program in the deepTools package (39) version 3.5.0
193 with the following parameters: "--binSize 1 --normalizeUsing RPKM --ignoreForNormalization Mt
194 Pt". The read normalized bigwig files were visualized in JBrowse 2 (40). For motif discovery, the
195 top 1000 binding events for each TF were obtained by sorting the GEM output narrowPeak files
196 first by increasing q-value then by decreasing signal value. Sequences of 200 bp centered at
197 the peak summits were extracted from the TAIR10 genome sequence and used as input for
198 MEME motif discovery (41) with a Markov background model of order 2 computed from the
199 peak sequences and parameters "-mod zoops -nmotifs 5 -minw 5 -maxw 15 -dna -searchsize 0
200 -revcomp -csites 1000". The resulting PWM motifs were imported into R by the universalmotif
201 package (42), aligned by the DiffLogo package (43), and plotted by the ggseqlogo package (44).
202 The maize DAP-seq peak dataset, with blacklist regions already removed, was obtained from
203 Zenodo under DOI 10.5281/zenodo.14991916 (45). The wheat DAP-seq data were obtained
204 from NCBI GEO accession GSE188699 (26). Adapter sequences were trimmed from the pair-
205 end FASTQ files by Trim Galore (version 0.6.6) and Cutadapt version 3.1 (34). The trimmed
206 pair-end reads were mapped to the *T. aestivum cv. Chinese Spring v2.1* genome sequence
207 assembly (46) using Bowtie2 (35) version 2.2.9 with default parameters. Aligned reads were
208 filtered by MAPQ score of at least 20 using SAMtools (36) version 1.11. Peak calling was done
209 by the MACS3 peak caller (38) version 3.0.0a5 with input library as the control and parameters
210 "-f BAMPE -g 11453938932.0 --keep-dup auto --nomodel --call-summits -q 0.05". Blacklist
211 regions of artifactual read enrichment were created by calling peaks on two replicates of the
212 input libraries individually, merging the peaks and computing the regions called in both

replicates of the input. DAP-seq peaks overlapping with blacklist regions were removed prior to all downstream analysis.

Arabidopsis DAP-seq sample clustering

MANorm2_utils (version 1.0.0) (47) were first used to find the number of DAP-seq reads of all *Arabidopsis* AtSPL DAP-seq samples contained in the GEM peak regions that occurred in at least one replicate. MAnorm2 R package (47) 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-variance curve was then fitted using parametric regression and occupied regions only. Peak regions ranked among the top 2,000 intervals by normalized signal in more than one sample were used to compute pairwise distances between TFs using the distBioCond function. The pairwise distance matrix was then used in hierarchical clustering of TFs with complete linkage method.

Peak annotation and 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 (48) with the protein coding genes in Araport11 (49). Genes that had a DAP-seq peak within -1000 bp upstream and 500 bp downstream from the TSS were designated potential target genes. The R package clusterProfiler (50) was used to identify the top 10 most enriched GO terms for the target genes of each TF in the Biological Process ontology annotated in the org.At.tair.db database, and the enrichment P-values of these GO terms for all the TFs were computed. The enrichment P-values were corrected by the Benjamini & Hochberg method and a matrix of −log10 adjusted P-values was computed with the enriched GO terms on the rows and DAP-seq TFs on the columns and plotted as a heatmap by the ComplexHeatmap package (51). 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 (52). Using the R package DESeq2 (53), 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 wild-type (WT) in control condition were

246 extracted from the contrast for the genotype factor between *spl7* and WT, with adjusted P-value
247 threshold of 0.1. The genes were then ranked by the log2 fold change between WT and *spl7*
248 and compared to the DAP-seq predicted target genes by the Gene Set Enrichment Analysis test
249 (GSEA) function in clusterProfiler (50).

250 **Analysis of differential binding between *Arabidopsis* AtSPL9 and AtSPL15**

251 Differential binding analysis between *Arabidopsis* AtSPL9 and AtSPL15 DAP-seq samples was
252 performed by the R package DiffBind (54) version 3.6.5. The samples were normalized by the
253 native DESeq2 normalization method RLE using reads in peaks, and differentially bound peaks
254 were calculated by DESeq2 with the AtSPL9 as the reference and q-value threshold of 0.05.
255 The significantly differentially bound peaks that have had a positive fold change were
256 designated as AtSPL15-preferred and those that had a negative fold change were designated
257 as AtSPL9-preferred. Sequences were extracted from the TAIR10 reference genome for the top
258 1000 AtSPL15-preferred or AtSPL9-preferred peaks sorted by adjusted P-values for motif
259 discovery using MEME-ChIP (55) version 5.3.0 with the parameters "-meme-mod anr -meme-
260 searchsize 0 -minw 5". Genes were associated with the top 3000 AtSPL15-preferred or AtSPL9-
261 preferred peaks and GO enrichment was performed as described above.

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

263 DAP-seq peaks for maize and wheat were annotated using the R package ChIPseeker (48),
264 with the Zm00001eb.1 genome annotation for maize (56) and the IWGSC RefSeq v2.1
265 annotation for wheat (46). Genes with DAP-seq peaks located within 10,000 bp upstream and
266 500 bp downstream of the TSS were designated as potential target genes. Homologs between
267 maize and *Arabidopsis* and between wheat and *Arabidopsis* were taken from the
268 Best.hit.arabi.name column in the PhytozomeV13 annotation_info file for maize and wheat,
269 respectively (57). Conserved target genes between maize and *Arabidopsis* were defined as
270 genes associated with peaks shared between ZmSBP8 and ZmSBP30 that are homologous to
271 genes associated with peaks shared between AtSPL9 and AtSPL15. For wheat and
272 *Arabidopsis*, conserved targets were defined as genes associated with peaks shared by at least
273 two of TaSPL7A/B/D and at least two of TaSPL13A/B/D that were homologous to those genes
274 associated with peaks shared between AtSPL9 and AtSPL15. To identify DNA-binding motifs,
275 sequences within the DAP-seq peaks associated with conserved target genes were extracted
276 from their respective genomes. Motif discovery was then performed using MEME-ChIP with the
277 parameters "-ccut 0 -meme-mod anr -minw 4 -meme-nmotifs 5".

278 **Structure modeling of SPL-DNA interaction**

279 To investigate the structural basis of differential DNA motif binding among SPL homologs from
280 *Arabidopsis*, maize and wheat, we used AlphaFold3 (AF3) (58) to predict structures of the SPL
281 TFs interacting with DNA sequences containing the different motif types. Our goal was to
282 determine whether PPIs could explain why SPL homologs in different species prefer distinct
283 DNA motifs, i.e., GTAC for *Arabidopsis* SPLs and GTASTAC for wheat SPLs, by evaluating
284 differences in TF-TF and TF-DNA interaction potential. To this end, we modeled both one-copy
285 and two-copy TF-DNA complexes for each protein in combination with either its species-
286 preferred binding motif or the motif preferred by an orthologous SPL in a contrasting species.

287 For protein inputs, we used full-length amino acid sequences corresponding to the isoforms
288 analyzed in the DAP-seq experiments (26, 45). For DNA inputs, we focused on DAP-seq peaks
289 associated with conserved SPL target genes across species. Within these peaks, we scanned
290 for motif matches corresponding to the preferred motif of the SPL in question, for example,
291 GTAC in *Arabidopsis* SPL peaks and GTASTAC in wheat SPL peaks. We selected the ten
292 highest-scoring motif matches for each protein as representative binding sites. For each
293 selected site, we extracted a 20 bp DNA sequence centered on the motif match, yielding a set
294 of DNA fragments we refer to as “species-preferred motif sequences”. To create a comparison
295 set, we then searched genomic regions nearby (within 10 kb of the preferred motif) for high-
296 scoring matches to the cross-species motif, i.e., the motif preferred by a homologous SPL in
297 another species. Importantly, these cross-species motif matches were selected to not overlap
298 any DAP-seq peaks, to reduce the likelihood that they are bound by the SPL under study. We
299 extracted 20 bp DNA fragments centered on these motifs to create the “cross-species motif
300 sequences”. Thus, for each SPL protein, we generated two matched DNA input sets: one
301 representing motifs preferred and bound by this SPL and one representing motifs not preferred
302 or bound by this SPL but preferred by a homologous SPL in another species.

303 Each protein-DNA combination was then modeled using AF3 with at least six replicate
304 prediction runs using different random seeds. Each run used either a one-copy or two-copy
305 version of the TF sequence paired with one of the twenty 20 bp-long DNA sequences including
306 the forward and reverse complement (RC) strands, as well as two zinc ions for each copy of the
307 protein. The resulting structures were visualized using ChimeraX v1.10.1 (59). Confidence in
308 structural predictions was assessed using AF3’s standard metrics: predicted template modelling
309 (pTM) score, the interface predicted template modeling (ipTM) score, and predicted aligned
310 error (PAE). Local Interaction Score (LIS) calculations were performed by modifying the Jupyter

311 notebook provided using the recommended parameters (60). For each combination of protein
312 (one- or two-copy) and motif (species-preferred or cross-species motif) condition, we computed
313 the median LIS score across all replicate runs for each of the ten DNA sequence instances.
314 Wilcoxon rank-sum tests were used to compare the LIS score distributions of the ten motif
315 matches in the species-preferred motif sequences versus the cross-species motif sequences.

316 **RESULTS**

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

318 We applied DAP-seq to all 16 members of the SPL family in *Arabidopsis,* successfully
319 identifying between 854 to 34,301 peaks (read enrichment q-value ≤ 0.01) for 14 members
320 (Figure 1A). By using ampDAP-seq, which assays protein binding on a PCR-amplified DNA
321 library, we obtained 2,189 peaks for AtSPL16 (Figure 1A). With the exception of AtSPL11 and
322 AtSPL13, all other AtSPLs had thousands of peaks, indicating strong DNA binding activities. In
323 contrast, AtSPL11 (854 peaks) or AtSPL13 (no peaks) may have potential DNA binding
324 activities that are less detectable under the tested experimental conditions in DAP-seq, or they
325 might require cofactors to enhance their DNA binding activity. The genome-wide binding site
326 maps of the AtSPL members showed both conserved and variable binding signals at the
327 promoter regions of the well-known downstream targets, for example, the flowering
328 development gene *SOC1* and an auxin response gene *ARF8* (Figure 1B and 1C). The
329 distribution of binding events relative to genomic features was similar for all the AtSPL
330 members, with around 30-40% of the binding events located in proximal promoter regions of -
331 1000 bp to 500 bp from the transcription start sites (TSS) (Figure 1D).

332 To gain insight into the potential biological functions of the AtSPL family members based on
333 their DNA binding profiles, we performed Gene Ontology (GO) enrichment analysis of genes
334 associated with the top 3000 binding events ranked by DAP-seq binding site scores. The top ten
335 enriched biological processes across family members revealed their shared and distinct
336 functions (Figure 1E). Notably, hormonal responses, such as response to auxin and abscisic
337 acid (ABA), were highly enriched for AtSPL15 and AtSPL9. Hypoxia and oxygen-related
338 responses were more enriched for AtSPL15, AtSPL1, AtSPL2, AtSPL4 and AtSPL12. In
339 contrast, the GO term “photosynthesis, light harvesting in photosystem I” was uniquely enriched
340 for AtSPL8, suggesting a potential role in capturing light energy. Interestingly, metal ion
341 homeostasis processes, such as cellular copper ion homeostasis and transition metal ion
342 transport, were predominantly enriched for AtSPL7 DAP-seq targets, consistent with its

343 previously characterized function in regulating copper uptake and distribution, which is essential
344 for plants' adaptation to low copper conditions (13, 52, 61, 62).

345 We next evaluated the potential for DAP-seq reported binding events to regulate gene
346 expression by testing the association between the DAP-seq predicted target genes and
347 differentially expressed genes (DEG) in *spl* mutants, in particular, in the root of *spl7* mutant from
348 a published dataset (52). Using Gene Set Enrichment Analysis (GSEA) tests (63, 64), we found
349 that among the AtSPL members that had strong GO term enrichment, DAP-seq targets of
350 AtSPL7 were the most enriched for down-regulated DEGs between *spl7* mutant and wild-type
351 Col-0 (up-regulated comparing wild-type and mutant; Figure 1F), suggesting that DAP-seq can
352 predict transcriptionally regulated genes by the relevant TFs for further functional studies.

353 **Genome-wide binding and DNA sequence motifs support two major classes in the AtSPL**
354 **family**

355 To investigate the relationships between the *Arabidopsis* SPL family members in terms of their
356 DNA binding specificities, we first clustered the AtSPLs based on their genome-wide binding
357 patterns reported by DAP-seq. To do this, we first determined the regions of the highest DAP-
358 seq read signals among all the AtSPL binding profiles and used the read enrichment signal in
359 these regions to calculate the pairwise distances between the individual TFs (47). Hierarchical
360 clustering of the distance matrix revealed two major subgroups (Figure 2A): Class A containing
361 four SPLs (AtSPL2, 10, 7 and 11) and Class B containing the rest of the SPLs (AtSPL9, 15, 1,
362 12, 4, 14, 5, 6, 3, 8). We then performed *de novo* motif discovery for each AtSPL using
363 sequences in the 1000 peaks with the highest significance of DAP-seq read enrichment in each
364 experiment. We found that the most significant motifs recognized by the Class A members
365 shared a core GTAC motif, flanked by a weaker T base on either side, while Class B members
366 preferred a core GTACGG motif (Figure 2B). These results suggest Class A and Class B
367 members could bind to distinct sets of genes. For example, the well-known gene *HSL2* (also
368 known as *VAL2*), which promotes vegetative phase change (65), showed strong binding signals
369 at its promoter region from all Class B members that preferred the GTACGG motif. In contrast,
370 relatively low binding signals were observed for Class A members that preferred the GTAC motif
371 (Figure 2C). Additionally, in Class B, motifs for AtSPL9 and AtSPL15 (close relatives derived
372 from a duplication event) contained weaker GG consensus sequence adjacent to the GTAC
373 core, suggesting these two paralogs bind DNA with more flexibility outside of the core motif than
374 other Class B members.

375 **DAP-seq revealed differential DNA binding specificities between paralogs AtSPL9 and**
376 **AtSPL15**

377 To investigate the functional coherence or divergence of the evolutionarily conserved AtSPL9
378 and AtSPL15, we conducted differential binding analysis (54) between the AtSPL9 and AtSPL15
379 DAP-seq data (54). We identified 5,066 AtSPL15-prefered and 3,122 AtSPL9-prefered binding
380 sites at FDR threshold of 0.05 (Figure 3A; Supplementary Table S2), about 18% and 15% of the
381 binding sites for these two SPLs, indicating potential functional distinctions between them
382 (Figure 3A). Strikingly, distinct biological functions were enriched for genes associated with the
383 AtSPL15- and AtSPL9-preferred binding events (Figure 3B). AtSPL15-preferred gene targets
384 were strongly enriched in processes related to ion homeostasis, positive regulation of
385 transcription, response to oxygen and hypoxia, and auxin response. In contrast, AtSPL9-
386 preferred gene targets were mainly enriched in secondary cell wall biogenesis, responses to
387 photooxidative stress and regulation of the abscisic acid (ABA)-activated signaling pathway.
388 These results suggest that beyond the conserved functions, AtSPL15 and AtSPL9 could have
389 distinct roles in regulating different biological processes, and DNA binding preferences in the
390 genome could mediate their specialized functions in plant development and stress responses.

391 To understand the mechanism underlying the differential DNA binding preferences between
392 AtSPL9 and AtSPL15, we performed *de novo* motif discovery using the sequences in the 1000
393 differentially bound peaks with the highest increase in binding by AtSPL15 and AtSPL9,
394 respectively (Figure 3C). We found that the most significant motifs in the AtSPL15-preferred
395 binding sites contained the consensus sequences GTACgg and TGTACT, while AtSPL9
396 preferred motifs contained the consensus sequences GNCGG and GTAC. As an example, the
397 differentially bound peak at the promoter region of the gene *DROUGHT-INDUCED 21* (*AtDI21*;
398 AT4G15910) showed strong binding by AtSPL9 but not by AtSPL15, and it contained a binding
399 site sequence GTAC (Figure 3D). To validate that the differential binding events could influence
400 gene expression, we performed reporter assays (66) by cloning a 758 bp sequence from the
401 promoter region of *DI21*, including the 200 bp AtSPL9-preferred DAP-seq peak, into the reporter
402 plasmid *pUC-GUS* (*DI21pro::GUS*) to drive the GUS reporter gene expression. When we
403 transiently expressed AtSPL9 and AtSPL15 by transfecting SPL9 and SPL15 separately, we
404 found that AtSPL9 induced significantly higher reporter activation compared to AtSPL15 (Figure
405 3E). These results support our hypothesis that AtSPL15 and AtSPL9 have distinct binding
406 preferences to DNA, resulting in differences in target gene regulation.

407 **Comparative analysis of SPL target genes in *Arabidopsis*, maize, and wheat suggests**
408 **conserved functions**

409 To characterize the DNA binding specificities among SPL homologs across plant species, we
410 compared the conserved genes associated with DAP-seq peaks - referred to as “target genes” -
411 of AtSPL9/15 in *Arabidopsis*, ZmSBP8/30 in maize (67) and TaSPL7/13 in wheat (including
412 three paralogs from the A/B/D subgenome) (26) from DAP-seq experiments conducted on their
413 respective genomes (26, 67). By comparing the target genes of maize ZmSBP8/30 to those of
414 *Arabidopsis* AtSPL9/15, we identified 4,557 unique target genes for AtSPL9/15, 1,125 unique
415 target genes for ZmSBP8/30, and 492 conserved target genes shared between the two species
416 (Figure 4A; Supplementary Table S3). Similarly, when comparing wheat TaSPL7/13 target
417 genes with the target genes of AtSPL9/15 in *Arabidopsis*, we found 4,017 unique target genes
418 for *Arabidopsis* AtSPL9/15, 2,709 unique target genes for wheat TaSPL7/13, and 939
419 conserved target genes (Figure 4B; Supplementary Table S4).

420 To gain insight into the function of these conserved target genes, we performed GO enrichment
421 analysis and found four major clusters of enriched biological processes (Figure 4C). Cluster 2
422 and Cluster 4 highlighted functions conserved across the three species, primarily related to cell
423 wall biogenesis, transcriptional regulation, and pattern specification. Cluster 1 also revealed
424 functions that were conserved among all three species, such as xylem and cell wall
425 development, while also suggesting *Arabidopsis*-wheat-specific conserved processes such as
426 cell wall modification, secondary metabolite biosynthesis, and lignin metabolic processes. These
427 GO terms suggest the potential role of TaSPL7/13 in specific aspects of wheat architecture,
428 possibly related to structural integrity and stress responses. Cluster 3 GO terms were
429 significantly enriched for SPL targets conserved between *Arabidopsis* and maize, including
430 processes associated with different developmental stages, such as floral organ development,
431 cotyledon development, and post-embryonic development. Additionally, we observed processes
432 such as plant organ specification and gibberellin response, which are closely related to
433 flowering time and overall plant development. Overall, these results suggest that different
434 biological functions and developmental processes are targeted by the SPL shared target genes
435 in different species, offering insights into the evolutionary trajectories of the SPL homologs in
436 the three species spanning monocot to dicot.

437 **SPL homologs bind to variations of the GTAC motifs**

438 Since conservation of binding motifs is often used as the basis for cross-species prediction of
439 *cis*-regulatory elements, we analyzed the enriched DNA sequence motifs in the DAP-seq peaks
440 associated with the conserved target genes in the three species. Unexpectedly, we identified
441 three different types of DNA binding motifs, although they all shared or partially shared the core
442 GTAC sequence (Figure 4D-E, red highlights). In the *Arabidopsis* genome, AtSPL9/15
443 recognizes motifs with a GTAC core adjacent to a GG dinucleotide. In the maize genome,
444 ZmSBP8/30 recognizes a back-to-back repeat of the GTAC sequence, GTACGTRC, although
445 one of the GTAC elements is weak. In the wheat genome, TaSPL7/13 binds to a motif
446 containing two strong GTAC elements overlapping by one base pair, resulting in the consensus
447 GTASTAC (GTACTAC or GTAGTAC). These different yet related motif architectures suggest
448 differences in genome-wide DNA recognition mechanisms among the SPL homologs.
449 Importantly, the more complex back-to-back or overlapping motifs in maize and wheat may
450 contribute to enhancing binding specificities in the much larger and complex genomes of maize
451 (2.3 Gbp) and wheat (17 Gbp) relative to *Arabidopsis* (135 Mbp).

452 **Structure modeling links protein-protein interaction preferences to DNA binding motif**
453 **variation across SPL homologs**

454 To explore the potential structural mechanisms underlying the motif variation among SPL
455 homologs, we performed structural prediction of the SPL-DNA binding complexes using
456 AlphaFold3 (AF3) (58). We began by modeling the interaction between full-length SPL proteins
457 and DNA sequences containing their species-preferred binding motifs. These sequences were
458 drawn from DAP-seq peaks associated with conserved target genes, representing binding
459 events that are conserved across species and therefore likely to be functionally important. As
460 shown in Figure 5A for the representative homologs, TaSPL13A interacting with the TaSPL-
461 preferred GTASTAC motif and AtSPL9 interacting with the AtSPL-preferred GTAC motif, the
462 predicted structures had low overall confidence scores (ipTM: 0.51-0.59; pTM: 0.29-0.31), likely
463 due to high fractions of disordered regions outside the conserved SBP DNA-binding domain
464 (fraction disordered: 0.80-0.81). Nevertheless, when we superimposed the AF3-predicted
465 structures with the experimentally solved X-ray crystal structure of AtSPL5 (PDB 8J49 Chain A
466 (68)), we observed strong structural similarity in the SBP domains, including two well-conserved
467 zinc-binding pockets (RMSD 0.685-0.672 Å, 63 Cα atoms; Figure 5B). Furthermore, the
468 predicted structures consistently positioned the zinc finger regions in the C-terminal portion of
469 the SBP domain in proximity to the DNA (Figure 5B), consistent with prior solution studies
470 showing that this zinc-binding site is required for DNA binding (69). While these results

471 confirmed structural conservation of the SBP domain across species, they did not readily
472 explain the observed motif divergence across species.

473 We next hypothesized that the wheat SPLs' strong preference for the overlapping GTAC motifs
474 (GTASTAC) may arise from enhanced protein-protein interactions (PPIs) between SPL
475 monomers. This hypothesis was motivated by three observations. First, DAP-seq has been
476 shown to capture dimeric protein binding to DNA for several TF families in multiple species (45,
477 70, 71). Secondly, DNA-guided TF interaction is known to produce overlapping or composite
478 motifs that partially reflect specificities of individual TFs (70, 72). Third, dimerization has been
479 observed in several SPLs in different species. In *Arabidopsis*, structural and biochemical studies
480 show that the SBP domains of *Arabidopsis* SPLs bind DNA as monomers (69, 73, 74), while
481 full-length proteins like AtSPL7 and AtSPL9 are capable of dimerization, though the functional
482 role of this dimerization in DNA binding remains unclear (75, 76). In particular, AtSPL7
483 dimerization may modulate its regulatory activity *in vivo*, possibly by preventing its translocation
484 from the cytoplasm to the nucleus (75). In maize, the SBP-box TF ZmTGA1 forms both
485 monomeric and dimeric complexes to bind DNA. Notably, the dimeric form of ZmTGA1 exhibits
486 higher stability than that of its teosinte ortholog, and this increased dimer stability was implicated
487 as a key molecular mechanism underlying the naked kernel phenotype that distinguishes
488 domesticated maize from its wild progenitor, teosinte (77). Together, these observations
489 suggest that PPI-driven DNA binding could underlie the emergence of complex binding motifs in
490 SPLs.

491 To test this hypothesis, we used AF3 to predict SPL-DNA complex structures with two copies of
492 each SPL interacting with their preferred binding motif (species-preferred motif) or with a motif
493 preferred by SPL homologs from a different species (cross-species motif) (see MATERIALS
494 AND METHODS). These DNA sequences were drawn from genomic regions associated with
495 conserved SPL target genes and centered on high-scoring motif matches either within a SPL's
496 own DAP-seq peaks (species-preferred motif sequences) or from nearby unbound regions
497 containing the motif of the SPL homolog in a different species (cross-species motif sequences).
498 This design allowed us to compare predicted structural interactions in sequence contexts that
499 are biologically relevant while minimizing confounding effects from local genomic features.
500 Figure 5C shows representative results for TaSPL13A, comparing its interaction with the
501 species-preferred GTASTAC motif (structure i) versus the cross-species GTAC motif preferred
502 by AtSPLs (structure ii). In the predicted aligned error (PAE) matrices, a strong low-error block
503 was present between the two TaSPL13A monomers in the GTASTAC-bound complex but

504 absent in the GTAC-bound complex (cyan arrows), suggesting a more stable dimer interface
505 between the two TaSPLA monomers in the presence of the species-preferred motif.

506 Since AF3 is better at predicting the overall geometry of TF-TF-DNA complexes than localizing
507 the precise TF-TF or TF-DNA contacts (70, 78), we did not analyze the direct contacts in the
508 predicted structures. Instead, we calculated the local interaction score (LIS) between the
509 subunits in the predicted structures (60). LIS has improved sensitivity in detecting PPIs,
510 particularly for those that involve flexible and small interfaces where traditional pTM scores are
511 less informative (60). The LIS heatmaps shown in Figure 5C confirmed that both monomer-
512 monomer (cyan boxes) and monomer-DNA (blue boxes) interactions were stronger in the
513 TaSPL13A complex bound to the GTASTAC motif (structure i) compared to that bound to the
514 GTAC motif (structure ii). In contrast, control structures predicted using a single copy of
515 TaSPL13A (structures iii and iv; Figure 5C) did not show differences in LIS values for the
516 monomer-DNA interaction (purple boxes) that are consistent with the motif preferences,
517 supporting the idea that PPIs between TaSPL13A monomers stabilize the binding to the
518 GTASTAC motif.

519 To generalize these results, we similarly modeled one- and two-copy SPL-DNA complexes for
520 AtSPL9, AtSPL15, six wheat TaSPL7/13 paralogs, and two maize ZmSBPs ZmSBP8 and
521 ZmSBP30. For each protein, we predicted structures with both the species-preferred motif
522 sequences and cross-species motif sequences (see MATERIALS AND METHODS), and
523 computed the LIS values for the monomer-monomer and monomer-DNA interactions. The
524 resulting LIS distributions across ten motif instances for species-preferred and cross-species
525 motifs are shown as dotplots and boxplots in Figure 5D. Statistical comparisons between
526 predictions for species-preferred versus cross-species motifs (Wilcoxon rank-sum test, $p < 0.05$;
527 orange background panels in Figure 5D) revealed a consistent and striking pattern. *Arabidopsis*
528 AtSPL9 and AtSPL15 showed significantly higher LIS values for the GTAC motif only in the one-
529 copy models. In contrast, most wheat SPL paralogs exhibited significantly stronger monomer-
530 monomer and monomer-DNA interaction scores when modeled with the GTASTAC motif, but
531 only in the two-copy structures. These results support a model in which PPIs between TaSPL
532 monomers enhance binding to the overlapping GTASTAC motif, whereas AtSPLs bind DNA as
533 monomers with a preference for the short GTAC motif.

534 Interestingly, applying the same analysis to maize ZmSBP8 and ZmSBP30 did not show
535 consistent preferences between one-copy and two-copy structures for either the GTAC or

536 GTACGTRC motifs (AtSPL-preferred or ZmSBP-preferred) (Figure S1). This may reflect the
537 flexibility of maize SBPs in forming monomeric or dimeric complexes, such as those found for
538 ZmTGA1 (77).

539 **DISCUSSION**

540 TFs regulate gene expression by binding to specific DNA sequences, and variations in their
541 binding sites can lead to differential regulation of genes within a TF family in a species and
542 across different species. Our analysis of the DAP-seq binding profiles of the *Arabidopsis* SPL
543 family revealed two major classes of binding site motifs (Figure 2), yet the enriched GO terms of
544 the target genes near the most enriched binding sites were much more diverse among the
545 family members (Figure 1E). This suggests that even closely related TFs, such as AtSPL9 and
546 AtSPL15, can develop distinct regulatory roles through subtle differences in their DNA-binding
547 preferences. By analyzing this comprehensive cistrome dataset for SPL TF family in
548 *Arabidopsis*, our study offers insights into how specific DNA-binding characteristics contribute to
549 the functional diversity within the SPL family of TFs.

550 Gene duplication is a key feature of genomic architecture, especially within TF families, leading
551 to opportunities for subfunctionalization and neofunctionalization (79–82). Some duplicated TFs
552 retain ancestral functions by maintaining their original DNA binding preferences, while others
553 acquire mutations that allow them to bind different DNA sequences and thus regulate different
554 sets of genes (2, 83, 84). Such changes can enhance a species' ability to respond to broader
555 environmental challenges including biotic or abiotic stresses, or activate new developmental
556 pathways for adaptation. DAP-seq offers a powerful approach to investigate this functional
557 divergence mediated by DNA binding, allowing for detailed scrutiny of DNA binding motifs and
558 target genes among paralogous TFs, even when genetic redundancy makes functional studies
559 challenging. For example, our analysis revealed that even slight changes in DNA binding motifs
560 between the duplicated homologs AtSPL9 and AtSPL15 led to distinct target gene sets, with
561 AtSPL15 showing preferences for genes involved in auxin responses, while AtSPL9 favoring
562 targets related to ABA signaling. These variations help differentiate the roles of closely related
563 TFs in regulating downstream targets and ensuring functional specificity. We expect that the
564 same analysis can be applied to other TF homologs within the SBP/SPL family or extend to
565 other TF families to gain a broader understanding of TF functional diversity.

566 The evolution of TF-DNA interactions is a complex and dynamic process that involves multiple
567 factors, including changes in the TF proteins, their DNA-binding motifs, and the genomic context

568 of the binding sites such as DNA shape and chemical modifications (3). In this study, we
569 observed that the conserved functions of the SPL/SBP homologs based on their DNA binding
570 targets were distributed differently in *Arabidopsis*, maize, and wheat: while some functions
571 maintained across all three species, others were shared only between specific pairs, such as
572 *Arabidopsis* and maize or *Arabidopsis* and wheat. For example, common regulatory functions
573 related to developmental processes were found between *Arabidopsis* and maize, while
574 *Arabidopsis* and wheat shared functions in stress response and structural regulation. Analysis of
575 DAP-seq binding sequences revealed species-specific preferences for related but distinct motif
576 variants, which may contribute to the divergence of regulatory networks across species. These
577 findings provide insights into how the SPL/SBP TFs have evolved to balance the core,
578 conserved functions with species-specific adaptations, highlighting the power of DAP-seq to
579 shed light on the evolutionary trajectories of TF networks across plant species.

580 A particularly interesting finding emerged when we compared the DNA binding motifs among
581 the SPL/SBP homologs across species, revealing that the SPL/SBP homologs in maize and
582 wheat bind to longer motif types that are variations of the core GTAC sequences recognized by
583 the *Arabidopsis* protein. Specifically, the GTACGTRC motif in maize is a back-to-back
584 palindromic sequence of two GTAC units, while the GTASTAC motif in wheat is a palindromic
585 sequence with one base pair overlap between two GTAC units (Figure 4E). Palindromic motifs
586 have also been observed for dimers of the bZIP family of TFs, suggesting such motifs can be a
587 general mechanism for refining TF-DNA recognition (64). Many TF families, including
588 AP2/ERF, bZIP, ARF, homeodomain and MADS-box, are known to form dimers or higher-order
589 complexes to bind DNA and regulate gene expression (28, 85, 86). Using AF3 structural
590 modeling, we found that TaSPL7/13 in wheat may recognize the palindromic GTASTAC motif
591 through DNA-guided dimerization, enabling binding to a more complex motif configuration than
592 the GTAC motif preferred by AtSPLs. The ability to recognize longer, more complex motifs
593 through PPI could be a mechanism that allows SBP/SPL TFs to increase DNA binding
594 specificities to achieve precise gene regulation, particularly in the context of complex plant
595 genomes like those of maize and wheat. Our findings underscore the potential role of PPI in the
596 evolution of DNA recognition mechanisms by TFs, offering insights into how TFs adapt to
597 regulate diverse regulatory programs in plants with highly divergent genome complexities.

598 **AUTHOR CONTRIBUTIONS**

S.C.H. and M.L.: Conceptualization, Formal analysis, Visualization, Methodology, Validation, Writing-original draft. W.L. and Y.Z.: Validation. M.G. and A.G.: Formal analysis. T.Y. and J.C.: Methodology, Validation. L.F.D.C.: Investigation, Methodology. S.C.H. and A.G.: Funding acquisition. All authors: review & editing.

ACKNOWLEDGEMENTS

We thank Joseph Ecker for the gift of TF expression plasmids. We also thank Aurelia Li for technical assistance. We acknowledge the Zegar Family Foundation for their generous support. 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 was also in part based on 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; Biological and Environmental Research Program under Award Number ERKPA88; and the Oak Ridge National Laboratory Director's R&D (DRD) Program 12093. 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.

CONFLICT OF INTEREST

The authors declare they have no competing interests.

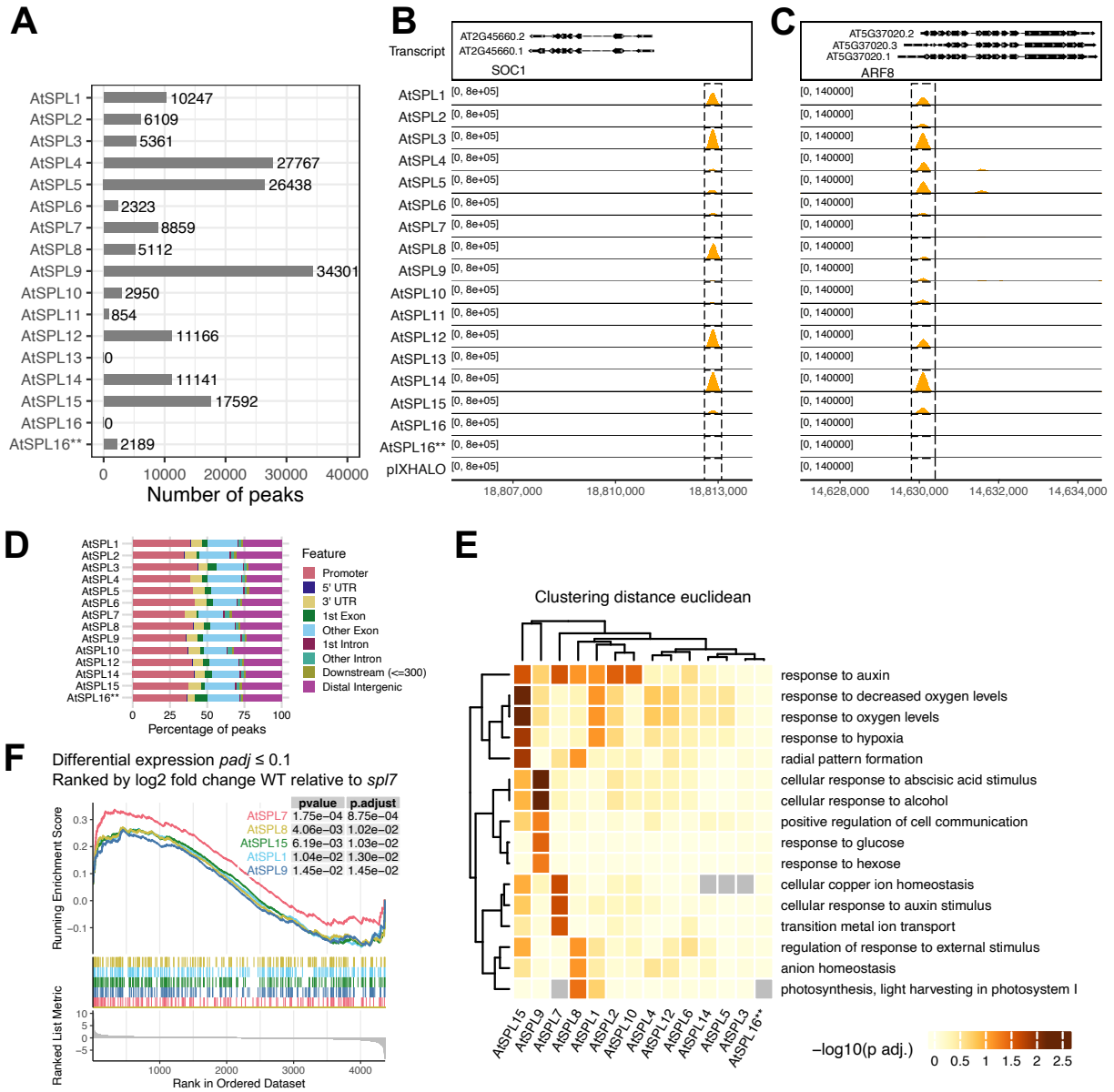


Figure 1: DAP-seq identifies genome-wide binding events for SPL TFs that are biologically relevant. (**A**) Number of DAP-seq peaks for all SPL members in *Arabidopsis*. ** indicates ampDAP-seq. (**B**) *Arabidopsis* AtSPL DAP-seq binding signal at the promoter region of a known target gene *SOC1*. (**C**) *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 for DAP-seq predicted targets of *Arabidopsis* SPLs. (**F**) GSEA of AtSPL DAP-seq targets and the differentially expressed genes in *spl7* mutant vs. Col-0 wild type (WT).

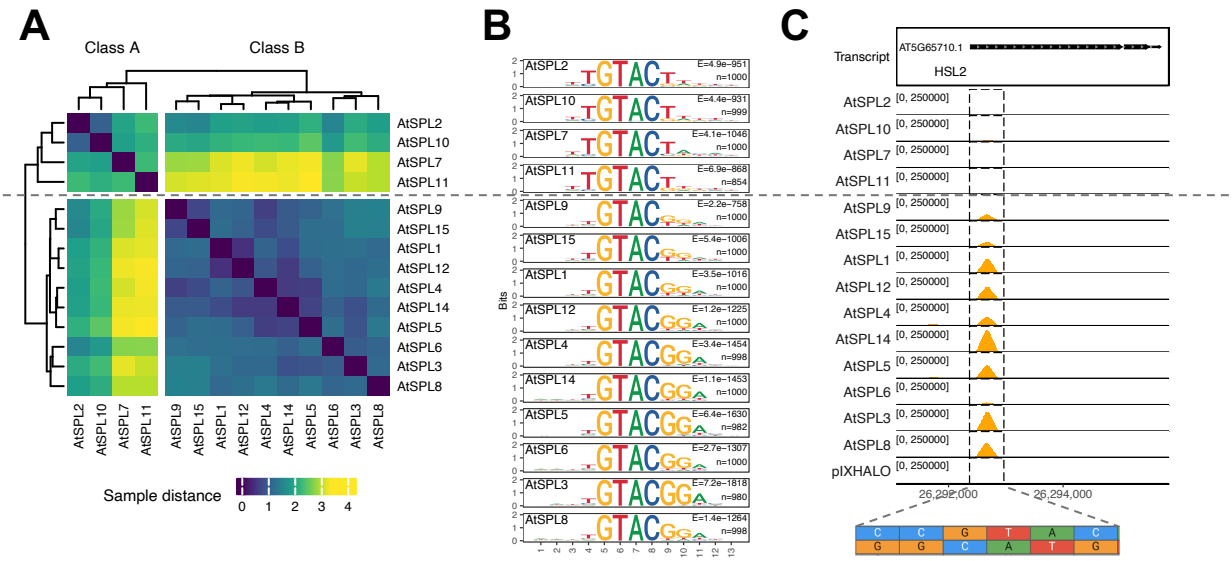


Figure 2: Genome-wide binding profile comparison and DNA motifs of the *Arabidopsis* SPL family. (**A**) Hierarchical clustering of AtSPL DAP-seq binding profiles splits the family members into Class A and Class B. (**B**) PWM models of the enriched motifs from the top 1000 DAP-seq peaks for each AtSPL.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

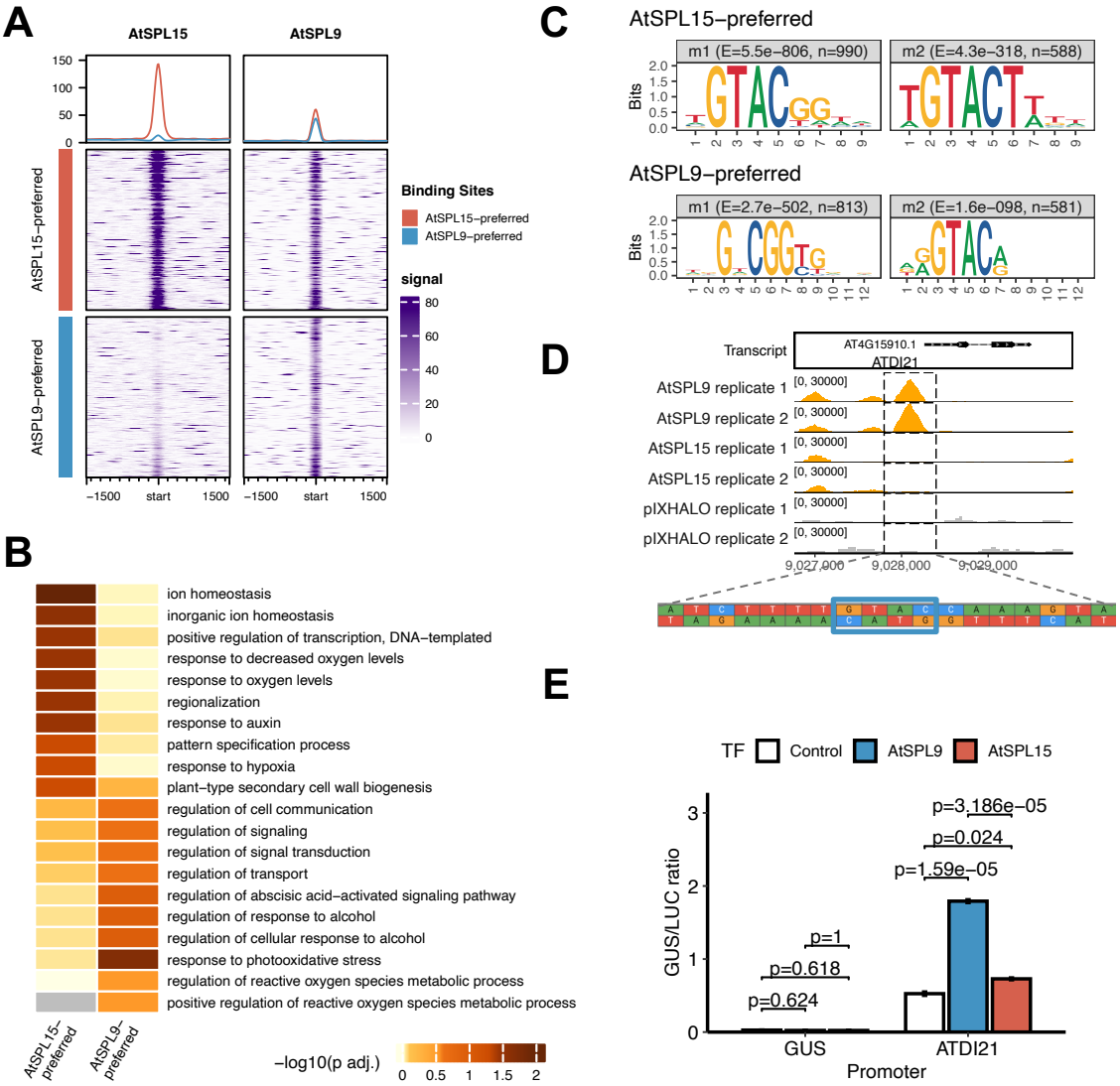


Figure 3: Differential binding events, target genes and motifs comparing *Arabidopsis* AtSPL9 versus AtSPL15. (**A**) DAP-seq read signals at differentially bound peaks of AtSPL15 and AtSPL9. (**B**) Top 10 enriched GO Biological Process terms for genes associated with AtSPL15-preferred and AtSPL9-preferred peaks. (**C**) PWM motif models enriched in AtSPL15- and AtSPL9-preferred binding sites. (**D**) AtSPL9-preferred binding site upstream of gene *AtDI21*. (E) Transient expression of AtSPL9 and AtSPL15 induced significantly different expressions of the reporter AtDI21::GUS. Error bars are standard error of the mean (n=3) and P-values were computed by two sample t-tests followed by Bonferroni correction.

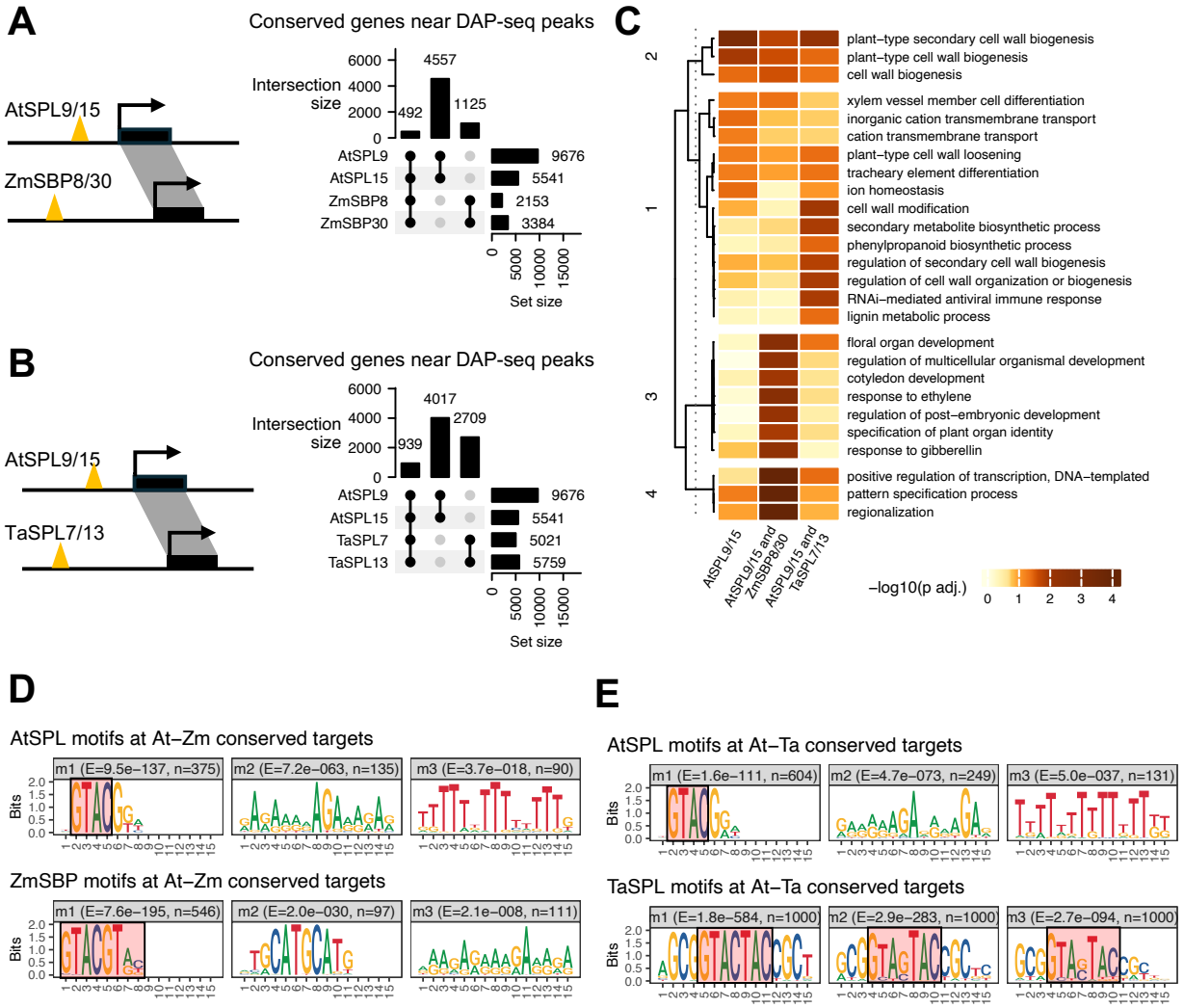


Figure 4: Comparative analysis of DAP-seq target genes and binding site sequences revealed variation of DNA binding mechanism among SPL homologs. (**A**) Upset plot comparing the target gene overlap between *Arabidopsis* AtSPL9/15 and maize ZmSBP8/30. (**B**) UpSet plot comparing the target gene overlap between *Arabidopsis* AtSPL9/15 and wheat TaSPL7/13. (**C**) GO enrichment analysis showing the biological processes enriched for targets conserved between SPL homologs between *Arabidopsis* and maize and between *Arabidopsis* and wheat. (**D**) DNA binding motifs discovered by MEME for DAP-seq peaks associated with conserved SPL targets in *Arabidopsis*, maize and wheat.

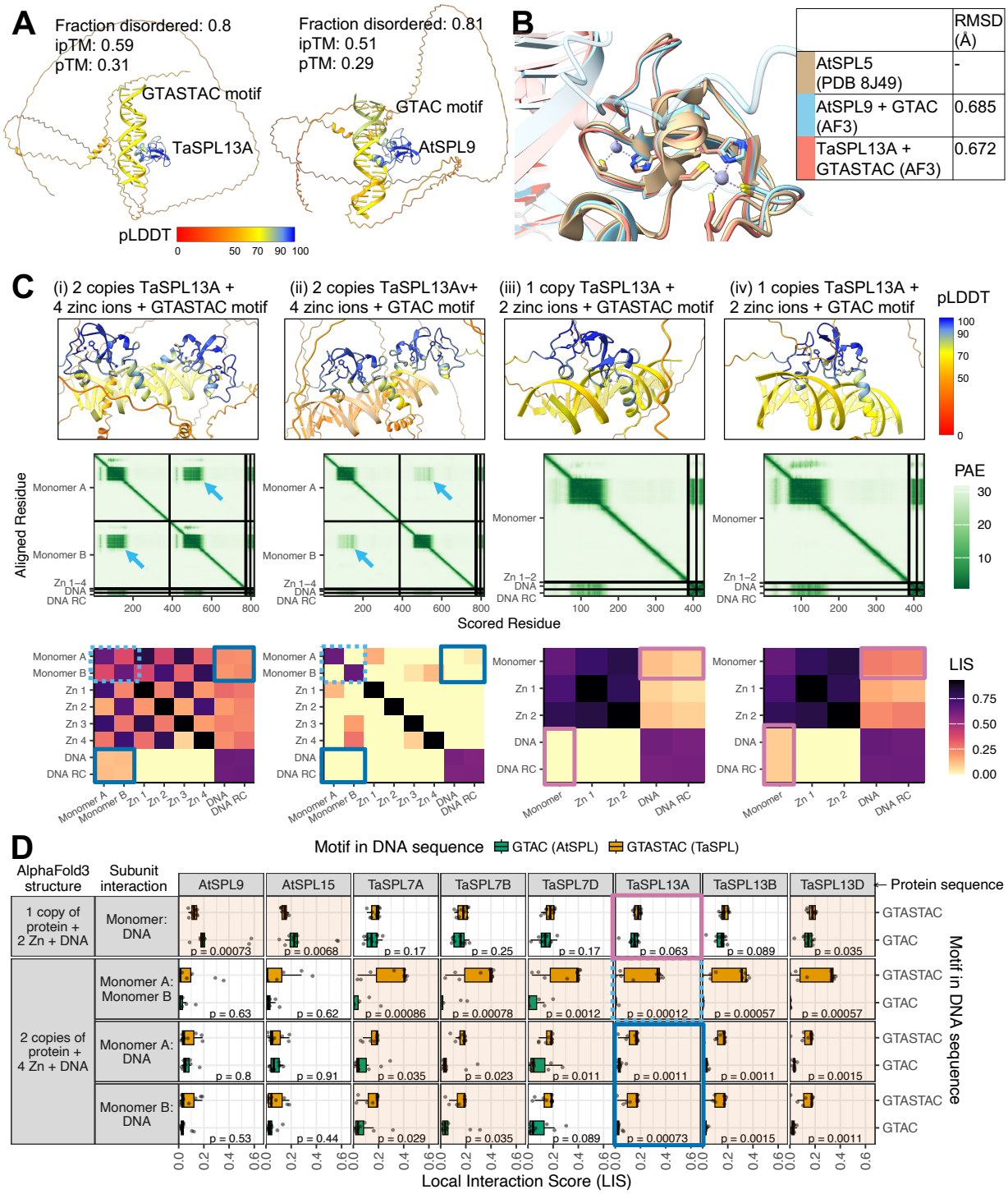


Figure 5: AlphaFold3 (AF3) structure modeling of DNA binding by *Arabidopsis* and wheat SPL proteins. (**A**) Representative AF3-predicted structures of TaSPL13A binding to its preferred DNA motif GTASTAC (left) and AtSPL9 binding to its preferred motif GTAC (right). (**B**) Superposition of the SBP domains from the predicted structures in (A) with the X-ray crystal structure of AtSPL5 (PDB: 8J49), highlighting the conserved zinc-binding regions. (**C**) AF3-

661 predicted complex structures, Predicted aligned error (PAE) matrices, and Local Interaction
662 Score (LIS) heatmaps for one or two copies of TaSPL13A interacting with either the species-
663 preferred GTASTAC motif or the cross-species GTAC motif (see MATERIALS AND
664 METHODS). Colored boxes highlight predicted monomer-monomer (cyan), monomer-DNA
665 (blue) interactions in two-copy protein structures, and monomer-DNA interactions (purple) in
666 one-copy protein structures. RC: Reverse Complement. (**D**) Summary of LIS values across
667 predicted SPL-DNA structures. Each point represents the median LIS from n≥6 replicate
668 predictions for a given motif instance. For *Arabidopsis* SPLs (AtSPL9 and AtSPL15), structures
669 were modeled with one or two protein copies interacting with either GTAC motifs (species-
670 preferred motifs bound in DAP-seq) or GTASTAC motifs (cross-species motifs, unbound). For
671 wheat SPLs (TaSPL7A/B/D and TaSPL13A/B/D), structures were modeled analogously with
672 GTASTAC (species-preferred motifs bound in DAP-seq) and GTAC (cross-species motifs,
673 unbound) motifs. Boxplots summarize monomer-monomer and monomer-DNA interaction
674 scores for each configuration. P-values were computed by two sample Wilcoxon rank-sum tests
675 between species-preferred and cross-species motif conditions and orange-shaded panels
676 indicate statistically significant differences ($p < 0.05$).
677
678

679 **REFERENCES**

680 1. Panchy,N., Lehti-Shiu,M. and Shiu,S.-H. (2016) Evolution of Gene Duplication in Plants. *Plant*
681 *Physiol.*, **171**, 2294–2316.

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

685 3. Rogers,J.M. and Bulyk,M.L. (2018) Diversification of transcription factor-DNA interactions and
686 the evolution of gene regulatory networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **10**,
687 e1423.

688 4. Birkenbihl,R.P., Jach,G., Saedler,H. and Huijser,P. (2005) Functional dissection of the plant-
689 specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *J.*
690 *Mol. Biol.*, **352**, 585–596.

691 5. Klein,J., Saedler,H. and Huijser,P. (1996) A new family of DNA binding proteins includes
692 putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene
693 SQUAMOSA. *Mol. Gen. Genet.*, **250**, 7–16.

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

696 7. Xu,M., Hu,T., Smith,M.R. and Poethig,R.S. (2016) Epigenetic Regulation of Vegetative Phase
697 Change in Arabidopsis. *Plant Cell*, **28**, 28–41.

698 8. Wang,J., Zhou,L., Shi,H., Chern,M., Yu,H., Yi,H., He,M., Yin,J., Zhu,X., Li,Y., *et al.* (2018) A
699 single transcription factor promotes both yield and immunity in rice. *Science*, **361**, 1026–
700 1028.

701 9. Zheng,C., Ye,M., Sang,M. and Wu,R. (2019) A Regulatory Network for miR156-SPL Module
702 in. *Int. J. Mol. Sci.*, **20**.

703 10. Yu,N., Niu,Q.-W., Ng,K.-H. and Chua,N.-H. (2015) The role of miR156/SPLs modules in
704 Arabidopsis lateral root development. *Plant J.*, **83**, 673–685.

705 11. The miR156/SPL Module, a Regulatory Hub and Versatile Toolbox, Gears up Crops for
706 Enhanced Agronomic Traits (2015) *Mol. Plant*, **8**, 677–688.

707 12. Jerome Jeyakumar,J.M., Ali,A., Wang,W.-M. and Thiruvengadam,M. (2020) Characterizing
708 the Role of the miR156-SPL Network in Plant Development and Stress Response.
709 *Plants*, **9**.

710 13. Yamasaki,H., Hayashi,M., Fukazawa,M., Kobayashi,Y. and Shikanai,T. (2009) SQUAMOSA
711 Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in
712 Arabidopsis. *Plant Cell*, **21**, 347–361.

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

717 15. Nagpal,P., Ellis,C.M., Weber,H., Ploense,S.E., Barkawi,L.S., Guilfoyle,T.J., Hagen,G.,
718 Alonso,J.M., Cohen,J.D., Farmer,E.E., *et al.* (2005) Auxin response factors ARF6 and
719 ARF8 promote jasmonic acid production and flower maturation. *Development*, **132**,
720 4107–4118.

721 16. Wu,G., Park,M.Y., Conway,S.R., Wang,J.-W., Weigel,D. and Poethig,R.S. (2009) The
722 sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis.
723 *Cell*, **138**, 750–759.

724 17. Hu,L., Qi,P., Peper,A., Kong,F., Yao,Y. and Yang,L. (2023) Distinct function of SPL genes in
725 age-related resistance in Arabidopsis. *PLoS Pathog.*, **19**, e1011218.

726 18. Wang,J.-W., Schwab,R., Czech,B., Mica,E. and Weigel,D. (2008) Dual Effects of miR156-
727 Targeted SPL Genes and CYP78A5/KLUH on Plastochron Length and Organ Size in
728 Arabidopsis thaliana. *Plant Cell*, **20**, 1231–1243.

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

731 20. Preston,J.C. and Hileman,L.C. (2013) Functional Evolution in the Plant SQUAMOSA-
732 PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Front. Plant Sci.*, **4**, 80.

733 21. Schwarz,S., Grande,A.V., Bujdoso,N., Saedler,H. and Huijser,P. (2008) The microRNA
734 regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis.

735 *Plant Mol. Biol.*, **67**, 183–195.

736 22. Hyun,Y., Richter,R., Vincent,C., Martinez-Gallegos,R., Porri,A. and Coupland,G. (2016)
737 Multi-layered Regulation of SPL15 and Cooperation with SOC1 Integrate Endogenous
738 Flowering Pathways at the Arabidopsis Shoot Meristem. *Dev. Cell*, **37**, 254–266.

739 23. Song,X., Lu,Z., Yu,H., Shao,G., Xiong,J., Meng,X., Jing,Y., Liu,G., Xiong,G., Duan,J., *et al.*
740 (2017) IPA1 functions as a downstream transcription factor repressed by D53 in
741 strigolactone signaling in rice. *Cell Res.*, **27**, 1128–1141.

742 24. Chuck,G.S., Brown,P.J., Meeley,R. and Hake,S. (2014) Maize SBP-box transcription factors
743 unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral
744 primordia initiation. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 18775–18780.

745 25. Wei,H., Zhao,Y., Xie,Y. and Wang,H. (2018) Exploiting SPL genes to improve maize plant
746 architecture tailored for high-density planting. *J. Exp. Bot.*, **69**, 4675–4688.

747 26. Pei,H., Teng,W., Gao,L., Gao,H., Ren,X., Liu,Y., Jia,J., Tong,Y., Wang,Y. and Lu,Z. (2023)
748 Low-affinity SPL binding sites contribute to subgenome expression divergence in
749 allohexaploid wheat. *Sci. China Life Sci.*, **66**, 819–834.

750 27. Bartlett,A., O'Malley,R.C., Huang,S.-S.C., Galli,M., Nery,J.R., Gallavotti,A. and Ecker,J.R.
751 (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat.*
752 *Protoc.*, **12**, 1659–1672.

753 28. O'Malley,R.C., Huang,S.-S.C., Song,L., Lewsey,M.G., Bartlett,A., Nery,J.R., Galli,M.,
754 Gallavotti,A. and Ecker,J.R. (2016) Cistrome and Epicistrome Features Shape the
755 Regulatory DNA Landscape. *Cell*, **165**, 1280–1292.

756 29. Li,M. and Huang,S.-S.C. (2022) DNA Affinity Purification Sequencing (DAP-Seq) for
757 Mapping Genome-Wide Transcription Factor Binding Sites in Plants. *Accelerated*
758 *Breeding of Cereal Crops*, 10.1007/978-1-0716-1526-3_15.

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

761 31. Broad Institute Picard Tools. *Broad Institute, GitHub Repository*.
762 http://broadinstitute.github.io/picard/.

763 32. Galanti,L., Shasha,D. and Gunsalus,K.C. (2021) Pheniqs 2.0: accurate, high-performance
764 Bayesian decoding and confidence estimation for combinatorial barcode indexing. *BMC*
765 *Bioinformatics*, **22**, 359.

766 33. New York University GENEFLOW. *New York University Center for Genomics and System*
767 *Biology Genomics Core, GitHub Repository*.
768 https://github.com/gencorefacility/GENEFLOW.

769 34. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing
770 reads. *EMBnet.journal*, **17**, 10–12.

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

773 36. Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A.,
774 Keane,T., McCarthy,S.A., Davies,R.M., *et al.* (2021) Twelve years of SAMtools and
775 BCFtools. *Gigascience*, **10**.

776 37. Guo,Y., Mahony,S. and Gifford,D.K. (2012) High resolution genome wide binding event
777 finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS*
778 *Comput. Biol.*, **8**, e1002638.

779 38. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C.,
780 Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based Analysis of ChIP-Seq (MACS).
781 *Genome Biol.*, **9**, 1–9.

782 39. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S.,
783 Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-
784 sequencing data analysis. *Nucleic Acids Res.*, **44**, W160-5.

785 40. Diesh,C., Stevens,G.J., Xie,P., De Jesus Martinez,T., Hershberg,E.A., Leung,A., Guo,E.,
786 Dider,S., Zhang,J., Bridge,C., *et al.* (2023) JBrowse 2: a modular genome browser with
787 views of synteny and structural variation. *Genome Biol.*, **24**, 74.

788 41. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to
789 discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.

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

792 43. Nettling,M., Treutler,H., Grau,J., Keilwagen,J., Posch,S. and Grosse,I. (2015) DiffLogo: a
793 comparative visualization of sequence motifs. *BMC Bioinformatics*, **16**, 1–9.

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

796 45. Galli,M., Chen,Z., Ghandour,T., Chaudhry,A., Gregory,J., Feng,F., Li,M., Schleif,N.,
797 Zhang,X., Dong,Y., *et al.* (2025) Transcription factor binding divergence drives
798 transcriptional and phenotypic variation in maize. *Nat Plants*, **11**, 1205–1219.

799 46. Zhu,T., Wang,L., Rimbert,H., Rodriguez,J.C., Deal,K.R., De Oliveira,R., Choulet,F., Keeble-
800 Gagnère,G., Tibbits,J., Rogers,J., *et al.* (2021) Optical maps refine the bread wheat
801 Triticum aestivum cv. Chinese Spring genome assembly. *Plant J.*, **107**, 303–314.

802 47. Tu,S., Li,M., Chen,H., Tan,F., Xu,J., Waxman,D.J., Zhang,Y. and Shao,Z. (2021) MAnorm2
803 for quantitatively comparing groups of ChIP-seq samples. *Genome Res.*, **31**, 131–145.

804 48. Wang,Q., Li,M., Wu,T., Zhan,L., Li,L., Chen,M., Xie,W., Xie,Z., Hu,E., Xu,S., *et al.* (2022)
805 Exploring Epigenomic Datasets by ChIPseeker. *Curr Protoc*, **2**, e585.

806 49. Cheng,C.-Y., Krishnakumar,V., Chan,A.P., Thibaud-Nissen,F., Schobel,S. and Town,C.D.
807 (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference
808 genome. *Plant J.*, **89**, 789–804.

809 50. Xu,S., Hu,E., Cai,Y., Xie,Z., Luo,X., Zhan,L., Tang,W., Wang,Q., Liu,B., Wang,R., *et al.*
810 (2024) Using clusterProfiler to characterize multiomics data. *Nat. Protoc.*,
811 10.1038/s41596-024-01020-z.

812 51. Gu,Z., Eils,R. and Schlesner,M. (2016) Complex heatmaps reveal patterns and correlations
813 in multidimensional genomic data. *Bioinformatics*, **32**, 2847–2849.

814 52. Kastoori Ramamurthy,R., Xiang,Q., Hsieh,E.-J., Liu,K., Zhang,C. and Waters,B.M. (2018)
815 New aspects of iron-copper crosstalk uncovered by transcriptomic characterization of
816 Col-0 and the copper uptake mutant spl7 in Arabidopsis thaliana. *Metallomics*, **10**, 1824–
817 1840.

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

820 54. Ross-Innes,C.S., Stark,R., Teschendorff,A.E., Holmes,K.A., Ali,H.R., Dunning,M.J.,
821 Brown,G.D., Gojis,O., Ellis,I.O., Green,A.R., *et al.* (2012) Differential oestrogen receptor
822 binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389–393.

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

825 56. Hufford,M.B., Seetharam,A.S., Woodhouse,M.R., Chougule,K.M., Ou,S., Liu,J., Ricci,W.A.,
826 Guo,T., Olson,A., Qiu,Y., *et al.* (2021) De novo assembly, annotation, and comparative
827 analysis of 26 diverse maize genomes. *Science*, **373**, 655–662.

828 57. Goodstein,D.M., Shu,S., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W.,
829 Hellsten,U., Putnam,N., *et al.* (2012) Phytozome: a comparative platform for green plant
830 genomics. *Nucleic Acids Res.*, **40**, D1178-86.

831 58. Abramson,J., Adler,J., Dunger,J., Evans,R., Green,T., Pritzel,A., Ronneberger,O.,
832 Willmore,L., Ballard,A.J., Bambrick,J., *et al.* (2024) Accurate structure prediction of
833 biomolecular interactions with AlphaFold 3. *Nature*, **630**, 493–500.

834 59. Meng,E.C., Goddard,T.D., Pettersen,E.F., Couch,G.S., Pearson,Z.J., Morris,J.H. and
835 Ferrin,T.E. (2023) UCSF ChimeraX: Tools for structure building and analysis. *Protein*
836 *Sci*, **32**, e4792.

837 60. Kim,A.-R., Hu,Y., Comjean,A., Rodiger,J., Mohr,S.E. and Perrimon,N. (2024) Enhanced
838 Protein-Protein Interaction Discovery via AlphaFold-Multimer. *bioRxiv*,
839 10.1101/2024.02.19.580970.

840 61. Araki,R., Mermod,M., Yamasaki,H., Kamiya,T., Fujiwara,T. and Shikanai,T. (2018) SPL7
841 locally regulates copper-homeostasis-related genes in Arabidopsis. *J. Plant Physiol.*,
842 **224–225**, 137–143.

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

847 63. Korotkevich,G., Sukhov,V., Budin,N., Shpak,B., Artyomov,M.N. and Sergushichev,A. (2021)
848 Fast gene set enrichment analysis. *bioRxiv*, 10.1101/060012.

849 64. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A.,
850 Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S., *et al.* (2005) Gene set enrichment

851 analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
852 *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15545–15550.

853 65. Fouracre,J.P., He,J., Chen,V.J., Sidoli,S. and Poethig,R.S. (2021) VAL genes regulate
854 vegetative phase change via miR156-dependent and independent mechanisms. *PLoS*
855 *Genet.*, **17**, e1009626.

856 66. Yoo,S.-D., Cho,Y.-H. and Sheen,J. (2007) Arabidopsis mesophyll protoplasts: a versatile
857 cell system for transient gene expression analysis. *Nat. Protoc.*, **2**, 1565–1572.

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

861 68. Yan,X., Yuan,X., Lv,J., Zhang,B., Huang,Y., Li,Q., Ma,J., Li,Y., Wang,X., Li,Y., *et al.* (2024)
862 Molecular basis of SAP05-mediated ubiquitin-independent proteasomal degradation of
863 transcription factors. *Nat Commun*, **15**, 1170.

864 69. Yamasaki,K., Kigawa,T., Inoue,M., Yamasaki,T., Yabuki,T., Aoki,M., Seki,E., Matsuda,T.,
865 Tomo,Y., Terada,T., *et al.* (2006) An Arabidopsis SBP-domain fragment with a disrupted
866 C-terminal zinc-binding site retains its tertiary structure. *FEBS Lett*, **580**, 2109–2116.

867 70. Xie,Z., Sokolov,I., Osmala,M., Yue,X., Bower,G., Pett,J.P., Chen,Y., Wang,K., Cavga,A.D.,
868 Popov,A., *et al.* (2025) DNA-guided transcription factor interactions extend human gene
869 regulatory code. *Nature*, **641**, 1329–1338.

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

873 72. Rodríguez-Martínez,J.A., Reinke,A.W., Bhimsaria,D., Keating,A.E. and Ansari,A.Z. (2017)
874 Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *Elife*, **6**.

875 73. Yamasaki,K., Kigawa,T., Inoue,M., Tateno,M., Yamasaki,T., Yabuki,T., Aoki,M., Seki,E.,
876 Matsuda,T., Nunokawa,E., *et al.* (2004) A novel zinc-binding motif revealed by solution
877 structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. *J*
878 *Mol Biol*, **337**, 49–63.

879 74. Liang,X., Nazarenus,T.J. and Stone,J.M. (2008) Identification of a consensus DNA-binding
880 site for the Arabidopsis thaliana SBP domain transcription factor, AtSPL14, and binding
881 kinetics by surface plasmon resonance. *Biochemistry*, **47**, 3645–3653.

882 75. Garcia-Molina,A., Xing,S. and Huijser,P. (2014) Functional characterisation of Arabidopsis
883 SPL7 conserved protein domains suggests novel regulatory mechanisms in the Cu
884 deficiency response. *BMC Plant Biol.*, **14**, 231.

885 76. Li,R.-J., Li,L.-M., Liu,X.-L., Kim,J.-C., Jenks,M.A. and Lü,S. (2019) Diurnal Regulation of
886 Plant Epidermal Wax Synthesis through Antagonistic Roles of the Transcription Factors
887 SPL9 and DEWAX. *Plant Cell*, **31**, 2711–2733.

888 77. Wang,H., Studer,A.J., Zhao,Q., Meeley,R. and Doebley,J.F. (2015) Evidence That the
889 Origin of Naked Kernels During Maize Domestication Was Caused by a Single Amino

890 Acid Substitution in tga1. *Genetics*, **200**, 965–974.

891 78. Lin,P.-Y., Huang,S.-C., Chen,K.-L., Huang,Y.-C., Liao,C.-Y., Lin,G.-J., Lee,H. and Chen,P.-
892 Y. (2025) Analysing protein complexes in plant science: insights and limitation with
893 AlphaFold 3. *Botanical Studies*, **66**, 1–12.

894 79. Flagel,L.E. and Wendel,J.F. (2009) Gene duplication and evolutionary novelty in plants.
895 *New Phytol.*, **183**, 557–564.

896 80. Magadum,S., Banerjee,U., Murugan,P., Gangapur,D. and Ravikesavan,R. (2013) Gene
897 duplication as a major force in evolution. *J. Genet.*, **92**, 155–161.

898 81. Rastogi,S. and Liberles,D.A. (2005) Subfunctionalization of duplicated genes as a transition
899 state to neofunctionalization. *BMC Evol. Biol.*, **5**, 28.

900 82. Ohno,S. (2013) Evolution by Gene Duplication Springer Science & Business Media.

901 83. Shiu,S.-H., Shih,M.-C. and Li,W.-H. (2005) Transcription factor families have much higher
902 expansion rates in plants than in animals. *Plant Physiol.*, **139**, 18–26.

903 84. Gera,T., Jonas,F., More,R. and Barkai,N. (2022) Evolution of binding preferences among
904 whole-genome duplicated transcription factors. *Elife*, **11**.

905 85. Strader,L., Weijers,D. and Wagner,D. (2022) Plant transcription factors - being in the right
906 place with the right company. *Curr. Opin. Plant Biol.*, **65**, 102136.

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