The ALOG domain defines a new family of plant-specific Transcription Factors acting during Arabidopsis flower development

Philippe Rieu, Veronica Maria Beretta, Francesca Caselli, Emmanuel Thévénon, Jérémy Lucas, Mahmoud Rizk, Emanuela Franchini, Elisabetta Caporali, Chiara Paleni, Max Nanao, et al.

PNAS RESEARCH ARTICLE | PLANT BIOLOGY

The ALOG domain defines a family of plant-specific transcription factors acting during Arabidopsis flower development

Philippe Rieu[a,1,2], Veronica Maria Beretta[b,1], Francesca Caselli[b], Emmanuel Thévénon[a], Jérémy Lucas[a], Mahmoud Rizk[c], Emanuela Franchini[b], Elisabetta Caporali[b], Chiara Paleni[b], Max H. Nanao[c], Martin M. Kater[b], Renaud Dumas[a], Chloe Zubieta[a], François Parcy[a,3], and Veronica Gregis[b,3]

Edited by Roberto Solano, Centro Nacional de Biotecnologia, Madrid, Spain; received June 29, 2023; accepted December 5, 2023 by Editorial Board Member Joseph J. Kieber

The ALOG (Arabidopsis LIGHT-DEPENDENT SHORT HYPOCOTYLS 1 (LSH1) and Oryza G1) proteins are conserved plant-specific Transcription Factors (TFs). They play critical roles in the development of various plant organs (meristems, inflorescences, floral organs, and nodules) from bryophytes to higher flowering plants. Despite the fact that the first members of this family were originally discovered in Arabidopsis, their role in this model plant has remained poorly characterized. Moreover, how these transcriptional regulators work at the molecular level is unknown. Here, we study the redundant function of the ALOG proteins LSH1,3,4 from Arabidopsis. We uncover their role in the repression of bract development and position them within a gene regulatory network controlling this process and involving the floral regulators LEAFY, BLADE-ON-PETIOLE, and PUCHI. Next, using in vitro genome-wide studies, we identified the conserved DNA motif bound by ALOG proteins from evolutionarily distant species (the liverwort *Marchantia polymorpha* and the flowering plants Arabidopsis, tomato, and rice). Resolution of the crystallographic structure of the ALOG DNA-binding domain in complex with DNA revealed the domain is a four-helix bundle with a disordered NLS and a zinc ribbon insertion between helices 2 and 3. The majority of DNA interactions are mediated by specific contacts made by the third alpha helix and the NLS. Taken together, this work provides the biochemical and structural basis for DNA-binding specificity of an evolutionarily conserved TF family and reveals its role as a key player in Arabidopsis flower development.

ALOG domain | plant transcription factors | flower development

Significance

Transcription Factors (TFs) are key proteins that bind specific genomic regions and regulate the expression of associated genes. Not all organisms possess the same set of TFs and some, like the ALOG [Arabidopsis LIGHT-DEPENDENT SHORT HYPOCOTYLS 1 (LSH1) and Oryza G1] TFs, are specific to plants. ALOG TFs have been shown to play important roles from bryophytes to flowering plants, but it was not known what DNA motif they recognize and how they bind DNA. Here, we identify this DNA motif and elucidate the structural basis of its binding by ALOG TF. We also reveal that several ALOG genes from Arabidopsis share a redundant function within the genetic network underlying correct floral meristem development.

The control of gene expression by Transcription Factors (TFs) is of key importance for all living organisms. Some TF types are widely shared among diverse organisms whereas others have emerged in specific groups. Plants, in particular, possess a diversity of TFs that are absent from other kingdoms (1). Many of these plant-specific TF families were born around the time of plant emergence from the water (in charophytes algae or bryophytes) and have later expanded and been co-opted in various processes along plant evolution (2). Most of these TFs have been characterized at the biochemical and structural levels. However, a few important regulators of plant physiology or development are suspected to act as TFs, but the characterization of their DNA-binding specificity or the structural basis for DNA recognition is still missing. This is the case for the *ALOG* [*Arabidopsis LIGHT-DEPENDENT SHORT HYPOCOTYLS 1* (*LSH1*) *and Oryza G1*] gene family. These genes, first described in *Arabidopsis thaliana* (3), are present in land plants as well as in some algae (4, 5), usually as gene families. For example, the Arabidopsis genome contains 10 *ALOG* genes (3) (called *LSH* in Arabidopsis).

ALOG genes have been shown to play crucial developmental roles in a wide variety of species. In the liverwort *Marchantia polymorpha*, *LATERAL ORGAN SUPPRESSOR 1* (*LOS1*) and *LOS2* are implicated in meristem maintenance and lateral organ development (5, 6). In tomato, the *ALOG* gene *TERMINATING FLOWER* (*TMF*) controls flowering by preventing the precocious expression of *ANANTHA* [*AN*; the tomato ortholog of the Arabidopsis *UNUSUAL FLORAL ORGANS* (*UFO*) gene] (7). In rice, the *ALOG* genes *LONG STERILE LEMMA1* (*G1*) (8), *TRIANGULAR HULL 1* (*TH1*) (9), *TAWAWA1* (*TAW1*) (10), and *G1-LIKE 1* and *G1-LIKE 2* (*G1L1* and *G1L2*) (11) have been shown to play diverse roles, such as controlling the development of the panicle and several floral organs. In pea, the *ALOG* gene *SYMMETRIC PETALS 1* is a regulator of floral organ internal asymmetry (12) and in *Medicago truncatula* *LSH1/LSH2* control nodule identity (13). In *Torenia fournieri*, *ALOG3* regulates the formation of the petal corolla neck (14). In Arabidopsis, LSH8 is involved in ABA signaling (15), LSH9 genetically interacts with the clock protein EARLY FLOWERING 3 (16) and LSH10 acts in a co-repressor complex (17). Arabidopsis mutants have also been isolated for *LSH1*, *LSH3*,

Author contributions: P.R., M.M.K., F.P., and V.G. designed research; P.R., V.M.B., F.C., E.T., J.L., E.F., E.C., R.D., C.Z., and V.G. performed research; P.R., V.M.B., F.C., M.R., C.P., M.H.N., C.Z., F.P., and V.G. analyzed data; and P.R., V.M.B., M.M.K., C.Z., F.P., and V.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. R.S. is a guest editor invited by the Editorial Board.

[1]P.R. and V.M.B. contributed equally to this work.

[2]Present address: Structural Plant Biology Laboratory, Department of Botany and Plant Biology, University of Geneva, Geneva 1211, Switzerland.

[3]To whom correspondence may be addressed. Email: francois.parcy@cnrs.fr or veronica.gregis@unimi.it.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2310464121/-/DCSupplemental.

Published February 27, 2024.

and *LSH4,* but none display any obvious phenotype (18, 19). However, constitutive expression of *LSH4* (and to a lesser extent *LSH3*) results in the inhibition of leaf growth and the production of extra and chimeric floral organs or shoots within flowers (19). *LSH3* and *LSH4* are expressed in the boundary regions of shoot organs under the transcriptional control of the CUP-SHAPED COTYLEDON1&2 TFs (19, 20). Overall, these effects suggest that Arabidopsis LSH3 and LSH4 might be involved in the control of organ formation in boundary regions of floral meristems.

Despite this increasing body of evidence that ALOG proteins play important roles in various plant species, little is known about the molecular and structural properties of these TFs. Several ALOG proteins were shown to directly interact with promoter regions and act as transcriptional repressors (17, 21, 22). A well-studied example is the repression of the tomato *AN* gene by the ALOG protein TMF, involving phase separation on the *AN* promoter (21, 23). At the molecular level, ALOG proteins share a conserved domain (named the ALOG domain) (4), flanked by non-conserved disordered regions of variable lengths. This ALOG domain likely originates from the DNA-binding domain (DBD) of bacterial recombinases found in mobile elements (4), into which a putative zinc ribbon was inserted. However, the DNA-binding mode and specificity determinants were unknown. Here, we provide a functional characterization of ALOG proteins from Arabidopsis, determine their DNA-binding specificity, compare this with ALOG proteins from early diverging plants, and demonstrate the importance of both the unstructured NLS and core alpha helix in DNA sequence recognition and binding.

Results

***LSH1*, *LSH3,* and *LSH4* Functional Characterization in *A. thaliana* Reproductive Meristems.** The role of *ALOG* genes in flower and inflorescence development has been documented across various species, such as rice, tomato, Torenia, and several legumes (7, 10, 12–14). However, their functions in the reproductive structures of Arabidopsis are poorly characterized. Among the 10 *LSH* genes, only *LSH1*, *LSH3,* and *LSH4* are expressed in the inflorescence meristems (IMs), flower meristems (FMs) and Stage-3 flower tissues (*SI Appendix*, Fig. S1). In order to further characterize their spatial expression profiles, we performed in situ hybridization. Consistent with *LSH3* and *LSH4* mRNA and LSH4-GFP published expression data (19, 24), all three genes were expressed in the boundary region between the IM and the FM. *LSH1* and *LSH3* were also expressed in the cryptic bract region while *LSH1* expression further extends in the IM (Fig. 1 *A*–*F*). The overlapping expression patterns of *LSH1*, *LSH3,* and *LSH4* genes suggest possible functional redundancy among them. This hypothesis is consistent with the lack of any discernible phenotypic effects in single *lsh1*, *lsh3*, or *lsh4* mutants (18–20).

To unravel the role of *LSH1*, *LSH3,* and *LSH4* in Arabidopsis, we generated mutant alleles for these genes using the CRISPR-Cas9 gene editing system. For all three genes, we obtained mutations with a premature stop codon leading to truncated proteins lacking the ALOG domain (*SI Appendix*, Fig. S2). None of the single (*lsh1*, *lsh3,* and *lsh4*) nor double mutants (*lsh1 lsh3*, *lsh1 lsh4*, and *lsh3 lsh4*) display any visible aberrant phenotype. We thus analyzed the *lsh1 lsh3 lsh4* triple mutant and observed that flowers were subtended by a well-developed bract-like organ emerging from the pedicel (Fig. 1 *M*–*O*). Such organs do not normally develop in wild-type (WT) Arabidopsis plants. The triple mutant flowers show no other strong aberrations: The floral organs had normal numbers, shapes, and dimensions and the flowers were fully fertile. Scanning electron microscopy (SEM) analysis of *lsh1 lsh3 lsh4* bracts showed that these organs display sepal-like cell types and occasionally develop structures resembling stigmatic papillae (Fig. 1 *O*–*R*).

Positioning LSH Proteins within the Gene Regulatory Network Controlling Bract Repression. In order to understand the role of *LSH* genes, we aimed at positioning them in the gene regulatory network controlling bract repression. In Arabidopsis, the genes *LEAFY* (*LFY*), *UFO*, *BLADE-ON-PETIOLE1* & *2* (*BOP*), *PUCHI*, *JAGGED* (*JAG*), and *JAGGED-like* are known to contribute to this process (25–29).

Since the master regulator LFY is expressed in the earliest stages of flower development (30) and contributes to bract repression (31), we wondered whether LFY could act upstream of *LSH* genes. Examination of published LFY ChIP (chromatin immunoprecipitation)-seq data revealed intense LFY binding to *LSH1* and *LSH3* promoter regions (*SI Appendix*, Fig. S3) (32, 33). We thus investigated *LSH* gene expression patterns in the *lfy-2* mutant (Fig. 1 *G*–*L*) (34) and found alterations in *LSH1*, *LSH3,* and *LSH4* expression. Specifically, *LSH1* expression became almost undetectable in the *lfy-2* mutant (compare Fig. 1 *A* and *B* with Fig. 1 *G* and *H*), whereas *LSH3* (compare Fig. 1 *C* and *D* with Fig. 1 *I* and *J*) and *LSH4* (compare Fig. 1 *E* and *F* with Fig. 1 *K* and *L*) expressions became more diffuse, ectopically present in regions such as the FM and the IM, where they are normally absent. Therefore, LFY activity is required for the localized expression of *LSH* genes in the meristem boundary and cryptic bract region (including for *LSH4* that is not directly bound by LFY in available ChIP-seq data).

To identify proteins that could interact directly with LSHs, we systematically analyzed interactions between LSHs and the different proteins acting in bract repression using yeast-two-hybrid (Y2H) assays (*SI Appendix*, Table S1). Among all tested interactions, we found that LSH1, LSH3, and LSH4 interacted with the BOP1 and BOP2 proteins (*SI Appendix*, Fig. S4). Similar interactions were previously observed in tomato, Torenia, and pea (12, 35, 36). Thus, our findings further support the conservation of the BOP-ALOG interaction in Arabidopsis.

Finally, we wished to identify direct target genes downstream of LSHs. We first used RNA-seq on young inflorescence tissues to establish the list of genes with altered expression in the *lsh* triple mutant (Dataset S1*A* and *SI Appendix*, Fig. S5). Next we performed amplified DNA Affinity Purification sequencing (ampDAP-seq) (37) to identify genomic regions potentially bound by LSH1 and LSH3 proteins. We found a strong overlap between regions bound by these two TFs, consistent with their redundant function in planta (Dataset S1 *B* and *C* and *SI Appendix*, Fig. S5). Since the list of genes both bound and regulated did not yield any gene known to repress bracts (Dataset S1*D*), we individually examined the ampDAP-seq data for the genomic regions of known bract repressors (Fig. 2*A* and *SI Appendix*, Fig. S6). We identified two LSH peaks in the *PUCHI* promoter (regions I and III). To determine whether LSH could bind these regions in vivo, ChIP was performed using a *pLSH4:LSH4-GFP* transgenic line (24). We found that LSH4-GFP specifically bound regions I and III, while region II, lacking the LSH binding site, did not exhibit any binding (Fig. 2*B*). This result thus corroborates ampDAP-sep data. We then assayed the expression of *PUCHI* by in situ hybridization in the *lsh1 lsh3 lsh4* triple mutant. In contrast to its specific expression on the adaxial side of the WT flower meristem (Fig. 2*C*) (27), we observed that, in the *lsh1 lsh3 lsh4* mutant, *PUCHI* expression was localized at the abaxial side of floral primordia; this was true both for stage 1 primordia when bract just starts emerging and for stage 2 primordia when the bract is clearly visible (stages according to ref. 38; Fig. 2 *D* and *E*). This result confirms that

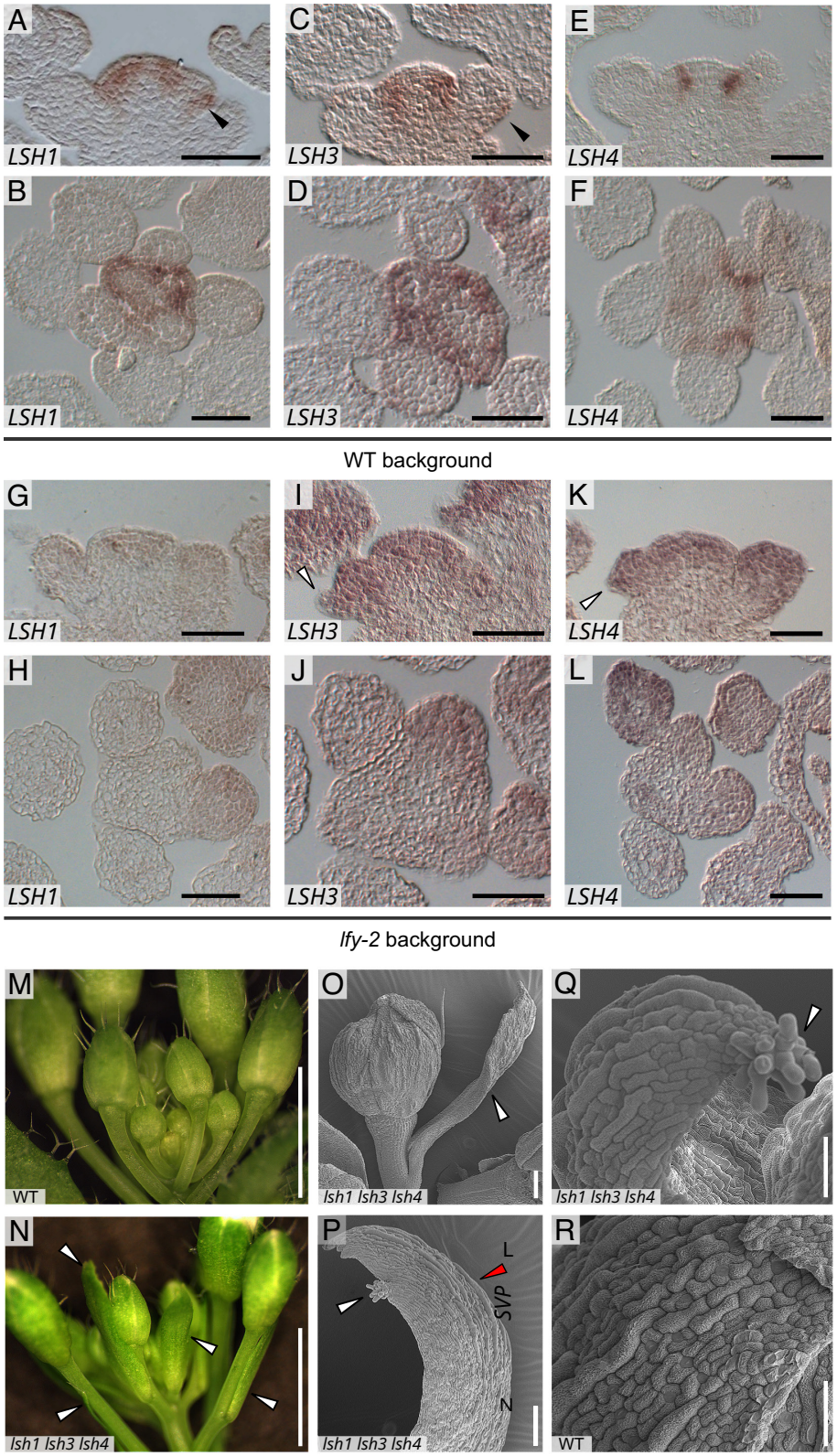

Fig. 1. Expression pattern of *LSHs* (*A*–*L*) and analysis of the *lsh1 lsh3 lsh4* (*M*–*R*) triple mutant. Expression profile of *LSH1* (*A*, *B*, *G*, and *H*), *LSH3* (*C*, *D*, *I*, and *J*), and *LSH4* (*E*, *F*, *K*, and *L*) in reproductive tissues analyzed by in situ hybridization in WT (*A*–*F*) and *lfy-2* (*G*–*L*) backgrounds. (*A*, *C*, *E*, *G*, *I*, and *K*) are longitudinal and (*B*, *D*, *F*, *H*, *J*, and *L*) are transversal sections. (Scale bar, 50 μm.) Black arrowheads in (*A* and *C*) indicate signal in the cryptic bract region, white arrowheads in (*I* and *K*) indicate bract primordia. See *SI Appendix*, Fig. S13 for negative controls. (*M* and *N*) Stereo microscope pictures of WT (*M*) and *lsh1 lsh3 lsh4* triple mutant (*N*) inflorescences. The white arrowheads in (*N*) indicate bracts. (Scale bar, 3 mm.) (*O*–*R*) SEM pictures of WT and *lsh1 lsh3 lsh4* plants. (*O*), floral bud, the white arrowhead indicates the bract. (Scale bar, 100 μm.) *lsh1 lsh3 lsh4* triple mutant bract (*P* and *Q*). The white arrowhead indicates the stigmatic papillae. The red arrowhead indicates the sepal-like cells. (Scale bars: *P* = 100 μm; *Q* = 50 μm.) WT sepal (*R*, Scale bar, 50 μm).

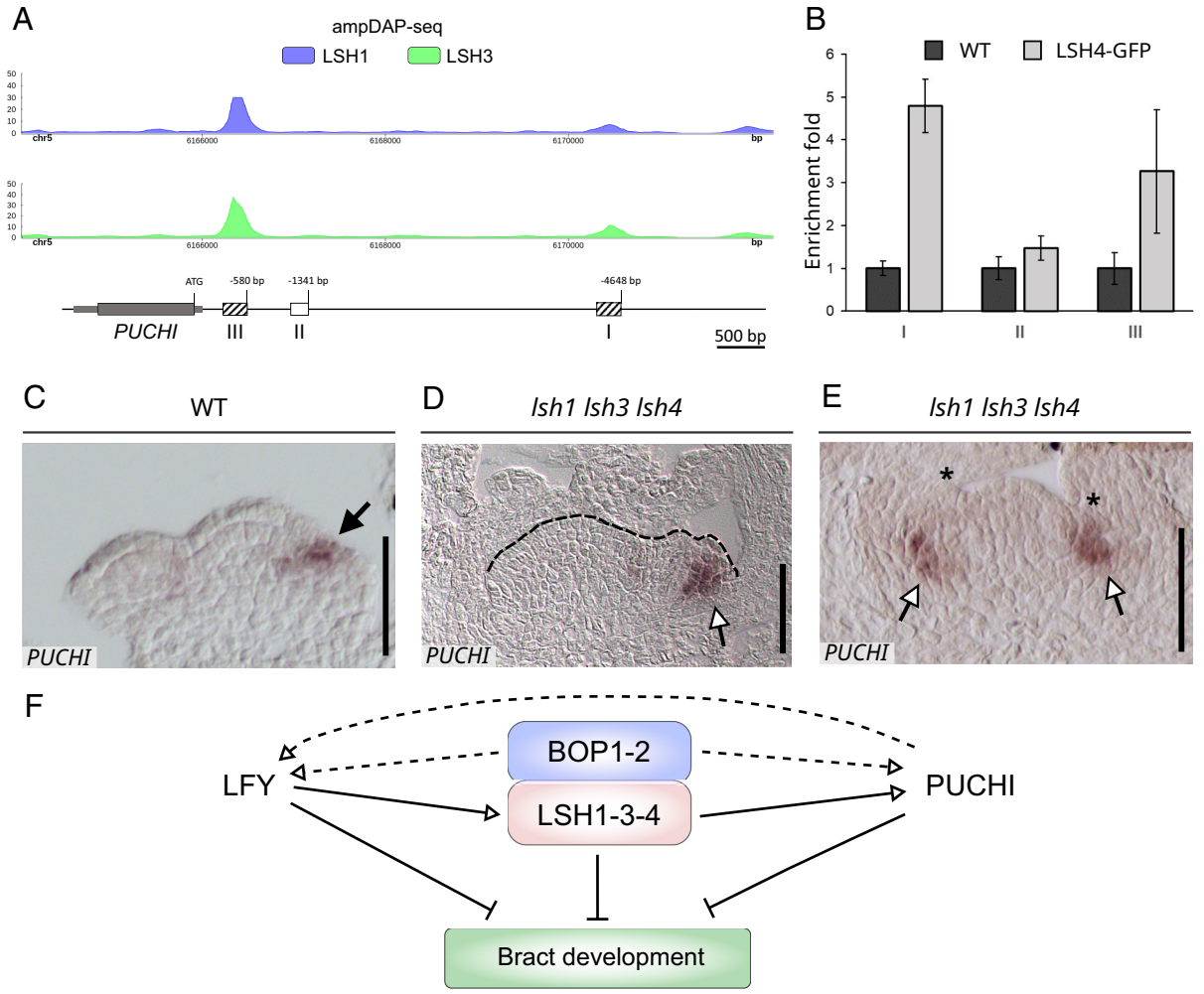

Fig. 2. LSHs regulate PUCHI expression. (*A*) Genome browser view showing ampDAP-seq binding of LSH proteins to the *PUCHI* promoter. Regions I to III analyzed by ChIP in (*B*) are presented below. *Y* axis indicates count per millions of reads mapped. (*B*) ChIP assays analyzed by real-time PCR show that LSH4-GFP bound region I and III of PUCHI promoter but not region II. Error bars represent the propagated error value using three technical replicates. ChIP results of one representative experiment out of two independent biological replicates is shown. (*C*–*E*) Analysis of *PUCHI* expression profile by in situ hybridization. WT or mutant backgrounds are indicated below pictures. (Scale bar, 50 μm.) Asterisks indicate the bracts that subtend the primordia and the arrow points to the *PUCHI* expression domain, adaxial to the flower primordia in WT and abaxial in the *lsh1 lsh3 lsh4* mutant. (*F*) Model showing the interactions involved in bract repression. Arrows indicate regulations (with solid arrows when direct) and the interaction between BOP and LSH proteins is shown.

LSH acts upstream of *PUCHI* and suggests that the altered *PUCHI* expression could contribute to the triple *lsh1 lsh3 lsh4* mutant bract phenotype. We also checked the expression of *LFY* and *BOP1/2* genes but none of them showed a strongly altered expression pattern in the *lsh1 lsh3 lsh4* triple mutant (*SI Appendix*, Fig. S7). Overall, our findings allow proposing a gene regulatory network involved in the repression of bract development (Fig. 2*F*).

Determination of the ALOG DNA-Binding Specificity. Next, we aimed at determining ALOG DNA-binding specificity. The search for overrepresented DNA motifs in LSH1/3-bound regions from the ampDAP-seq experiment identified the same 7-bp YACTGTW (Y = T/C, W = A/T) motif for both proteins (Fig. 3*A* and *SI Appendix*, Fig. S8*A*). This motif displays several high-information positions, strongly suggesting specific contacts between LSH residues and DNA bases at these locations. The position weight matrix (PWM) corresponding to this motif reliably predicted LSH1/3 DNA binding (Fig. 3*B* and *SI Appendix*, Fig. S8*A*). Next, we used electrophoretic mobility shift assay (EMSA) to validate LSH binding to the YACTGTW motif. We found that both the in vitro-produced full-length proteins and the isolated ALOG domains specifically interact with this motif (*SI Appendix*, Fig. S8*B*), showing that the ALOG domain is the DBD and fully confers the binding specificity. The ALOG domain was thus used in the rest of this work. EMSA using a DNA probe of optimal affinity according to LSH3 PWM showed a single shifted band with LSH1 or LSH3 DBDs (Fig. 3*C* and *SI Appendix*, Fig. S8*C*) with apparent Kds for LSH1 and LSH3 DBDs estimated below 50 nM, in the range of other affinities found for TF/DNA (39). DNA bases from this motif were systematically mutated and, overall, most mutations at the high-information positions of the motif strongly reduced the binding of LSH1 and LSH3 DBDs (Fig. 3*D*). These experiments validated the YACTGTW motif as the LSH binding site.

The LSH motif showed little symmetry, suggesting it is bound by a single DBD monomer (20, 24). Co-immunoprecipitation (*SI Appendix*, Fig. S9*A*), Y2H (*SI Appendix*, Fig. S9*B*), and EMSA (*SI Appendix*, Fig. S9*C*) experiments consistently confirmed that Arabidopsis LSH proteins do not dimerize as opposed to observations made on proteins from other species (21, 35).

ALOG DBD Structure Defines a specific Class of Plant TF. Next, we used a crystallographic approach to obtain structural information on how ALOG DBD binds DNA. Deletion experiments first

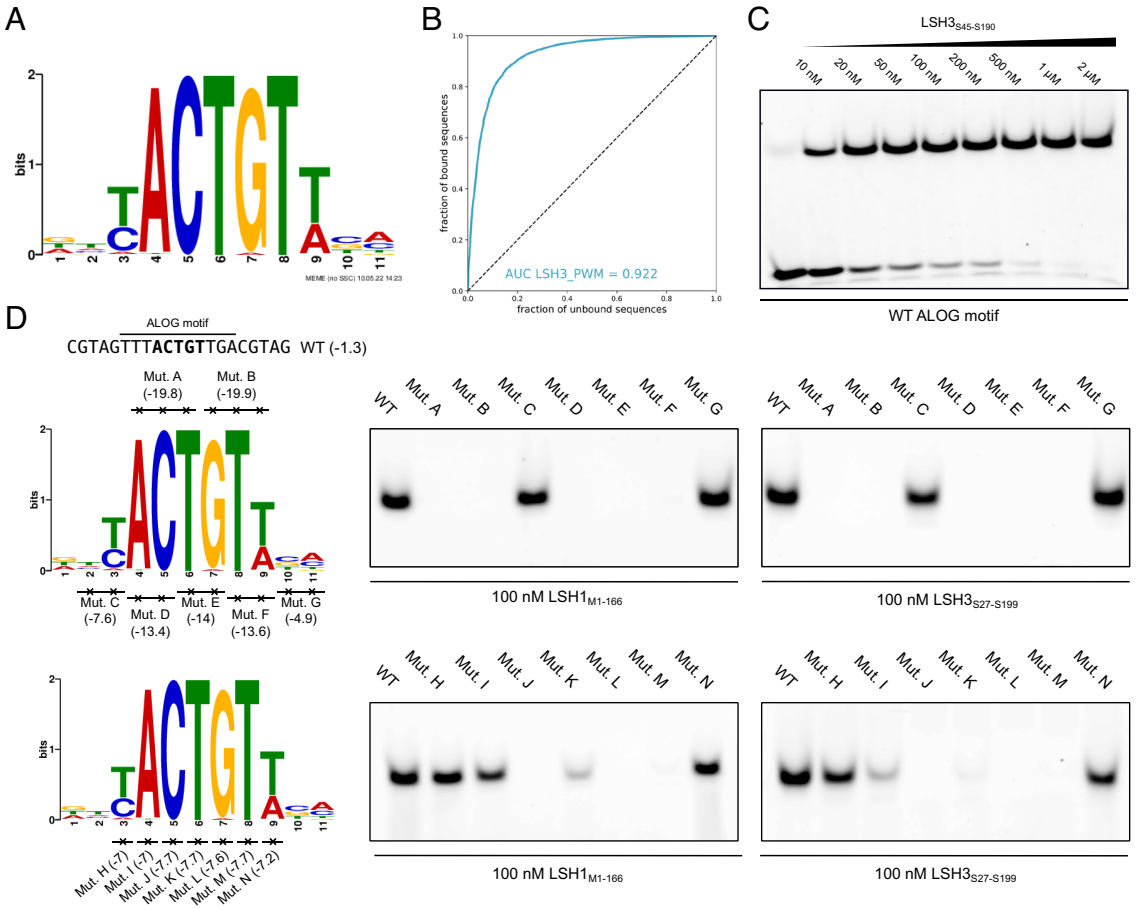

Fig. 3. Determination of ALOG binding specificity by ampDAP-seq. (*A*) Logo obtained for LSH3 in ampDAP-seq, using the 600 peaks with the strongest signal. (*B*) Receiver operating characteristic (ROC) curve for LSH3 using all peaks except those used to build the logo. The value of the Area Under the Curve (AUC) is indicated. (*C*) EMSA with ALOG highest-score sequence DNA probe (WT ALOG motif) and $LSH3_{S45\text{-}S190}$. Based on the analysis of three independent EMSAs, we found an apparent Kd of 28 nM for LSH3-DBD/DNA. (*D*) EMSA with $LSH1_{M1\text{-}L166}$, $LSH3_{S27\text{-}S199}$ and indicated DNA probes. The WT ALOG probe (*Top Left*) was mutated at positions indicated on the LSH3 logos (*Bottom Left*). Scores between brackets were obtained by scanning each DNA probe sequence with the LSH3 PWM (the best binding sites correspond to the less negative score values). EMSA with described DNA probes (*Right*). Uncropped gels are provided in *SI Appendix*, Fig. S14.

allowed us to define the minimal LSH3 DBD and, unexpectedly, showed that the conserved nuclear localization signal (NLS; Fig. 4*A*) was required for DNA binding (*SI Appendix*, Fig. S10 *A* and *B*). Crystals obtained with LSH3 minimal DBD (LSH3S45-S190) bound to the YACTGTW motif (TAGTT**TACTGTT**GACGT DNA molecule) yielded a structure of the complex at 2.1 Å resolution (Table 1). We found that the ALOG domain from LSH3 is made of four helices and bears a high degree of structural similarity (conservation of the core arrangement of alpha helices) with the DBDs from LoxP recombinase (PDB 1NZB), tyrosine recombinase (PDB 5HXY) and the apoptosis regulator protein BCL-2 (PDB 6YLD) (40), as previously proposed (4). Furthermore, this crystallographic structure revealed that the DNA base-contacting residues originate from helix 3 but also from the C-terminal extension that includes the NLS (Fig. 4 *B*–*E*). Finally, the LSH3-DBD structure elucidated the structural role of the 24-amino acid loop inserted between helices 2 and 3: this loop provides the three cysteines and the histidine residues (103-HxxxC...119-CxC) that chelate a Zn^{2+} ion, forming a non-canonical zinc ribbon-like arrangement similar to a Gag knuckle (Fig. 4*G*) (41). Interestingly, except for one residue at its very base, this region does not engage directly in DNA interaction as previously hypothesized (4).

We then mutated the residues from helix 3 (D131S and R136S) that establish direct contact with DNA bases (Fig. 4 *C* and *D*) and observed a significant reduction in DNA-binding affinity (Fig. 4*F*). Mutations of the residues from helices 1 and 4 that contact the phosphate backbone but are not involved in direct base read-out also somewhat reduced DNA binding (*SI Appendix*, Fig. S10 *C*–*E*). Mutating the Zn^{2+}-coordinating residues (Fig. 4*H*) or deleting LSH1 or TMF zinc ribbons to mimic recombinases also abolished binding to the YACTGTW motif (*SI Appendix*, Fig. S10*F*), confirming the importance of the Zn ribbon as a structural component of the protein (21). Based on structural comparisons to all other plant TF-DNA complexes, we propose that the ALOG domain defines a specific family of plant TFs within the Zn^{2+} coordinating DBD superclass and the HC3 class of Plant-TFClass (1, 42).

Once established the ALOG-DBD DNA structure, we used Alphafold2 ColabFold (AFC) to analyze the interaction between ALOG and BOP proteins. From the diverse models generated, involving distinct BOP domains interacting with ALOG, we have chosen one that demonstrates robustness when considering pairs of proteins from different species. We tested this model using Y2H experiments with Arabidopsis proteins BOP2 and LSH3 (*SI Appendix*, Fig. S11). Mutations in residues from the ankyrin-repeat domain of BOP or ALOG DBD (opposite to DNA-contacting surface) strongly reduced their interaction, corroborating the AFC prediction. Such data will be useful to understand how BOP proteins affect ALOG activity.

The Properties of ALOG TFs Are Conserved during Plant Evolution. Finally, we examined the properties of ALOG proteins outside of Arabidopsis. The sequence conservation of ALOG

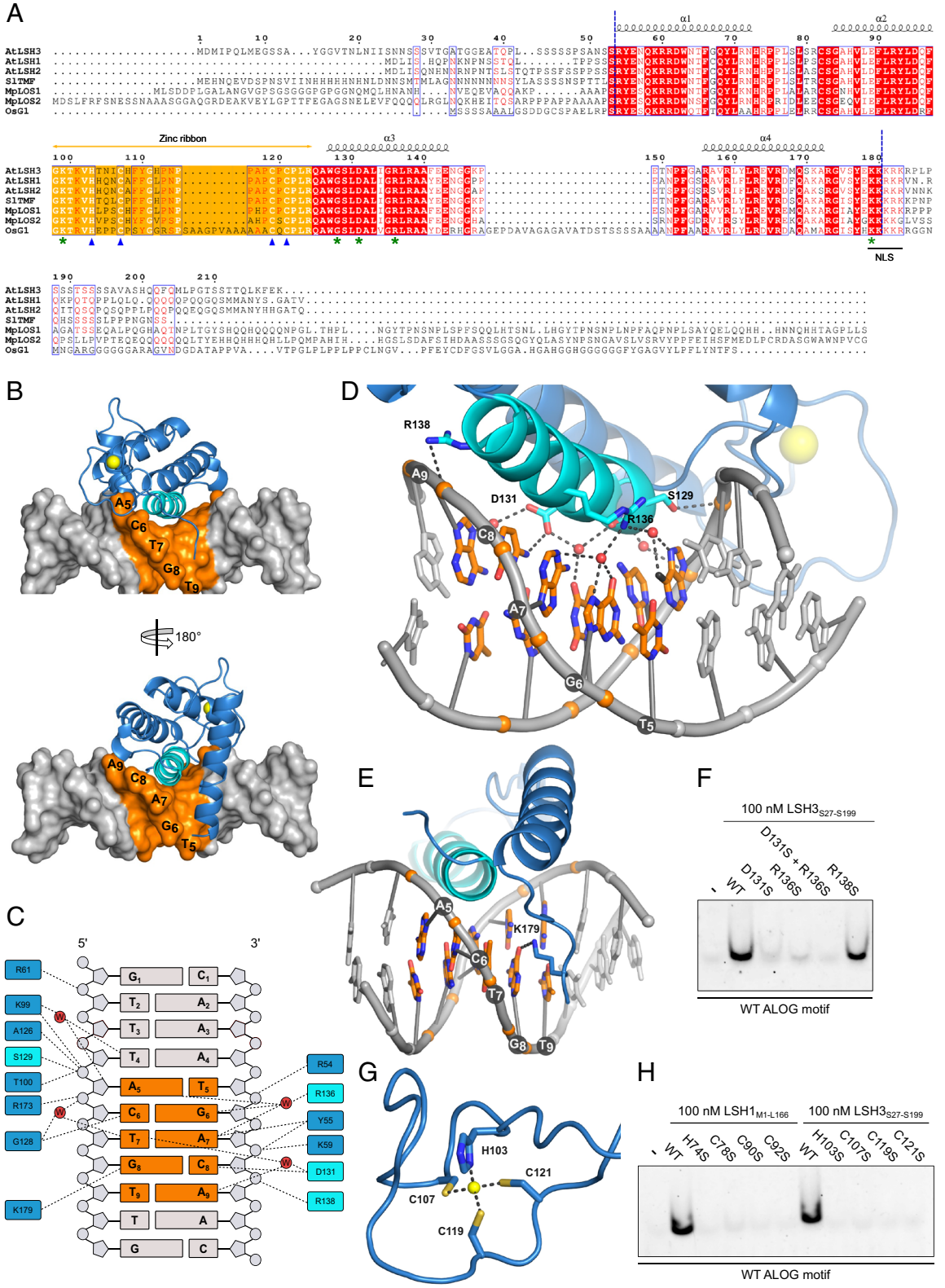

Fig. 4. Structure of LSH3 DBD in complex with DNA. (*A*) Alignment of studied ALOG proteins. Blue dotted lines indicate the limits of the ALOG domain. The zinc ribbon is highlighted in orange, with the key histidine and cysteine residues indicated by blue triangles. Green stars denote residues in direct contact with DNA bases. NLS = Nuclear Localization Signal. Numbers are relative to AtLSH3. At = *Arabidopsis thaliana*, Sl = *Solanum lycopersicum*, Mp = *Marchantia polymorpha*, Os = *Oryza Sativa*. (*B*) ALOG-DBD/DNA complex. Throughout this figure, LSH3 DBD is shown in blue except helix 3 colored in cyan. The DNA duplex is depicted in gray with bases with the highest information content in orange. The Zn^{2+} ion is shown in yellow. A 180° rotation along the *y*-axis was applied to obtain the bottom picture. (*C*) Protein-DNA interactions. (*D*) Ribbon diagram of LSH3 DBD bound to its cognate DNA, with a focus on helix 3. Interactions are indicated by black dashed lines. For clarity, helix 1 was removed. (*E*) Close-up view of LSH3 C-terminal tail in contact with DNA. (*F*) EMSA with ALOG highest-score sequence DNA probe (WT ALOG motif) and indicated $LSH3_{S27\text{-}S190}$ versions. (*G*) Close-up view of the zinc ribbon. Side chains of the residues coordinating the zinc ion are represented. (*H*) EMSA with WT ALOG motif DNA probe and indicated $_{LSH3S27\text{-}S190}$ and $LSH1_{M1\text{-}L166}$ versions. Uncropped gels are provided in *SI Appendix*, Fig. S14.

Table 1. Data collection and refinement statistics

	ALOG-DNA
Data collection	
Space group	$P4_322$
Cell dimensions	
a, *b*, *c* (Å)	79.80, 79.80, 127.01
α, β, γ (°)	90, 90, 90
Resolution (Å)	67.-2.14 (2.19-2.14)*
R_{sym} or R_{merge} (%)	4.5 (223)
$I/\sigma I$	21.7 (0.9)
Completeness (%)	99.8 (99.1)
Redundancy	7.0 (6.9)
CC(1/2)	100 (32.5)
Refinement	
Resolution (Å)	50.-2.14 (2.19-2.14)
No. reflections	43,003
R_{work}/R_{free}	21.4/25.2 (40.1/42.85)
No. atoms	1,809
Protein	1,402
Water	51
DNA	345
Zn^{2+}	1
B-factors	
Protein	67
Water	67
DNA	92
Zn^{2+}	59
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.54

*Refers to the highest resolution shell.

domains across land plant species (Fig. 4*A*) suggests that the DNA-binding specificity of these factors might be conserved. To test this hypothesis, we performed ampDAP-seq with ALOG proteins from Marchantia, tomato, and rice (including the rice G1 protein that bears an additional insertion absent from other members of this family; Fig. 4*A*). We found that all tested proteins have similar DNA-binding specificities, revealing a high degree of conservation from bryophytes to flowering plants (*SI Appendix*, Fig. S12*A*). Consistent with a functional role of the ALOG domain, several of the mutations identified genetically in rice or tomato fall within key residues of the ALOG domain, within helices 1 and 3 and the zinc ribbon (*SI Appendix*, Fig. S12*B*). As the DNA-binding specificity is highly conserved, it is likely that the structural features described for the LSH3-DNA complex are valid for all ALOG proteins. The diversity of ALOG functions in plants would have to be explained by variations in other features (such as changes in *cis*-elements in their target genes).

Discussion

Since the initial identification of the *LSH1* gene in Arabidopsis (3) and of the *G1* gene in rice (8), which defined the ALOG family, there has been a growing body of evidence that ALOG genes play important roles in many land plants such as Marchantia, Arabidopsis, tomato, rice, pea, or Medicago. However, fundamental questions have remained unanswered regarding the DNA sequence they recognize, their DNA-binding mechanism, and their general and conserved roles in diverse species. In this work, we resolved several of these key questions.

First, we performed a set of genetic experiments showing that the Arabidopsis *lsh1 lsh3 lsh4* triple mutant flowers display a well-developed and chimeric bract, an organ that is normally absent in WT flowers. This phenotype, absent from single and double mutants, shows that these factors act redundantly in Arabidopsis. Analyzing other genes involved in bract repression, we unraveled an intricate gene regulatory network downstream of the master floral regulator LFY which directly controls the proper expression pattern of *LSH* genes (Fig. 2*F*). As also reported in tomato, pea, and Torenia (12, 35, 36), we showed that Arabidopsis LSHs interact with BOP proteins and we identified key residues for this interaction. This LSH-BOP complex likely acts by controlling the expression of the bract repressor PUCHI (27). Indeed, *PUCHI* expression is no longer confined to the adaxial side of the boundary region in the *lsh1 lsh3 lsh4* triple mutant (as also observed to a lesser extent in the *bop1 bop2* double mutant) (27). The regulation of *PUCHI* by LSH appears direct since the interaction between LSH proteins and PUCHI promoter was observed in ampDAP-seq experiments (for LSH1 and LSH3) and in ChIP experiment (for LSH4), and several ALOG binding sites are conserved in the *PUCHI* promoter (*SI Appendix*, Fig. S6*B*). These observations suggest that LSH-BOP complexes are required for the boundary identity by confining the expression of *PUCHI* at the adaxial boundary of FM and thereby preventing bract development. Whether Arabidopsis LSHs recruit BOPs to the nucleus forming a transcriptional complex as reported in tomato, is still an open question (35). Since the *lsh1 lsh3 lsh4* triple mutant bract-like organs show chimeric features like sepaloid and carpelloid cells, LSHs may also be involved in repressing floral organ identity genes. Further genetic and molecular analyses will be needed to fully explain the *lsh* mutant phenotype and to determine whether parts of the regulatory network unraveled in Arabidopsis also exist in other flowering plants. Considering the similarity of the phenotypic defect induced by mutations in LFY and/or ALOG genes in rice, it is plausible that the regulatory connection identified here between LFY and the ALOG genes may exhibit conservation.

Next, we provided answers regarding how ALOG proteins perform their function as TFs. Using ampDAP-seq, we identified the DNA motif bound by ALOG proteins from Marchantia to flowering plants (Arabidopsis, tomato, and rice). This result confirmed ALOGs as sequence-specific DNA-binding proteins, with a bound motif conserved over a large evolutionary scale. Structural studies reveal the molecular and atomic determinants of DNA-binding specificity. The organization of the helical bundle and the approximate position of the DNA-contacting surface had been roughly anticipated based on the comparison with recombinases and with lambda and integron integrases. These proteins bind DNA using helices 1, 3, and 4, with a diversity of residues (positions and types) in direct contact with DNA. The LSH3-DNA structure established that base-protein contacts are made by ALOG residues located on helix 3 (not from helix 1 or 4), but also at the base of the zinc ribbon and, more surprisingly, a residue coming from the NLS. In contrast with previous predictions (4), the main role of the zinc ribbon is not to provide additional contacts with DNA but rather to help in stabilizing the DNA-contacting helices. The ALOG-DBD/DNA structure does not resemble any classified TF/DNA structure (1, 42), and thus defines a specific family in the TFClass reference classification.

While this work helps clarify both the physiological role of ALOG proteins and their mechanism of DNA-binding and

recognition, many ALOG properties remain to be investigated. For example, it is known that ALOG proteins are not functionally interchangeable: The ALOG proteins G1 and TAW1 from rice cannot complement a Marchantia *los1* mutant while LOS1 complements a rice *g1* mutant (5). Thus, despite having highly conserved DNA-binding preferences, other features contribute to ALOG proteins' function in vivo. A first explanation could be that each ALOG protein interacts with specific partners to fulfill certain functions. Another possibility is that each ALOG protein has specific biochemical properties. In fact, it was reported in tomato that ALOG proteins undergo phase separation (21, 23). This property depends on redox conditions that control the formation of intermolecular disulfide bonds, and also on intrinsic ALOG characteristics independent of the DBD, notably their disordered regions. Cysteine residues from the zinc ribbon were proposed to contribute to the formation of TMF intermolecular disulfide bonds (21). However, our structural data suggest that mutating TMF cysteines rather disrupts the zinc ribbon and abolishes DNA binding. Further work is thus required to understand the implication of each ALOG protein domain (notably non-conserved disordered regions) in their physiological function, including their ability to undergo phase separation.

Overall, our work provides essential information on how ALOG TFs interact with DNA. An important remaining challenge is the determination of larger transcriptional complexes ALOGs are able to form. This likely underlies the differences in ALOG physiological function and explains how a highly conserved DBD module can be functionally diversified via additional non-conserved and disordered domains.

Materials and Methods

Plant Material and Growth Conditions. All experiments were performed in *A. thaliana* accession Columbia-0 (Col-0) except the ChIP experiment (*pLSH4:LSH4-GFP* line in the Landsberg *erecta* background) (24). Plants were grown in a controlled environment at 20 to 22 °C either under long-day conditions (16 h light/8 h dark) or under short-day (8 h light/16 h dark) conditions for 4 wk after germination and then transferred to long-day conditions. When necessary, seeds of Arabidopsis were germinated on MS plates (Murashige & Skoog, supplemented with 1% sucrose and solidified with 0.7% Plant Agar) with the appropriate selection. The *lfy-2* mutant is described in ref. 34.

CRISPR-Cas9 Mutant Generation. For the generation of *lsh1*, *lsh3,* and *lsh4* single knock-out mutants, 20-bp specific protospacers (Dataset S2) were selected for each gene using CRISPR-P v2 database (43) and cloned into BbsI site of pENChimera entry vector under the Arabidopsis *U6-26* promoter and then combined into the pDe-CAS9 destination vector containing the Cas9 (44) by single site Gateway® LR reaction.

Scanning Electron Microscopy (SEM). SEM samples were prepared as previously described (45) by gold coating them using a sputter coater (SEMPREP2; Nanotech) followed by observation with a FESEM SIGMA SEM (Zeiss).

In Situ Hybridization. In situ hybridization analyses were performed as described in ref. 46. Gene fragments were amplified using the primer pairs listed in Dataset S2 to generate RNA antisense probes. Evaluation of the expression profile in the inflorescence and FMs, which was previously published for *BOP1*, *BOP2*, *PUCHI,* and *LFY* was used as a positive control (27, 47, 48). The antisense and sense RNA probes for *LSH1*, *LSH3,* and *LSH4* were transcribed from pGEM®-T Easy with T7/SP6 RNA polymerase (Promega) according to the manufacturer's instructions. WT reproductive meristems transversal sections hybridized with sense probes for *LSH1*, *LSH3,* and *LSH4* are shown in *SI Appendix*, Fig. S13.

Y2H. Yeast transformation was performed as previously described (49). Empty pGADT7 and pGBKT7 vectors were used as controls for the autoactivation. AH109 yeast cells transformed with the pGADT7 vectors were mated with Y187 yeast cells transformed with the pGBKT7 constructs. The protein–protein interaction assays were performed on selective yeast synthetic dropout medium lacking Leu, Trp, Ade, and His, and supplemented with indicated concentrations of 3-aminotriazole. Plates were grown for 5 d at 28 °C before taking pictures. Previously published REM35/REM35 and REM34/REM34 interactions were used as positive and negative controls, respectively (46).

Transcriptome Analysis. Tissue enriched in IM, FM, and flowers primordia until stage 8 was collected for both *lsh1 lsh3 lsh4* and WT plants. Four biological replicates were collected, each replicate consisting of seven inflorescences. RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). Sequencing libraries were prepared by Novogene with poly-A enrichment and sequenced with Illumina (paired end, 2 × 150 bp mode). Reads were trimmed with fastq-mcf v1.05 with options -l 50 -q 30 and adapters specified by Novogene. Quality was assessed with FastQC v0.11.5 and MultiQC v1.12 (50). *A. thaliana* genome assembly TAIR10 with annotation Araport11 (51) was downloaded from the JGI data portal. Rsem v1.2.31 (52) was used to prepare the reference sequences for mapping from the transcript file and transcript-to-gene information in order to obtain gene-level abundance estimates. Rsem and Bowtie2 v2.3.4.1 were used to map the reads. Mapping statistics and differential gene expression analysis were conducted in R version 4.0.2 (R Core Team). The Bioconductor environment was used (Biobase v2.48.0) (53). Package Tximport v1.16.1 (54) was used to import count tables, and differential gene expression analysis was performed with DESeq2 v1.28.1 (44). *P*-values for differential expression were adjusted with the Benjamini–Hochberg method and the cutoffs for differential expression were set at $|\log2(\text{Fold Change})| > 1$ and padj < 0.05. Plots were obtained with ggplot2 v3.3.5. RNA-seq data have been deposited at GEO (GSE247357; link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE247357).

Cloning. All genes were amplified from genomic DNA with a Phusion high fidelity polymerase (NEB) or a platinum SuperFi II polymerase (ThermoFisher) when the GC content was over 70%. All clonings were performed by Gibson Assembly and clones were checked by Sanger sequencing. Mutations were introduced by Gibson Assembly. Primers and plasmids used in this study are provided in Datasets S2 and S3, respectively. For the Y2H assay, coding sequences of interest were cloned in the GAL4 system Gateway® vectors (pGADT7 and pGBKT7; Clontech) using the Q5® High-Fidelity DNA Polymerase (NEB) and the primers listed in Dataset S2. The mutated clones of BOP1 ($BOP1_{D326S\text{-}D333S\text{-}E381S\text{-}E393S}$) and LSH3 ($LSH3_{H73S\text{-}R81S}$) were purchased from Azenta (Genewiz) and cloned from pDNR207 into the pGADT7 and pGBKT7, respectively.

AmpDAP-seq. All coding sequences were cloned in the pTNT-5xmyc vector (43). We used the ampDAP-seq libraries described in ref. 55. ampDAP-seq experiments were performed in triplicates following a previously described protocol (56). The analysis of ampDAP-seq triplicates (peak finding, motif search, and evaluation) was performed using the procedure described in ref. 57. The ampDAP-seq data have been deposited at GEO and are publicly available as of the date of publication (GSE235674; link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235674). For the comparison between differentially expressed genes (DEG, *lsh1 lsh3 lsh4* vs. WT) and genes bound in ampDAP-seq, a gene was considered bound when it contained at least one of the best 3,000 ampDAP-seq peaks within its gene region (1 kb upstream of the transcription start site and 500 pb downstream of the transcription stop site).

ChIP. ChIP assays were performed as described previously (58). Dynabead Protein G for Immunoprecipitation (Invitrogen) conjugated with Anti-GFP antibody (Roche) was employed for the immunoprecipitation. Quantitative real-time PCR assays were performed to determine the enrichment of the fragments. Detection was performed in triplicate using SYBR Green Super Mix 2X (Bio-Rad) and a Bio-Rad C1000TM thermal cycler. ACTIN7 was used to normalize the fold change and calculations were performed as described previously (59). Primers are listed in Dataset S2.

ALOG Recombinant Protein Production and Purification from Bacteria. All genes were cloned in the pETM-11 vector containing a N-terminal 6xHis tag and a TEV cleavage site (60). Plasmids were transformed into *Escherichia coli* Rosetta2 (DE3) cells (Novagen). Bacteria were grown in LB medium at 37 °C up to an OD_{600nm} of 0.6. Cells were then shifted to 20 °C and 0.4 mM isopropyl b-D-1-thiogalactopyranoside (IPTG) was added. After a 3 h incubation at 20 °C, cells were collected by centrifugation and sonicated in Buffer 1 (25 mM Tris pH 8, 600 mM NaCl, 1 mM TCEP) supplemented with one EDTA-free Pierce Protease Inhibitor Tablets (ThermoFisher). Lysed

cells were then centrifuged for 30 min at 15,000 rpm. Supernatant was mixed with Ni Sepharose High-Performance resin (Cytiva) previously equilibrated with Buffer 1. Resin was washed with Buffer 1 containing 35 mM imidazole and bound proteins were eluted with Buffer 1 containing 300 mM imidazole. Eluted proteins were mixed with TEV protease (0.01% w/w) and dialyzed overnight at 4 °C against Buffer 1. The following day, elution was loaded again on Ni Sepharose High-Performance resins (Cytiva) to remove tags and contaminants. Contaminant DNA was removed by passing proteins on Q Sepharose High-Performance resin (Cytiva) pre-equilibrated with Buffer 1. DNA-free proteins (260/280 ratio below 0.6) were recovered in the flow-through and further purified by Size Exclusion Chromatography (SEC) with a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated with Buffer 2 (25 mM Tris pH 8, 150 mM NaCl, 1 mM TCEP).

EMSA. Complementary oligos (listed in Dataset S2) were annealed overnight in annealing buffer (10 mM Tris pH 7.5, 150 mM NaCl, and 1 mM EDTA). We used either TAMRA-labeled oligos (Macrogen) or complementary oligos with an overhanging G labeled with Cy5-dCTP. For this, 4 pmol of double-stranded DNA was labeled with 1 unit of Klenow fragment polymerase (NEB) and 8 pmol Cy5-dCTP (Cytiva) in Klenow buffer during 1 h at 37 °C. Enzymatic reaction was then stopped with a 10-min incubation at 65 °C. Uncropped gels are provided in *SI Appendix*, Fig. S14.

Binding reactions were performed in 20 µL with different binding buffers as indicated in each figure legend. Buffer A (10 mM HEPES pH 7.5, 300 µg/mL BSA, 140 ng/µL fish sperm DNA (Sigma-Aldrich), 100 µM spermidine, 0.25% CHAPS, 1.5 mM TCEP, 0.8% glycerol) was used if unspecified. Otherwise, buffer B (25 mM Tris pH 8, 150 mM NaCl, 1 mM TCEP, 140 ng/µL fish sperm DNA, 0.8% glycerol) was used as indicated in each figure legend with indicated additives. Proteins were added at indicated concentrations. All untagged proteins were purified by SEC. Binding reactions were incubated for 20 min on ice and then loaded on a 6% native polyacrylamide gel. Gels were electrophoresed at 90 V for 75 min at 4 °C and revealed with an Amersham ImageQuant 800 imager (Cytiva).

Estimations of Kd were based on the quantification of binding experiments from Fig. 3*C* and *SI Appendix*, Fig. S8*C* and two other independent EMSAs for each protein. Kds were estimated with the Kaleidagraph software using a Michaelis–Menten model.

co-immunoprecipitation (co-IP). Myc- and FLAG-tagged versions of the different ALOG proteins were produced using the Quick Coupled Transcription/Translation System (TnT; Promega). 25 µL of TnT reactions producing indicated proteins were mixed with Buffer 1 to reach 150 µL and rotated at 4 °C during 1 h. 10 µL of pre-washed anti-myc beads were then added and incubated with the proteins for 1 h at 4 °C on a rotating wheel. Beads were then washed 4 times with Buffer 1. 1× protein Blue was then added to the beads, and beads were boiled for 5 min at 95 °C. Western Blots were then performed and revealed with HRP-conjugated anti-myc (Invitrogen; Cat# R951-25; 1:5,000 dilution) and anti-FLAG (Sigma-Aldrich, Cat# A8592; 1:1,000 dilution) antibodies. Uncropped gels are provided in *SI Appendix*, Fig. S14.

Protein Crystallization, Data Collection, and Refinement. For crystallization, $LSH3_{S45\text{-}S190}$ was purified as described above except that SEC was performed in Buffer 1 supplemented with 1 mM spermidine (Alfa Aesar). Protein was then concentrated to 5.3 mg/mL. Blunt-end complementary HPLC-purified oligos (oPR799-oPR800, see Dataset S2) were resuspended to 10 mM in 1× annealing buffer (25 mM Tris pH8, 150 mM NaCl) and annealed. Protein and DNA duplexes were directly mixed in a molar ratio of 1.1:1. ALOG-DNA was mixed at a 1:1 ratio with 0.1M HEPES, pH 7.5, 0.1M NaCl, and 1.6M ammonium sulfate and crystallized using the hanging drop method. The protein-DNA complex yielded crystals after 7 d at 4 °C. Glycerol was added to the drop to ~20% final concentration as cryoprotectant and the crystals were then flash frozen in $N_{2(l)}$.

Diffraction data were collected at 100 K at the European Synchrotron Radiation Facility, Grenoble, France, on ID23-2 at a wavelength of 0.873 Å. Indexing was performed using MXCube (61), and the default optimized oscillation range and collection parameters were used for data collection in helical collection mode to minimize radiation damage. Data were automatically processed by XDS within the Grenades pipeline (62). The data was integrated and scaled using the programs XDS and XSCALE (63). Phasing (Zn^{2+}) and initial model building was performed using CRANK2 (64, 65). Model building was performed using Coot (66) and all refinements were carried out in Phenix (67). Data collection and refinement statistics are given in Table 1. The structure is deposited under PDB 8P5Q.

Data, Materials, and Software Availability. AmpDAP-seq data (GEO link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235674) (68); RNA-seq data (GEO link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE247357) (69). The structure is deposited under PDB 8P5Q (70) (see report)].

ACKNOWLEDGMENTS. We thank Miguel Blazquez for *M. polymorpha* genomic DNA and advice, Michel Hernould for tomato genomic DNA, Martin F. Yanofsky for providing the *AP1* coding sequence, Robert Sablowski for providing the *pLSH4:LSH4:GFP* line, Luca Tadini for the valuable discussion and Israr Ud Din, Andrea Finocchio, Bao Ngan Tu and Stefano Buratti for technical support. Part of this work was carried out at NOLIMITS, an advanced imaging facility established by the Università degli Studi di Milano. This work was supported by the ANR-18-CE12-0014 ChromAuxi project to R.D., the ANR-17-CE20-0014-01 Ubiflor and ANR-21-CE20-0024 Beflore projects to F.P., the Grenoble Alliance for Integrated Structural & Cell Biology Labex financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003), a PhD Fellowship from Commissariat à l'énergie atomique et aux énergies alternatives to P.R., a post-doctoral fellowship from the University of Milan to F.C., a PhD fellowships from the Doctorate School in Molecular and Cellular Biology, Università degli Studi di Milano to V.M.B. and E.F. and a post-doctoral fellowship MUR–PRIN2020/2020RX4NWM to V.M.B.

Author affiliations: [a]Laboratoire Physiologie Cellulaire et Végétale, Université Grenoble Alpes, Centre national de la recherche scientifique, Commissariat à l'énergie atomique et aux énergies alternatives, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, Département de Biologie Structurale et Cellulaire intégrée, Grenoble F-38054, France; [b]Dipartimento di Bioscienze, Università degli Studi di Milano, Milano 20133, Italy; and [c]Structural Biology Group, European Synchrotron Radiation Facility, Grenoble 38000, France

1. R. Blanc-Mathieu, R. Dumas, L. Turchi, J. Lucas, F. Parcy, Plant-TFClass: A structural classification for plant transcription factors. *Trends Plant Sci.* **29**, 40–51 (2023).
2. P. K. I. Wilhelmsson, C. Mühlich, K. K. Ullrich, S. A. Rensing, Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in streptophyte algae. *Genome Biol. Evol.* **9**, 3384–3397 (2017).
3. L. Zhao *et al.*, Overexpression of LSH1, a member of an uncharacterised gene family, causes enhanced light regulation of seedling development. *Plant J.* **37**, 694–706 (2004).
4. L. M. Iyer, L. Aravind, ALOG domains: Provenance of plant homeotic and developmental regulators from the DNA-binding domain of a novel class of DIRS1-type retroposons. *Biol. Direct* **7**, 39 (2012).
5. S. Naramoto, Y. Hata, J. Kyozuka, The origin and evolution of the ALOG proteins, members of a plant-specific transcription factor family, in land plants. *J. Plant Res.* **133**, 323–329 (2020).
6. S. Naramoto *et al.*, A conserved regulatory mechanism mediates the convergent evolution of plant shoot lateral organs. *PLoS Biol.* **17**, e3000560 (2019).
7. C. A. MacAlister *et al.*, Synchronization of the flowering transition by the tomato TERMINATING FLOWER gene. *Nat. Genet.* **44**, 1393–1398 (2012).
8. A. Yoshida, T. Suzaki, W. Tanaka, H. Y. Hirano, The homeotic gene long sterile lemma (G1) specifies sterile lemma identity in the rice spikelet. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20103–20108 (2009).
9. X. Li *et al.*, TH1, a DUF640 domain-like gene controls lemma and palea development in rice. *Plant Mol. Biol.* **78**, 351–359 (2012).
10. A. Yoshida *et al.*, TAWAWA1, a regulator of rice inflorescence architecture, functions through the suppression of meristem phase transition. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 767–772 (2013).
11. V. M. Beretta *et al.*, The ALOG family members OsG1L1 and OsG1L2 regulate inflorescence branching in rice. *Plant J.* **115**, 351–368 (2023).
12. L. He *et al.*, SYMMETRIC PETALS 1 encodes an ALOG domain protein that controls floral organ internal asymmetry in pea (Pisum sativum L.). *Int. J. Mol. Sci.* **21**, 4060 (2020).
13. K. Schiessl *et al.*, Light-sensitive short hypocotyl genes confer symbiotic nodule identity in the legume *Medicago truncatula*. *Curr. Biol.* 34, 1–16 (2024). https://doi.org/10.1101/2023.02.12.528179.
14. W. Xiao, S. Su, T. Higashiyama, D. Luo, A homolog of the ALOG family controls corolla tube differentiation in Torenia fournieri. *Development* **146**, dev177410 (2019).
15. J. Zou *et al.*, Arabidopsis LSH8 positively regulates ABA signaling by changing the expression pattern of ABA-responsive proteins. *Int. J. Mol. Sci.* **22**, 10314 (2021).
16. M. O. Press, C. Queitsch, Variability in a short tandem repeat mediates complex epistatic interactions in *Arabidopsis thaliana*. *Genetics* **205**, 455–464 (2017).
17. M. S. Vo Phan, I. Keren, P. T. Tran, M. Lapidot, V. Citovsky, Arabidopsis LSH10 transcription factor and OTLD1 histone deubiquitinase interact and transcriptionally regulate the same target genes. *Commun. Biol.* **6**, 58 (2023).
18. M. Lee *et al.*, Molecular characterization of Arabidopsis thaliana LSH1 and LSH2 genes. *Genes Genomics* **42**, 1151–1162 (2020).
19. S. Takeda *et al.*, CUP-SHAPED COTYLEDON1 transcription factor activates the expression of LSH4 and LSH3, two members of the ALOG gene family, in shoot organ boundary cells. *Plant J.* **66**, 1066–1077 (2011).
20. E. Cho, P. C. Zambryski, ORGAN BOUNDARY1 defines a gene expressed at the junction between the shoot apical meristem and lateral organs. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2154–2159 (2011).

21. X. Huang *et al.*, ROS regulated reversible protein phase separation synchronizes plant flowering. *Nat. Chem. Biol.* **17**, 549–557 (2021).
22. P. Peng *et al.*, The rice TRIANGULAR HULL1 protein acts as a transcriptional repressor in regulating lateral development of spikelet. *Sci. Rep.* **7**, 13712 (2017).
23. X. Huang *et al.*, Heterotypic transcriptional condensates formed by prion-like paralogous proteins canalize flowering transition in tomato. *Genome Biol.* **23**, 78 (2022).
24. S. Bencivenga, A. Serrano-Mislata, M. Bush, S. Fox, R. Sablowski, Control of oriented tissue growth through repression of organ boundary genes promotes stem morphogenesis. *Dev. Cell* **39**, 198–208 (2016).
25. J. R. Dinneny, R. Yadegari, R. L. Fischer, M. F. Yanofsky, D. Weigel, The role of JAGGED in shaping lateral organs. *Development* **131**, 1101–1110 (2004).
26. M. D. Wilkinson, G. W. Haughn, UNUSUAL FLORAL ORGANS controls meristem identity and organ primordia fate in Arabidopsis. *Plant Cell* **7**, 1485–1499 (1995).
27. M. R. Karim, A. Hirota, D. Kwiatkowska, M. Tasaka, M. Aida, A role for Arabidopsis PUCHI in floral meristem identity and bract suppression. *Plant Cell* **21**, 1360 (2009).
28. M. Norberg, M. Holmlund, O. Nilsson, The BLADE ON PETIOLE genes act redundantly to control the growth and development of lateral organs. *Development* **132**, 2203–2213 (2005).
29. C. K. Ohno, G. V. Reddy, M. G. B. Heisler, E. M. Meyerowitz, The Arabidopsis JAGGED gene encodes a zinc finger protein that promotes leaf tissue development. *Development* **131**, 1111–1122 (2004).
30. F. Parcy, O. Nilsson, M. A. Busch, I. Lee, D. Weigel, A genetic framework for floral patterning. *Nature* **395**, 561–566 (1998).
31. D. Weigel, J. Alvarez, D. R. Smyth, M. F. Yanofsky, E. M. Meyerowitz, *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859 (1992).
32. K. Goslin *et al.*, Transcription factor interplay between LEAFY and APETALA1/CAULIFLOWER during floral initiation. *Plant Physiol.* **174**, 1097–1109 (2017).
33. C. Sayou *et al.*, A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nat. Commun.* **7**, 11222 (2016).
34. V. Grandi, V. Gregis, M. M. Kater, Uncovering genetic and molecular interactions among floral meristem identity genes in Arabidopsis thaliana. *Plant J.* **69**, 881–893 (2012).
35. C. Xu, S. J. Park, J. Van Eck, Z. B. Lippman, Control of inflorescence architecture in tomato by BTB/POZ transcriptional regulators. *Genes Dev.* **30**, 2048–2061 (2016).
36. S. Su *et al.*, A BLADE-ON-PETIOLE orthologue regulates corolla differentiation in the proximal region in Torenia fournieri. *Nat. Commun.* **14**, 4763 (2023).
37. R. C. O'Malley *et al.*, Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016).
38. D. R. Smyth, J. L. Bowman, E. M. Meyerowitz, Early flower development in Arabidopsis. *Plant Cell* **2**, 755–767 (1990).
39. S. J. Maerkl, S. R. Quake, A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
40. L. Holm, A. Laiho, P. Törönen, M. Salgado, DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, e4519 (2023).
41. S. S. Krishna, I. Majumdar, N. V. Grishin, Structural classification of zinc fingers: Survey and summary. *Nucleic Acids Res.* **31**, 532–50 (2003).
42. E. Wingender, T. Schoeps, M. Haubrock, J. Dönitz, TFClass: A classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* **43**, D97–D102 (2015).
43. H. Liu *et al.*, CRISPR-P 2.0: An improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant* **10**, 530–532 (2017).
44. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
45. M. Di Marzo *et al.*, MADS-box and bHLH transcription factors coordinate transmitting tract development in *Arabidopsis thaliana*. *Front. Plant Sci.* **11**, 526 (2020).
46. F. Caselli *et al.*, REM34 and REM35 control female and male gametophyte development in Arabidopsis thaliana. *Front. Plant Sci.* **10**, 1351 (2019).
47. S. R. Hepworth, Y. Zhang, S. McKim, X. Li, G. W. Haughn, BLADE-ON-PETIOLE–dependent signaling controls leaf and floral patterning in Arabidopsis. *Plant Cell* **17**, 1434–1448 (2005).
48. S. Torti *et al.*, Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *Plant Cell* **24**, 444–462 (2012).
49. S. de Folter, R. G. H. Immink, Yeast protein–protein interaction assays and screens. *Methods Mol. Biol.* **754**, 145–165 (2011).
50. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
51. C. Cheng *et al.*, Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).
52. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).
53. W. Huber *et al.*, Orchestrating high-throughput genomic analysis with bioconductor. *Nat. Methods* **12**, 115–121 (2015).
54. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2016).
55. X. Lai *et al.*, The LEAFY floral regulator displays pioneer transcription factor properties. *Mol. Plant* **14**, 829–837 (2021).
56. A. Bartlett *et al.*, Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* **12**, 1659–1672 (2017).
57. P. Rieu *et al.*, The F-box protein UFO controls flower development by redirecting the master transcription factor LEAFY to new cis-elements. *Nat. Plants* **9**, 315–329 (2023).
58. V. Gregis, A. Sessa, C. Dorca-Fornell, M. M. Kater, The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. *Plant J.* **60**, 626–637 (2009).
59. R. Petrella *et al.*, BPC transcription factors and a polycomb group protein confine the expression of the ovule identity gene *SEEDSTICK* in Arabidopsis. *Plant J.* **102**, 582–599 (2020).
60. A. Dümmler, A. M. Lawrence, A. de Marco, Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb. Cell Fact.* **4**, 34 (2005).
61. J. Gabadinho *et al.*, MxCuBE: A synchrotron beamline control environment customized for macromolecular crystallography experiments. *J. Synchrotron. Radiat.* **17**, 700–707 (2010).
62. S. Monaco *et al.*, Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *J. Appl. Crystallogr.* **46**, 804–810 (2013).
63. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
64. N. S. Pannu *et al.*, Recent advances in the *CRANK* software suite for experimental phasing. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 331–337 (2011).
65. P. Skubák *et al.*, A new MR-SAD algorithm for the automatic building of protein models from low-resolution X-ray data and a poor starting model. *IUCrJ* **5**, 166–171 (2018).
66. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
67. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
68. P. Rieu, J. Lucas, F. Parcy, C. Zubieta, The ALOG domain defines a new family of plant-specific transcription factors acting during Arabidopsis flower development. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235674. Deposited 23 June 2023.
69. V. M. Beretta, F. Caselli, C. Paleni, V. Gregis, M. M. Kater, The ALOG domain defines a new family of plant-specific Transcription Factors acting during Arabidopsis flower development. NCBI GEO. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE247357. Deposited 8 November 2023.
70. C. Zubieta, M. H. Nanao, P. Rieu, Structure of an ALOG domain from Arabidopsis thaliana in complex with DNA. Protein Data Bank. https://www.rcsb.org/structure/8P5Q. Deposited 24 May 2023.