Evolution of plant cell-type-specific *cis*-regulatory elements

Haidong Yan[1,+], John P. Mendieta[1,+], Xuan Zhang[1], Alexandre P. Marand[1#], Yan Liang[2], Ziliang Luo[1], Mark A.A. Minow[1], Thomas Roulé[3], Doris Wagner[3], Xiaoyu Tu[4], Yonghong Wang[2,5], Silin Zhong[6], Susan R. Wessler[7], and Robert J. Schmitz[1,*]

[1]Department of Genetics, University of Georgia, Athens, GA, 30602, USA
[2]College of Life Sciences, Shandong Agricultural University, Taian, Shandong 271018, China
[3]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Joint Center for Single Cell Biology, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China
[5]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
[6]State Key Laboratory of Agrobiotechnology, School of Life Sciences, The Chinese University of Hong Kong
[7]Department of Botany and Plant Sciences, University of California, Riverside, CA, 92521, USA

[+]These authors contributed equally.
[#] Current address: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, 48109, USA
*Corresponding author

Abstract:

Cis-regulatory elements (CREs) are critical in regulating gene expression, and yet our understanding of CRE evolution remains a challenge. Here, we constructed a comprehensive single-cell atlas of chromatin accessibility in *Oryza sativa*, integrating data from 104,029 nuclei representing 128 discrete cell states across nine distinct organs. We used comparative genomics to compare cell-type resolved chromatin accessibility between *O. sativa* and 57,552 nuclei from four additional grass species (*Zea mays*, *Sorghum bicolor*, *Panicum miliaceum*, and *Urochloa fusca*). Accessible chromatin regions (ACRs) had different levels of conservation depending on the degree of cell-type specificity. We found a complex relationship between ACRs with conserved noncoding sequences, cell-type specificity, conservation, and tissue-specific switching. Additionally, we found that epidermal ACRs were less conserved compared to other cell types, potentially indicating that more rapid regulatory evolution has occurred in the L1 epidermal layer of these species. Finally, we identified and characterized a conserved subset of ACRs that overlapped the repressive histone modification H3K27me3, implicating them as potentially critical silencer CREs maintained by evolution. Collectively, this comparative genomics approach highlights the dynamics of cell-type-specific CRE evolution in plants.

Main

Cis-regulatory elements (CREs) function as pivotal hubs, facilitating the binding of transcription factors (TFs) and recruitment of chromatin-modifying enzymes, thereby fine-tuning gene expression in a spatiotemporal-specific manner[1]. CREs play important roles in developmental and environmental processes, and their functional divergence frequently drives evolutionary change[2,3]. Prior studies highlighted the dynamic nature of CREs throughout evolution and their involvement in regulating gene expression via distinct chromatin pathways[4-9]. Across diverse cell types, gene expression is intricately regulated by multiple distinct CREs, each exerting control within specific cell or tissue type or particular developmental stage and environmental cue[10-12]. Environmental sensing and adaptation, in plants, relies heavily upon epidermal cells[13]. For example epidermal bulliform cells in grasses change their turgor pressure to roll the leaf to slow water loss under stressful conditions, with the TF, ZINC FINGER HOMEODOMAIN 1 (ZHD1), modulating leaf rolling by influencing rice (*Oryza sativa*) bulliform cell development[14,15]. Several studies have identified CREs functioning in a cell-type-specific manner within various plant species[16-23]. Despite these findings, our understanding of CREs exhibiting evolutionarily conserved or divergent cell-type-specific activities remains limited.

Through single-cell assay for transposase accessible chromatin sequencing (scATAC-seq), we constructed an expansive single-cell reference atlas of accessible chromatin regions (ACRs) within rice. We then leveraged these data in tandem with four additional scATAC-seq leaf datasets from diverse grasses (*Zea mays*, *Sorghum bicolor*, *Panicum miliaceum*, and *Urochloa fusca*)[24] allowing us to compare ACRs across species and cell types. We quantified the proportion of ACRs that were conserved in these monocots and found high rates of cell-type-specific ACR turnover, particularly in epidermal cells. This indicates that the CREs associated with specific cell types are rapidly changed by evolution. Finally, we used both conserved non-coding sequences (CNS) and H3K27me3 to find a series of conserved ACRs and the candidate CREs within them that are potentially important for recruitment of Polycomb-mediated gene silencing.

Construction of an ACR atlas at single-cell resolution in rice

To create a cell-type-resolved ACR rice atlas, we conducted scATAC-seq across a spectrum of nine organs in replicate (Fig. 1a and b). Data quality metrics, such as correlation between biological replicates, transcription start site enrichment, fraction of reads in peaks, fragment size distribution, and organelle content, revealed excellent data quality (Supplementary Fig. 1 and 2). Following strict quality control filtering, we identified 104,029 high-quality nuclei, with an average of 41,701 unique Tn5 integrations per nucleus. Based on a nine step annotation strategy, which included RNA *in situ* and spatial-omic (slide-seq) validation of cell-type specificity, we identified a total of 128 cell states, encompassing 60 main cell types across various developmental stages from all the organs sampled (Fig. 1b; Extended Data Fig. 1a; Supplementary Note 1; Supplementary Fig. 3-12; Supplementary Tables 1-5).

By analyzing cell-type-aggregated chromatin accessibility profiles we identified a total of 120,048 ACRs (Extended Data Fig. 1b and c). Among these ACRs, 42,649 were categorized as 'cell-type-specific ACRs', exhibiting accessible chromatin in less than 17% (10/60) of the main cell types, whereas approximately 77,399 were classified as 'broad ACRs' with chromatin accessibility in more than 17% of the cell types (Fig. 1c and d; Extended Data Fig. 1d and e). When analyzing their proximity to genomic features in the rice genome, about half of the ACRs were gene proximal (52%; located within 2 kb of genes; Fig. 1e). These proximal ACRs had higher but less variable chromatin accessibility than genic and distal ACRs (Extended Data Fig. 1f). In contrast, about 19% of the ACRs overlapped genes, mostly in introns, and the remaining 29% were categorized as distal (Fig. 1e; situated more than 2 kb away from genes). The greater chromatin accessibility variance in non-proximal ACRs, suggests these regions may be important in different cellular contexts or tissues.

We were interested in leveraging the *O. sativa* atlas to understand how ACRs change during grass evolution. The *O. sativa* ACRs were overlapped with syntenic regions defined by their relationship to four different grass species *Z. mays*, *S. bicolor*, *P. miliaceum*, and *U. fusca* that have single-cell ATAC sequencing using combinatorial indexing (sciATAC-seq) data from leaves[24]. The analysis revealed that 34% (40,477) of the *O. sativa* ACRs were within 8,199 syntenic regions (~86 Mb in *O. sativa*) between *O. sativa* and at least one of four examined grass species, whereas the majority of ACRs (66%; 79,571) were located in non-syntenic regions. Notably, these non-syntenic ACRs were enriched (p = 6e-248 to 0.0031; Fisher's exact test) for cell-type specificity, particularly for proximal and distal ACRs (Fig. 1f; Extended Data Fig. 1g). This reveals that most of the ACRs in the grass species examined occurred in non-syntenic regions.

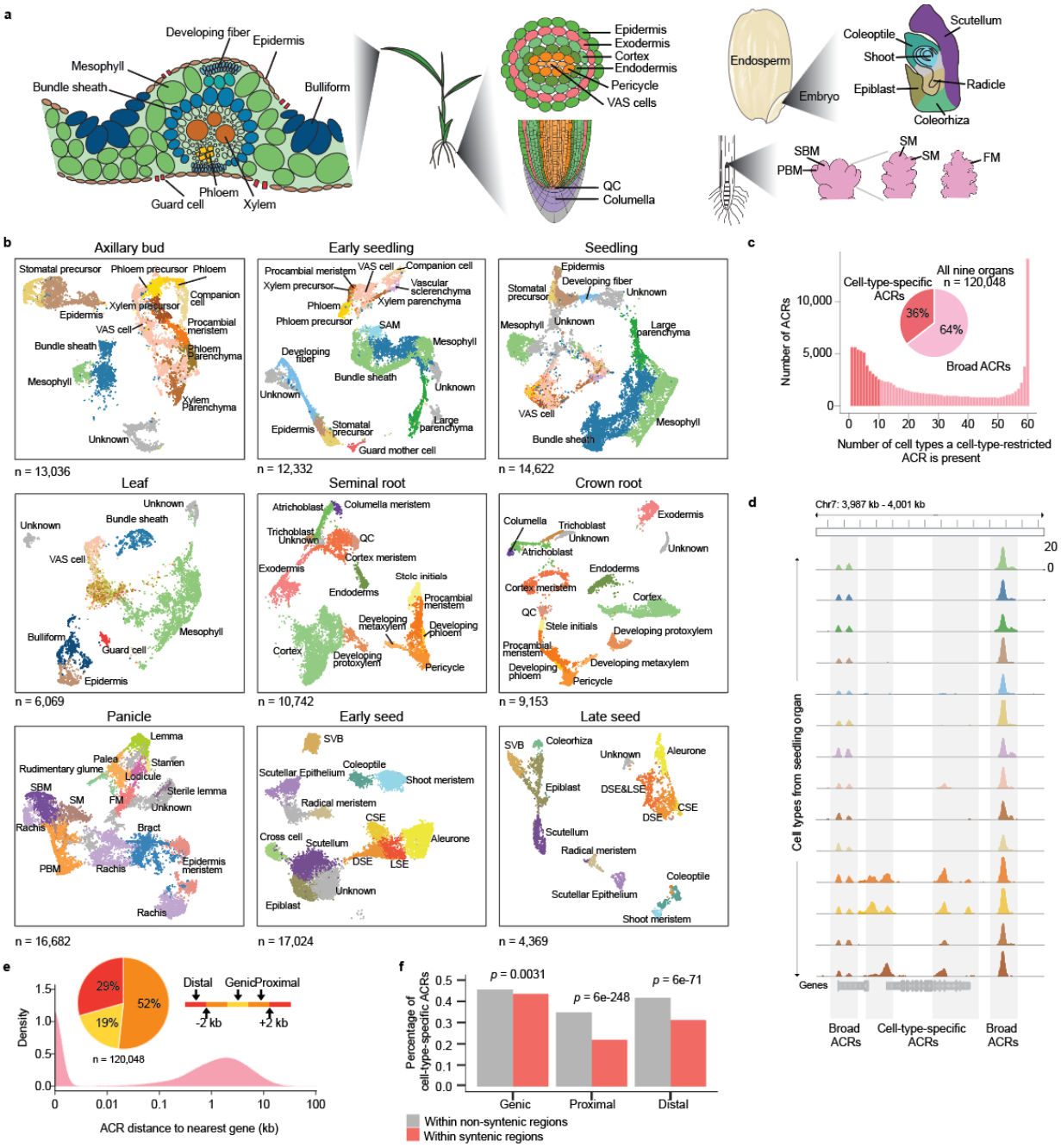
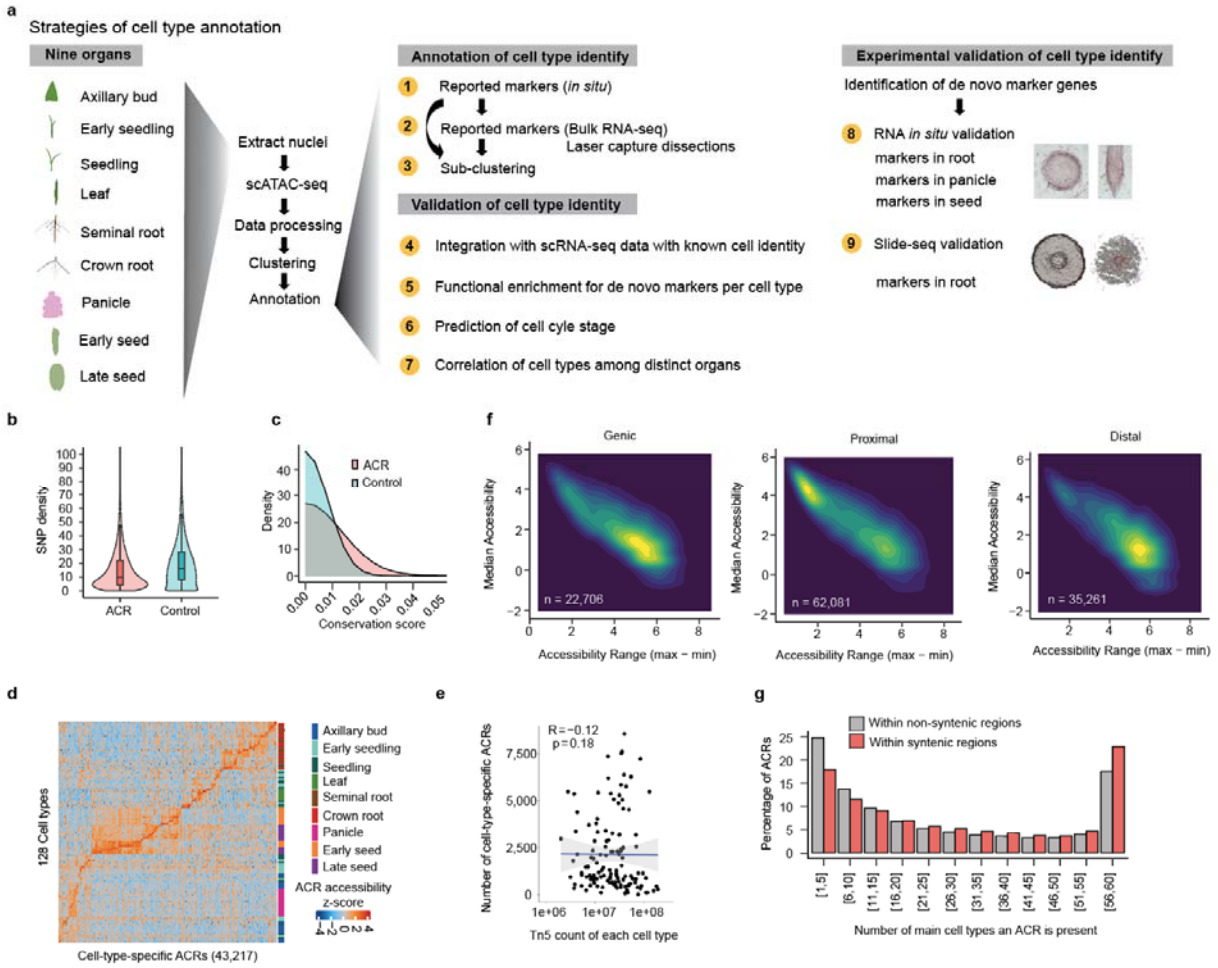


Fig. 1. Identifying cell types and characterizing ACRs in rice using scATAC-seq data. a, Overview of cell types in leaf, root, seed, and panicle organs. VAS cells: vasculature-related cells. QC: Quiescent Center. SBM: Secondary Branch Meristem. PBM: Primary Branch Meristem. SM: Spikelet Meristem. FM: Floral Meristem. **b,** UMAP projection of nuclei, distinguished by assigned cell-type labels in axillary bud, early seedling (7 days after sowing), seedling (14 days after sowing), leaf (V4 stage; four leaves with visible leaf collars), seminal root, crown root, panicle, early seed development (6 DAP; days after pollination), and late seed development (10 DAP). SAM: Shoot Apical Meristem. CSE: Central Starchy Endosperm. DSE: Dorsal Starchy Endosperm. LSE: Lateral Starchy Endosperm. SVB: Scutellar Vascular Bundle. **c,** Evaluation of proportions of ACRs that are cell-type specific versus broad. **d,** A screenshot illustrates the examples of cell-type-specific and broad ACRs. **e,** Accessible chromatin regions (ACRs) show a bimodal distribution of distance to the nearest gene. The ACRs were categorized into three major

groups based on their locations to the nearest gene: genic ACRs (overlapping a gene), proximal ACRs (located within 2 kb of genes), and distal ACRs (situated more than 2 kb away from genes). **f,** The enrichment of cell-type-specific ACRs in non-syntenic regions as opposed to syntenic regions. Significance testing was performed using a two-sided Fisher's exact test.



Extended Data Fig. 1. Pipeline of cell identity annotation and quality control of rice ACRs. a, A pipeline of cell identity annotation corresponding to nine strategies shown in Supplementary Note 1. **b-c,** A comparison between ACRs (n = 120,048) and control regions (n = 120,048) in terms of SNP density (**b**) and conservation scores (**c**). The SNP data were downloaded from Rice SNP-Seek Database[25]. The SNP density is the number of SNPs per 1,000 bp within the ACRs or control regions. The ACRs exhibited lower SNP density and higher conservation scores compared to a control composed of non-ACR DNA. The control set was created to match each ACR by excluding exons, ACRs, and unmappable regions, and they possess a similar GC content (< 5% average difference) compared to the positive set. **d,** Row normalized chromatin accessibility z-scores for ACRs across cell types. **e,** Spearman correlation is shown between the total number of Tn5 insertions and number of cell-type-specific ACRs across all cell types. The identification of cell-type-specific ACRs was independent of sequencing depth, as evidenced by the lack of correlation between Tn5 insertions and the number of cell-type-specific ACRs per cell type. **f,** A two-dimensional density plot illustrates the median chromatin accessibility across 128 cell types for 120,048 ACRs, along with the range of chromatin accessibility (calculated as the difference between the maximum and minimum values). **g,** Percentage of ACRs within main cell types in *O. sativa*, distributed

across syntenic and non-syntenic regions shared with four other grass species (*Z. mays*, *S. bicolor*, *P. miliaceum*, and *U. fusca*).

The landscape of cell-type-specific ACRs across grass species

To determine to what degree ACR number, genomic position and cell-type specificity differs amongst grasses, we compared the composition and distribution of leaf ACRs across the five species (Fig. 2a). Using previous cell-type annotations[24], we calculated the proportion of both broad and cell-type-specific ACRs across all species. An average of ~53,000 ACRs were identified across the five species, with 15-35% of the ACRs classified as cell-type specific (Fig. 2b). Broad and cell-type-specific ACRs were equivalent in their distributions around promoters, distal, and genic regions (Fig. 2c; Extended Data Fig. 2a-c).

Prior hypotheses have suggested that large scale regulatory rewiring could play a key role in cell-type adaptation to environmental conditions[8,26]. To explore instances where divergent TF activity in orthologous cell types occurred, we associated gene body chromatin accessibility of TFs with their cognate TF motifs across different species and cell types. Approximately 71% of the TFs examined exhibited a positive correlation between the local chromatin accessibility of their gene body and global enrichment of their cognate binding motifs within ACRs across all cell types (Extended Data Fig. 3a). This finding suggests a relationship between the local TF gene chromatin accessibility and TF protein activity in the same cell. Moreover, the genomic sequences from all ACRs discovered in all species and cell types exhibited enrichment of TF motifs compared to the control set of sequences (Extended Data Fig. 3b). Furthermore, TF motif enrichment analysis revealed known TF-cell-type specificities (Fig. 2d). For example, the HOMEODOMAIN GLABROUS 1 (HDG1) TF is critical for epidermis and cuticle development[27] and its motif was enriched in epidermis cells in all five species (Extended Data Fig. 3c). We also observed motif enrichments of WRKY, HOMEODOMAIN-LEUCINE ZIPPER (HD-ZIP), and PLANT ZINC FINGER (PLINC) in epidermal cells across all five species examined (Fig. 2e; Extended Data Fig. 3d; Supplementary Table 6). This result indicates that these TFs play a conserved role in the development of the grass epidermis. Phloem companion, sieve element cell, and bundle sheath cell TFs exhibited similar enrichments across the species (Fig. 2e). However, species-specific motif patterns were also observed, with *O. sativa* being the most different. For example, the DNA-BINDING ONE ZINC FINGER (DOF) TF family exhibited higher enrichment scores in four species (*Z. mays*, *S. bicolor*, *P. miliaceum*, and *U. fusca*) that are C_4 photosynthesizing species, as opposed to *O. sativa*, which uses C_3 photosynthesis (Fig. 2e; Extended Data Fig. 3e and f). The DOF TF family regulates the formation and function of vascular tissues in *Arabidopsis thaliana*[28], and has been implicated in the switch from C_3 to C_4[23,29,30] (Fig. 2e; Extended Data Fig. 3e and f). These results suggest that distinct DOF TF motif enrichment in phloem and bundle sheath cells might fulfill different vasculature roles in C_3 and C_4 grass leaves[31,32]. Taken together, our findings demonstrate the power of scATAC-seq data in a comparative framework to explore regulatory evolution, both

based on the relationship of ACRs to TF motifs, as well as the relationship between TFs and their corresponding motifs.

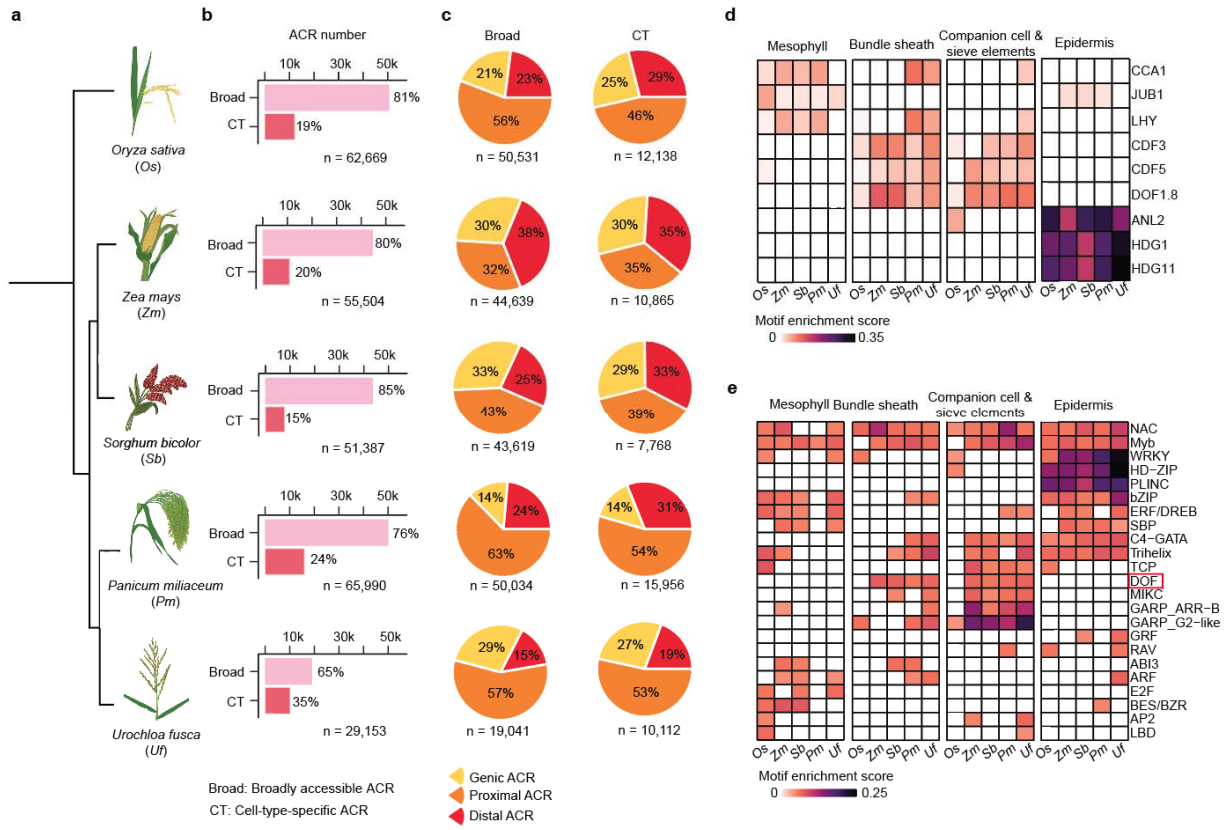
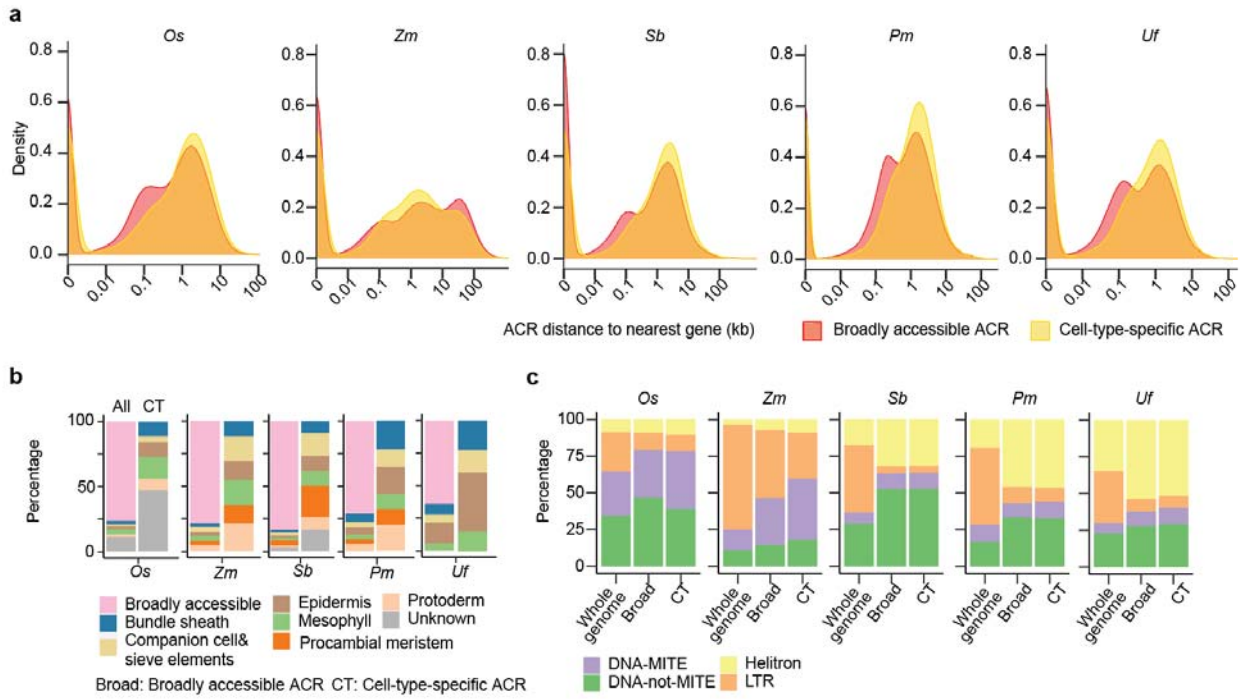
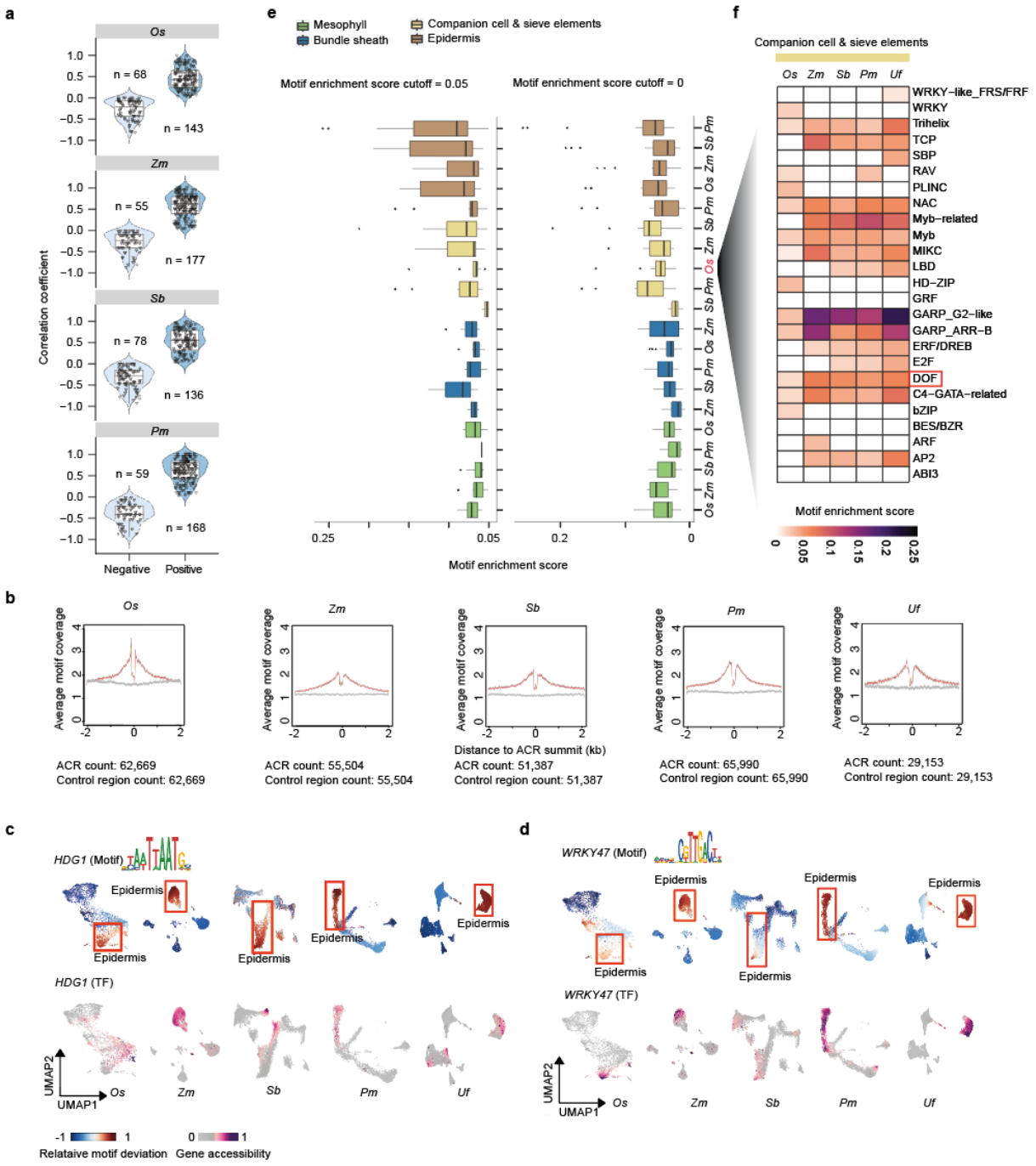


Fig. 2. Position and motif enrichment of cell-type-specific ACRs across species. a, A phylogenetic tree illustrates five species under examination. **b,** The count of broad and cell-type-specific ACRs. **c,** Broad and cell-type-specific ACRs were classified into three main groups based on their proximity to the nearest gene: genic ACRs (overlapping a gene), proximal ACRs (located within 2 kb of genes), and distal ACRs (situated more than 2 kb away from genes). *O. sativa*, *P. miliaceum*, and *U. fusca* showed a higher percentage of proximal ACRs, but a lower percentage of distal ACRs compared to *Z. mays* and *S. bicolor*, likely reflecting differences in intergenic space and overall genome sizes. **d**, A heatmap illustrates nine TF motif enrichments, consistent with the known TF dynamics among cell types [CIRCADIAN CLOCK ASSOCIATED 1 (CCA1), LATE ELONGATED HYPOCOTYL (LHY) and JUNGBRUNNEN1 (JUB1): mesophyll[33]; CYCLING DOF FACTOR 3 (CDF3) and CDF5: companion cells[34]; DOF1.8: vascular-related cells[35]; ANTHOCYANINLESS2 (ANL2), HOMEODOMAIN GLABROUS 1 (HDG1), and HDG11: epidermis[27,36,37]]. **e,** A heatmap illustrates collapsed TF motif enrichment patterns into super motif families across various species for each cell type. The motif enrichment score cutoff was set to 0.05. The DOF TF motif family was highlighted by a red frame. To mitigate the impact of substantial variations in cell numbers across species or cell types, we standardized (down-sampled) the cell counts by randomly selecting 412 cells per cell type per species. This count represents the lowest observed cell count for a given cell type across all species (See Methods: Linear-model based motif enrichment analysis).



Extended Data Fig. 2. Characterization of ACRs across five grass species. a, Distribution of distance of ACRs to their nearest genes in broadly accessible and cell-type-specific ACRs across five species. **b,** The distribution of cell-type-specific ACRs across different cell types was characterized by a similar percentage representation. **c,** The compositions of the family of transposable elements (TEs) overlapping with broad and cell-type-specific ACRs illustrate comparable compositions across TE super families, distinct from the whole genome.



Extended Data Fig. 3. Examination of the chromatin accessibility of motifs enriched in specific cell types across species. a, Spearman correlation between chromatin accessibility of TF genes and motif deviation. Note that for *U. fusca* species, the analysis was omitted due to the limited number of cell types (four) available for spearman correlation, making it statistically unfeasible. **b,** Mean motif coverage centered on ACRs (red line) and control regions (gray line) for each species, with 95% confidence intervals shown as a shaded polygon. **c,** The UMAP panels highlight one TF motif (HDG1) enriched with ACRs in epidermis, where their cognate TFs are known to accumulate. **d,** The UMAP panels highlight one representative TF motif underlying the WRKY family enriched with ACRs in epidermis cells, where their cognate TFs are known to accumulate. **e-f,** Presenting box (**e**) and heatmap (**f**) plots illustrating the

collapsed motif enrichment patterns into TF motif families across various species for each cell type. Each dot within the box represents a TF motif family. The DOF TF motif family was highlighted by a red frame in the heatmap. The reduced motif enrichment cutoff setting reveals additional motif families enriched in companion cells and sieve elements, compared to the higher beta cutoff setting as shown in Fig. 2e. It is worth noting that we still observed lower enrichment scores in *O. sativa* compared to other species.

Species-specific evolution of cell-type-specific ACRs

To understand how cell-type-specific and broad ACRs changed over evolution, we examined ACRs within syntenic regions among the studied species. To compare ACRs, we devised a synteny-based BLASTN pipeline that allowed us to compare sequences directly (See Methods: Identification of syntenic regions; Fig. 3a; Extended Data Fig. 4a; Supplementary Table 7). Using *O. sativa* ACRs as a reference, we identified three classes of cross-species ACR conservation: 1) ACRs with matching sequences that are accessible in both species (shared ACRs), 2) ACRs with matching sequences, but are only accessible in one species (variable ACRs), and 3) ACRs where the sequence is exclusive to a single species (species-specific ACRs; Fig. 3b). The shared ACR BLASTN hits were often small syntenic sequences, highlighting the large divergence of grass ACRs sequences. However, the majority (92-94%) of these shared BLASTN sequences encoded known TF motifs (Supplementary Fig. 13), indicating that shared ACRs are conserved regulatory regions. In contrast, variable ACRs represent a blend of conserved and divergent regulatory elements, and species-specific ACRs likely indicate novel regulatory loci. We found that, on average, shared ACRs were enriched (p = 9e-55 to 4e-17; Fisher's exact test) for broad ACRs, whereas the variable (p = 2e-16 to 2e-04; Fisher's exact test) and species-specific (p = 2e-06 to 2e-03; Fisher's exact test) classes were enriched for cell-type specificity (Fig. 3c; Extended Data Fig. 4b and c). Moreover, we observed that the genomic distribution of shared ACRs were biased towards proximal ACRs (Extended Data Fig. 4d). This contrasts with cell-type-specific ACRs, which are overrepresented in the species-specific class, and tend to reside in distal genomic regions (Extended Data Fig. 4e).

We further investigated whether the cell-type-specific ACRs were conserved in their cell-type specificity by evolution. Pairwise comparison between *O. sativa* and the other grasses revealed that a low number (128 to 299) of shared ACRs retained cell-type specificity in both species (Extended Data Fig. 4f). Of these few shared cell-type specific ACRs, the majority (62%-68%), were accessible in the identical cell-type in both *O. sativa* and the corresponding species (Fig. 3d; Extended Data Fig. 4f). For example, the promoter ACR associated with *LATERAL ROOT DEVELOPMENT 3 (LRD3)*, a gene critical in companion cell and sieve element development[38], showed sequence conservation between *O. sativa* and *S. bicolor* (Fig. 3d). Interestingly, ACRs which were mesophyll specific in *O. sativa* changed their cell-type specificity to bundle sheath 17%-41% of the time, while bundle sheath ACRs changed to mesophyll 9%-25% of the time (Fig. 3d; Extended Data Fig. 4g). This result is likely due to the functional divergence associated with the shift from C_3 (*O. sativa*) to C_4 (all other species sampled) photosynthesis[23,24]. Of all the classes of cross-species ACR conservation, species-specific ACRs were the most predominant in

every cell type (Fig. 3e). These findings suggest a dynamic and rapid evolution of cell-type-specific ACRs within the examined species. Notably, ACRs in L1 layer (epidermis and protoderm) cells exhibited the highest proportion of species-specific ACRs (Fig. 3e; Extended Data Fig. 5a-c), suggesting that L1 ACRs are more evolutionarily divergent compared to the other cell types.

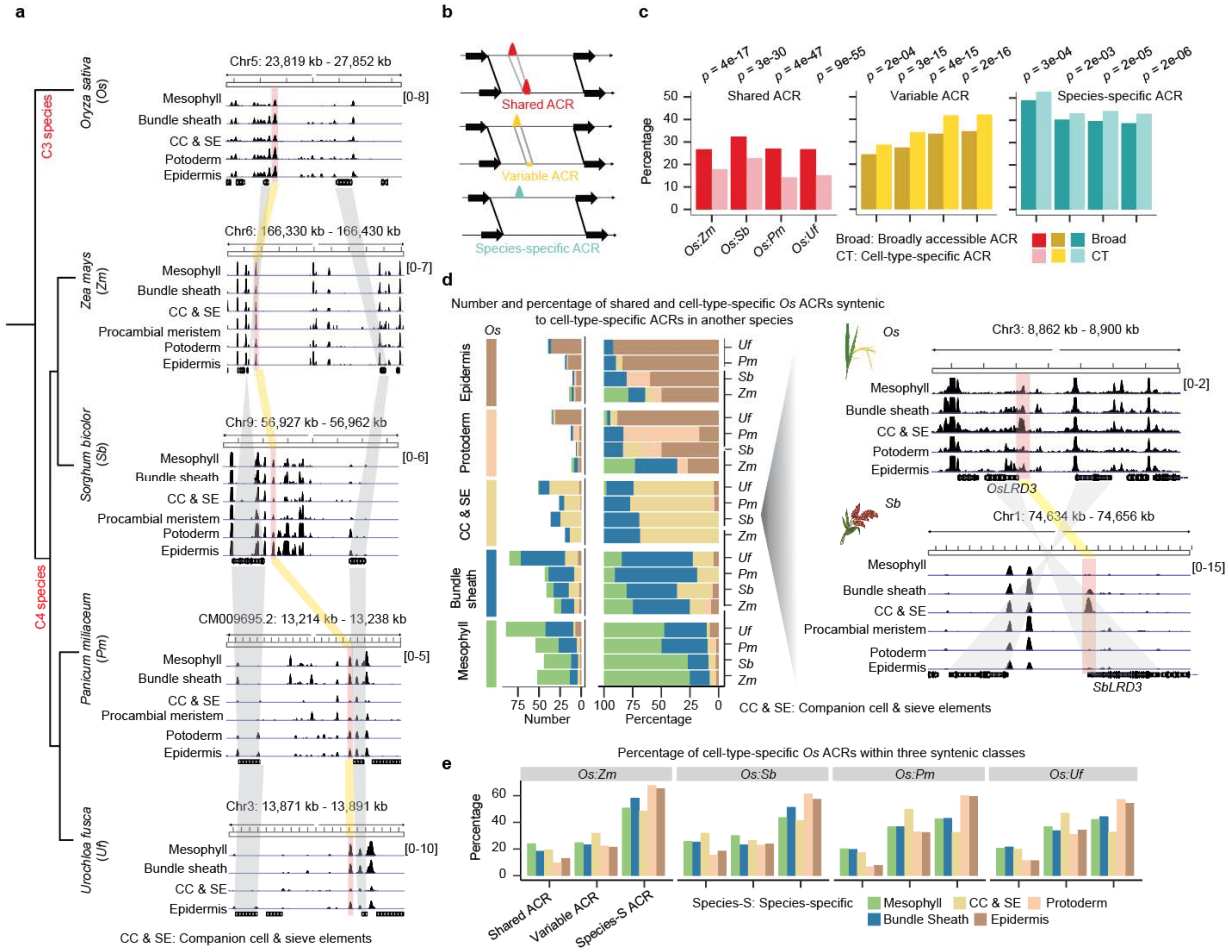
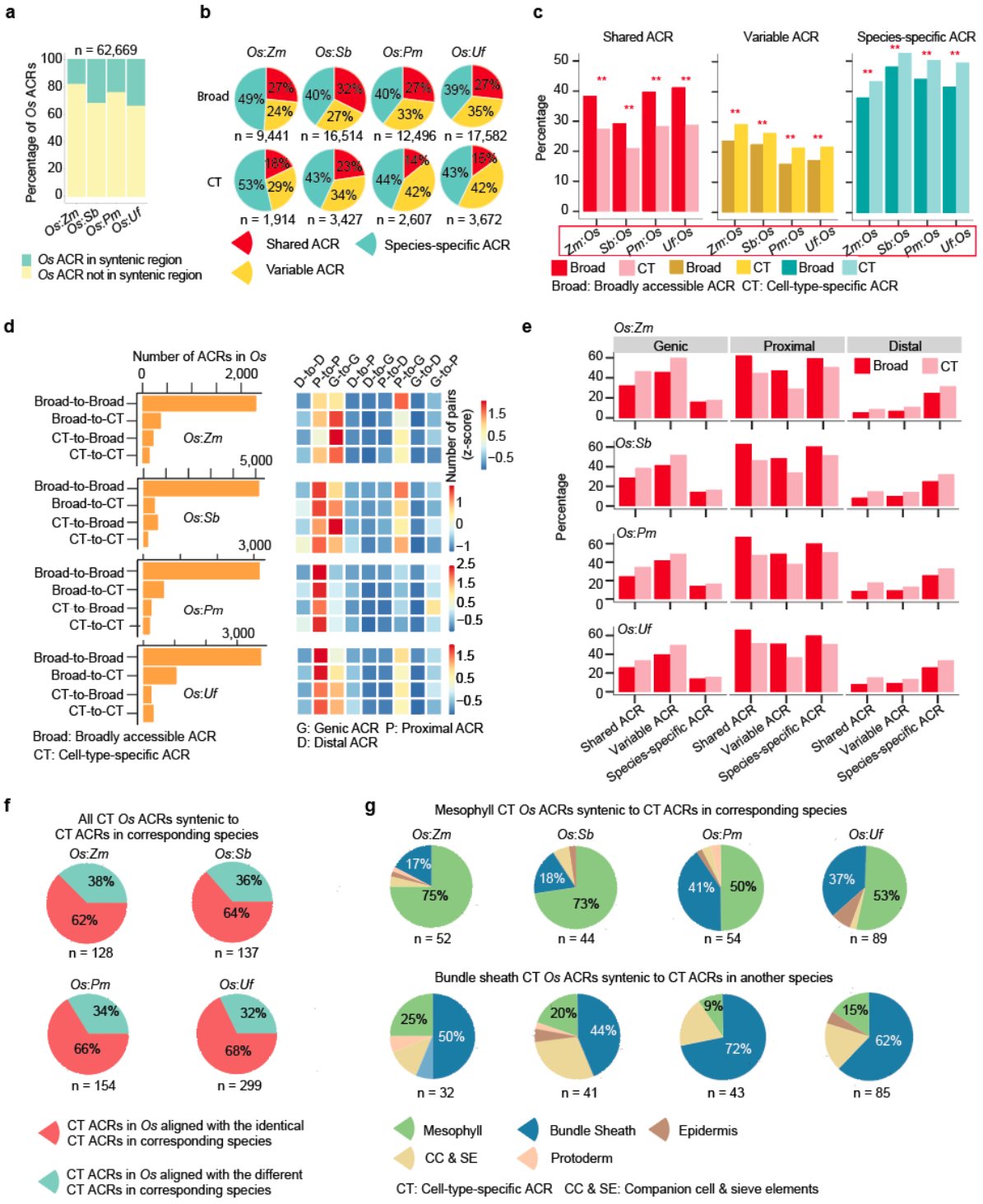


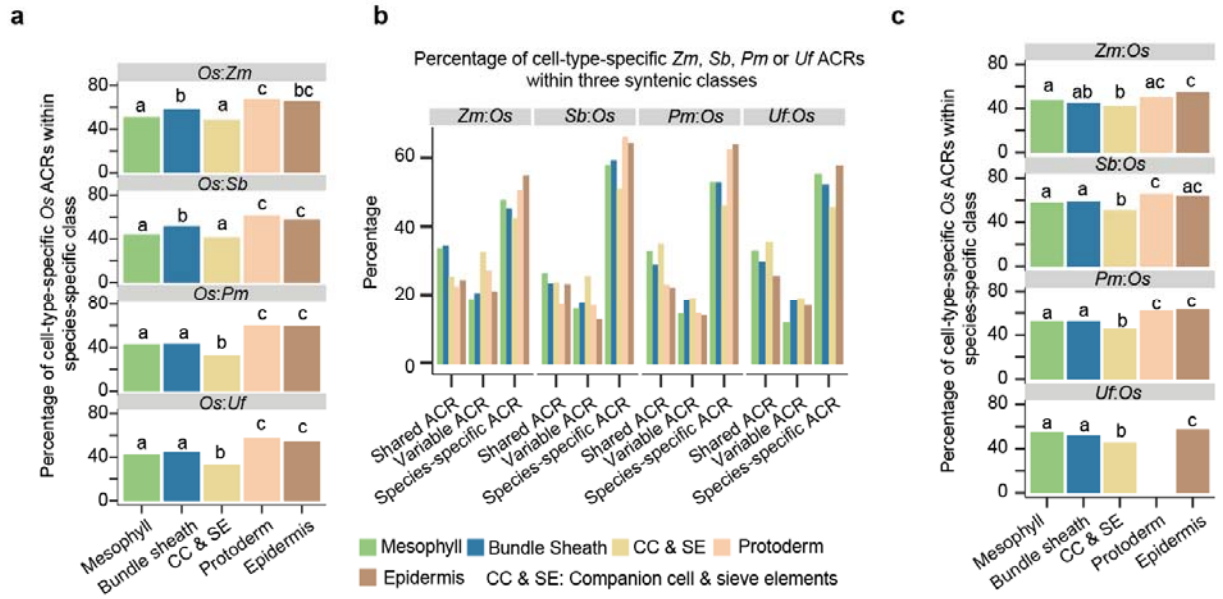
Fig. 3. Cell-type-specific ACRs are frequently species-specific. a, A screenshot illustrating syntenic regions capturing shared ACRs across five species. The red bars denote syntenic ACRs within regions flanked by corresponding syntenic gene pairs, while the gray color highlights these syntenic gene pairs. **b,** Three classes depicting variations in ACR conservation between two species. 'Shared ACRs': ACRs with matching sequences that are accessible in both species; 'Variable ACRs': ACRs with matching sequencesbut are only accessible in one species; 'Species-specific ACRs': ACRs where the sequence is exclusive to a single species. **c,** The percentage of broad and cell-type-specific ACRs underlying three classes shown in panel **b**. The significance test was done by using the Fisher's exact test (alternative = 'two.sided'). **d, Left**, the number and percentage of *O. sativa* shared ACRs that retain or change cell-type specificity amongst the other four species. **Right**, a screenshot of a *O. sativa* phloem specific ACR that retains phloem specificity in *S. bicolor*. This ACR is situated at the promoter region of *LRD3* which is specifically expressed in companion cell and phloem sieve elements (Supplementary Table 2). The gray shaded region highlights the syntenic gene pair. **e,** The percentage of cell-type-specific ACRs identified

across all cell types within each species pair split into three classes shown in panel **b**. The percentage for each cell type within the three classes collectively sum to 100%.



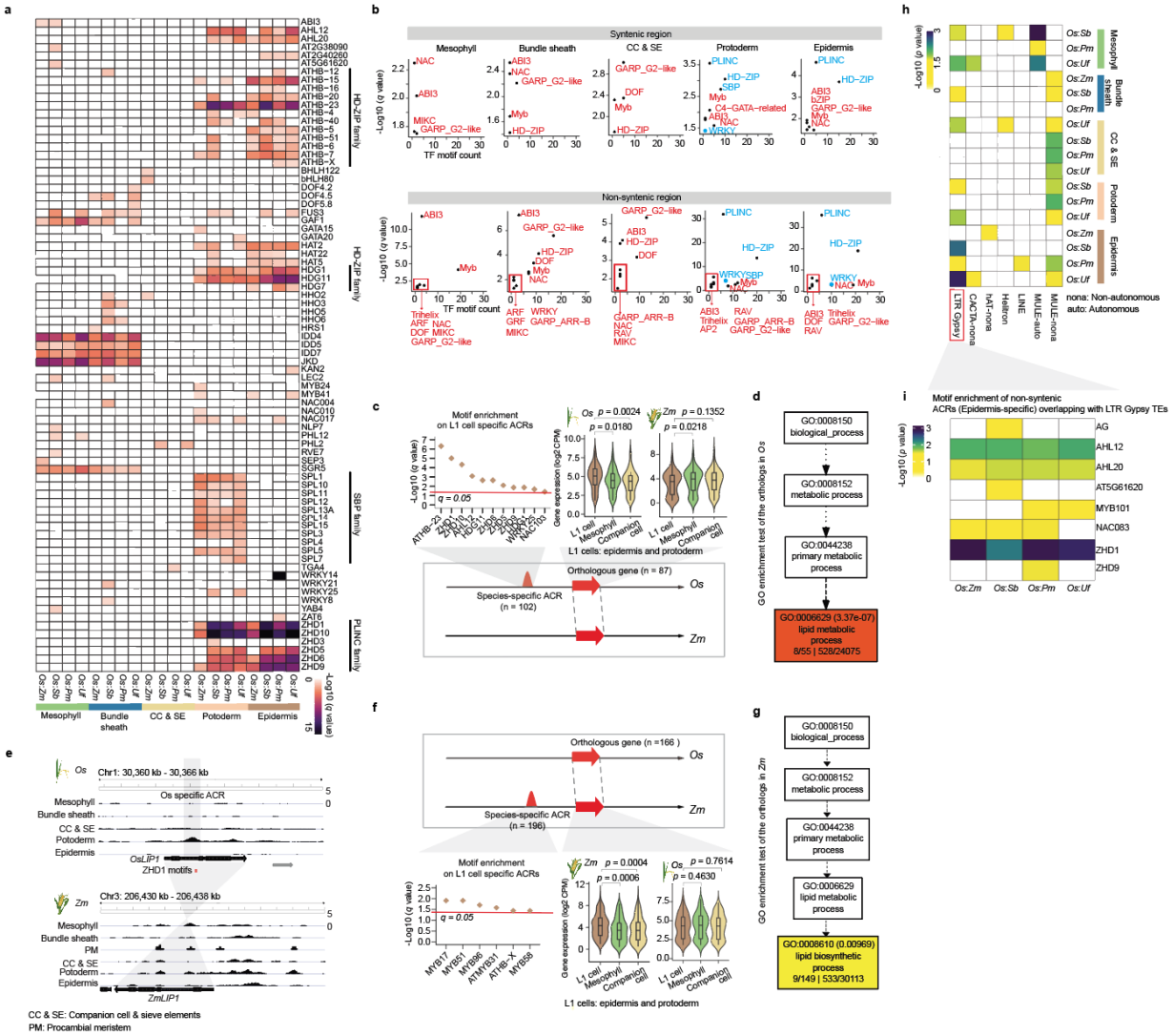
Extended Data Fig. 4. Characterization of ACRs within syntenic regions. a, The bar plot illustrates the percentage of *O. sativa* ACRs within syntenic and non-syntenic regions when compared to another

species. **b,** The pie charts depict percentages of broadly accessible (Broad) and cell-type-specific (CT) ACRs within three classes of ACR conservation in Fig. 3b. The total number of ACR in *O. sativa* is below the pie charts. **c,** The percentage of broad and cell-type-specific ACRs using *Z. mays*, *S. bicolor*, *P. miliaceum*, or *U. fusca* as the baseline underlying three classes shown in Fig. 3b. The x-axis labels, highlighted by a red frame, are the inverse to x-axis labels in Fig. 3c by using *O. sativa* as a baseline. The significance test was done by using the Fisher's exact test (alternative = 'two.sided'). '**' denotes a *p* value < 0.01, and '*' denotes a *p* value < 0.05. **d,** Characterization of ACRs under 'shared ACR' class and their proximity categorizations. **Left,** the number of ACRs under the 'shared ACR' class, categorized into four combinations based on broadly accessible and cell-type-specific ACRs. **Right,** a heatmap of ACR categories based on their proximity to genes: genic ACRs overlapping genes, proximal ACRs (within 2 kb of genes), and distal ACRs (more than 2 kb away from the TSS). **e,** The percentage of shared, variable, and species-specific ACR classes in genic, proximal, and distal manners for each species pair. The percentage for each class within genic, proximal, and distal groups collectively sum to 100%. e.g. Genic, proximal and distal broad shared ACRs sum to 100%. **f,** The pie charts illustrate percentage of cell-type-specific *O. sativa* ACRs syntenic to cell-type-specific ACRs in corresponding species. **g,** The pie charts illustrate mesophyll and bundle sheath specific *O. sativa* ACRs syntenic to cell-type-specific ACRs in corresponding species.



Extended Data Fig. 5. Epidermal cell specific ACRs are less conserved in sequence than other examined cell types. a, The analysis demonstrates an enrichment of cell-type-specific ACRs based on Fisher's exact test. This test assesses whether cell-type-specific ACRs are more likely to be situated in the 'species-specific' class, compared to other cell types. Statistically significant differences (*p* value < 0.05) in all pairwise comparisons are denoted by distinct letters, determined using Fisher's exact test with the alternative set to 'two.sided'. Bars sharing the same letter indicate that they are not significantly (*p* value > 0.05) different from each other. **b,** The proportion of cell-type-specific ACRs identified collectively across all cell types within each species pair under three syntenic classes. The percentage for each cell type within the three classes collectively sum to 100%. **c,** This panel has the same meaning as panel **a**, and corresponds to an enrichment test on panel **b**. *U. fusca*-to-*O. sativa* did not include a bar for protoderm as this cell type has not been identified in *U. fusca* scATAC-seq dataset.

Delving deeper into the L1 TF families, we focused on species-specific L1 ACRs. Within the species-specific syntenic ACRs, we observed a predominance of TF motifs for the HD-ZIP, SQUAMOSA PROMOTER BINDING PROTEIN (SBP), PLINC families (Extended Data Fig. 6a). Many of these, such as HDG1, ZHD1, ATHB-20, SPL3, SPL4, and SPL5, function in epidermal cell development[15,39-42]. The predominance of these motifs in the species-specific ACR class suggests that although L1 TF motif sequences are well conserved (Fig. 2d and e), yet the positions of these motifs are not conserved in grass genomes. Upon comparing TF-motif enrichment in syntenic and non-syntenic ACRs, we observed the presence of these epidermal motif families in both groups (Extended Data Fig. 6b; Supplementary Table 8), indicating their essential roles in both plant epidermal cell development and rapid gene-regulatory co-option in species-specific sequences. Notably, some TF-motif families such as WRKY in epidermal cells, were more enriched in non-syntenic ACRs (Extended Data Fig. 6b), further supporting that this family may drive phenotypic innovation of the epidermal layer.



Extended Data Fig. 6. Identification of TF motifs and TEs associating with epidermal cell divergence. a, TF motif enrichment tests were performed in species-specific ACRs per species pair using *O. sativa* as the baseline (See Methods: Binomial test-based motif enrichment analysis). The TF motif with a *q* value less than 0.05 was indicated by filling them with a white color. **b,** This panel utilized the same method as panel **a** to conduct TF motif enrichment tests in syntenic and non-syntenic ACRs in the *O. sativa*-to-*Z. mays* species pair (The results from other species pairs were shown in Supplementary Table 8). The '-Log10 (*q* value)' of each TF family was calculated by averaging '-Log10 (*q* value)' across all individual TF motifs within this family. TF family names were marked besides each dark dot, while the PLINC, HD-ZIP, SBP, and WRKY family were highlighted using blue color. The x-axis indicates the number of TF motifs detected to be enriched (*q* value < 0.05) in the relative TF family. **c,** TF motif enrichment tests were performed in species-specific ACRs neighboring *O. sativa* ortholog exhibiting higher expression levels in epidermis based on Binomial test (See Methods: Binomial test-based motif enrichment analysis). Significance testing in violin plot was performed using the t-test (alternative = 'two.sided'). **d,** GO enrichment test was performed in *O. sativa* orthologous genes based on agriGO[43]. **e,** A screenshot of *LIP1* accessibility in *O. sativa* and *Z. mays* L1 cells which contains an *O. sativa*

epidermal specific and species-specific ACR with two ZHD1 motif sites. No corresponding ZHD1 motifs were found in *ZmLIP1*. **f-g,** These two plots have the same meanings as panel **d** and **e**, but focusing on species-specific ACRs in *Z. mays*. **h,** Enrichment of TE family on ACRs within non-syntenic regions relative to those in syntenic regions. Significance testing was performed using Fisher's exact test (alternative = 'two.sided'). The TE with a *p* value less than 0.05 was indicated by filling them with a white color. **i,** TF motif enrichment tests were performed on the epidermis specific ACRs overlapping with LTR *Gypsy* TEs based on Binomial test (See Methods: Binomial test-based motif enrichment analysis). The TE with a *p* value less than 0.05 was indicated by filling them with a white color.

To look for derived species-specific ACRs associated with the altered expression of surrounding gene orthologs in epidermal cells, we performed a single-nucleus RNA sequencing (snRNA-seq) analysis in *O. sativa* (Supplementary Fig. 7; Supplementary Table 9). By integrating with a snRNA-seq data from *Z. mays*[20], we identified 87 orthologous genes, irrespective of syntenic regions, which exhibited elevated L1 *O. sativa* expression compared to *Z. mays* (Extended Data Fig. 6c; Supplementary Table 10). A gene ontology enrichment test for these 87 genes revealed eight genes that were significantly enriched in the lipid metabolic process (Extended Data Fig. 6d), likely related to epidermal cuticle metabolism. Among the eight genes, one orthologous gene to *Arabidopsis GDSL LIPASE GENE* (*LIP1*) has been shown to be specifically expressed in the epidermis[44]. We further identified 102 L1 cell-type-specific ACRs from *O. sativa* that were the closest to the 87 orthologous genes and observed 11 TF motifs were enriched (q = 3e-10 to 5e-04; Binomial test) in these ACRs (Extended Data Fig. 6c). These included TF family motifs known for their roles in epidermal cell development such as ZHD1[15], HDG11[45], ZHD5[46], HDG1[27], and WRKY25[47]. For example, within the *OsLIP1* intron, we identified two ZHD1 motifs within a species-specific ACR that was specifically accessible in L1 cells (Extended Data Fig. 6e). We also flipped this comparison by identifying 166 orthologs with elevated maize epidermal expression compared to rice (Supplementary Table 10). This revealed 196 L1 cell-type-specific ACRs in *Z. mays*. Within these ACRs, the most enriched (q = 0.0129 to 0.0392; Binomial test) TF motif was MYELOBLASTOSIS 17 (MYB17; Extended Data Fig. 6f and g). This R2R3 MYB family TF is associated with epidermal cell development, specifically for its involvement in the regulation of epidermal projections (Brockington et al. 2013). Furthermore, we hypothesized that some of these novel motifs could be related to *O. sativa* transposable element (TE) expansion. Notably, we found long terminal repeat retrotransposon (LTR)-associated ACRs from the *Gypsy* family were enriched (p = 0.0012 to 0.0415; Fisher's exact test) in *O. sativa* epidermal cell-type-specific ACRs (Extended Data Fig. 6h). The ZHD1 motif was enriched within these *Gypsy*-associated ACRs (p = 0.0011 to 0.0052; Binomial test) (Extended Data Fig. 6i). By linking snRNA-seq to scATAC-seq data, we tied gene-proximal CRE changes to elevated epidermal expression of conserved orthologs 50 million years derived[48]. These CRE changes drive variance in species-specific L1 layer development, potentially contributing to species differences in environmental adaptation.

CNS are enriched in cell-type-specific ACRs

To augment our syntenic ACR BLASTN approach, we sought to identify CNS through comparative genomics[6,49-51]. Outside of UTRs, CNS are thought to encompass CREs that drive processes too critical to be lost during evolution[52]. To our knowledge, no plant study has overlapped CNS with cell-type-specific chromatin accessibility measurements, leaving it unknown whether CNS retain the same cell-type accessibility in different species. 53,253 *O. sativa* and 284,916 *Z. mays* published CNS overlapped with the leaf ACRs identified in this study[53]. Excluding CNS overlapping with untranslated regions, 30.8% and 21.3% of CNS overlapped with the leaf-derived ACRs in *O. sativa* and *Z. mays*, respectively (Fig. 4a). Using all ACRs in the *O. sativa* atlas, this ratio increased to 65.0% (Fig. 4a), indicating that a significant portion of these CNS likely function in specific cell types and tissues. One common assumption is that CNS that overlap with ACRs (CNS ACRs) retain the exact same function between species. We observed 39% to 51% of total CNS ACRs within the 'shared CNS ACR' class (Extended Data Fig. 7a and b), suggesting these ACRs have conserved functions and conserved cellular contexts between *O. sativa* and *Z. mays.* Within syntenic regions, ACRs with CNS are more cell-type specific than those without (Fig. 4b). This enrichment was consistent for all classes of ACRs we identified, except for species-specific ACRs which were equivalently cell-type specific and broad (Fig. 4b). The enrichment of CNS in cell-type-specific ACRs stress the importance of rare cell-type function, as the cell-type-specific CREs are preferentially retained during evolution.

However, the majority (49-61%) of all CNS ACRs differed in cell-type specificity between *O. sativa* and *Z. mays* (Extended Data Fig. 7a and b). Specifically, we examined the CNS found in *Z. mays* ACRs that did not have a corresponding leaf ACR in *O. sativa*. Leveraging the *O. sativa* atlas, we identified these sequences had divergent cellular or tissue chromatin accessibilities. 249 (75%) of the *Z. mays* leaf variable CNS ACRs were accessible in non-leaf cell states (Fig. 4c-e; Extended Data Fig. 7c), highlighting the instability of the cellular context in which CNS acts. This suggests frequent co-option of CNS CREs into different tissues or cell types. Investigating the CNS ACRs that lost leaf cell-type specificity, we observed that these ACRs were accessible in many non-leaf cell types, uniformly distributed amongst the atlas cell annotations (Extended Data Fig. 7c-g). Consistent with our findings that epidermal-specific ACRs tend to have the most species-specific ACRs in syntenic regions (Fig. 3e; Extended Data Fig. 5a-c), L1 cells showed a significantly lower ratio of non-syntenic CNS ACRs compared to other cell types. This lower ratio demonstrates the frequent loss of L1 CNS, further supporting the rapid evolution of epidermal transcriptional regulation.

Interestingly, we noticed a pattern where some CNS within ACRs also overlapped domains of H3K27me3[6] (Fig. 4f). H3K27me3 is a histone modification associated with facultative heterochromatin established by the POLYCOMB REPRESSIVE COMPLEX 2 (PRC2)[54-56]. Genes silenced by PRC2 and H3K27me3 are important regulators that are only expressed in

narrow developmental stages or under specific environmental stimuli, where they often initiate important transcriptional changes[55,57]. This importance makes the identification of key CREs controlling H3K27me3 silencing especially interesting. Upon closer examination of the ACRs overlapping domains of H3K27me3 to ACRs away from H3K27me3, we observed that H3K27me3 ACRs were significantly enriched for CNS (Fig. 4g). This enrichment supports that some of these CNS underpin conserved, and critical components, of H3K27me3 silencing.

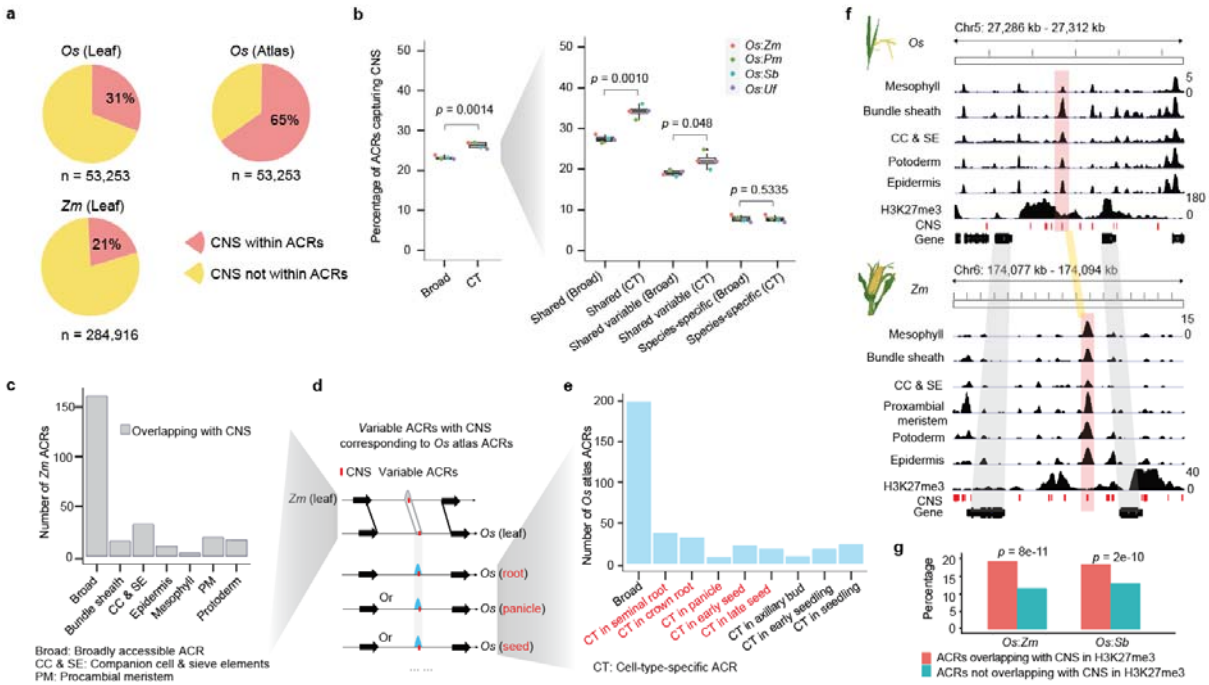
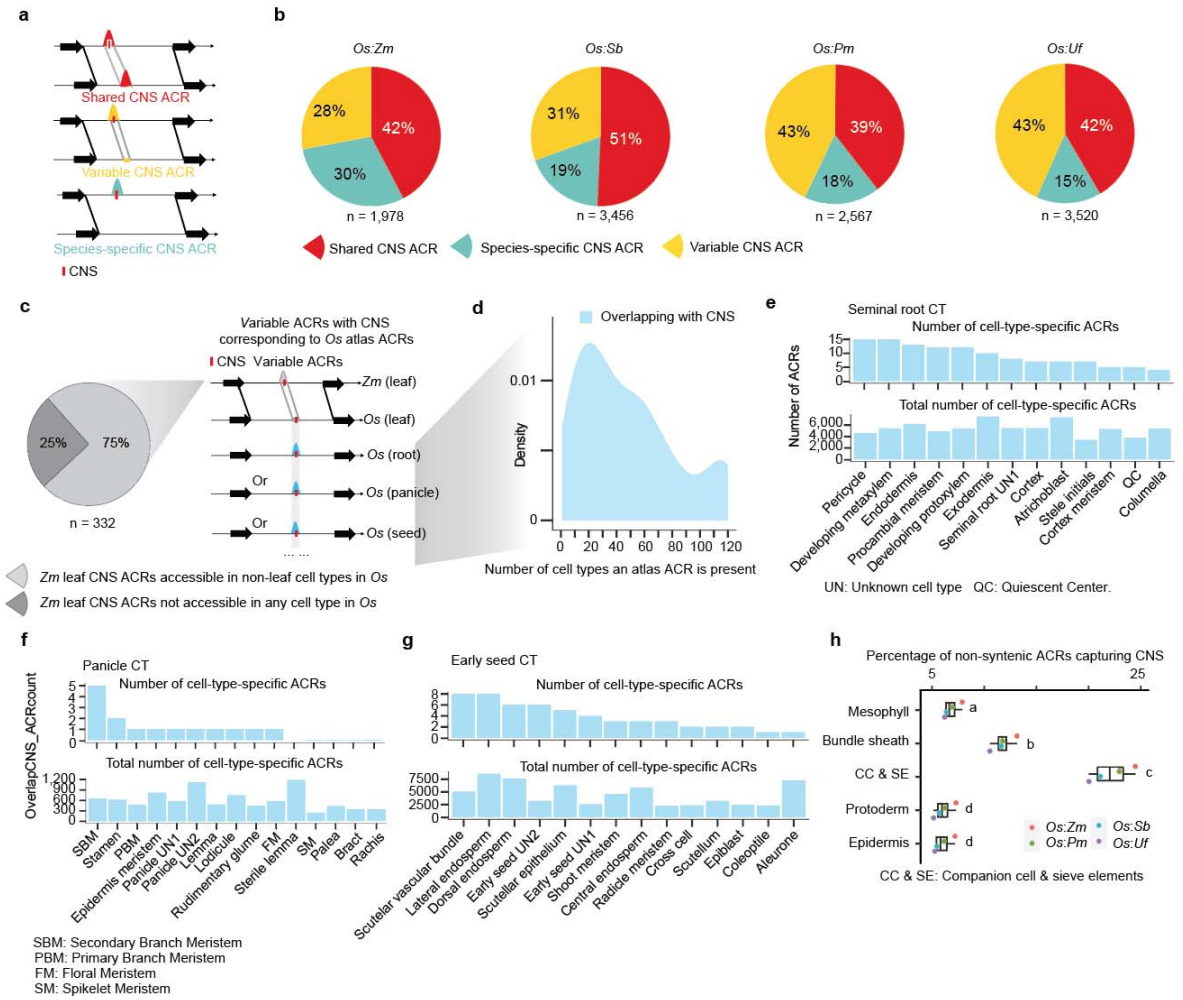


Fig. 4. Cell-type-specific ACRs exhibit an enrichment of CNS. a, The percentage of CNS overlapping with ACRs. 'n' indicates the number of CNS. 'Atlas' means the ACRs were from the *O. sativa* atlas in Fig. 1b. **b, Left,** the percentage of broad and cell-type-specific ACRs within syntenic and non-syntenic regions overlapping with the CNS. **Right,** thispanel presented similar meaning as the left panel but focusing on three classes within syntenic regions shown in Fig. 3b. Broad: Broadly accessible ACR; CT: Cell-type-specific ACR. Significance testing was performed using the t-test (alternative = 'two.sided'). **c,** The bar plot showcases the count of *Z. mays* variable ACRs accessible in leaf cell types. **d,** A sketch illustrating whether variable ACRs containing CNS in *Z. mays* capture ACRs derived from the *O. sativa* atlas. **e,** The bar plot represents the count of *O. sativa* atlas ACRs accessible in non-leaf cell types. **f,** An example of a syntenic block containing *O. sativa*-to-*Z. mays* conserved ACRs within a H3K27me3 region. CNS were highlighted using red color. **g,** The percentage of ACRs capturing CNS in and outside of H3K27me3 regions. The percentage for each group within H3K27me3 and not within H3K27me3 regions collectively sum to 100%. Significance testing was performed using Fisher's exact test (alternative = 'two.sided').



Extended Data Fig. 7. Examination of *O. sativa* atlas ACRs corresponding to variable *Z. mays* ACRs. a, Three classes depicting variations in ACR conservation between two species. 'Shared CNS ACRs': CNS ACRs with matching sequences that are accessible in both species; 'Variable CNS ACRs': CNS ACRs with matching sequences, but are only accessible in one species; 'Species-specific CNS ACRs': CNS ACRs where the sequence is exclusive to a single species. **b,** The pie charts depict percentages of CNS ACRs within three classes of ACR conservation in panel **a**. The total number of ACR in *O. sativa* is below the pie charts. **c,** A sketch illustrating whether variable ACRs containing CNS in *Z. mays* capture ACRs derived from the *O. sativa* atlas. The left pie chart panel represents the percentage of *Z. mays* leaf CNS ACRs that were accessible in non-leaf cell types in the *O. sativa* atlas. **d,** A density plot illustrating the number of non-leaf *O. sativa* cell-types in which an *O. sativa* ACR syntenic to *Z. mays* variable ACRs are accessible. **e-g,** The bar plots represent the count of *O. sativa* atlas cell-type-specific ACRs accessible in non-leaf cell states from seminal root, panicle, and early seed organs. These ACRs were derived from cell-type-specific ACRs from Fig. 4e. **h,** The percentage of non-syntenic ACRs capturing CNS. All pairwise comparisons are statistically significant (p value < 0.05) as indicated by different letters based on t-test (alternative = 'two.sided'). If two bars have the same letter, then they are equivalent (p value > 0.05).

Candidate silencer CREs are enriched in broad ACRs

To assess the stability and change of H3K27me3 related ACRs across grass lineages, we focused on comparing *O. sativa*, *Z. mays*, and *S. bicolor* using previously published H3K27me3 ChIP-seq data[6]. We considered ACRs near or within H3K27me3 regions and classified them into two groups: H3K27me3-broad, representing H3K27me3 associated ACRs with chromatin accessibility in many cell types and H3K27me3-cell-type specific, those H3K27me3 associated ACRs with chromatin accessibility in few cell types (Fig. 5a). The composition of H3K27me3-broad and H3K27me3-cell-type-specific ACRs was consistent across all species (Extended Data Fig. 8a). H3K27me3-broad ACRs exhibited a depletion of H3K27me3 at the ACR, consistent with the absence of nucleosomes in ACRs[58]. In contrast, H3K27me3 depletion was not observed in H3K27me3-cell-type-specific ACRs, with most cells in the bulk H3K27me3 measurement likely containing nucleosomes with H3K27me3 (Fig. 5b). This is consistent with the H3K27me3-cell-type-specific ACRs potentially acting after the removal of facultative heterochromatin in a specific cell type(s). However, the chromatin accessibility of the H3K27me3-broad ACRs appears to be concurrent with H3K27me3, suggesting these ACRs may regulate H3K27me3 maintenance and removal across the majority of cellular contexts.

To assess the transcriptional state of genes near the H3K27me3-broad ACRs, we evaluated snRNA-seq/single-cell RNA sequencing (scRNA-seq) from *O. sativa* seedling (Supplementary Fig. 7) and root[59], and *Z. mays* seedling. The results revealed significantly lower expression (p = 3e-34 to 4e-06; Wilcoxon test) for H3K27me3-broad ACRs associated genes across many cell types (Fig. 5c; Extended Data Fig. 8b and c; Supplementary Table 11). Moreover, 58 bulk RNA-seq libraries from *O. sativa* organs[60], demonstrated that the expression of genes near H3K27me3-broad ACRs were significantly lower (p = 2e-07 to 0.0265; Wilcoxon test) than genes near H3K27me3-absent broad ACRs (Extended Data Fig. 8d). To dissect the roles of H3K27me3-broad ACRs, we identified 2,164 H3K27me3-broad ACRs and measured neighboring gene expression in *O. sativa* cells (Supplementary Table 12). About 926 (~42.8%) of the H3K27me3-broad ACRs were associated with 838 genes that exhibited no expression across any sampled cell type (Extended Data Fig. 8e; Supplementary Table 12), consistent with these H3K27me3 proximal genes only being expressed under specific conditions. The 1,108 expressed genes associated with H3K27me3-broad ACRs were enriched (p < 2e-16; Fisher's exact test) for cell-type specificity compared to genes without H3K27me3 (Extended Data Fig. 8f). In summary, single-cell expression analysis revealed that the genes linked to H3K27me3-broad ACRs exhibited the hallmarks of facultative gene silencing.

We hypothesized that the H3K27me3-broad ACRs would be enriched for PRC2 silencer elements, as their consistent chromatin accessibility provides an avenue to recruit PRC2 to maintain H3K27me3 throughout development. Supporting the presence of silencer CREs with these ACRs, a known silencer CRE ~5.3 kb upstream of *FRIZZY PANICLE* was within a H3K27me3-broad ACR[61] (Extended Data Fig. 8g). To exploit the known Polycomb *A. thaliana*

targets, we used scATAC-seq[20] and H3K27me3[62] data from *A. thaliana* roots and annotated H3K27me3-broad ACRs. The *A. thaliana* H3K27me3-broad ACRs significantly (p < 2e-16; Binomial test) captured 53 of the 170 known Polycomb responsive elements compared to a control class of ACRs, supporting their putative silencer function[55] (Extended Data Fig. 8h). Furthermore, we implemented a *de novo* motif analysis on the 170 *A. thaliana* elements and identified all reported Polycomb response element (PRE) motifs (CTCC, CCG, G-box, GA-repeat, AC-rich, and Telobox)[55] (Fig. 5d). Using these motifs and our chromatin accessibility data, we predicted putative binding sites in *O. sativa*, *Z. mays*, and *S. bicolor*, and observed that five motifs were significantly (p = 2e-178 to 1e-05; Binomial test) enriched in the H3K27me3-broad ACRs compared to a genomic control (Fig. 5d). About 88.0% to 92.7% of H3K27me3-broad ACRs contained at least one PRE motif and 0.1% to 0.2% encompassing all the six of the major PRE classes (Extended Data Fig. 8i). EMF2b is an essential component of the PRC2 complex[55,63,64]. About 63.7% (4,053/6,364) of the H3K27me3-broad ACRs significantly (p < 2e-16; Binomial test) overlapped peaks from ChIP-seq targeting EMF2b compared to a control class of ACRs (Fig. 5e). EMF2b occupancy was enriched at H3K27me3-broad ACRs (Fig. 5f), supporting H3K27me3-broad ACRs acting as PRC2 recruiting silencers. Beyond PRE motifs, we observed significant enrichment (p = 1e-16 to 0.0376; Hypergeometric test) of four TF family motifs in H3K27me3-broad ACRs: APETALA2-like (AP2), basic Helix-Loop-Helix (bHLH), Basic Leucine Zipper (bZIP), and C2H2 zinc-finger (ZnF) (Extended Data Fig. 8j; Supplementary Table 13). AP2 and C2H2 are known to recruit PRC2[55]. In general, the motif discovery within the H3K27me3-broad ACRs supports them harboring H3K27me3 silencer CREs.

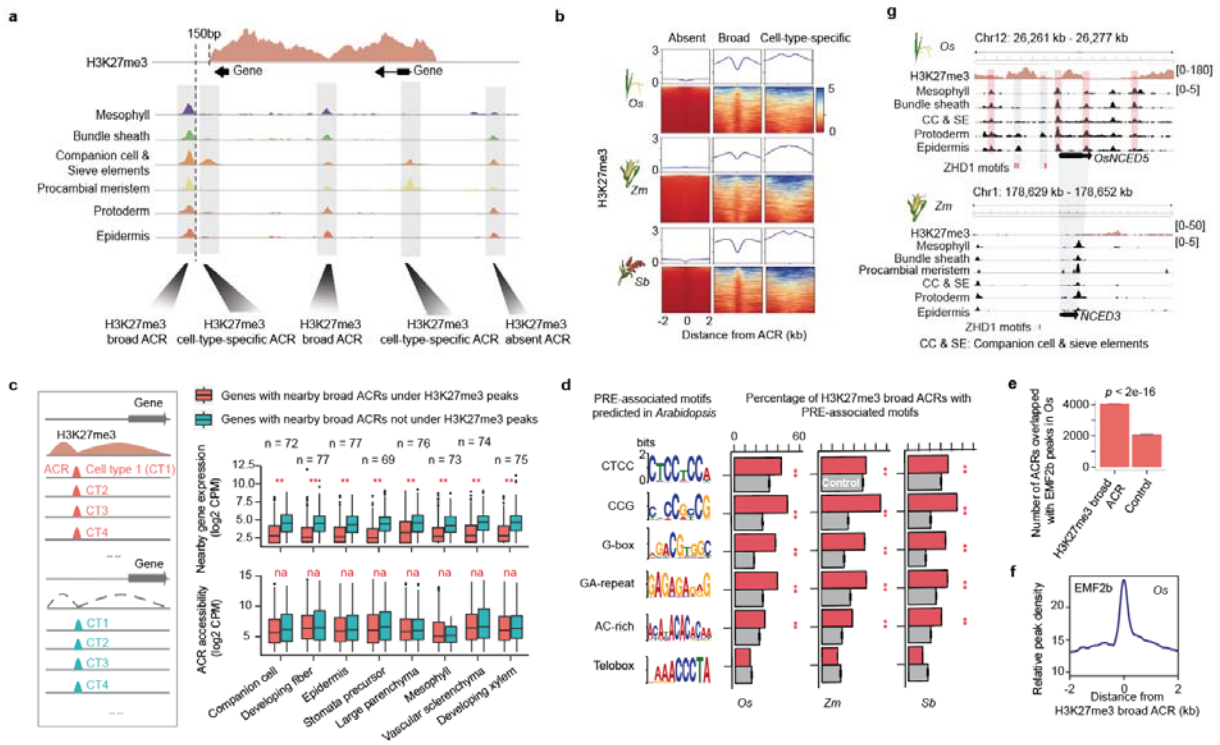
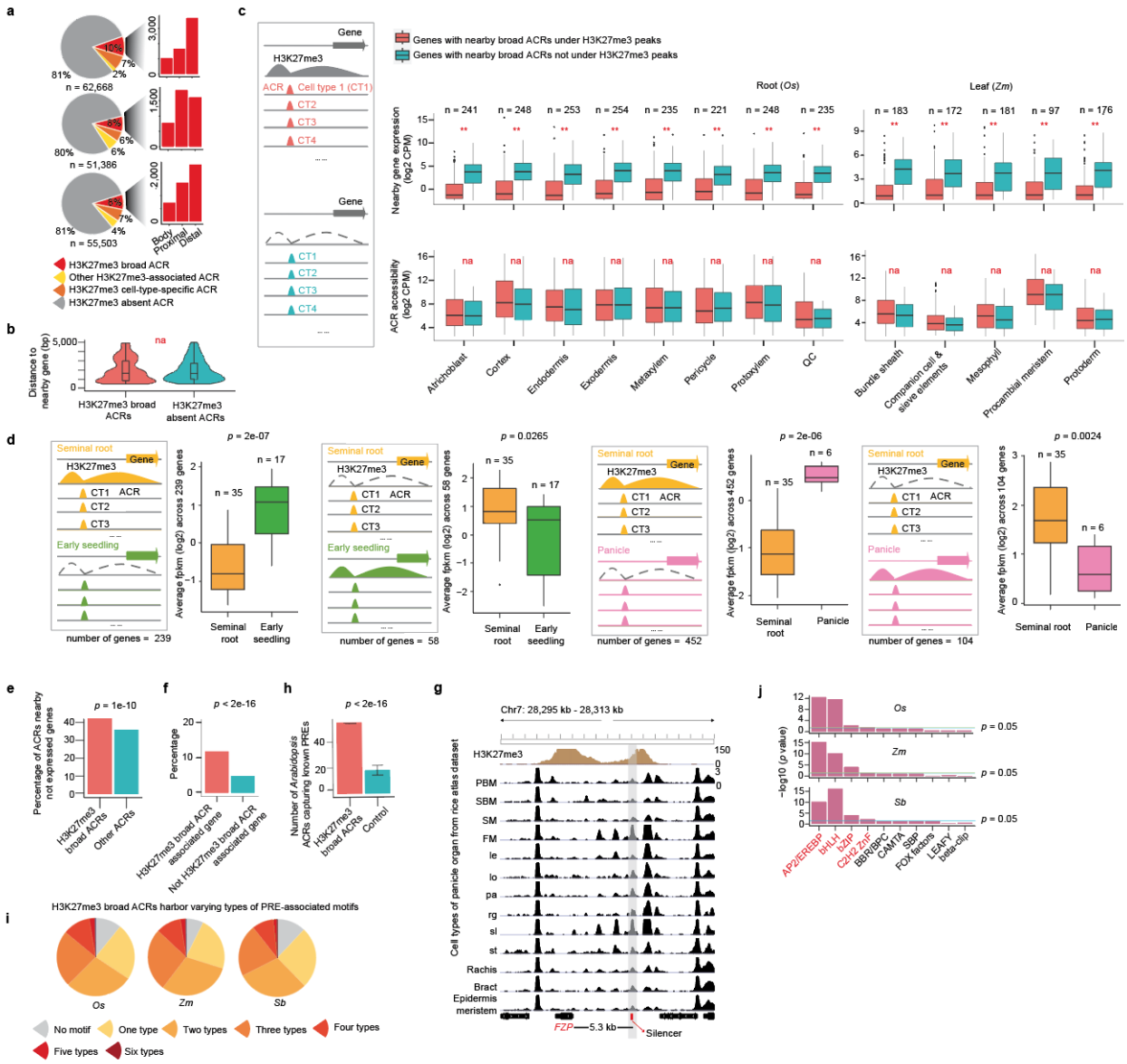


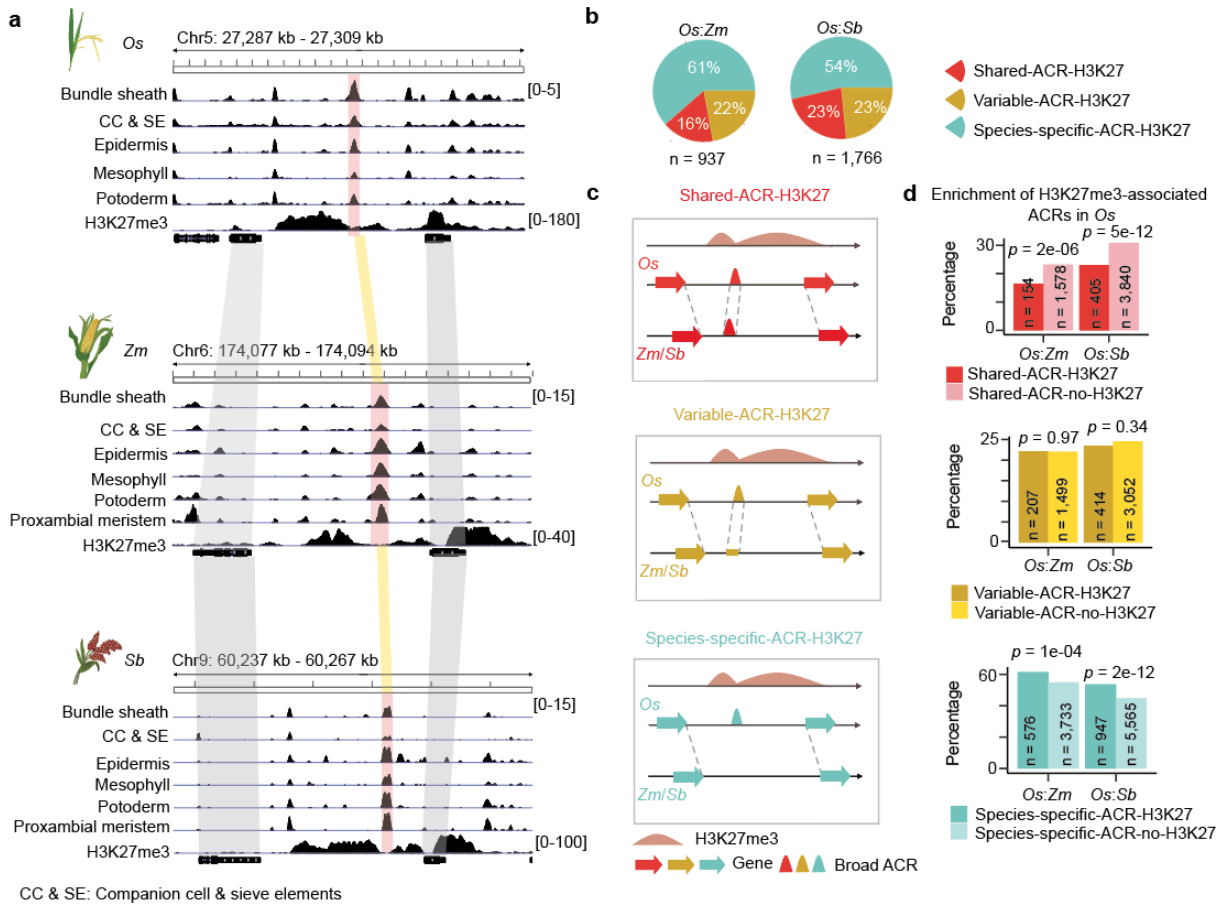
Fig. 5. Discovery of candidate silencer CREs across species. a, A sketch graph illustrates the classification of ACRs based on their proximity to H3K27me3 peaks. We classified the ACRs into two groups: H3K27me3-associated ACRs (found within or surrounding H3K27me3 peaks) and H3K27me3-absent ACRs. The H3K27me3-associated ACR were further divided into broad ACRs, characterized by chromatin accessibility in at least five cell types, and cell-type-specific ACRs, accessible in less than three out of six examined cell types across all the species. **b,** Alignment of H3K27me3 chromatin attributes at summits of distinct ACR groups. **c,** A comparative analysis of expression levels and chromatin accessibility of genes surrounding broad ACRs under and outside of H3K27me3 peaks. ** indicate p value < 0.01, which was performed by the Wilcoxon test (alternative = 'two.sided'). The broad ACRs where the H3K27me3 region overlapped >50% of the gene body were positioned within 500 to 5,000 bp upstream of the transcriptional start site of their nearest gene. **d,** Percentage of H3K27me3-broad ACRs in *O. sativa*, *Z. mays*, and *S. bicolor* capturing six known motifs enriched in PREs in *A. thaliana*. ** indicate p value < 0.01, which was performed by the Binomial test (alternative = 'two.sided'; See Methods: Construction of control sets for enrichment tests). **e**, The number of H3K27me3-broad-ACR capturing EMF2b ChIP-seq peak. The significance test was done by using the Binomial test (See Methods: Construction of control sets for enrichment tests; alternative = 'two.sided'). **f**, A metaplot of EMF2b ChIP-seq profile at summits of H3K27me3-broad ACRs identified from *O. sativa* leaf. **g,** a screenshot of *OsNCED5* accessibility in *O. sativa* and *NCED3* accessibility in *Z. mays* L1 cells which contains four H3K27me3-broad ACRs and two *O. sativa* epidermal specific and species-specific ACRs with three ZHD1 motif sites.



Extended Data Fig. 8. Characterization of H3K27me3-broad ACRs potentially enriched for silencer CREs. a, Pie charts showcasing the composition of different categories of ACRs, and bar plots displaying the number of ACRs within the H3K27me3-broad-ACR group. **b,** No significant difference in the distance to nearby genes was identified between H3K27me3-broad ACRs to H3K27me3-absent ACRs as shown in Fig. 5c. **c,** Comparative analysis of expression levels and chromatin accessibility of genes surrounding broad ACRs under and outside of H3K27me3 peaks. ** indicate *p* value < 0.01, which was performed by the Wilcoxon test (alternative = 'two.sided'). **d,** Four categories of gene sets overlapped with H3K27me3 peaks and comparisons of their expression between bulk RNA-seq data libraries from two different organs. The 'n' above the box indicates the number of libraries used in this comparison. Significance testing was performed using the Wilcoxon test (alternative = 'two.sided'). **e**, Percentage of H3K27me3-broad ACRs nearby genes exhibiting no expression across any cell type was significantly higher than the other ACRs associated with the not expressed genes. Significance testing was performed using Fisher's exact test (alternative = 'two.sided'). **f,** Percentage of H3K27me3-broad ACRs nearby

genes exhibiting cell-type-specific expression was significantly higher than the other genes not associating with H3K27me3-broad ACRs. Significance testing was performed using Fisher's exact test (alternative = 'two.sided'). **g,** A screenshot illustrates a H3K27me3-broad ACR harboring a reported silencer within an H3K27me3 peak in the panicle organ, located approximately 5.3 kb upstream of *FZP* gene. SBM: Secondary Branch Meristem. PBM: Primary Branch Meristem. SM: Spikelet Meristem. FM: Floral Meristem. le: Lemma. lo: Lodicule. pa: Palea. rg: Rudimentary glume. sl: Sterile lemma. **h,** The number of H3K27me3-broad-ACR capturing known 53 PREs is determined based on *A. thaliana* scATAC-seq data. The significance test was done by using the binomial test (See Methods: Construction of control sets for enrichment tests). **i,** Pie charts illustrate the comparison of different types of PRE-associated motifs harbored in H3K27me3-broad ACRs. **j**, Four TF families, highlighted in red, were significantly enriched in H3K27me3-broad ACRs. The motif data were collected from 568 TFs from *A. thaliana* belonging to 24 families within the JASPAR database[65]. The *p* value was computed using a hypergeometric test.

Mirroring our previous syntenic ACR analysis, we split the *O. sativa* H3K27me3-broad ACRs into three classes, shared, variable and species specific, to investigate the relationship between these candidate silencers and species divergence (Extended Data Fig. 9a-c; Supplementary Table 14). Between 54% to 61% of the H3K27me3-broad ACRs were present in the species-specific class, and the H3K27me3-broad ACRs were more likely to be species-specific compared to broad ACRs without H3K27me3 (Extended Data Fig. 9d). Since these H3K27me3-broad ACRs exhibit hallmarks of PRC2 recruitment, we suspect that altered silencer CREs use context to drive species-specific developmental and environmental responses.



Extended Data Fig. 9. Evolutionary dynamics of H3K27me3-broad ACRs across different species. **a,** An example of a syntenic region containing H3K27me3-broad ACRs that were conserved across *O. sativa*, *Z. mays*, and *S. bicolor*. **b,** The pie charts depict percentages of H3K27me3-broad ACRs within three classes of ACR conservation in Fig. 3b. **c,** Three categories of *O. sativa* H3K27me3-broad ACRs that are syntenic to regions from another species. **d,** Comparison of ACR percentage underlying three classes corresponding to panel **c** associated with H3K27me3 or not. The *p* values displayed in the bottom bar plot were computed using Fisher's exact test (alternative = 'two.sided').

Since we identified L1 cells as being enriched in species-specific ACRs, we sought to examine the changes in H3K27me3 regulation within this tissue. We examined our previously identified 87 rice-to-maize orthologs to see if these genes contained H3K27me3. We observed 18 of these 87 genes were close to H3K27me3-broad ACRs (Supplementary Table 15). For example, we identified four H3K27me3-broad ACRs, and three ZHD1 motifs within two species-specific ACRs surrounding *OsNCED5* that were specifically accessible in L1 cells (Fig. 5g). OsNCED5 is known to regulate tolerance to water stress and regulate leaf senescence in *O. sativa*[66]. These results highlight that H3K27me3 mediated silencing may play a critical role in divergent regulation in L1 layer development.

Discussion

Our comparison of *O. sativa* with four other grasses revealed patterns in the evolutionary dynamics of *O. sativa* ACRs within syntenic and non-syntenic regions and discovered that the grass L1 layer exhibits elevated rates of transcriptional regulatory divergence compared to other tissue/cell types. *O. sativa* had an increased prevalence of the ZHD1 TF motif, partly driven by past LTR TE expansion, that contributed to elevated ortholog L1 expression compared to *Z. mays* (Fig. 6a). This was driven by a contrasting dichotomy; the epidermal TF motifs were the most cell-type specific of all those studied, yet their cognate ACRs exhibited the strongest target divergence amongst the measured species. This duality highlights tandem conservation of core epidermal CRE and motif targets, and the rapid co-option of novel regulatory regions into these existing regulatory frameworks. This rapid regulatory evolution might relate to the dynamic environmental pressures the epidermis has evolved to withstand[13]. Although to a lesser extent than the epidermis, this interesting contrast, where the cell-type restricted TF motifs are conserved and the cell-type-specific chromatin accessibility of cognate CREs are not, extends to other cell types. This supports a larger pattern of novel CRE target evolution that co-opts established cell-type-specific TF networks.

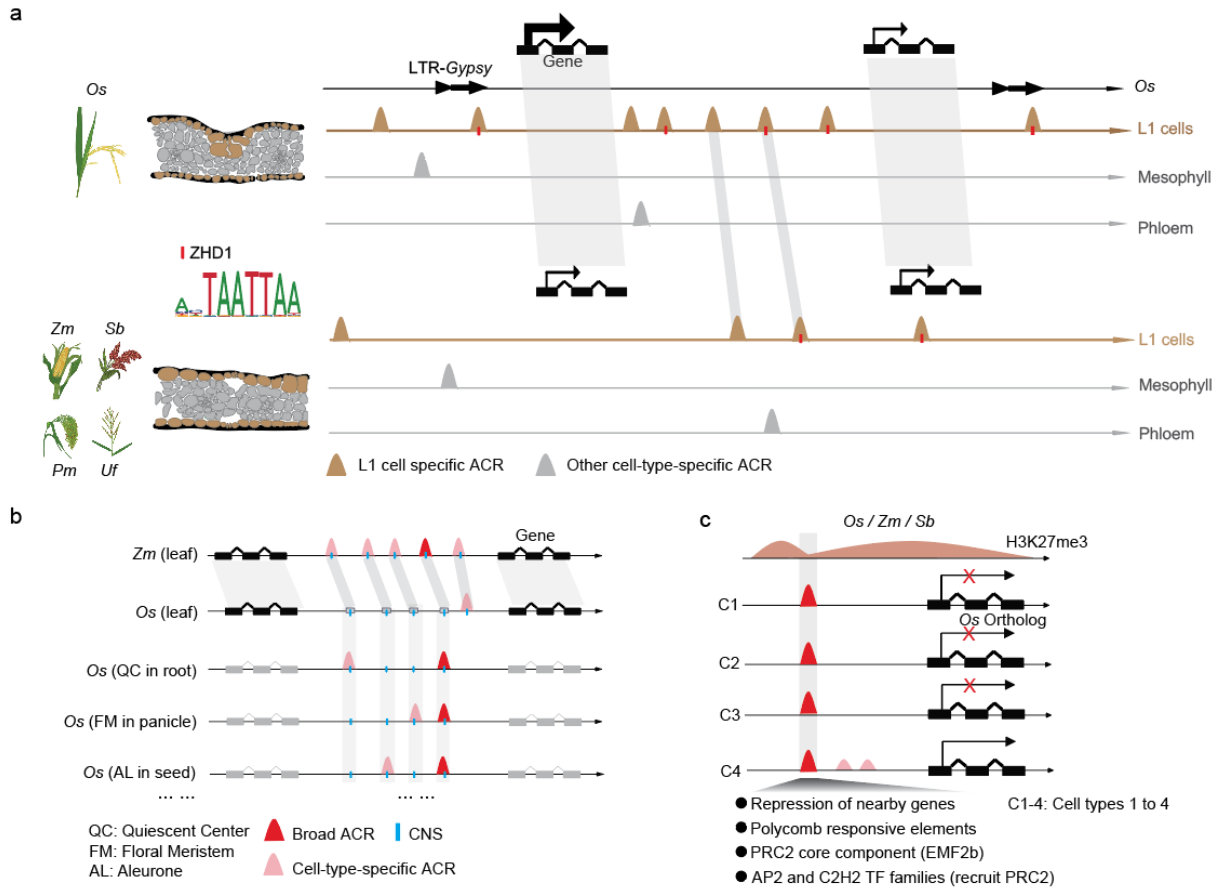


Fig. 6. Evolution of cell-type-specific ACRs and CREs. **a,** The analysis of leaf cell types across these species revealed an enrichment of cell-type-specific ACRs in species-specific regions. Notably, these species-specific ACRs were enriched within L1 cells compared to all others examined. Additionally, it was observed that epidermal cell-specific ACRs significantly overlapped with LTR-*Gypsy* TEs, which

were enriched for the ZHD1 motif known to regulate leaf curling. The L1 ACRs within the ZHD1 motif likely contribute to species-specific elevated expression of genes involved in L1 cell development. **b,** We found an enrichment of CNS in cell-type-specific ACRs. Although some CNS ACRs retained the same cell-type specificity between *O. sativa* and *Z. mays*, these CNS ACRs often switched tissue or cell-type accessibility between grass species. **c,** Despite being within facultative heterochromatin, H3K27me3-broad ACRs are accessible in many cell types, providing a physical entry point for PRC2 to bind. Several lines of evidence support that the H3K27me3-broad ACRs contain silencer CREs. Specifically, these ACRs are linked to transcriptionally silent genes, are enriched for PRE motifs, enriched for TF family motifs (AP2 and C2H2) reported to recruit PRC2, and enriched for PRC2 subunit (EMF2b) ChIP-seq peaks.

Highlighting the rapid rate of regulatory evolution, ACRs, and the CREs within them, underpin phenotypic variation within plant and mammalian species[10,20,67,68]. Despite the link between CREs and phenotypic variation, how these transcriptional regulatory circuits have changed during species divergence is challenging to address. This is partly due to rapid CRE changes occluding pairwise comparison; even closely related plant species share but a fraction of their ACR/CRE complements[6,69]. We used a comparative single-cell epigenomics approach to characterize the evolution of cell-type-specific ACRs and CREs in grasses. We demonstrate that grass cell-type-specific ACRs have changed tremendously over 50 million years[48], with relatively few (0.2-0.4%) cell-type-specific ACRs remaining conserved across the examined plant species. This contrasts with mammals, where ~17% (66,781/384,412) of cell-type-specific ACRs remain conserved over 87 million years[70]. The huge difference in ACR conservation highlights the speed at which plant CRE evolution takes place compared to mammals. The repeated whole genome duplications in plant lineages[71,72], and the functional redundancy they provide, may be the fuel for rapid CRE divergence driving plants' adaptation to diverse environments[73].

Integration of the *O. sativa* atlas with CNS revealed ~65% were accessible in at least one cell type. We expect that most CNS not captured by an ACR in our study are likely accessible in an unsampled cell type or developmental condition. This stresses the need for expanded accessible chromatin atlases using more tissues, segments of development, and environment. Our results also demonstrated that CNS ACRs switch cell-type specificity between species (Fig. 6b). This shows that although CNS can be conserved function, this function can sometimes be altered in different cellular contexts across lineages[74]. However, it remains possible that the main CNS function conserved between *O. sativa* and *Z. mays* occurs outside of leaves; i.e. these switching CNS may underpin conserved functions in conserved non-leaf contexts. Nonetheless, the switching of CNS leaf cell-type accessibility highlights the importance of merging chromatin accessibility data with CNS datasets, as the assumption of conserved CNS sequence equaling conserved CRE function is often invalid.

Much focus has been placed on enhancer CREs within ACRs, yet silencers are equally important, as they repress gene expression until the proper developmental or environmental cues. Our prior research uncovered that some ACRs flanking H3K27me3 are linked to the suppression of nearby genes[6,9]; however, the question remains whether these ACRs function as silencers. Our *O. sativa* cell-type ACR atlas allowed the identification of ACRs within H3K27me3 regions that were accessible in most cell types. Several lines of evidence support these H3K27me3-broad

ACRs as silencers of linked, transcriptionally repressed genes. Specifically, these putative silencers were enriched for CNS, PREs and related TF motifs, and PRC2 *in vivo* occupancy[64] (Fig. 6c). Future gene editing of these putative silencers will reveal more about their role in PRC2 recruitment and gene silencing. Similar H3K27me3-broad ACR putative silencers were also present in the epigenomic landscape of *O. sativa*, *Z. mays*, and *S. bicolor*, suggesting this is a conserved feature of grass genomes. H3K27me3 silencing is deeply conserved in eukaryotes[75], and we hypothesize that other single-cell comparative genomic investigations will find this pattern of broadly accessible silencers in all species with H3K27me3.

Our rice atlas of cell-type-specific ACRs and this cross-species analysis provides a useful resource to enhance our understanding of regulatory evolution more broadly. This resource, and these observations, will fuel research into identifying key CREs controlling specific genes by demarcating high-confident targets for genome editing.

Methods

Preparation of plant materials

Early seedlings, specifically seedling tissues above ground, were collected 7 and 14 days after sowing. Flag leaf tissue was harvested at the V4 stage, characterized by collar formation on leaf 4 of the main stem. Axillary buds were obtained from rice plants grown in the greenhouse at approximately the V8 stage. Rice seminal and crown root tips (bottom 2 cm) were gathered at the same stage as seedling tissues, 14 days after sowing. Panicle tissue was acquired from rice plants grown in the greenhouse. Inflorescence primordia (2-5 mm) were extracted from shoots harvested at the R1 growth stage, where panicle branches had formed. Early seeds were harvested at approximately six days after pollination (DAP), and late seeds at approximately ten DAP. All tissues were collected between 8 and 9 am, and all fresh materials were promptly utilized for library construction starting at 10 am.

Single-cell ATAC-seq library preparation

Nuclei isolation and purification were performed as described previously[76]. In brief, the tissue was finely chopped on ice for approximately 2 minutes using 600 μL of pre-chilled Nuclei Isolation Buffer (NIB, 10 mM MES-KOH at pH 5.4, 10 mM NaCl, 250 mM sucrose, 0.1 mM spermine, 0.5 mM spermidine, 1 mM DTT, 1% BSA, and 0.5% TritonX-100). After chopping, the entire mixture was passed through a 40-μm cell strainer and then subjected to centrifugation at 500 rcf for 5 minutes at 4°C. The supernatant was carefully decanted, and the pellet was reconstituted in 500 μL of NIB wash buffer, which consisted of 10 mM MES-KOH at pH 5.4, 10 mM NaCl, 250 mM sucrose, 0.1 mM spermine, 0.5 mM spermidine, 1 mM DTT, and 1% BSA. The sample was filtered again, this time through a 10-μm cell strainer, and then gently layered onto the surface of 1 mL of a 35% Percoll buffer, prepared by mixing 35% Percoll with 65% NIB wash buffer, in a 1.5-mL centrifuge tube. The nuclei were subjected to centrifugation at 500 rcf for 10 minutes at 4°C. Following centrifugation, the supernatant was carefully removed, and the pellets were resuspended in 10 μL of diluted nuclei buffer (DNB, 10X Genomics Cat# 2000207). About 5 μL of nuclei were diluted 10 times and stained with DAPI (Sigma Cat. D9542) and then the nuclei quality and density were evaluated with a hemocytometer under an

epifluorescence microscope. The original nuclei were diluted with a DNB buffer to a final concentration of 3,200 nuclei per uL. Finally, 5 uL of nuclei (16,000 nuclei in total) were used as input for scATAC-seq library preparation. scATAC-seq libraries were prepared using the Chromium scATAC v1.1 (Next GEM) kit from 10X Genomics (Cat# 1000175), following the manufacturer's instructions. (10xGenomics, CG000209_Chromium_NextGEM_SingleCell_ATAC_ReagentKits_v1.1_UserGuide_RevE). Libraries were sequenced with Illumina NovaSeq 6000 in dual-index mode with eight and 16 cycles for i7 and i5 index, respectively.

Single-nuclei RNA-seq library preparation and data analysis

The protocol for nuclei isolation and purification was adapted from the previously described scATAC-seq method. In summary, to minimize RNA degradation and leakage, the tissue was finely chopped on ice for approximately 1 minute using 600 μL of pre-chilled Nuclei Isolation Buffer containing 0.4U/μL RNase inhibitor (Roche, Protector RNase Inhibitor, Cat. RNAINH-RO) and a comparatively low detergent concentration of 0.1% NP-40. Following chopping, the entire mixture was passed through a 40-μm cell strainer and then subjected to centrifugation at 500 rcf for 5 minutes at 4°C. The supernatant was carefully decanted, and the pellet was reconstituted in 500 μL of NIB wash buffer, comprising 10 mM MES-KOH at pH 5.4, 10 mM NaCl, 250 mM sucrose, 0.5% BSA, and 0.2U/μL RNase inhibitor. The sample was filtered again, this time through a 10-μm cell strainer, and gently layered onto the surface of 1 mL of a 35% Percoll buffer. The Percoll buffer was prepared by mixing 35% Percoll with 65% NIB wash buffer in a 1.5-mL centrifuge tube. The nuclei were then subjected to centrifugation at 500 rcf for 10 minutes at 4°C. After centrifugation, the supernatant was carefully removed, and the pellets were resuspended in 50 μL of NIB wash buffer. Approximately 5 μL of nuclei were diluted tenfold and stained with DAPI (Sigma Cat. D9542). Subsequently, the nuclei's quality and density were evaluated with a hemocytometer under a microscope. The original nuclei were further diluted with a DNB buffer to achieve a final concentration of 1,000 nuclei per μL. Ultimately, a total of 16,000 nuclei were used as input for snRNA-seq library preparation. For snRNA-seq library preparation, we employed the Chromium Next GEM Single Cell 3'GEM Kit v3.1 from 10X Genomics (Cat# PN-1000123), following the manufacturer's instructions (10xGenomics, CG000315_ChromiumNextGEMSingleCell3-_GeneExpression_v3.1_DualIndex_RevB). The libraries were subsequently sequenced using the Illumina NovaSeq 6000 in dual-index mode with 10 cycles for the i7 and i5 indices, respectively.

The raw BCL files obtained after sequencing were demultiplexed and converted into FASTQ format using the default settings of the 10X Genomics tool cellranger mkfastq[77] (v7.0.0). The raw reads were processed with cellranger count[77] (v7.0.0) using the Japonica rice reference genome[57] (v7.0). Genes were kept if they were expressed in more than three cells and each cell having a gene expression level of at least 1,000 but no more than 10,000 expressed genes. Cells with over 5% mitochondria or chloroplast counts were filtered out. The expression matrix was normalized to mitigate batch effects based on global-scaling normalization and multicanonical correlation analysis in Seurat[78] (v4.0). The Scrublet tool[79] was employed to predict doublet cells in this dataset. SCTransform in Seurat was used to normalize the data and identify variable genes.The nearest neighbors were computed using FindNeighbors using 30 PCA dimensions. The clusters were identified using FindClusters with a resolution of 1. The cell types were annotated based on the marker gene list (Supplementary Table 2). To identify genes exhibiting

higher expression in a particular cell type than in the others, we utilized the 'cpm' function from edgeR[80] (v3.38.1), for normalizing the expression matrix. Genes within a specific cell type that displayed more than 1.5-fold change in log2 Counts Per Million (CPM) values compared to the average log2 CPM across all cell types were determined as specifically expressed genes in that particular cell type.

Slide-seq library preparation and data analysis

Root tissues from rice seedlings 14 days after sowing were used for the Slide-seq V2 spatial transcriptomics. The tissues were embedded in the Optimal Cutting Temperature (OCT) compound, snap-frozen in a cold 2-methylbutane bath, and cryosectioned into 10 um thick slices. The spatial transcriptome library was constructed following a published method[81,82]. In brief, the tissue slices were placed on the Slide-seq V2 puck and underwent the RNA hybridization and reverse transcription process. After tissue clearing and spatial bead collections, the cDNA was synthesized and amplified for a total of 14 cycles. The library was constructed using Nextera XT Library Prep Kit (Illumina, USA) following the manufacturer's instructions.

The reads alignment and quantification were conducted following the Slide-seq pipeline (https://github.com/MacoskoLab/slideseq-tools). The data processing is similar to the procedures applied in the analysis of snRNA-seq but setting a resolution of 0.7 for the FindClusters function in Seurat[78] (v4.0). The cell types were annotated based on the histology of cross-sectioned roots. The marker genes of each cluster were identified using the Wilcoxon Rank Sum test in FindAllMarkers.

RNA *in situ* Hybridization

The rice samples were put into the vacuum tissue processor (HistoCore PEARL, Leica) to fix, dehydrate, clear, and embed, and were subsequently embedded in paraffin (Paraplast Plus, Leica). The samples were sliced into 8 μm sections with a microtome (Leica RM2265). The cDNAs of the genes were amplified with their specific primer pairs in situ_F/in situ_R and subcloned into the pGEM-T vector (Supplementary Table 16). The pGEM-gene vectors were used as the template to generate sense and antisense RNA probes. Digoxigenin-labelled RNA probes were prepared using a DIG Northern Starter Kit (Roche) according to the manufacturer's instructions. Slides were observed under bright fields through a microscope (ZEISS) and photographed with an Axiocam 512 color charge-coupled device (CCD) camera.

Raw reads processing of scATAC-seq

Data processing was executed independently for each tissue and/or replicate. Initially, raw BCL files were demultiplexed and converted into FASTQ format, utilizing the default settings of the 10X Genomics tool cellranger-atac make-fastq[83] (v1.2.0). Employing the Japonica rice reference genome[84] (v7.0), the raw reads underwent processing using cellranger-atac count[83] (v1.2.0). These steps encompassed adaptor/quality trimming, mapping, and barcode attachment/correction. Subsequently to the initial processing, reads that were uniquely mapped with mapping quality > 10 and correctly paired were subjected to further refinement through SAMtools view (v1.7; -f 3 -q 10; Li et al. 2009). To mitigate the impact of polymerase chain reaction (PCR) duplicates, the Picard tool MarkDuplicates[85] (v2.16.0) was applied on a per-nucleus basis. To elevate data quality, a blacklist of regions was devised to exclude potentially spurious reads. The methodology involved the exclusion of regions displaying bias in Tn5

integration from Tn5-treated genomic DNA. Specifically, regions characterized by 1-kilobase windows with coverage exceeding four times the genome-wide median were eliminated. We further leveraged ChIP-seq input data[6], to filter out collapsed sequences in the reference using the same criteria. This blacklist also incorporated sequences of low-complexity and homopolymeric sequences through RepeatMasker[86] (v4.1.2). Moreover, nuclear sequences exhibiting homology surpassing 80% to mitochondrial and chloroplast genomes[87] (BLAST+; v2.11.0) were also included within the blacklist. Furthermore, BAM alignments were converted into BED format, wherein the coordinates of reads mapping to positive and negative strands were subjected to a shift by +4 and -5, respectively. The unique Tn5 integration sites per barcode were finally retained for subsequent analyses.

Identifying high-quality nuclei

To ensure the acquisition of high-quality nuclei, we harnessed the capabilities of the Socrates package for streamlined processing[20]. To gauge the fraction of reads within peaks, we employed MACS2[88] (v2.2.7.1) with specific parameters (genomesize = 3e8, shift= -75, extsize = 150, fdr = 0.05) on the bulk Tn5 integration sites. Subsequently, we quantified the number of integration sites per barcode using the callACRs function. Next, we estimated the proximity of Tn5 integration sites to genes, focusing on a 2 kb window surrounding the TSS. This estimation was achieved through the buildMetaData function, which culminated in the creation of a meta file. For further refinement of cell selection, we harnessed the findCells function, implementing several criteria: 1) A minimum read depth of 1,000 Tn5 integration sites was required. 2) The total number of cells was capped at 16,000. 3) The proportion of reads mapping to TSS sites was above 0.2, accompanied by a z-score threshold of 3. 4) Barcode FRiP scores were required to surpass 0.1, alongside a z-score threshold of 2. 5) We filtered out barcodes exhibiting a proportion of reads mapping to mitochondrial and chloroplast genomes that exceeded two standard deviations from the library mean. 6) We finally used the detectDoublets function to estimate doublet likelihood by creating synthetic doublets and conducting enrichment analysis. These multiple steps ensured the meticulous identification and selection of individual cells, facilitating a robust foundation for subsequent analyses.

Nuclei clustering

For the nuclei clustering, we leveraged all functions from the Socrates package[20]. We binned the entire genome into consecutive windows, each spanning 500 bp. We then tabulated the count of windows featuring Tn5 insertions per cell. Barcodes falling below one standard deviation from the mean feature counts (with a z-score less than 1) were excluded. Moreover, barcodes with fewer than 1,000 features were eliminated. We pruned windows that exhibited accessibility in less than 0.5% or more than 99.5% of all nuclei. To standardize the cleaned matrix, we applied a term frequency-inverse document frequency normalization function. The dimensionality of the normalized matrix underwent reduction through the utilization of non-negative matrix factorization, facilitated by the R package RcppML[89] (v0.3.7). We retained 50 column vectors from an uncentered matrix. Subsequently, we selected the top 30,000 windows that displayed the highest residual variance across all cells. This selection was based on fitting a model where the proportion of cells with accessibility served as an independent variable and variance as the dependent variable. To further reduce the dimensionality of the nuclei embedding, we employed the UMAP technique using umap-learn (k = 50, min_dist = 0.1, metrix = ‘euclidean’) in R[90] (v0.2.8.0). Furthermore, we clustered nuclei using the callClusters function within the Socrates

framework[20]. Louvain clustering was applied, with a setting of k=50 nearest neighbors at a resolution of 0.7. This process underwent 100 iterations with 100 random starts. Clusters with an aggregated read depth of less than 1 million and 50 cells were subsequently eliminated. To filter outlying cells in the UMAP embedding, we estimated the mean distance for each nucleus using its 50 nearest neighbors. Nuclei that exceeded 3 standard deviations from the mean distance were deemed outliers and removed from consideration.

Estimation of gene accessibility scores

To estimate the gene accessibility, we employed a strategy wherein the Tn5 insertion was counted across both the gene body region as well as a 500 bp extension upstream. Subsequently, we employed the SCTransform algorithm from the Seurat package[78] (v4.0) to normalize the count matrix that was then transformed into a normalized accessibility score, with all positive values scaled to 1.

Cell type validation

Upon completing the initial annotation process, which was based on a curated list of marker genes, we further expanded our marker repertoire by incorporating markers collected from published bulk RNA-seq data encompassing a diverse array of cell types, which were acquired *via* laser capture dissection. In brief, we collected the markers from several studies, including three distinct cell types within rice leaves[91] and ten cell types across rice seed organs[92,93]. From these sources, we selected the top 100 variably expressed markers for each cell type and employed them to compute cell identity enrichment scores. We undertook a comprehensive assessment of markers linked to the different cell types. Subsequently, we randomly drew 100 markers from this pool and repeated this procedure 1,000 times to construct a null distribution based on marker chromatin accessibility scores. For each target cell type marker, we compared their accessibility scores to this null distribution. This facilitated the derivation of an enrichment score per cell, delineating the marker's significance for each representative cell type. We next employed a MAGIC algorithm[94] to refine these enrichment scores. These scores were then mapped onto a UMAP plot, enhancing the cell identity annotation.

Furthermore, we undertook validation of our single-cell chromatin accessibility atlas through integration with published scRNA-seq from rice root tissue. This validation was achieved through two distinct approaches (Supplementary Fig. 8). In the first approach, we leveraged the marked enrichment technique, adapting the above mentioned methodology with the incorporation of the top 20 markers derived from marker identification using the 'FindMarkers' function in Seurat[78] (v4.0). Following the acquisition of a smoothed score for each cell type, individual cells were annotated to specific cell types based on the largest enrichment score within that cell type. A threshold was further set, requiring the maximum score to exceed 0.5 for confident labeling; otherwise, the cell was labeled as 'Unknown'. The second approach entailed employing a k-nearest neighbor (knn) strategy. This strategy commenced with the normalization of scRNA-seq datasets, mirroring the process applied to scATAC-seq datasets. The top 3,000 most variable genes within the scATAC-seq dataset were then identified using the Seurat function 'FindVariableFeatures', subsequently filtering to include only genes common to both datasets. By treating the scRNA-seq cells as a reference, a dimension reduction process was conducted to generate a loading matrix, which was then utilized to project the scATAC-seq cells onto the scRNA-seq cell embedding. The integration of these two datasets was achieved through

the Harmony algorithm[95] (v0.1.0). Within the dual embeddings, the 20 nearest neighbors of each scATAC-seq cell in the scRNA-seq dataset were computed. The most frequent label among these RNA neighbors (> 10 cells) was subsequently assigned as the label for each scATAC-seq cell or designated as NA if no label meeting this threshold was identified.

Cell cycle prediction

The prediction of cell cycle stages per nucleus was executed similarly to annotating cell identities based on the aforementioned enriched scores. In brief, we collected a set of 55 cell-cycle marker genes from a previous study[96]. For every cell-cycle stage, the cumulative gene accessibility score for each nucleus was computed. These resultant scores were subsequently normalized using the mean and standard deviation derived from 1,000 permutations of the 55 random cell-cycle stage genes, with exclusion of the focal stage. Z-scores corresponding to each cell-cycle stage were transformed into probabilities using the 'pnorm' function in R. Furthermore, the cell-cycle stage displaying the highest probability was designated as the most probable cell stage.

ACR identification

Upon segregating the comprehensive single-base resolution Tn5 insertion sites BED dataset into distinct subsets aligned with annotated cell types, we executed the MACS2 tool[88] (v2.2.7.1) for precise peak identification per cell type. Notably we employed non-default parameters, specifically: --extsize 150, --shift -75, -nomodel -keep-dup all. To mitigate potential false positives, a permutation strategy was applied, generating an equal number of peaks based on regions that were mappable and non-exonic. This approach encompassed the assessment of Tn5 insertion sites and density within both the original and permuted peak groups. By scrutinizing the permutation outcomes, we devised empirically derived false discovery rate (FDR) thresholds specific to each cell type. This entailed determining the minimum Tn5 density score within the permutation cohort where the FDR remained < 0.05. To further eliminate peaks that exhibited significant overlap with nucleosomes, we applied the NucleoATAC tool[97] (v0.2.1) to identify potential nucleosome placements. Peaks that featured over 50% alignment with predicted nucleosomes were systematically removed. The average fragment size of reads overlapping with the peaks were calculated and the peaks with the average fragment size > 150 bp were filtered out. Ultimately, the pool of peaks for each cell type was amalgamated and fine-tuned, yielding 500 bp windows that were centered on the summit of ACR coverage.

Identification of cell-type-specific ACRs in rice atlas

To identify cell-type-specific ACRs in the rice atlas, we implemented a series of cutoffs to determine whether the peak is accessible for a specific cell type. For each cell type, we first normalized the read coverage depth obtained from the MACS2 tool divided by total count of reads, and ensured that the maximum of normalized coverage within the peak exceeded a predefined threshold set at 2. Additionally, we calculated Tn5 integration sites per peak, filtering out peaks with fewer than 20 integration sites. Subsequently, we constructed a peak by cell type matrix with Tn5 integration site counts. This matrix underwent normalization using the 'cpm' function wrapped in edgeR[80] (v3.38.1) and 'normalize.quantiles' function wrapped within preprocessCore[98] (v1.57.1) in the R programming environment. To further refine our selection, a threshold of 2 was set for the counts per million value per peak per cell type. Peaks that satisfied these distinct cutoff criteria were deemed accessible in the designated cell types. We identified

these peaks as displaying chromatin accessibility in anywhere from one to ten of the primary cell types among the total pool of 60 main cell types.

Identification of cell-type-specific ACRs in leaf tissue across five species

To account for the lower cell number found in these data sets, cell-type-specific ACRs were identified using the same method as found in our prior study[24]. In brief we utilized a modified entropy metric combined with a bootstrapping approach[59]. For each cell-type and species a sample of 250 cells were taken 5,000 times with replacement and a specificity metric was calculated. This specificity metric was then compared against a series of 5,000 null distributions consisting of a random shuffle of 250 cells which were of mixed cell populations. Then, a nonparametric test was used to the median real bootstrap specificity score versus the null, and ACRs were labeled as cell-type-specific if they had a *p* value less than 0.0001.

Correlation between chromatin accessibility of TF genes and motif deviation

We sourced rice and *A. thaliana* TFs from PlantTFDB[99] (v4.0) database. To identify rice orthologs of *A. thaliana* TFs, we employed BLAST[87] (BLAST+; v2.11.0) by utilizing protein fasta alignments with an e-value threshold of 1e-5 used for significance. Alignments were restricted to fasta sequences categorized as TFs from either species. To further refine the putative orthologs, we applied filters based on functional similarity to *A. thaliana* TFs. Alignments with less than 15% identity were excluded, along with rice TFs associated with distinct families. From the remaining candidates, we selected the orthologs demonstrating the highest Pearson correlation coefficient concerning the motif deviation scores. Motif deviation scores of specific TF motifs within nuclei were computed *via* chromVAR[100] (v1.18.0).

Linear-model based motif enrichment analysis

We employed the FIMO tool from the MEME suite[101] (v5.1.1) with a significance threshold of *p* value $< 10^{-5}$ to predict motif locations. The motif frequency matrix used was sourced from the JASPAR plants motif database[65] (v9). Subsequently, we constructed a binarized peak-by-motif matrix and a motif-by-cell count matrix. This involved multiplying the peak-by-cell matrix with the peak-by-motif matrix. To address potential overrepresentation and computational efficiency, down-sampling was implemented. Specifically, we standardized the cell count by randomly selecting 412 cells per cell type per species. This count represents the lowest observed cell count for a given cell type across all species. For each cell type annotation, total motif counts were predicted through negative binomial regression. This involved two input variables: an indicator column for the annotation, serving as the primary variable of interest, and a covariate representing the logarithm of the total number of nonzero entries in the input peak matrix for each cell. The regression provided coefficients for the annotation indicator column and an intercept. These coefficients facilitated the estimation of fold changes in motif counts for the annotation of interest in relation to cells from all other annotations. This iterative process was conducted for all motifs across all cell types. The obtained *p* values were adjusted using the Benjamini-Hochberg procedure to account for multiple comparisons. Finally, enriched motifs were identified by applying a dual filter criterion: corrected *p* values < 0.01, fold-change of the top enriched TF motif in cell type-specific peaks for all cell types should be over 1, and beta (motif enrichment score) > 0.05 or beta > 0.

Binomial test-based motif enrichment analysis

To assess the enrichment of motifs in a target set of ACRs, we performed analysis for each specific motif. We randomly selected an equivalent number of ACRs as found in the target set, repeating this process 100 times. Notably, the randomly selected ACR set did not overlap with the actual target set of ACRs. Following this, we computed the average ratio of ACRs capturing the motif within the null distribution.

Subsequently, we executed an exact Binomial test[102], wherein we set this ratio as the hypothesized probability of success. The number of ACRs overlapping the motif in the target set was considered the number of successes, while the total number of ACRs in the target set represented the number of trials. The alternative hypothesis was specified as 'two.sided'. This meticulous approach allowed us to robustly evaluate and identify significant motif enrichments within the target set of ACRs.

Construction of control sets for enrichment tests

To perform comparative analysis of expression levels and chromatin accessibility of genes surrounding broad ACRs under and outside of H3K27me3 peaks, we sampled the same number of ACRs per cell type regarding the broad ACRs not under H3K27me3 peaks. This step is to make sure that their nearby gene chromatin accessibility exhibited similar values compared to the broad ACRs under the H3K27me3 peaks.

To check if the H3K27me3-broad-ACRs could significantly capture the known PREs and capture the EMF2b ChIP-seq peaks, we generated control sets by randomly selecting not-H3K27me3-broad-ACR instances 100 times, yielding a mean number value for the control sets. The Binomial test *p* value was calculated by comparing the mean ratio to the observed number of H3K27me3-broad ACRs overlapped with the PREs.

To test if H3K27me3-broad ACRs in *O. sativa*, *Z. mays*, and *S. bicolor* significantly capture six known motifs, we generated control sets by simulating sequences with the same length as ACRs 100 times, yielding a mean proportion for the control sets. The binomial test p value was calculated by comparing the mean ratio to the observed overlapping ratio of H3K27me3-broad ACRs capturing the motifs.

***De novo* motif analysis**

To identify position weight matrix of six known motifs within 170 *A. thaliana* PREs[55], we employed the streme function with default settings from the MEME suite[101] (v5.1.1). The control sequences were built up to match each PRE sequence by excluding exons, PREs, and unmappable regions, and they possess a similar GC content (< 5% average difference) and same sequence length compared to the positive set.

Identification of syntenic regions

Identification of syntenic gene blocks was done using the GENESPACE[103] (v1.4). In brief, to establish orthologous relationships between ACR sequences, ACRs in the *O. sativa* genome were extended to incorporate the two closest gene models for a 'query block' since GENESPACE only draws relationships between protein coding sequences. Then the GENESPACE function 'query_hits' was used with the argument 'synOnly = TRUE' to retrieve syntenic blocks. The resulting syntenic hits were further filtered to allow only a one-to-one relationship between

O.sativa and the corresponding species. The corresponding syntenic blocks were then named and numbered, and both the genes and genomic coordinates were recorded.

To further identify corresponding ACRs within these blocks we set up a BLASTN pipeline[104] (v2.13.0). For each comparison of species, using *O. sativa* as the reference the underlying nucleotide sequences of the syntenic regions were extracted using Seqkit, and used as the blast reference database[105] (v2.5.1). The sequences underlying the ACRs within the same syntenic region in a different species were then used as the query. The blast was done using the following parameters to allow for alignment of shorter sequences '-task blastn-short -evalue 1e-3 -max_target_seqs 4-word_size 7 -gapopen 5 -gapextend 2 -penalty -1 -reward 1 -outfmt 6'. This procedure was run for each syntenic region separately for all species comparisons. The resulting BLASTN files were combined, and then filtered using a custom script. Alignments were only considered valid if the e-value passed a stringent threshold of 1e-3, and the alignment was greater than 20 nucleotides with the majority of the shared ACRs (92% to 94%) containing the alignment regions including TF motif binding sites (Supplementary Fig. 13). The resulting filtered BLAST files, and the BED files generated from these BLAST files allowed us to draw our relationships between ACRs in the corresponding syntenic space.

Estimation of conservation scores

Conservation scores were predicted using PhyloP[106] (v1.0), where values are scaled between 0 to 1, with one being highly conserved and 0 being non-conserved. Phylogenies to train PhyloP were generated using PhyloFit[107] (v1.0), and neutral and conserved sequences were identified using the whole genome aligner progressive cactus.

ChIP-seq analysis

The clean reads of EMF2b were downloaded from a previous study[64]. The reads were mapped to the rice reference genome[84] (v7.0) using bowtie2[108] (v2.5.2) with the following parameters: '--very-sensitive --end-to-end'. Reads with MAPQ > 5 were used for the subsequent analysis. Aligned reads were sorted and duplicated reads were removed using SAMtools[109] (v1.7). Peak calling was performed using epic2[110] with the following parameters: '-fdr 0.01 --bin-size 150 --gaps-allowed 1'. The peak 'BED' and 'BIGWIG' files of H3K27me3 ChIP-seq data for leaf, root, and panicle rice organs were downloaded from RiceENCODE[111] (http://glab.hzau.edu.cn/RiceENCODE/).

GO enrichment test

The GO enrichment tests were performed based on the AgriGO[43] (v2) by setting the Chi-square statistical test and multi-test adjustment method is Hochberg (FDR).

Additional resources

Cell-type resolved data can be viewed through our public Plant Epigenome JBrowse Genome Browser[112] (http://epigenome.genetics.uga.edu/PlantEpigenome/index.html)

Data Availability

scATAC-seq data encompassing 18 libraries from nine organs were accessible in NCBI (PRJNA1007577/GSE252040; https://dataview.ncbi.nlm.nih.gov/object/PRJNA1007577?reviewer=kgarq48dii11vomg44kgr1jq66; PRJNA1052039; https://dataview.ncbi.nlm.nih.gov/object/PRJNA1052039?reviewer=flhu9sl84o5m999r1ph8tlmmbg)

Acknowledgements

This research was funded by the National Science Foundation (IOS-2134912) to SRW and RJS and the UGA Office of Research to RJS. APM and JPM were supported by the National Institutes of Health (K99GM144742) and (T32GM142623), respectively.

Contributions

R.J.S., H.Y., J.P.M., A.P.M., M.A.A.M., D.W., S.Z., and S.R.W. designed and conceived experiments and managed the project. X.Z., Y.L., Z.L., X.T., S.Z., Y.W., and H.Y. participated in material collection and sample processing. H.Y., J.P.M., and T.R. performed the bioinformatics analyses. H.Y. and J.P.M. wrote the manuscript. Y.L. Y.W. and Z.L. contributed to marker validation. R.J.S., J.P.M., A.P.M., M.A.A.M., D.W., S.Z., and S.R.W. edited the manuscript.

References

1 Preissl, S., Gaulton, K. J. & Ren, B. Characterizing cis-regulatory elements using single-cell epigenomics. *Nature Reviews Genetics* **24**, 21-43 (2023).

2 Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**, 59-69 (2012).

3 Marand, A. P., Eveland, A. L., Kaufmann, K. & Springer, N. M. cis-Regulatory elements in plant development, adaptation, and evolution. *Annual review of plant biology* **74**, 111-137 (2023).

4 Oka, R. *et al.* Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome biology* **18**, 1-24 (2017).

5 Maher, K. A. *et al.* Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *The Plant Cell* **30**, 15-36 (2018).

6 Lu, Z. *et al.* The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nature Plants* **5**, 1250-1259 (2019).

7 Reynoso, M. A. *et al.* Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* **365**, 1291-1295 (2019).
8 Kajala, K. *et al.* Innovation, conservation, and repurposing of gene function in root cell type development. *Cell* **184**, 3333-3348. e3319 (2021).
9 Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nature plants* **5**, 1237-1249 (2019).
10 Cusanovich, D. A. *et al.* A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309-1324. e1318 (2018).
11 Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
12 Lu, Z. *et al.* Tracking cell-type-specific temporal dynamics in human and mouse brains. *Cell* **186**, 4345-4364. e4324 (2023).
13 Javelle, M., Vernoud, V., Rogowsky, P. M. & Ingram, G. C. Epidermis: the formation and functions of a fundamental plant tissue. *New Phytologist* **189**, 17-39 (2011).
14 Kadioglu, A., Terzi, R., Saruhan, N. & Saglam, A. Current advances in the investigation of leaf rolling caused by biotic and abiotic stress factors. *Plant Science* **182**, 42-48 (2012).
15 Xu, Y. *et al.* Overexpression of OsZHD1, a zinc finger homeodomain class homeobox transcription factor, induces abaxially curled and drooping leaf in rice. *Planta* **239**, 803-816 (2014).
16 Dorrity, M. W. *et al.* The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. *Nature communications* **12**, 3334 (2021).
17 Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, J. & Libault, M. Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level. *Molecular Plant* **14**, 372-383 (2021).
18 Tu, X., Marand, A. P., Schmitz, R. J. & Zhong, S. A combinatorial indexing strategy for low-cost epigenomic profiling of plant single cells. *Plant Communications* **3** (2022).
19 Nobori, T. *et al.* Time-resolved single-cell and spatial gene regulatory atlas of plants under pathogen attack. *bioRxiv*, 2023.2004. 2010.536170 (2023).
20 Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**, 3041-3055. e3021 (2021).
21 Feng, D. *et al.* Chromatin accessibility illuminates single-cell regulatory dynamics of rice root tips. *BMC biology* **20**, 274 (2022).
22 Zhang, L. *et al.* Asymmetric gene expression and cell-type-specific regulatory networks in the root of bread wheat revealed by single-cell multiomics analysis. *Genome Biology* **24**, 65 (2023).
23 Swift, J. *et al.* Single nuclei sequencing reveals C4 photosynthesis is based on rewiring of ancestral cell identity networks. *bioRxiv*, 2023.2010. 2026.562893 (2023).
24 Mendieta, J. P. *et al.* Investigating the cis-Regulatory Basis of C3 and C4 Photosynthesis in Grasses at Single-Cell Resolution. *bioRxiv*, 2024.2001. 2005.574340 (2024).
25 Mansueto, L. *et al.* Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic acids research* **45**, D1075-D1081 (2017).
26 Lv, Z., Zhao, W., Kong, S., Li, L. & Lin, S. Overview of molecular mechanisms of plant leaf development: a systematic review. *Frontiers in Plant Science* **14** (2023).

27 Wu, R. *et al.* CFL1, a WW domain protein, regulates cuticle development by modulating the function of HDG1, a class IV homeodomain transcription factor, in rice and Arabidopsis. *The Plant Cell* **23**, 3392-3411 (2011).
28 Le Hir, R. & Bellini, C. The plant-specific Dof transcription factors family: new players involved in vascular system development and functioning in Arabidopsis. *Frontiers in Plant Science* **4**, 164 (2013).
29 Dai, X. *et al.* Chromatin and regulatory differentiation between bundle sheath and mesophyll cells in maize. *The Plant Journal* **109**, 675-692 (2022).
30 Yanagisawa, S. Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *The Plant Journal* **21**, 281-288 (2000).
31 Ueno, O., Kawano, Y., Wakayama, M. & Takeda, T. Leaf vascular systems in C3 and C4 grasses: a two-dimensional analysis. *Annals of Botany* **97**, 611-621 (2006).
32 Gao, Z., Shen, W. & Chen, G. Uncovering C4-like photosynthesis in C3 vascular cells. *Journal of Experimental Botany* **69**, 3531-3540 (2018).
33 Shimadzu, S., Furuya, T. & Kondo, Y. Molecular mechanisms underlying the establishment and maintenance of vascular stem cells in Arabidopsis thaliana. *Plant and Cell Physiology* **64**, 274-283 (2023).
34 Otero, S. & Helariutta, Y. Companion cells: a diamond in the rough. *Journal of experimental botany*, erw392 (2016).
35 Ramachandran, V. *et al.* Plant-specific Dof transcription factors VASCULAR-RELATED DOF1 and VASCULAR-RELATED DOF2 regulate vascular cell differentiation and lignin biosynthesis in Arabidopsis. *Plant molecular biology* **104**, 263-281 (2020).
36 Kubo, H., Kishi, M. & Goto, K. Expression analysis of ANTHOCYANINLESS2 gene in Arabidopsis. *Plant science* **175**, 853-857 (2008).
37 Amanda, D. *et al.* DEFECTIVE KERNEL1 (DEK1) regulates cell walls in the leaf epidermis. *Plant Physiology* **172**, 2204-2218 (2016).
38 Ingram, P., Dettmer, J., Helariutta, Y. & Malamy, J. E. Arabidopsis Lateral Root Development 3 is essential for early phloem development and function, and hence for normal root system development. *The Plant Journal* **68**, 455-467 (2011).
39 Chen, X. *et al.* SQUAMOSA promoter□binding protein□like transcription factors: Star players for plant growth and development. *Journal of integrative plant biology* **52**, 946-951 (2010).
40 Denyer, T. *et al.* Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. *Developmental cell* **48**, 840-852. e845 (2019).
41 Fang, J. *et al.* The URL1–ROC5–TPL2 transcriptional repressor complex represses the ACL1 gene to modulate leaf rolling in rice. *Plant Physiology* **185**, 1722-1744 (2021).
42 Horstman, A. *et al.* AIL and HDG proteins act antagonistically to control cell proliferation. *Development* **142**, 454-464 (2015).
43 Tian, T. *et al.* agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic acids research* **45**, W122-W129 (2017).
44 Rombolá-Caldentey, B., Rueda-Romero, P., Iglesias-Fernández, R., Carbonero, P. & Oñate-Sánchez, L. Arabidopsis DELLA and two HD-ZIP transcription factors regulate

GA signaling in the epidermis through the L1 box cis-element. *The Plant Cell* **26**, 2905-2919 (2014).

45 Yu, L. H. *et al.* Arabidopsis EDT 1/HDG 11 improves drought and salt tolerance in cotton and poplar and increases cotton yield in the field. *Plant biotechnology journal* **14**, 72-84 (2016).

46 Hong, S.-Y., Kim, O.-K., Kim, S.-G., Yang, M.-S. & Park, C.-M. Nuclear import and DNA binding of the ZHD5 transcription factor is modulated by a competitive peptide inhibitor in Arabidopsis. *Journal of Biological Chemistry* **286**, 1659-1668 (2011).

47 Rosado, D., Ackermann, A., Spassibojko, O., Rossi, M. & Pedmale, U. V. WRKY transcription factors and ethylene signaling modify root growth during the shade-avoidance response. *Plant Physiology* **188**, 1294-1311 (2022).

48 Wolfe, K. H., Gouy, M., Yang, Y.-W., Sharp, P. M. & Li, W.-H. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences* **86**, 6201-6205 (1989).

49 Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS biology* **3**, e7 (2005).

50 Babarinde, I. A. & Saitou, N. Genomic locations of conserved noncoding sequences and their proximal protein-coding genes in mammalian expression dynamics. *Molecular biology and evolution* **33**, 1807-1817 (2016).

51 Song, B. *et al.* Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome research* **31**, 1245-1257 (2021).

52 Nelson, A. C. & Wardle, F. C. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* **140**, 1385-1395 (2013).

53 Hendelman, A. *et al.* Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection. *Cell* **184**, 1724-1739. e1716 (2021).

54 Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043 (2002).

55 Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nature genetics* **49**, 1546-1552 (2017).

56 Schmitz, R. J., Grotewold, E. & Stam, M. Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *The plant cell* **34**, 718-741 (2022).

57 Ouyang, W. *et al.* Haplotype mapping of H3K27me3-associated chromatin interactions defines topological regulation of gene silencing in rice. *Cell Reports* **42** (2023).

58 Minnoye, L. *et al.* Chromatin accessibility profiling methods. *Nature Reviews Methods Primers* **1**, 10 (2021).

59 Zhang, T.-Q., Chen, Y., Liu, Y., Lin, W.-H. & Wang, J.-W. Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nature communications* **12**, 2053 (2021).

60 Yu, Y., Zhang, H., Long, Y., Shu, Y. & Zhai, J. Plant public RNA□seq database: a comprehensive online database for expression analysis of~ 45 000 plant public RNA□seq libraries. *Plant Biotechnology Journal* **20**, 806 (2022).

61 Bai, X. *et al.* Duplication of an upstream silencer of FZP increases grain yield in rice. *Nature Plants* **3**, 885-893 (2017).

62 Wang, Z. *et al.* AraENCODE: A comprehensive epigenomic database of Arabidopsis thaliana. *Molecular Plant* **16**, 1113-1116 (2023).

63 Tonosaki, K. & Kinoshita, T. Possible roles for polycomb repressive complex 2 in cereal endosperm. *Frontiers in plant science* **6**, 144 (2015).
64 Tan, F.-Q. *et al.* A coiled-coil protein associates Polycomb Repressive Complex 2 with KNOX/BELL transcription factors to maintain silencing of cell differentiation-promoting genes in the shoot apex. *The Plant Cell* **34**, 2969-2988 (2022).
65 Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic acids research* **50**, D165-D173 (2022).
66 Huang, Y. *et al.* OsNCED5, a 9-cis-epoxycarotenoid dioxygenase gene, regulates salt and water stress tolerance and leaf senescence in rice. *Plant science* **287**, 110188 (2019).
67 Andrews, G. *et al.* Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. *Science* **380**, eabn7930 (2023).
68 Engelhorn, J. *et al.* Phenotypic variation in maize can be largely explained by genetic variation at transcription factor binding sites. *bioRxiv*, 2023.2008. 2008.551183 (2023).
69 Zhao, T. & Schranz, M. E. Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proceedings of the National Academy of Sciences* **116**, 2165-2174 (2019).
70 Zemke, N. R. *et al.* Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390-402 (2023).
71 Clark, J. W. & Donoghue, P. C. Whole-genome duplication and plant macroevolution. *Trends in plant science* **23**, 933-945 (2018).
72 Del Pozo, J. C. & Ramirez-Parra, E. Whole genome duplications in plants: an overview from Arabidopsis. *Journal of Experimental Botany* **66**, 6991-7003 (2015).
73 Jump, A. S., Marchant, R. & Peñuelas, J. Environmental change and the option value of genetic diversity. *Trends in plant science* **14**, 51-58 (2009).
74 Ciren, D., Zebell, S. & Lippman, Z. Extreme restructuring of cis-regulatory regions controlling a deeply conserved plant stem cell regulator. *bioRxiv*, 2023.2012. 2020.572550 (2023).
75 Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Current opinion in genetics & development* **43**, 31-37 (2017).
76 Zhang, X., Marand, A. P., Yan, H. & Schmitz, R. J. Massive-scale single-cell chromatin accessibility sequencing using combinatorial fluidic indexing. *BioRxiv* (2023).
77 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049 (2017).
78 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33**, 495-502 (2015).
79 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell systems* **8**, 281-291. e289 (2019).
80 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics* **26**, 139-140 (2010).
81 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
82 Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature biotechnology* **39**, 313-319 (2021).

83 Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nature biotechnology* **37**, 925-936 (2019).
84 Ouyang, S. *et al.* The TIGR rice genome annotation resource: improvements and new features. *Nucleic acids research* **35**, D883-D887 (2007).
85 Institute, B. Picard toolkit. *Broad Institute, GitHub repository* (2019).
86 Smit, A. F. Repeat-Masker Open-3.0. *http://www. repeatmasker. org* (2004).
87 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10**, 1-9 (2009).
88 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, 1-9 (2008).
89 DeBruine, Z. J., Melcher, K. & Triche Jr, T. J. Fast and robust non-negative matrix factorization for single-cell experiments. *bioRxiv*, 2021.2009. 2001.458620 (2021).
90 McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).
91 Hua, L. *et al.* The bundle sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis. *The Plant Journal* **107**, 268-286 (2021).
92 Itoh, J.-I. *et al.* Genome-wide analysis of spatiotemporal gene expression patterns during early embryogenesis in rice. *Development* **143**, 1217-1227 (2016).
93 Wu, T.-Y., Müller, M., Gruissem, W. & Bhullar, N. K. Genome wide analysis of the transcriptional profiles in different regions of the developing rice grains. *Rice* **13**, 1-19 (2020).
94 Van Dijk, D. *et al.* Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716-729. e727 (2018).
95 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16**, 1289-1296 (2019).
96 Pettkó-Szandtner, A. *et al.* Core cell cycle regulatory genes in rice and their expression profiles across the growth zone of the leaf. *Journal of plant research* **128**, 953-974 (2015).
97 Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome research* **25**, 1757-1770 (2015).
98 Bolstad, B. M. & Bolstad, M. B. M. Package ‘preprocessCore’. (2013).
99 Jin, J. *et al.* PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic acids research*, gkw982 (2016).
100 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature methods* **14**, 975-978 (2017).
101 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).
102 Wagner□Menghin, M. M. Binomial test. *Encyclopedia of statistics in behavioral science* (2005).
103 Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11**, e78526 (2022).

104 Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic acids research* **49**, D10 (2021).
105 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS one* **11**, e0163962 (2016).
106 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* **20**, 110-121 (2010).
107 Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Molecular biology and evolution* **21**, 468-488 (2004).
108 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
109 Li, H. *et al.* The sequence alignment/map format and SAMtools. *bioinformatics* **25**, 2078-2079 (2009).
110 Stovner, E. B. & Sætrom, P. epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392-4393 (2019).
111 Xie, L. *et al.* RiceENCODE: A comprehensive epigenomic database as a rice Encyclopedia of DNA Elements. *Molecular Plant* **14**, 1604-1606 (2021).
112 Hofmeister, B. T. & Schmitz, R. J. Enhanced JBrowse plugins for epigenomics data visualization. *BMC bioinformatics* **19**, 1-6 (2018).