the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

*The Plant Journal* (2023) **113,** 160–173 doi: 10.1111/tpj.16045

# Coexpression network and *trans*-activation analyses of maize reproductive phasiRNA loci

**Junpeng Zhan[1], Lily O'Connor[1,2], D. Blaine Marchant[3], Chong Teng[1], Virginia Walbot[3] and Blake C. Meyers[1,4,*]**

[1]*Donald Danforth Plant Science Center, St Louis, MO, 63132, USA,*
[2]*Department of Biology, Washington University, St Louis, MO, 63130, USA,*
[3]*Department of Biology, Stanford University, Stanford, CA, 94305, USA, and*
[4]*Division of Plant Science and Technology, University of Missouri, Columbia, MO, 65211, USA*

Received 2 August 2022; revised 16 November 2022; accepted 21 November 2022; published online 28 November 2022.
*For correspondence (e-mail bmeyers@danforthcenter.org).

## SUMMARY

**The anther-enriched phased, small interfering RNAs (phasiRNAs) play vital roles in sustaining male fertility in grass species. Their long non-coding precursors are synthesized by RNA polymerase II and are likely regulated by transcription factors (TFs). A few putative transcriptional regulators of the 21- or 24-nucleotide phasiRNA loci (referred to as *21-* or *24-PHAS* loci) have been identified in maize (*Zea mays*), but whether any of the individual TFs or TF combinations suffice to activate any *PHAS* locus is unclear. Here, we identified the temporal gene coexpression networks (modules) associated with maize anther development, including two modules highly enriched for the *21-* or *24-PHAS* loci. Comparisons of these coexpression modules and gene sets dysregulated in several reported male sterile TF mutants provided insights into TF timing with regard to phasiRNA biogenesis, including antagonistic roles for OUTER CELL LAYER4 and MALE STERILE23. *Trans*-activation assays in maize protoplasts of individual TFs using bulk-protoplast RNA-sequencing showed that two of the TFs coexpressed with *21-PHAS* loci could activate several 21-nucleotide phasiRNA pathway genes but not transcription of *21-PHAS* loci. Screens for combinatorial activities of these TFs and, separately, the recently reported putative transcriptional regulators of *24-PHAS* loci using single-cell (protoplast) RNA-sequencing, did not detect reproducible activation of either *21-PHAS* or *24-PHAS* loci. Collectively, our results suggest that the endogenous transcriptional machineries and/or chromatin states in the anthers are necessary to activate reproductive *PHAS* loci.**

**Keywords: anther, reproductive phasiRNA, coexpression network, *trans*-activation assay, protoplast, single-cell RNA-seq.**

## INTRODUCTION

Maize anthers accumulate two classes of phased, small interfering RNAs (phasiRNAs) – 21-nucleotide phasiRNAs and 24-nucleotide phasiRNAs – derived from intergenic, long non-coding genomic loci. Such loci, referred to as *21-* and *24-PHAS* loci, also exist in the genomes of rice and numerous other angiosperm species, although the numbers of loci and accumulation patterns vary (Liu et al., 2020; Pokhrel et al., 2021). For example, there are approximately 463 *21-PHAS* and 176 *24-PHAS* loci in maize (Zhai et al., 2015), and 2338 *21-PHAS* and 172 *24-PHAS* loci in rice (*Oryza sativa*) (Tian et al., 2021). The *21-* and *24-PHAS* precursors are produced by RNA polymerase II, capped, polyadenylated, exported to the cytoplasm (Li et al., 2016), and then cleaved by an Argonaute (AGO) protein guided by the 22-nucleotide microRNA (miRNA) family miR2118 (triggers of 21-nucleotide phasiRNAs) or miR2275 (triggers of 24-nucleotide phasiRNAs). The cleaved *PHAS* precursors are converted by RNA-DEPENDENT RNA POLYMERASE6 (RDR6) to double-stranded RNA molecules, which are subsequently diced by endoribonucleases DICER-LIKE 4 (DCL4) or DCL5, yielding 21- and 24-nucleotide phasiRNAs (Liu et al., 2020).

In rice, mutations in a few *21-PHAS* loci cause temperature- and/or photoperiod-sensitive male sterility, an agriculturally important trait that is used for hybrid rice production (Ding et al., 2012; Fan et al., 2016). A recently-reported partial rice *mir2118* knockout mutant showed reduced 21-nucleotide phasiRNA abundance and male and female sterility, with developmental defects in anther

tapetal and middle layers (Araki et al., 2020). The rice *MEIOSIS ARRESTED AT LEPTOTENE1* (*MEL1*) gene (also known as *AGO5c*) encodes an AGO protein that loads a subset of 21-nucleotide phasiRNAs in rice anthers (Komiya et al., 2014); *mel1* loss-of-function mutants are male sterile and exhibit defects in both tapetal and pollen mother cells (Nonomura et al., 2007). The maize AGO5b and AGO5c proteins [also called MALE-ASSOCIATED ARGONAUTE1 (MAGO1) and MAGO2, respectively] have recently been shown to load 21-nucleotide phasiRNAs (Lee et al., 2021). The maize *AGO5c* gene, an ortholog of rice *MEL1*, is the causal gene of the classic maize male sterile mutant *male sterile28* (*ms28*), which also exhibits defects in tapetal development (Li, Huang, et al., 2021). The maize *dcl5* mutant, which lacks 24-nucleotide phasiRNAs almost entirely, shows temperature-sensitive male sterility and tapetal developmental defects (Teng et al., 2020).

Currently, the molecular roles of both classes of phasiRNAs are not well understood. Recent studies have shown that a number of rice 21-nucleotide phasiRNAs mediate cleavage of mRNAs of protein-coding genes (Jiang et al., 2020; Zhang et al., 2020), and a subset of the maize 21-nucleotide phasiRNAs mediate cleavage of their own precursors (i.e. *cis*-cleavage) and transcripts derived from retrotransposons (Lee et al., 2021; Tamim et al., 2018). In whole anther studies, maize 24-nucleotide phasiRNAs have been implicated in regulation of CHH (H = A/C/T) DNA methylation of their genomic loci (Zhang, Ma, et al., 2021); because 24-nucleotide phasiRNAs are synthesized in the tapetum but accumulate in prophase I meiocytes, they may also increase CHH methylation of their genomic loci in the germinal cells (Zhou et al., 2022). These published data in rice and maize indicate that both 21- and 24-nucleotide phasiRNAs play crucial roles in supporting anther development and male fertility in grasses, and the limited data on their functions indicate that each class probably acts through distinctive molecular mechanisms.

We previously characterized the patterns of phasiRNA accumulation in maize anthers (Zhai et al., 2015). Accumulation of each class of phasiRNAs is highly coordinated: the total abundance of 21-nucleotide phasiRNAs peak at the pre-meiotic phase of anther development (when anther length is in the range 0.4–0.7 mm), whereas the 24-nucleotide phasiRNAs peak when anther length is in the range 1.5–2.5 mm and the male germinal cells undergo meiosis. The total abundance of *21-PHAS* and *24-PHAS* precursors peaks concurrently or slightly before that of the corresponding phasiRNAs (Zhai et al., 2015). These published results strongly suggest that both precursor and phasiRNA accumulation are tightly controlled temporally by transcriptional regulators. The maize homeodomain leucine zipper type IV (HD-ZIP IV) transcription-factor (TF) gene *OUTER CELL LAYER4* (*OCL4*), preferentially expressed in epidermal tissues including the anther epidermis (Vernoud et al., 2009), has been proposed as a key regulator of *21-PHAS* and *miR2118* genes (Yadava et al., 2021). Loss-of-function *ocl4* mutants lack *21-PHAS* precursors and 21-nucleotide phasiRNAs (Zhai et al., 2015), and the mutant anthers produce an abnormal (partially duplicated) endothecial layer (Vernoud et al., 2009). Three basic helix–loop–helix (bHLH) family TFs – MALE STERILE23 (MS23), MS32, and bHLH122 – have each been shown necessary for 24-nucleotide phasiRNA biogenesis. A paralog of MS32, bHLH51, likely co-regulates downstream gene networks with MS23, MS32, and/or bHLH122 because it forms heterodimers with MS23 and bHLH122, and single mutants of each the four bHLHs exhibit defects in tapetal development (Nan et al., 2022).

In the present study, to characterize the temporal accumulation patterns of individual reproductive *PHAS* precursors and further understand the transcriptional regulation of *PHAS* loci and anther development in maize, we initially performed a coexpression network analysis on a previously published anther RNA-sequencing (RNA-Seq) data set (Zhai et al., 2015). We identified 10 temporal gene coexpression modules, including two modules that each contained many *21-PHAS* or *24-PHAS* loci, among other genes. RNA-Seq of maize leaf protoplasts ectopically expressing TFs that temporally coexpress with the *21-PHAS* loci demonstrated that a few individual TFs can activate genes associated with the phasiRNA pathways, but none were sufficient to activate any of the *PHAS* loci. To examine the combinatorial activities of putative transcriptional regulators of *PHAS* loci, we performed single-cell RNA-Seq of two sets of protoplasts ectopically expressing TFs with a putative role in regulating *21-* or *24-PHAS* loci. We did not detect reproducible activation of the *21-* or *24-PHAS* loci, but this method shows potential for future assays of target *trans*-activation by TFs.

## RESULTS

### Temporal gene coexpression networks in maize anthers

To characterize the temporal gene expression programs underlying maize anther development, we analyzed a previously published set of RNA-Seq data of male reproductive tissues of the maize inbred line W23 *bz2* (Zhai et al., 2015). The samples represent 10 anther developmental stages (spanning pre-meiotic to post-meiotic phases) plus mature pollen (Table S1), with biological replicates; although these data were generated more than 7 years ago, they still comprise a 'gold standard' data set for anther biology (Han et al., 2022; Li, Zhu, et al., 2021). RNA-Seq read counts of all expressed genes and reproductive *PHAS* loci (including 356 *21-PHAS* and 121 *24-PHAS*) were normalized to reads per kilobase per million mapped reads (RPKM) and the average RPKM of each gene was

calculated for samples with replicates (Table S2). A Spearman correlation coefficient (SCC) analysis and a principal component analysis (PCA) both showed that the pollen sample clustered distinctly from the anther samples, consistent with the fact that the anther samples comprise more complex tissues than the pollen and that maize pollen expresses a unique suite of genes (Nelms & Walbot, 2022). Both analyses separated the 10 anther developmental stages into three groups/clusters, corresponding to the pre-meiotic, meiotic, and post-meiotic phases of anther development (Figure 1a,b) (Kelliher & Walbot, 2011), indicating dynamic transcriptome reprogramming during the transition to and from meiosis.

To identify the gene regulatory networks controlling the dynamic transcriptome of the developing anther, and the putative transcriptional regulators of reproductive *PHAS* loci, we applied the weighted gene coexpression network analysis (WGCNA) tool (Langfelder & Horvath, 2008) to the normalized RPKM data from the anther samples. By excluding genes that showed low or relatively stable expression levels across the 10 anther developmental stages (see Experimental Procedures), 4235 of the expressed genes, 269 *21-PHAS*, and 84 *24-PHAS* loci were selected for WGCNA. Our analysis identified 10 coexpression modules, containing 194–1244 genes, including 2–60 TF genes (Figure 2a,b; Figure S4; Table S3). Strikingly, all the *21-PHAS* loci were assigned to Module 4 (M4) and 80 of the 84 *24-PHAS* loci were assigned to M7, indicating that the majority of each class of *PHAS* loci is highly coordinated and thus possibly regulated by a common set of upstream regulators. These observations are consistent with the temporally coordinated accumulation of *PHAS* precursors (based on total abundance of each type) and reproductive phasiRNAs derived from individual *PHAS* loci (Zhai et al., 2015). An analysis of chromosomal locations of the *PHAS* loci assigned to M4 and M7, versus those that were not, did not detect dramatic differences, except that the *21-PHAS* loci on chromosome 8 were all coexpressed in M4 (Figure S5). The module eigengenes (i.e. first principal component of the $\log_2$RPKM data of each module) of the 10 coexpression modules separated into three distinct clusters, corresponding to pre-meiotic, meiotic, and post-meiotic stages of anther development (Figure 2a). M4 represents a set of genes that are preferentially expressed in pre-meiotic (0.2–1.0 mm) anthers. M7 includes 229 genes that are preferentially expressed during early meiotic stages (1.5 and 2.0 mm), peaking at 1.5 mm. M2 (299 genes) and M6 (648 genes) are preferentially expressed when anther length is in the range 1.5–3.0 mm, with slightly different patterns of expression at stages prior to 1.5 mm and after 3.0 mm. M2 and M6 each included one *24-PHAS* locus, and both modules clustered with M7 (Figure 2a,b; Table S3), suggesting an association of these two modules with the gene regulatory programs activating *24-PHAS* loci. The remaining modules (M1, M3, M5, M8, M9, and M10) were preferentially expressed in 4- and 5-mm anthers, when meiosis has completed, and the microspore undergoes two mitotic divisions to give rise to pollen grains (Zhai et al., 2015).

To uncover the biological processes associated with each module, we performed Gene Ontology (GO) (http://geneontology.org) enrichment analyses, and detected significantly enriched GO terms [false discovery rate (FDR) < 0.05] in eight of the 10 modules (Figure S1;

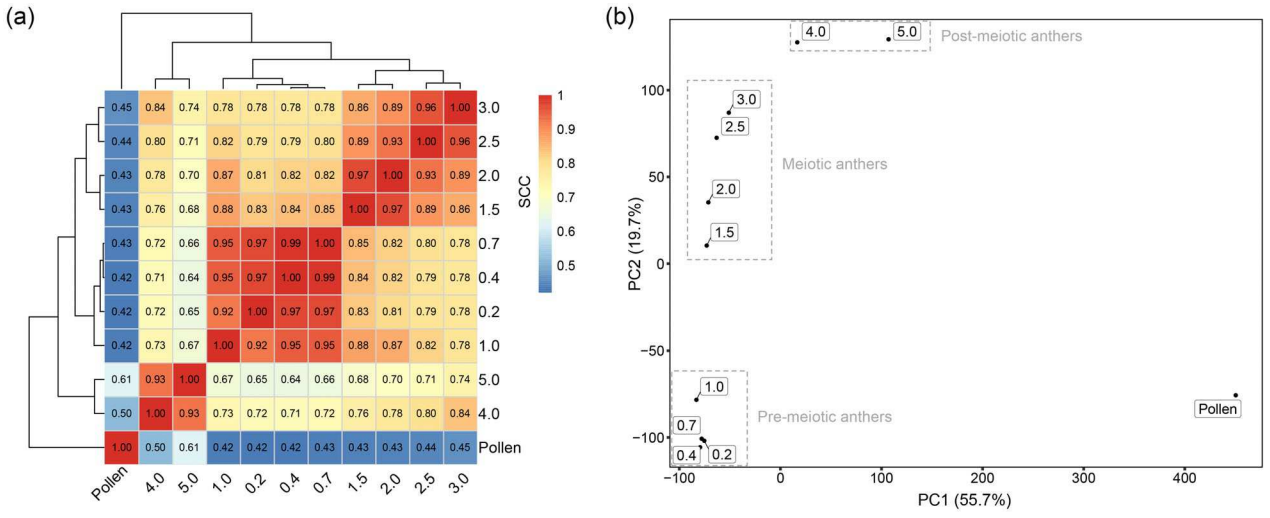


**Figure 1.** Relationships among the mRNA transcriptomes of 10 developmental stages of maize anthers and mature pollen.
(a) SCC heatmap of anther and pollen transcriptomes. The hierarchical cluster dendrogram was generated with (1 – SCC) as the distances.
(b) PCA of anther and pollen transcriptomes. The percentage of variance explained by each principal component is given in parentheses. Numbers in small rectangles are anther lengths (in mm).

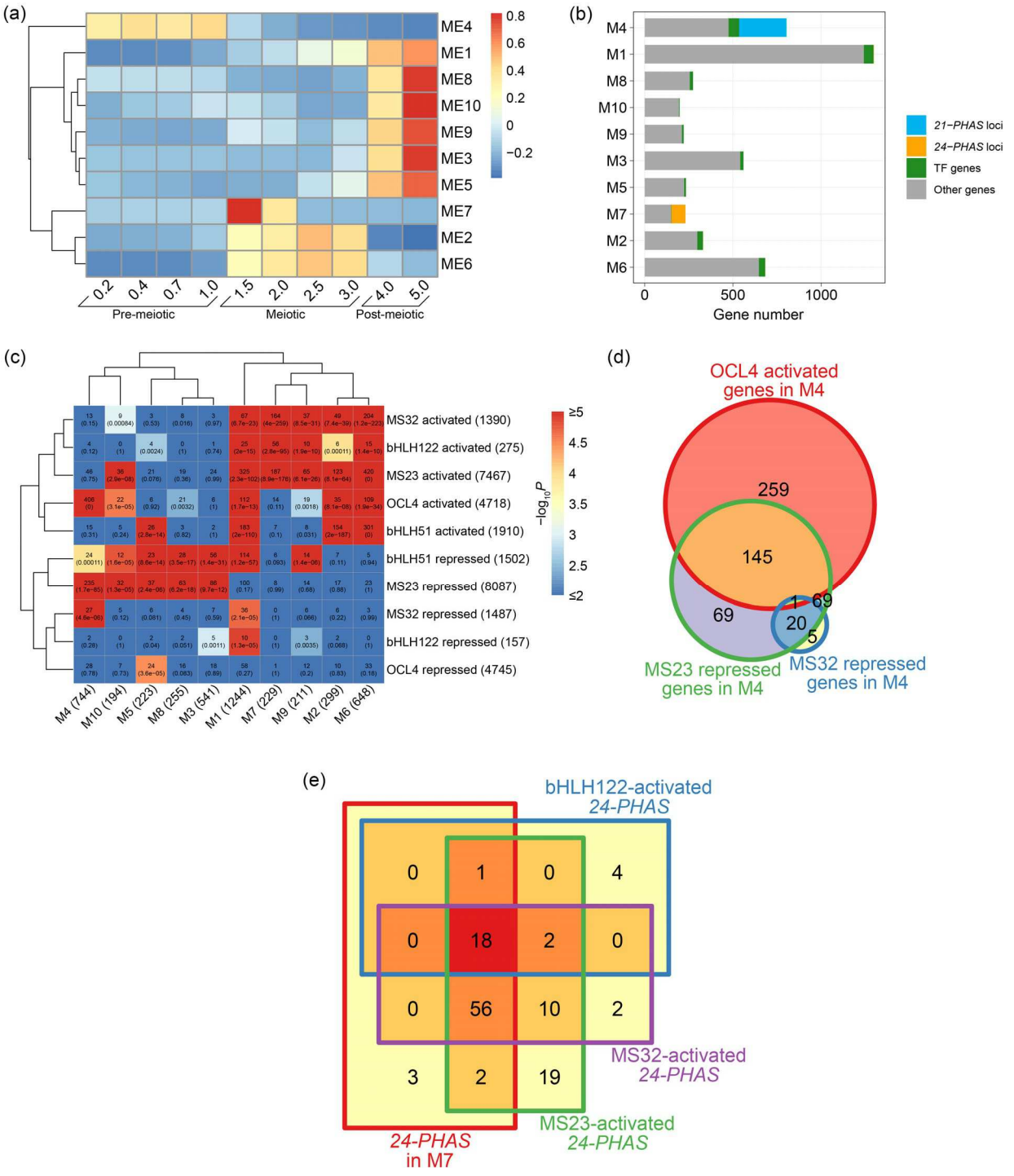


**Figure 2.** Temporal coexpression networks in maize anthers.

(a) Heat map of module eigengenes (MEs) of the 10 coexpression modules. Anther developmental stages (based on lengths, in mm) and phases (pre-meiotic, meiotic, and post-meiotic) are indicated on the *x*-axis. The heatmap was clustered on Euclidean distance.

(b) Numbers of *PHAS* loci, TF genes, and other genes in each module.

(c) Relationships of the coexpression modules and genes modulated by putative transcriptional regulators of *PHAS* loci. The heatmap denotes *P* values ($-\log_{10}$) of hypergeometric tests of overlapping genes in a given pair of gene sets. Boxes contain the numbers of overlapping genes and the associated *P* values (in parentheses). The total number of genes in each gene set is indicated in parentheses on the axes. The heatmap was hierarchically clustered on Euclidean distance.

(d) Venn diagram of M4 genes that are activated by OCL4 or repressed by MS23/MS32.

(e) Venn diagram of *24-PHAS* loci in M7 and the *24-PHAS* loci that are activated by MS23/MS32/bHLH122.

Table S4). M7, which includes genes preferentially expressed during meiosis, was enriched for 'meiotic chromosome segregation' and 'chiasma assembly', two biological processes specific to meiosis I. This is consistent with the anther developmental stages (1.5 and 2.0 mm) when M7 genes show the highest level of expression. M4 was significantly enriched for 'regulation of transcription, DNA-templated' and 'auxin-activated signaling pathway'; upon closer inspection of this module, we detected 60 TF genes (including *OCL4*) and many auxin signaling pathway genes including *Auxin import carrier1* (*AIC1*), *Auxin response factor4* (*ARF4*), *Pin-form4* (*PIN4*), and *Aux/IAA-transcription factor12* (*IAA12*), plus several *small auxin upregulated RNA* (*SAUR*)-like genes (Table S3), suggesting that TFs and auxin signaling pathway(s) play important roles in the regulation of pre-meiotic anther development. Notably, M6, which includes genes primarily expressed when anther length is in the range 1.5–3.0 mm, was enriched for 'sporopollenin biosynthetic process'. The genes associated with this GO term included two genes known to be essential for male fertility, *Ms10* (also known as *Abnormal pollen vacuolation1* [*APV1*]) and *Ms26*, both of which have been implicated in pollen exine formation (Chen et al., 2017; Djukanovic et al., 2013; Wan et al., 2020). Together, the GO term enrichments suggest that the coexpression networks accurately reflect the gene regulatory mechanisms underlying well-established anther developmental processes and are useful starting points for identifying TF regulators of anther development and *PHAS* loci expression.

An analysis of 36 previously reported maize male-sterility-associated genes (Albertsen et al., 2014; An et al., 2019; Chaubal et al., 2003; Chen et al., 2017; Cigan et al., 2001; Djukanovic et al., 2013; Evans, 2007; Fox et al., 2017; Han et al., 2022; Jiang et al., 2021; Liu et al., 2022; Moon et al., 2013; Nan et al., 2017, 2022; Qi et al., 2022; Somaratne et al., 2017; Tian et al., 2017; Vernoud et al., 2009; Wang et al., 2012, 2013, 2019; Xie et al., 2018; Zhang et al., 2018, 2021; Zhu et al., 2019) detected 22 of those genes in M1, M2, M4, M6 or M7 (Table S5). Among the putative regulators of *24-PHAS* loci and/or tapetal development, *Ms23* was assigned to M2, whereas *bHLH51* and *bHLH122* were assigned to M6 (Table S3), modules that cluster with M7 (Figure 2a) and each contains one *24-PHAS* locus. The prolonged expression patterns of the three bHLHs compared to the *24-PHAS* loci assigned to M7 support the previous observation that the TFs are involved in regulation of numerous protein-coding genes beyond *24-PHAS* loci (Nan et al., 2022). To determine whether any of the coexpression modules are (perhaps partly) regulated by OCL4 and/or the four bHLHs, we identified the genes and *PHAS* loci dysregulated in mutants of each gene using two recently published RNA-Seq data sets (Tables S6 and S7) (Nan et al., 2022; Zhai et al., 2015) and tested the differential expressed gene sets for enrichments of genes assigned to each coexpression module (Figure 2c). M4, the pre-meiotic module, was most significantly enriched ($P = 0$, hypergeometric test) for OCL4-activated genes, which we defined as genes that are downregulated in a male-sterile *ocl4* mutant compared to its fertile siblings. Moreover, 258 of the 269 *21-PHAS* loci assigned to M4 were downregulated in the *ocl4* mutant (Table S6). These data support the hypothesis that OCL4 regulates 21-nucleotide phasiRNA biogenesis. Interestingly, M4 was also significantly enriched ($P < 10^{-5}$, hypergeometric test) for genes that are repressed by MS23 or MS32 (i.e. genes upregulated in the *ms23/ms32* mutants). This is in line with the previous observation that an *ms23* mutant exhibited a defect in pre-meiotic anther development (i.e. excessive periclinal divisions in the cell layer that differentiates into tapetum) (Nan et al., 2017, 2022). A closer inspection of OCL4-activated or MS23/MS32-repressed genes in M4 showed that 146 of M4 genes were activated by OCL4 and repressed by MS23, whereas only one of the OCL4-activated genes was repressed by MS32 (Figure 2d). One interpretation of this pattern is that the earlier-acting OCL4-mediated processes must be repressed at the next stage of development by MS23 to facilitate normal anther developmental progression. These results confirm that OCL4, MS23, and MS32 play essential roles in regulating the gene networks active during pre-meiotic anther development, and suggest that OCL4 and MS23 may regulate a subset of those networks antagonistically.

The meiosis-associated modules M2, M6, and M7, preferentially expressed when anther length is in the range 1.5–3.0 mm, were enriched for genes activated by MS23, MS32, and bHLH122, whereas only M2 and M6 were enriched for bHLH51-activated genes (Figure 2c). This is consistent with the previous report that bHLH51 acts downstream and later than the other three bHLHs (Nan et al., 2022). Seventy-seven of the 80 *24-PHAS* loci assigned to M7 were downregulated in at least one of the *bhlh* single mutants, and the largest number of downregulated *24-PHAS* loci were detected in the *ms23* and *ms32* mutants (77 and 74 loci, respectively, in contrast to 0 and 19 loci downregulated in *bhlh51* and *bhlh122*, respectively) (Figure 2e), supporting major roles for MS23 and MS32 among the four bHLHs in activating reproductive *24-PHAS* loci (Nan et al., 2022). Notably, M2, M6, and the 'post-meiosis-specific' M1 were also enriched for OCL4-activated genes (Figure 2c), suggesting that the OCL4-regulated gene network remains active until anther development progresses beyond the pre-meiotic phase, and is likely associated with maintenance of epidermal cell identity (Vernoud et al., 2009; Zhai et al., 2015). Together, these results demonstrate complex interplay among the gene networks regulated by OCL4 and the bHLHs.

### *Trans*-activation assays of putative transcriptional regulators of pre-meiotic anther development and *21-PHAS* loci by bulk-protoplast RNA-Seq

We focused on M4 to identify putative transcriptional regulators of early anther development and *21-PHAS* loci activation. M4 includes OCL4 and 59 additional TFs from 25 known protein families (Table S3). To test whether any single TF is sufficient to activate the *21-PHAS* loci, or other genes in the module, we transiently expressed 23 TFs from 16 families, individually – including 22 with a GFP fused to the N-terminus (p35S::GFP-TF) and one with a C-terminal GFP (p35S::TF-GFP) (Table S8) – in maize leaf protoplasts and performed bulk-protoplast RNA-Seq to identify genes activated by each TF. We chose most of the TFs based on relatively high expression levels in pre-meiotic anthers, and also included a few that are expressed at lower levels (Figure S2). An analysis of the spatial expression patterns of these TF genes in a laser-capture-microdissection-derived RNA-Seq data set of 1.5- and 2.0-mm W23 *bz2* anthers (Zhou et al., 2022) showed that transcripts of 21 of the genes were detected in the data set, and 15 of the genes (except *ZHD5*, *GBP15*, and *NAC119*, *TCP23*, *DOF7*, and *MYB131*) are preferentially expressed in 'other somatic layers' (including epidermis, endothecium, and middle layer) compared to tapetal cells and meiocytes (Figure S3). The fact that a few of the genes were not detected in other somatic layers of 1.5- or 2.0-mm anthers was likely because the two stages are well beyond the stages (0.2–1.0 mm) when the *21-PHAS* loci and their putative transcriptional regulators are preferentially expressed. Therefore, we speculate that the majority, if not all, of the tested TF genes are expressed in the anther epidermis, consistent with the model that the *21-PHAS* loci are activated in the anther epidermis (Zhai et al., 2015). Differential expression analyses showed that, compared to the control construct expressing GFP alone (p35S::GFP), 20 of the TF constructs activated 0 to 4 genes [fold change (FC) > 1.5; FDR < 0.05] (Figure 3a; Figure S6; Table S9). For example, OCL4 activated three genes, none of which were a member of M4. Notably, two of those genes – GRMZM2G043295 (encoding a putative anthocyanidin 5,3-*O*-glucosyltransferase) and GRMZM2G076818 (encoding an uncharacterized protein) – were also identified as OCL4-activated genes based on RNA-Seq of the *ocl4* mutant (Table S6). These data suggest that the 20 TFs may require endogenous cofactors to activate the full suite of downstream genes or that they need to be post-translationally modified by anther stage-specific factors to become functional TFs.

The other three TFs – EREB69, MYB131, and NAC29 – activated larger numbers of genes (205, 439, and 72, respectively) in protoplasts (Figure 3a; Figure S6; Table S9). Five of the MYB131-activated genes (*DOF7*, *TCP23*, GRMZM2G071846, GRMZM2G129304, and GRMZM2G369703), and one of the NAC29-activated genes (GRMZM2G105987) were members of M4. Comparison of the EREB69/MYB131/NAC29-activated gene sets with the genes expressed in the W23 *bz2* anthers/pollen indicated that the majority of the genes (114 EREB69-activated, 398 MYB131-activated, and 68 NAC29-activated) were expressed in anthers or pollen (Figure 3b), suggesting that the activated genes are endogenous targets of the three TFs in anthers. By GO enrichment analysis, the EREB69-activated genes are putatively involved in 'borate transmembrane transport', whereas the MYB131-activated genes are involved in 'cinnamic acid biosynthetic process' (Table S10). Borate is required for pollen germination (Rawson, 1996), whereas cinnamic acid is an essential component of the pollen coat (Battat et al., 2019), supporting the important roles of EREB69 and MYB131 in regulating pollen development. Notably, MYB131 strongly activated *AGO18b* ($\log_2$FC = 5.2), which encodes an Argonaute protein that loads 21-nucleotide phasiRNAs, 24-nucleotide phasiRNAs, and several miR2275-family miRNAs in maize anthers (Sun et al., 2019), whereas NAC29 activated *AGO5a* ($\log_2$FC = 3.9), which encodes a paralog of the MALE-ASSOCIATED ARGONAUTE-1 (MAGO1; aka AGO5b) and MAGO2 (AGO5c) proteins; AGO5b and AGO5c load 21-nucleotide phasiRNAs in maize anthers (Lee et al., 2021). These data suggest that MYB131 and NAC29 are involved in regulating the stabilization or functions of 21- and 24-nucleotide phasiRNAs. Together, the results of our *trans*-activation assays using bulk leaf protoplasts suggest that most of the tested TFs require endogenous transcriptional coregulators or processes such as post-translational modifications to fully activate the downstream gene networks, and also that a few of the TFs are involved in the regulation of the biogenesis and activities of phasiRNAs or their miRNA triggers.

### Screen for TF combinations co-regulating reproductive *PHAS* loci by single-cell RNA-Seq

To determine whether any combinations of the pre-meiotic TFs can co-activate *21-PHAS* loci or genes associated with the 21-nucleotide phasiRNA pathway, we transiently expressed in maize leaf protoplasts a pool of all constructs used in the protoplast RNA-Seq analysis described above, including 23 TF constructs and the p35S::GFP plasmid. GFP-positive protoplasts were isolated using a Namocell Hana single-cell dispenser, and 96 of the single protoplasts – referred to as PS1 (protoplast set 1) hereafter – were processed for single-cell (protoplast) RNA-Seq. To determine whether any combinations of MS23, MS32, bHLH51, and bHLH122 can co-activate *24-PHAS* loci or 24-nucleotide phasiRNA pathway genes, we also performed a single protoplast RNA-Seq experiment of another set of 96 protoplasts

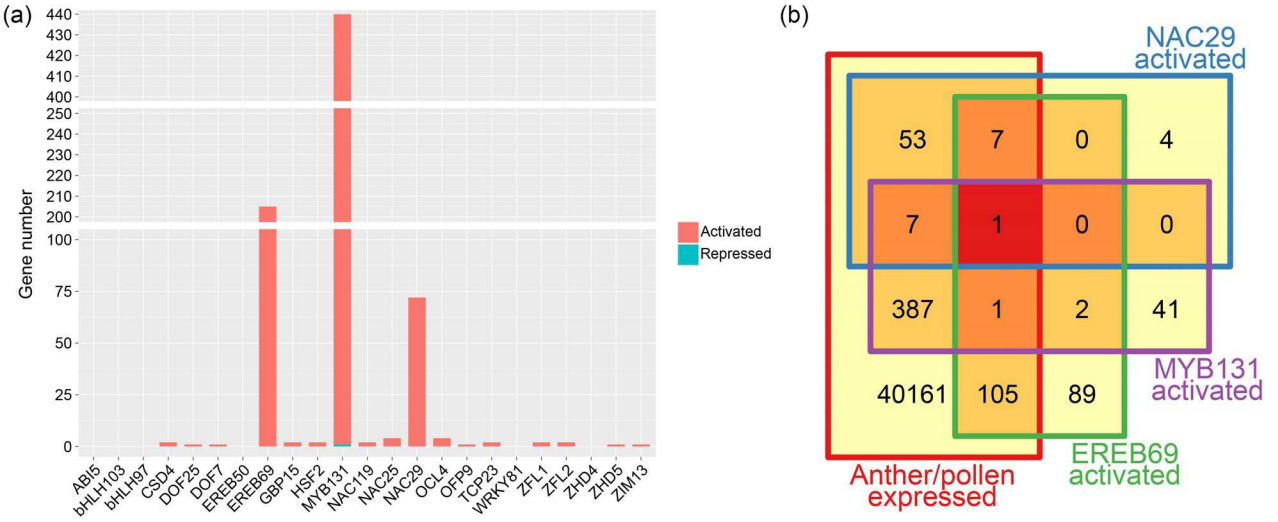


**Figure 3.** Differential gene expression analysis of bulk protoplasts expressing pre-meiotic TFs.
(a) Bar graphs of numbers of genes that are dysregulated in protoplasts expressing individual TFs. All TFs except ZEA FLORICAULA LEAFY1 (ZFL1) are members of M4. ZFL1 is a paralog of ZFL2.
(b) Venn diagram of the anther/pollen-expressed gene set and gene sets activated by EREB69/MYB131/NAC29.

(protoplast set 2; PS2), transfected with a pool of eight constructs expressing the four bHLHs, each as N- and C-terminal fusions with GFP, as well as p35S::GFP (Table S11). A PCA of all 192 protoplasts showed that the two cell populations were well separated on the second principal component (Figure 4a), indicating distinct gene expression programs activated in protoplasts by the two pools of TF constructs. Within each set of protoplasts, the protoplasts expressing the larger numbers of constructs are generally distinct from those with no detectable construct-specific transcripts (Figure 4a), further supporting the notion that the diverged transcriptome profiles of individual protoplasts result from ectopic expression of the TFs.

An examination of construct-specific transcripts in PS1 indicated that each protoplast expressed 0–23 TF constructs, whereas protoplasts in PS2 expressed one to eight TF constructs; seven PS1 and five PS2 protoplasts expressed the p35S::GFP plasmid [transcripts per million (TPM) > 0] (Figure 4b; Table S12). The percentages of transfected TF constructs expressed in individual protoplasts are significantly higher in PS2 than PS1 ($P < 2.2 \times 10^{-16}$, Student's *t*-test) (Figure 4c). Because PS2 was transfected with generally larger copy numbers of individual plasmids compared to PS1 (Table S8), we analyzed all protoplasts in the two sets for correlation between plasmid copy numbers and the numbers of protoplasts that express them. Our result indicates that the correlation was strong and significant (SCC = 0.69; $P = 7.6 \times 10^{-6}$) (Figure 4d). In both pools, we included a much lower copy number of the p35S::GFP plasmid than the TF plasmids (Table S8). As a result, the number of protoplasts in each pool expressing p35S::GFP were low compared to other constructs (Figure 4d; Table S13). These data suggest that the number of protoplasts that express a given construct is at least in part determined by the copy number of the plasmid in the transfection cocktail.

An analysis of the abundance of individual *PHAS* precursors in single protoplasts detected modest activation of several *21-PHAS* and *24-PHAS* loci in a subset of the PS1 and PS2 protoplasts compared to the protoplasts expressing no TFs (Figure 4e,f). We further examined the activation of individual genes associated with the 21- and/or 24-nucleotide phasiRNA pathways, including *DCL4*, *DCL5*, *AGO5b*, *AGO5c*, and *AGO18b* (and *AGO18a*, a paralog of *AGO18b*). Each of these genes are activated in one or more of the protoplasts (Figure S7); however, the activation of these individual genes was not well replicated and could not be correlated with the presence or absence of specific TFs. Furthermore, an analysis of the genes detected as activated by EBEB69, MYB131, or NAC29 in our bulk protoplast RNA-Seq assays showed that some of them are expressed in the single protoplasts expressing each of the three TFs (often together with other TFs; Figure S8). Similar to the *PHAS* loci, we did not observe any correlation between the presence of each of the three TFs and activation of the downstream genes. Thus, we conclude that the TF combinations expressed in the individual protoplasts cannot activate their downstream gene networks in a reproducible manner, at least in this system when assessed via single-cell RNA-Seq.

## DISCUSSION

We previously characterized the temporal accumulation patterns of reproductive phasiRNAs, their precursors, and

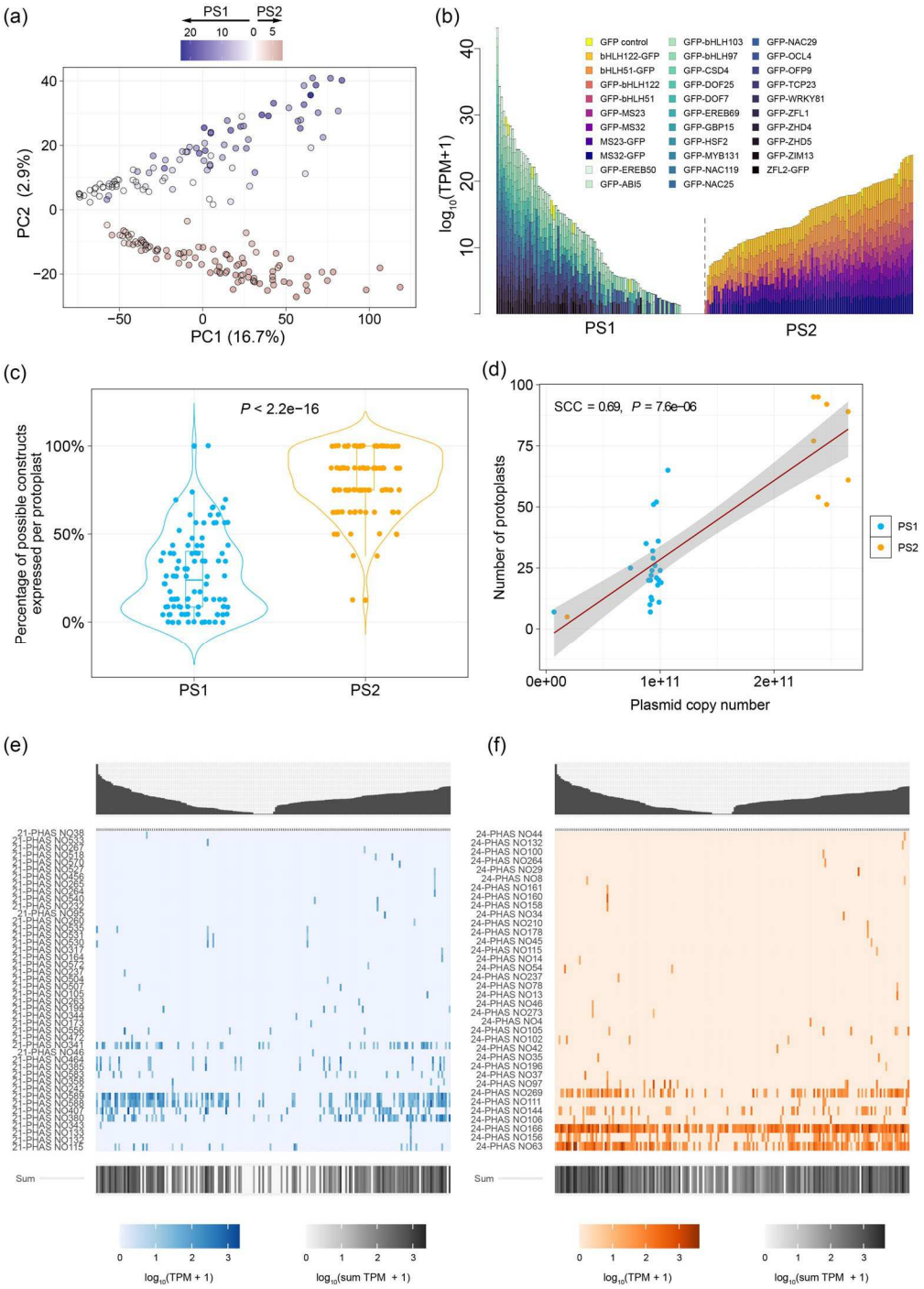


**Figure 4.** Single-cell RNA-Seq analysis of protoplasts transfected with plasmid pools.
(a) Principal component analysis plot of all protoplasts in PS1 and PS2. The percentage of variance explained by each principal component is given in parentheses. Color key, number of constructs expressed in individual protoplasts of each set.
(b) Abundance of construct-specific transcripts in each protoplast. On the *x*-axis, protoplasts in each set were ordered based on the total abundance of construct-specific transcripts (in decreasing order for PS1 but in increasing order for PS2).
(c) Violin plot showing percentages of TF constructs expressed in each protoplast. The percentages are provided in Table S12.
(d) Scatterplot and SCC analysis of plasmid copy numbers versus the number of protoplasts expressing each plasmid. The regression line was calculated using the generalized linear regression method, and the shaded area represents the SE. Input data for this plot are provided in Table S13.
(e) Expression levels of detected *21-PHAS* loci in the single protoplasts.
(f) Expression levels of detected *24-PHAS* in the single protoplasts. In (e, f), the columns representing individual protoplasts are ordered the same way as in (b), with PS1 on the left.

miRNA triggers in maize anthers (Zhai et al., 2015). Here, to identify putative transcriptional regulators of anther development and *PHAS* locus activation, we began by analyzing the previously generated RNA-Seq data set. Our analyses highlight extensive transcriptional reprogramming at the pre-meiotic stages, as indicated by a strong enrichment of TFs in the pre-meiotic module (M4) (Figure S1; Table S4). M4 was also enriched for auxin signaling pathway genes (Figure S1; Table S4), suggesting a role for auxin signaling in pre-meiotic anther development. In Arabidopsis (*Arabidopsis thaliana*), auxin plays a key role in regulating germinal cell specification (Zheng et al., 2021); whether the associated molecular mechanisms are conserved between Arabidopsis and maize is a question that is worthy of further exploration. In addition to the coexpression modules associated with pre-meiotic or meiotic anther development and *PHAS* loci activation, our network analyses identified several modules (M1, 3, 5, 8, 9, and 10) that are preferentially expressed at the post-meiotic stages of anther development (Figure 2a), when the microspore undergoes two mitoses to give rise to pollen grains. Several genes in M1, including TF genes *Lateral organ boundaries domain10* (*LBD10*) and *Ms7*, have been shown to be necessary for male fertility (Table S5). Thus, further exploration of these modules may uncover transcriptional networks that are active during gametogenesis and are crucial for male fertility.

A significant previous finding was the discovery of the coordinated accumulation patterns of reproductive phasiRNAs and their precursors (Zhai et al., 2015); here, we confirm that the majority of individual *21-PHAS* or *24-PHAS* precursors detectable by RNA-Seq are highly coordinated (Figure 2a; Figure S4), suggesting that the *21-PHAS* and *24-PHAS* loci are each regulated by a common set of upstream regulators. Our coexpression network analysis detected many TFs that are temporally coexpressed with *21-PHAS* loci in M4 (Figure 2b; Table S3), including OCL4, which had previously been implicated in regulating *21-PHAS* loci and pre-meiotic anther development. By contrast, only two TF genes – *Homeobox126* (*HB126*) and *Golden2-like33* (*GLK33*) – were assigned to the same module as the majority of expressed *24-PHAS* loci (Table S3), suggesting strongly that these two TFs regulate *24-PHAS* loci. Nonetheless, three of the previously described putative regulators of *24-PHAS* loci were assigned to M2 (*Ms23*) and M6 (*bHLH51* and *bHLH122*), which are modules with distinguishable yet similar expression profiles compared to M7 (Figure 2a; Table S3), highlighting the possibility that M2 and/or M6 include additional TF regulators of *24-PHAS* loci. Furthermore, 356 of the 463 *21-PHAS* loci were detected as expressed (and 269 met the criteria to be selected for WGCNA), whereas 121 of the 176 *24-PHAS* were expressed (and 84 were selected for WGCNA) (Tables S2 and S3); the latter is consistent with a recent study reporting that a subset of the *24-PHAS* loci show either low expression levels or low variability in expression levels across anther developmental stages (Nan et al., 2022). Together, these data suggest that *21-* and *24-PHAS* loci are each regulated by distinct mechanisms, and the loci detected by WGCNA are more coordinately regulated than the others.

Further analyses of enrichments of genes regulated by OCL4, or each of the four bHLHs, in the coexpression modules showed that M4 was significantly enriched for OCL4-activated genes, and M2, M6, and M7 were enriched for bHLH-regulated genes (Figure 2c). The latter three modules presumably contain direct targets of the bHLHs. The differential enrichments of genes regulated by individual bHLH TFs in the coexpression modules support their overlapping, yet diverged, regulatory roles (Nan et al., 2022). Interestingly, we also identified a set of genes that are antagonistically regulated by OCL4 and MS23, and documented that several modules (M1, M2, and M6) are preferentially expressed in anthers larger than 1.0 mm and were enriched for OCL4-activated genes (Figure 2c). This suggests that the regulatory role of OCL4 remains active beyond the pre-meiotic phase, and that the gene networks regulated by OCL4 and the bHLHs are likely interconnected.

Our bulk-protoplast RNA-Seq analyses of 23 individual putative transcriptional regulators of M4 genes described that three TFs can activate a notable number of downstream genes ectopically, including a few *AGO* genes that have been implicated in 21-nucleotide (and 24-nucleotide) phasiRNA pathway(s) (Figure 3a; Table S9). None of the TFs (including OCL4, the previously reported putative regulator of *21-PHAS* loci) could activate *21-PHAS* loci in protoplasts. Among the reported putative transcriptional regulators of *24-PHAS* loci, MS23 MS32, and bHLH122 have been shown to co-activate *DCL5* in a synergistic manner in maize leaf protoplasts, whereas none of the individual bHLHs or their combinations could activate a select *24-PHAS* locus (Nan et al., 2022). Together, these results suggest that TF regulators may need to form complexes, which probably do not form properly in leaf protoplasts, to activate *PHAS* loci. Alternatively, but not mutually exclusively, the activation of the *PHAS* loci may require chromatin modification or post-translational TF modification processes that occur specifically in anthers, although evidence directly supporting this notion has yet to be obtained.

Our single-protoplast RNA-Seq data showed that none of the TF combinations could activate *PHAS* loci reproducibly (Figure 4e,f; Table S11), suggesting that either a larger number of protoplasts are necessary to yield reproducible results, or that, as discussed above, the endogenous TF complexes, post-translational events, and/or chromatin states are necessary for reproducible activation of *PHAS* loci. We included p35S::GFP in the two plasmid pools with the goal to obtain cells expressing only p35S::

GFP, which would be informative for differential gene expression analyses, but no protoplasts exhibited this status. As such, differential expression analyses to detect *trans*-activation of individual genes are not feasible in the present study. We also noted that several of the single protoplasts in PS1 (i.e. protoplasts 1, 3, 4, 8, 9, 18, 39, 54, 81, 86, and 91) did not accumulate detectable transcripts derived from the exogenous constructs (Table S12). This was surprising because all single protoplasts were dispensed based on fluorescence. The total unique molecular identifier (UMI) counts of these single protoplasts are significantly lower than the other protoplasts in PS1 ($P = 0.027$, Student's *t*-test) (Figure S9), suggesting that the low abundances of construct-specific transcripts could reflect low sequencing depth.

Future optimization by adjusting the amount of plasmid DNA, increasing the number of protoplasts to be sequenced, and/or increasing sequencing depths may enable meaningful differential gene expression analyses. In support of this idea, our analysis of the correlation between plasmid copy numbers used for transfection and the number of protoplasts expressing them (Figure 4d) suggests that reducing plasmid copy numbers appears to be promising for yielding protoplasts expressing smaller numbers of constructs, such as only the p35S::GFP construct, or one to a few of the TF constructs. For TF–target interactions that require endogenous chromatin states or other anther stage-specific features, generating protoplasts from the corresponding tissues would be crucial for the success of *trans*-activation assays. For example, protoplasts prepared from pre-meiotic or meiotic anther tissues may be ideal for assays of the *21-PHAS* and *24-PHAS* loci. We also note that, because 2132 TF genes are expressed in maize anthers/pollen (Table S2), a much larger-scale *trans*-activation assays of many TFs may be necessary to identify the potential transcriptional regulators of *PHAS* loci or other anther genes; this is becoming increasingly feasible, reflecting rapid developments of technologies such as single-cell RNA-Seq. Based on these insights, we assert that the single-protoplast RNA-Seq approach has great potential for future *trans*-activation assays of TF combinations and the discovery of gene regulatory networks.

In summary, our analyses provide valuable insights into the gene networks controlling the major developmental events occurring in maize anthers. The results demonstrate that the activation of reproductive *PHAS* loci likely requires TF complexes, post-translational TF modifications, and/or chromatin modification processes that are active in maize anthers, although certain phasiRNA pathway genes such as *AGOs* can be activated ectopically. Furthermore, our single protoplast analyses suggest that the single-cell RNA-Seq approach has the potential to be applied to high-throughput analyses of target activation by individual or multiple TFs.

## EXPERIMENTAL PROCEDURES

### Plant materials and growth

Seeds of B73 maize were germinated and grown in constant darkness at 25°C and 50–70% relative humidity. Leaves were collected from more than 10 etiolated seedlings 10–12 days after germination for each protoplast isolation.

### Maize leaf protoplast transfection

Plasmid constructs for expressing individual TFs as N- or C-terminal fusions with GFP (p35S::GFP-TF or p35S::TF-GFP) were generated using previously described methods (Nan et al., 2022; Yang et al., 2017). The sources of coding DNA sequence (CDS) clones are provided in Table S8. ZFL1 was accidentally amplified with the ZFL2 primers and subsequently included in downstream experiments to potentially characterize the redundancy between the two paralogs. The p35S::GFP plasmid was used as the negative control for the bulk-protoplast RNA-Seq analyses, and was included in both plasmid pools for single protoplast RNA-Seq. Leaf protoplasts were transfected using a previously described poly(ethylene glycol)/calcium-mediated method (Gomez-Cano et al., 2019; Nan et al., 2022). For the bulk-protoplast RNA-Seq experiments, 15 ng of plasmid DNA was used for each transfection (of approximately 0.25 million protoplasts), and three transfections were pooled to yield one biological replicate. Two biological replicates, each derived from one protoplast isolation, were performed for each construct. Concurrently, one biological replicate (three pooled transfections) with the p35S::GFP plasmid was performed for every protoplast isolation to use as controls for differential expression analyses (Table S1). For the single protoplast RNA-Seq experiments, 625 ng of each of the 23 pre-meiotic TF constructs was mixed with 31.25 ng of the p35S::GFP plasmid to comprise plasmid pool 1, and 1666.7 ng of each of the 8 bHLH constructs plus 83.3 ng of the p35S::GFP plasmid constituted pool 2. The two plasmid pools were used for transfection of two protoplast samples (approximately 0.25 million protoplasts each) derived from a single isolation. Using a Hana single-cell dispenser (Namocell Inc., Mountain View, CA, USA), GFP-positive protoplasts were dispensed into 96-well plates (one protoplast per well), each well containing 0.8 μl Primer Master Mix [0.225% Triton X-100 (Sigma-Aldrich, St Louis, MO, USA), 1.6 mM dNTP mix (New England Biolabs, Ipswich, MA, USA), 1.875 μM barcoded oligo(dT) CEL-seq2 primers (1 sec up to 96 sec)] (Table S14) and frozen at –80°C. Protoplast transformation efficiencies were approximately 10–30% as determined using an AxioZoom microscope (Zeiss, Oberkochen, Germany) with green-light filters.

### RNA-Seq library preparation and sequencing

For each bulk-protoplast RNA-Seq sample (biological replicate), approximately 0.75 million protoplasts (three transfections pooled) per sample were subjected to RNA extraction using the TRI reagent (Sigma-Aldrich), yielding approximately 1.4–2.8 μg of total RNA. After treatment with DNAase I (New England Biolabs) and cleaned using the Zymo RNA Clean and Concentrator kit (Zymo Research, Irvine, CA, USA), 300–400 ng RNA per sample was used for RNA-Seq library preparation using the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs) in accordance with the manufacturer's instructions for use with the NEBNext Poly(A) mRNA magnetic isolation module. Libraries were sequenced on a NextSeq 550 instrument (Illumina, Inc., San Diego, CA, USA) at the University of Delaware DNA Sequencing and Genotyping Center to produce 76-nucleotide single-end reads.

Libraries for the single protoplast assays were prepared following a modified CEL-seq2 protocol (Hashimshony et al., 2016). Plates of protoplasts were incubated at 65°C for 3 min, spun, then incubated again at 65°C for 3 min then placed on ice. To each well, 0.7 μl of reverse transcription mix (8:2:1:1 of Superscript IV 5× Buffer, 100 mM DTT, RNase Inhibitor, Superscript IV; Thermo Fisher Scientific, Waltham, MA, USA) was added, the plates were spun and incubated at 42°C for 2 min, 50°C for 15 min, and 55°C for 10 min, and then placed on ice. The samples were pooled by row into eight-strip tubes (eight pooled samples per plate) and excess primers were digested with the addition of 4.6 μl of exonuclease I mix (2.5 μl of 10× Exonuclease I Buffer, 2.1 μl of Exonuclease I; New England Biolabs) and incubated at 37°C for 20 min and 80°C for 10 min, and then placed on ice. The pooled samples were purified with 44.28 μl (1.8× volume) of pre-warmed RNAClean XP beads (Beckman Coulter Life Sciences, Indianapolis, IN, USA), washed twice with 80% ethanol, and the RNA eluted from the beads with 7 μl of RNase-free water. Second strand synthesis was initiated with the addition of 3 μl of second strand synthesis mix (2.31 μl Second Strand Reaction dNTP-free Buffer, 0.23 μl of 10 mM dNTPs, 0.08 μl of DNA ligase, 0.3 μl of DNA polymerase I, 0.08 μl of RNase H; New England Biolabs) and then incubated at 16°C for 4 h. The RNAClean XP beads were removed from the pooled samples with a magnetic rack and the eight pooled samples per original plate were further pooled into a single tube. The pooled samples were purified with 96 μl (1.2× volume) of AMPure XP beads (Beckman Coulter Life Sciences) followed by two washes of 80% ethanol, and eluted with 6.4 μl of RNase-free water. *In vitro* transcription was initiated with the addition of 9.6 μl of MegaScript T7 IVT mix (1:1:1:1:1:1 of CTP solution, GTP solution, UTP solution, ATP solution, 10× Reaction Buffer, T7 Enzyme Mix; Thermo Fisher Scientific) to the pooled samples, followed by incubation at 37°C overnight. The leftover AMPure XP beads were removed with a magnetic rack and 28.8 μl (1.8× volume) of prewarmed RNAClean XP beads (Beckman Coulter Life Sciences) were added and washed with ethanol, as previously described, then eluted in 6.5 μl of RNase-free water. The amplified RNA quality and quantity for both pooled samples were analyzed with an RNA 6000 Pico chip on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The RNAClean XP beads were removed from the pooled samples and 1.5 μl of priming mix (9:5:1 of RNase-free water, 10 mM dNTPs, 1 M tagged random hexamer primer) (Table S14) was added and incubated at 65°C for 5 min, and then placed on ice. A second round of reverse transcription was initiated with the addition of 4 μl of reverse transcription mix (4:2:1:1 of First Strand Buffer, 0.1 M DTT, RNaseOUT, SuperScript II; Thermo Fisher Scientific) to each pooled sample then incubated at 25°C for 10 min, 42°C for 1 h, and 70°C for 10 min, before being placed on ice. For the final polymerase chain reaction (PCR), 5.5 μl of pooled sample was added to 21 μl of PCR master mix with Illumina TruSeq Small RNA PCR primer (RP1) and Index Adaptors (RPI1 and RPI2) (Table S14) (6.5 μl of RNase-free water, 12.5 μl of Ultra II Q5 Master Mix, 1 μl of 10 μM RP1, 1 μl of 10 μM RPI 'X'). Libraries were amplified with 13 rounds of PCR (98°C for 30 sec, then 13 cycles of 98°C for 10 sec, 65°C for 15 sec, and 72°C for 30 sec, and finished with 72°C for 3 min). The final PCR products were purified with 26.5 μl (1.0× volume) of AMPure XP beads (Beckman Coulter Life Sciences), then washed twice with ethanol and eluted with 25 μl of RNase-free water, and purified again with 25 μl (1.0× volume) of AMPure XP beads (Beckman Coulter Life Sciences), washed twice, and eluted into 26 μl of RNase-free water. The libraries were assessed with an Agilent High Sensitivity DNA chip (Agilent Technologies) and sequenced on a NextSeq 550 instrument at the University of Delaware DNA Sequencing and Genotyping Center to produce 2 × 150-nucleotide paired-end reads with 20% PhiX.

### RNA-Seq data analysis

The previously published RNA-Seq data of (i) wild-type W23 *bz2* and the *ocl4* mutant (Zhai et al., 2015) and (ii) the *ms23/m32/bhlh51/bhlh122* mutants (Nan et al., 2022) were obtained from the Meyers Laboratory Next-Gen Sequence Databases (https://mpss.meyerslab.org/) (Nakano et al., 2020); information about individual libraries is provided in Table S1. Those raw reads, and the bulk-protoplast RNA-Seq reads generated in the present study, were mapped to the maize reference genome (B73 RefGen_v3; Schnable et al., 2009) using HISAT, version 2.1.0 (Kim et al., 2015), with intron sizes set to 30–8000 bp (Zhan et al., 2015). Reads mapped to each gene were counted using the featureCounts program of the Subread package, version 1.6.3 (Liao et al., 2019). Read counts of the expressed genes, which we define as those with raw counts per million (CPM) > 1 in (i) at least 2 of the 50 bulk-protoplast RNA-Seq libraries or (ii) at least one library for the previously published data sets, were normalized using EDGER, version 3.20.9 (Robinson et al., 2010). The SCC analysis and PCA were performed with $\log_2$-transformed RPKM data of the expressed genes. Differential expression analyses were also performed using EDGER with a generalized linear model-based method; FC >1.5 and FDR < 0.05 were used as the cutoff criteria to call differentially expressed genes. Coexpression network analysis of the anther RNA-Seq data was performed using WGCNA, version 1.64, on the $\log_2$-transformed RPKM values of expressed genes as described previously (Zhan et al., 2015) with the following modifications: (i) the matrix of pairwise SCCs between genes was transformed into a connection strength matrix by raising the correlation matrix to the power of 16 and (ii) after the hierarchical clustering tree was cut with the Dynamic Tree Cut algorithm (Langfelder et al., 2007), modules with fewer than 100 genes were merged into their closest neighboring module to generate the final modules. Functional and GO annotations of maize genes were obtained from Ensembl Plants (Bolser et al., 2017), Gramene (Tello-Ruiz et al., 2022), MaizeGDB (Portwood et al., 2019), Plant TFDB (Tian et al., 2020), and Grassius (Yilmaz et al., 2009). GO term enrichment analyses were performed using BLAST2GO, version 3.0.11 and OMICSBOX, version 2.0.36 (Götz et al., 2008). Annotations of *21-PHAS* and *24-PHAS* loci were obtained from Zhai et al. (2015) and the genomic coordinates were converted from B73 RefGen_v2 to B73 RefGen_v3 using the Ensembl Plants Assembly Converter.

The raw CEL-seq2 reads were demultiplexed based on cell-specific barcodes (Table S14) using Fastq-Multx (Aronesty, 2013). The UMI sequences from read 1 were added to the read 2 sequence names and then filtered and trimmed with FASTP (parameters: -y -x -3 -f 6) (Chen et al., 2018). The clean reads were mapped to B73 RefGen_v3 with HISAT, and the UMIs quantified with SAMTOOLS (Li et al., 2009) and UMI-TOOLS (Smith et al., 2017). The counts of UMIs mapped to each gene/construct were normalized to TPM (Table S11). Construct-specific transcripts were only counted for a 4-bp region at the junction between the CDS of a given TF (or GFP for the p35S::GFP construct) and the vector backbone. PCA was performed with $\log_{10}$(TPM + 1) of the expressed genes (raw UMI count > 0 in at least one library).

## ACKNOWLEDGEMENTS

This work is supported by funding and resources from the Donald Danforth Plant Science Center and by National Science Foundation (NSF) grant IOS-1754097. DBM is funded by NSF postdoctoral

fellowship PRFB IOS-1907220. We thank Rachel Agoglia and Kyle Schramer (Namocell Inc.) for assistance with single-cell dispensation, Brewster Kingham and Mark Shaw (University of Delaware DNA Sequencing & Genotyping Center) for assistance with Illumina sequencing, Mayumi Nakano for assistance with data handling, and Joanna Friesner for assistance with editing.

## AUTHOR CONTRIBUTIONS

JZ, CT, and BCM designed the project. JZ, LO, DBM, and CT performed experiments. JZ, LO, and DBM analyzed data. JZ, LO, and DBM drafted the manuscript. All authors edited the manuscript.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The bulk and single protoplast RNA-Seq data reported in this article have been deposited in the NCBI Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE210330 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE210330). GEO accession numbers of individual libraries are provided in Table S1.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** GO term enrichments in the WGCNA-identified coexpression modules. Only biological process terms are shown. Gray indicates no significant enrichment. The heat map was generated using data from Table S4.

**Figure S2.** Expression patterns of M4 TFs during anther development. Arrows indicate genes chosen for *trans*-activation assays. Raw RNA-Seq data were from Zhai et al. (2015). Anther developmental stages (in mm) are indicated on the *x*-axis.

**Figure S3.** Expression patterns of M4 TFs based on laser-capture-microdissection-derived RNA-Seq data from Zhou et al. (2022). Numbers on the *x*-axis represent anther developmental stages (in mm). FA, fixed anthers; ME, meiocytes; OSC, other somatic cells; TAP, tapetum.

**Figure S4.** *Z*-score plots showing scaled expression profiles of genes in the coexpression modules identified using WGCNA. Gray lines represent individual genes, and red lines the median. Anther developmental stages (in mm) are indicated on the *x*-axes.

**Figure S5.** Circos plot showing genomic positions of expressed *21-* and *24-PHAS* loci. Circular tracks from outside to inside: genomic positions by chromosomes, expressed *21-PHAS* that are not assigned to M4, *21-PHAS* in M4, expressed *24-PHAS* that are not assigned to M7, and *24-PHAS* in M7. Height of dots represents maximal expression levels of *PHAS* loci (in RPKM).

**Figure S6.** Mean-difference plots of all the genes tested for differential expression in bulk protoplasts. Differentially expressed genes are represented by red dots, and the other genes black. Horizontal blue lines indicate 1.5-fold changes (i.e. $|\log_2 FC| = 0.59$).

**Figure S7.** Expression patterns of known/putative phasiRNA pathway genes in the single protoplasts. Columns representing individual protoplasts are in the same order as Figure 4(b).

**Figure S8.** Expression levels of EREB69/MYB131/NAC29-activated genes in the single protoplasts. Rows of heatmaps represent individual genes activated by EREB69 (a), MYB131 (b) or NAC29 (c).

**Figure S9.** Violin plot showing distribution of total UMI counts ($\log_{10}$ transformed) in the PS1 protoplasts with zero construct-specific transcripts compared to all other protoplasts in PS1.

**Table S1.** Summary of RNA-Seq data generated/analyzed in the present study.

**Table S2.** Normalized expression levels in RPKMs (mean of biological replicates) and functional annotations of genes expressed in the W23 *bz2* anthers (0.2–5.0 mm) and mature pollen.

**Table S3.** Module assignments, module memberships, and functional annotations of genes selected for coexpression network analysis.

**Table S4.** GO term enrichments of the coexpression modules.

**Table S5.** Summary of previously reported male-sterility-associated genes in maize.

**Table S6.** Genes dysregulated in the *ocl4* mutant.

**Table S7.** Genes dysregulated in the *bhlh* mutants.

**Table S8.** Sources of CDS clones, sequences of primers, and plasmid copy numbers used in the present study.

**Table S9.** Results of differential gene expression analyses of the bulk protoplasts expressing individual transcription factors.

**Table S10.** GO term enrichments of the gene sets activated by EREB69, MYB131 or NAC29.

**Table S11.** Normalized expression levels in TPMs of genes expressed in single protoplasts.

**Table S12.** Normalized abundance of construct-specific transcripts in TPMs.

**Table S13.** Copy numbers of plasmid constructs and the number of protoplasts expressing them.

**Table S14.** Primer sequences for CEL-seq2 library construction.

## REFERENCES

**Albertsen, M.C., Fox, T.W., Leonard, A.L., Li, B., Loveland, B.R. & Trimnell, M.** (2014) *Cloning and use of the ms9 gene from maize*. World Patent. Available at: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2014152447 [Accessed 20th July 2022].

**An, X., Dong, Z., Tian, Y., Xie, K., Wu, S., Zhu, T. *et al.*** (2019) *ZmMs30* encoding a novel gdsl lipase is essential for male fertility and valuable for hybrid breeding in maize. *Molecular Plant*, **12**, 343–359.

**Araki, S., Le, N.T., Koizumi, K., Villar-Briones, A., Nonomura, K.-I., Endo, M. *et al.*** (2020) miR2118-dependent U-rich phasiRNA production in rice anther wall development. *Nature Communications*, **11**, 3115.

**Aronesty, E.** (2013) Comparison of sequencing utility programs. *Open Bioinformatics Journal*, **7**, 1–8.

**Battat, M., Eitan, A., Rogachev, I., Hanhineva, K., Fernie, A., Tohge, T. *et al.*** (2019) A MYB triad controls primary and phenylpropanoid metabolites for pollen coat patterning. *Plant Physiology*, **180**, 87–108.

**Bolser, D.M., Staines, D.M., Perry, E. & Kersey, P.J.** (2017) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. *Methods in Molecular Biology*, **1533**, 1–31.

**Chaubal, R., Anderson, J.R., Trimnell, M.R., Fox, T.W., Albertsen, M.C. & Bedinger, P.** (2003) The transformation of anthers in the *msca1* mutant of maize. *Planta*, **216**, 778–788.

**Chen, S., Zhou, Y., Chen, Y. & Gu, J.** (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.

**Chen, X., Zhang, H., Sun, H., Luo, H., Zhao, L., Dong, Z. *et al.*** (2017) IRREGULAR POLLEN EXINE1 is a novel factor in anther cuticle and pollen EXINE formation. *Plant Physiology*, **173**, 307–325.

**Cigan, A.M., Unger, E., Xu, R.-J., Kendall, T. & Fox, T.W.** (2001) Phenotypic complementation of *ms45* maize requires tapetal expression of MS45. *Sexual Plant Reproduction*, **14**, 135–142.

**Ding, J., Shen, J., Mao, H., Xie, W., Li, X. & Zhang, Q.** (2012) RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. *Molecular Plant*, **5**, 1210–1216.

**Djukanovic, V., Smith, J., Lowe, K., Yang, M., Gao, H., Jones, S. *et al.*** (2013) Male-sterile maize plants produced by targeted mutagenesis of the cytochrome P450-like gene (*MS26*) using a re-designed I-*Cre*I homing endonuclease. *The Plant Journal*, **76**, 888–899.

**Edgar, R., Domrachev, M. & Lash, A.E.** (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, **30**, 207–210.

**Evans, M.M.S.** (2007) The *indeterminate gametophyte1* gene of maize encodes a LOB domain protein required for embryo sac and leaf development. *Plant Cell*, **19**, 46–62.

**Fan, Y., Yang, J., Mathioni, S.M., Yu, J., Shen, J., Yang, X. *et al.*** (2016) *PMS1T*, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 15144–15149.

**Fox, T., DeBruin, J., Haug Collet, K., Trimnell, M., Clapp, J., Leonard, A. *et al.*** (2017) A single point mutation in *Ms44* results in dominant male sterility and improves nitrogen use efficiency in maize. *Plant Biotechnology Journal*, **15**, 942–952.

**Gomez-Cano, L., Yang, F. & Grotewold, E.** (2019) Isolation and efficient maize protoplast transformation. *Bio-Protocol*, **9**, e3346.

**Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J. *et al.*** (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–3435.

**Han, Y., Hu, M., Ma, X., Yan, G., Wang, C., Jiang, S. *et al.*** (2022) Exploring key developmental phases and phase-specific genes across the entirety of anther development in maize. *Journal of Integrative Plant Biology*, **64**, 1394–1410.

**Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L. *et al.*** (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology*, **17**, 77.

**Jiang, P., Lian, B., Liu, C., Fu, Z., Shen, Y., Cheng, Z. *et al.*** (2020) 21-nt phasiRNAs direct target mRNA cleavage in rice male germ cells. *Nature Communications*, **11**, 5191.

**Jiang, Y., An, X., Li, Z., Yan, T., Zhu, T., Xie, K. *et al.*** (2021) CRISPR/Cas9-based discovery of maize transcription factors regulating male sterility and their functional conservation in plants. *Plant Biotechnology Journal*, **19**, 1769–1784.

**Kelliher, T. & Walbot, V.** (2011) Emergence and patterning of the five cell types of the *Zea mays* anther locule. *Developmental Biology*, **350**, 32–49.

**Kim, D., Langmead, B. & Salzberg, S.L.** (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.

**Komiya, R., Ohyanagi, H., Niihama, M., Watanabe, T., Nakano, M., Kurata, N. *et al.*** (2014) Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *The Plant Journal*, **78**, 385–397.

**Langfelder, P. & Horvath, S.** (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

**Langfelder, P., Zhang, B. & Horvath, S.** (2007) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, **24**(5), 719–720.

**Lee, Y.-S., Maple, R., Dürr, J., Dawson, A., Tamim, S., del Genio, C. *et al.*** (2021) A transposon surveillance mechanism that safeguards plant male fertility during stress. *Nature Plants*, **7**, 34–41.

**Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.*** (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

**Li, S., Le, B., Ma, X., Li, S., You, C., Yu, Y. *et al.*** (2016) Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. *eLife*, **5**, e22750.

**Li, Y., Huang, Y., Pan, L., Zhao, Y., Huang, W. & Jin, W.** (2021) *Male sterile 28* encodes an ARGONAUTE family protein essential for male fertility in maize. *Chromosome Research*, **29**, 189–201.

**Li, Z., Zhu, T., Liu, S., Jiang, Y., Liu, H., Zhang, Y. *et al.*** (2021) Genome-wide analyses on transcription factors and their potential microRNA regulators involved in maize male fertility. *The Crop Journal*, **9**, 1248–1262.

**Liao, Y., Smyth, G.K. & Shi, W.** (2019) The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, **47**, e47.

**Liu, X., Zhang, S., Jiang, Y., Yan, T., Fang, C., Hou, Q. *et al.*** (2022) Use of CRISPR/Cas9-based gene editing to simultaneously mutate multiple homologous genes required for pollen development and male fertility in maize. *Cell*, **11**, 439.

**Liu, Y., Teng, C., Xia, R. & Meyers, B.C.** (2020) PhasiRNAs in plants: their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *Plant Cell*, **32**, 3059–3080.

**Moon, J., Skibbe, D., Timofejeva, L., Wang, C.J., Kelliher, T., Kremling, K. *et al.*** (2013) Regulation of cell divisions and differentiation by MALE STERILITY32 is required for anther development in maize. *The Plant Journal*, **76**, 592–602.

**Nakano, M., McCormick, K., Demirci, C., Demirci, F., Gurazada, S.G.R., Ramachandruni, D. *et al.*** (2020) Next-generation sequence databases: RNA and genomic informatics resources for plants. *Plant Physiology*, **182**, 136–146.

**Nan, G.-L., Teng, C., Fernandes, J., O'Connor, L., Meyers, B.C. & Walbot, V.** (2022) A cascade of bHLH-regulated pathways program maize anther development. *Plant Cell*, **34**, 1207–1225.

**Nan, G.-L., Zhai, J., Arikit, S., Morrow, D., Fernandes, J., Mai, L. *et al.*** (2017) MS23, a master basic helix-loop-helix factor, regulates the specification and development of the tapetum in maize. *Development*, **144**, 163–172.

**Nelms, B. & Walbot, V.** (2022) Gametophyte genome activation occurs at pollen mitosis I in maize. *Science*, **375**, 424–429.

**Nonomura, K.-I., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H. *et al.*** (2007) A germ cell-specific gene of the *ARGONAUTE* family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell*, **19**, 2583–2594.

**Pokhrel, S., Huang, K., Bélanger, S., Zhan, J., Caplan, J.L., Kramer, E.M. *et al.*** (2021) Pre-meiotic 21-nucleotide reproductive phasiRNAs emerged in seed plants and diversified in flowering plants. *Nature Communications*, **12**, 4941.

**Portwood, J.L., 2nd, Woodhouse, M.R., Cannon, E.K. *et al.*** (2019) MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Research*, **47**, D1146–D1154.

**Qi, X., Guo, S., Wang, D., Zhong, Y., Chen, M., Chen, C. *et al.*** (2022) *ZmCOI2a* and *ZmCOI2b* redundantly regulate anther dehiscence and gametophytic male fertility in maize. *The Plant Journal*, **110**, 849–862.

**Rawson, H.M.** (1996) The developmental stage during which boron limitation causes sterility in wheat genotypes and the recovery of fertility. *Functional Plant Biology*, **23**, 709–717.

**Robinson, M.D., McCarthy, D.J. & Smyth, G.K.** (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

**Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S. *et al.*** (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

**Smith, T., Heger, A. & Sudbery, I.** (2017) UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Research*, **27**, 491–499.

**Somaratne, Y., Tian, Y., Zhang, H., Wang, M., Huo, Y., Cao, F. *et al.*** (2017) ABNORMAL POLLEN VACUOLATION1 (APV1) is required for male fertility by contributing to anther cuticle and pollen exine formation in maize. *The Plant Journal*, **90**, 96–110.

**Sun, W., Chen, D., Xue, Y., Zhai, L., Zhang, D., Cao, Z. *et al.*** (2019) Genome-wide identification of AGO18b-bound miRNAs and phasiRNAs in maize by cRIP-seq. *BMC Genomics*, **20**, 656.

**Tamim, S., Cai, Z., Mathioni, S.M., Zhai, J., Teng, C., Zhang, Q. *et al.*** (2018) *Cis*-directed cleavage and nonstoichiometric abundances of 21-nucleotide reproductive phased small interfering RNAs in grasses. *The New Phytologist*, **220**, 865–877.

**Tello-Ruiz, M.K., Jaiswal, P. & Ware, D.** (2022) Gramene: a resource for comparative analysis of plants genomes and pathways. *Methods in Molecular Biology*, **2443**, 101–131.

**Teng, C., Zhang, H., Hammond, R., Huang, K., Meyers, B.C. & Walbot, V.** (2020) *Dicer-like 5* deficiency confers temperature-sensitive male sterility in maize. *Nature Communications*, **11**, 2912.

**Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. & Gao, G.** (2020) PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Research*, **48**, D1104–D1113.

**Tian, P., Zhang, X., Xia, R., Liu, Y., Wang, M., Li, B. *et al.*** (2021) Evolution and diversification of reproductive phased small interfering RNAs in *Oryza* species. *The New Phytologist*, **229**, 2970–2983.

**Tian, Y., Xiao, S., Liu, J., Somaratne, Y., Zhang, H., Wang, M. *et al.*** (2017) *MALE STERILE6021* (*MS6021*) is required for the development of anther cuticle and pollen exine in maize. *Scientific Reports*, **7**, 16736.

**Vernoud, V., Laigle, G., Rozier, F., Meeley, R.B., Perez, P. & Rogowsky, P.M.** (2009) The HD-ZIP IV transcription factor OCL4 is necessary for trichome patterning and anther development in maize. *The Plant Journal*, **59**, 883–894.

**Wan, X., Wu, S., Li, Z., An, X. & Tian, Y.** (2020) Lipid metabolism: critical roles in male fertility and other aspects of reproductive development in plants. *Molecular Plant*, **13**, 955–983.

**Wang, C.-J.R., Nan, G.-L., Kelliher, T., Timofejeva, L., Vernoud, V., Golubovskaya, I.N. *et al.*** (2012) Maize *multiple archesporial cells 1* (*mac1*), an ortholog of rice *TDL1A*, modulates cell proliferation and identity in early anther development. *Development*, **139**, 2594–2603.

**Wang, D., Skibbe, D.S. & Walbot, V.** (2013) Maize *male sterile 8* (*Ms8*), a putative β-1,3-galactosyltransferase, modulates cell division, expansion, and differentiation during early maize anther development. *Plant Reproduction*, **26**, 329–338.

**Wang, Y., Liu, D., Tian, Y., Wu, S., An, X., Dong, Z. *et al.*** (2019) Map-based cloning, phylogenetic, and microsynteny analyses of *ZmMs20* gene regulating male fertility in maize. *International Journal of Molecular Sciences*, **20**, 1411.

**Xie, K., Wu, S., Li, Z., Zhou, Y., Zhang, D., Dong, Z. *et al.*** (2018) Map-based cloning and characterization of *Zea mays male sterility33* (*ZmMs33*) gene, encoding a glycerol-3-phosphate acyltransferase. *Theoretical and Applied Genetics*, **131**, 1363–1378.

**Yadava, P., Tamim, S., Zhang, H., Teng, C., Zhou, X., Meyers, B.C. *et al.*** (2021) Transgenerational conditioned male fertility of HD-ZIP IV transcription factor mutant *ocl4*: impact on 21-nt phasiRNA accumulation in pre-meiotic maize anthers. *Plant Reproduction*, **34**, 117–129.

**Yang, F., Li, W., Jiang, N., Yu, H., Morohashi, K., Ouma, W.Z. *et al.*** (2017) A maize gene regulatory network for phenolic metabolism. *Molecular Plant*, **10**, 498–515.

**Yilmaz, A., Nishiyama, M.Y., Jr., Fuentes, B.G., Souza, G.M., Janies, D., Gray, J. *et al.*** (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiology*, **149**, 171–180.

**Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.-L., Walbot, V. *et al.*** (2015) Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 3146–3151.

**Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M. *et al.*** (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell*, **27**, 513–531.

**Zhang, D., Wu, S., An, X., Xie, K., Dong, Z., Zhou, Y. *et al.*** (2018) Construction of a multicontrol sterility system for a maize male-sterile line and hybrid seed production based on the *ZmMs7* gene encoding a PHD-finger transcription factor. *Plant Biotechnology Journal*, **16**, 459–471.

**Zhang, M., Ma, X., Wang, C., Li, Q., Meyers, B.C., Springer, N.M. *et al.*** (2021) CHH DNA methylation increases at *24-PHAS* loci depend on 24-nt phased small interfering RNAs in maize meiotic anthers. *The New Phytologist*, **229**, 2984–2997.

**Zhang, S., Wu, S., Niu, C., Liu, D., Yan, T., Tian, Y. *et al.*** (2021) *ZmMs25* encoding a plastid-localized fatty acyl reductase is critical for anther and pollen development in maize. *Journal of Experimental Botany*, **72**, 4298–4318.

**Zhang, Y.C., Lei, M.Q., Zhou, Y.F., Yang, Y.W., Lian, J.P., Yu, Y. *et al.*** (2020) Reproductive phasiRNAs regulate reprogramming of gene expression and meiotic progression in rice. *Nature Communications*, **11**, 6031.

**Zheng, Y., Wang, D., Ye, S., Chen, W., Li, G., Xu, Z. *et al.*** (2021) Auxin guides germ-cell specification in Arabidopsis anthers. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e2101492118.

**Zhou, X., Huang, K., Teng, C., Abdelgawad, A., Batish, M., Meyers, B.C. *et al.*** (2022) 24-nt phasiRNAs move from tapetal to meiotic cells in maize anthers. *The New Phytologist*, **235**, 488–501.

**Zhu, T., Wu, S., Zhang, D., Li, Z., Xie, K., An, X. *et al.*** (2019) Genome-wide analysis of maize *GPAT* gene family and cytological characterization and breeding application of *ZmMs33*/*ZmGPAT6* gene. *Theoretical and Applied Genetics*, **132**, 2137–2154.