Article

Ginger genome reveals the *SMPED1* gene causing sex-phase synchrony and outcrossing in a flowering plant

Jian-Li Zhao [1,2,9] ✉, Yang Dong [3,4,9], Ao-Dan Huang [1,9],
Sheng-Chang Duan [3,4,9], Xiao-Chang Peng [1,9], Hong Liao [1],
Jiang-Hua Chen [2], Yin-Ling Luo[5], Qin-Ying Lan[2], Ya-Li Wang[1], Wen-Jing Wang[1],
Xin-Meng Zhu[1], Pei-Wen Luo[1], Xue Xia[1], Bo Li[1], W. John Kress [6], Jia-Jia Han [1] ✉,
Spencer C. H. Barrett [7] ✉, Wei Chen [3,4] ✉ & Qing-Jun Li[1,8]

In many flowering plants, male and female reproductive organs mature at different times to avoid self-pollination, a phenomenon termed dichogamy. Most dichogamous species are either protandrous or protogynous, making this strategy difficult to study genetically. However, in the ginger *Alpinia mutica*, protandrous and protogynous floral morphs co-occur within populations, and the synchronized rhythmic movement of styles and dehiscence of stamens promotes cross-pollination between morphs. Here we demonstrate that a single Mendelian locus with a dominant allele governing protogyny controls sexual polymorphism. We used haplotype-resolved genomes and population genomics to identify the dichogamy-determining region, revealing a large deletion in the protandrous morphotype. We found that the key gene *SMPED1*, located adjacent to the deletion, governs the timing of anther dehiscence and style movement. *SMPED1* is widespread among angiosperms and probably has conserved function. Our findings represent a new genetic characterization of a key mating system gene controlling the synchrony of sex organs in flowering plants.

The spectacular floral diversity of angiosperms[1] is associated with myriad floral adaptations promoting cross-pollination and avoiding self-fertilization and the harmful genetic consequences of inbreeding[2–4]. Most flowers are hermaphroditic, with male and female reproductive organs. Differences in the timing of female and male sexual maturity function to promote outcrossing, especially in self-compatible species. Outcrossing facilitates adaptation to changing environments by promoting genetic diversity and heterozygosity[5]. A widespread mechanism promoting cross-pollination is dichogamy, in which anther dehiscence and stigma receptivity occur at different times. Indeed, over 70% of angiosperm species possess dichogamy[6–8], which occurs in two basic forms: protandry,

with male organs maturing before female ones; and protogyny, in which female organs mature first (Fig. 1a). Darwin was fascinated by the diversity of plant sexual systems, including dichogamy, and interpreted them as elaborate mating strategies for promoting outcrossing and avoiding selfing[9,10]. In contrast to several other mating strategies, including dioecy (populations with male and female individuals)[11], andromonoecy (male and hermaphrodite flowers on the same plant)[12] and heterostyly (populations in which floral morphs differ in stigma and anther height)[13,14], the genes controlling sex-organ function in dichogamous species are largely unknown. Molecular analyses have been thwarted because most dichogamous species are fixed for either protandry or protogyny.

Heterodichogamy (Fig. 1a) is a type of dichogamy in which populations are sexually polymorphic for protandry and protogyny and a single diallelic locus controls the two sexual phases in some species[3,15–17]. Plants with this genetic polymorphism provide a valuable system to elucidate the molecular mechanisms governing dichogamy. A unique form of heterodichogamy is found in *Alpinia* spp. (Zingiberaceae) (Fig. 1a)[18]. Here, the protandrous (PA) and protogynous (PG) morphs exhibit remarkable daily synchronization of reciprocal style movement and anther dehiscence. This synchronization promotes bee-mediated cross-pollination between the floral morphs and avoids self-pollination (Fig. 1b,c, Extended Data Fig. 1 and Supplementary Videos 1 and 2)[18–20]. Briefly, the style of the PG morph curves downward in early morning to receive pollen via pollinators from the PA morph, whose styles are erect and pollen sacs are dehiscent. After noon, the styles of PA plants curve downward and receive pollen from PG plants, whose styles have moved upward and whose anthers are now dehiscent. The reciprocal spatiotemporal switching of sex-organ position occurs synchronously and reversibly within the 24-hour lifespan of the flowers (Fig. 1c and Extended Data Fig. 1)[18–22]. This complex mating strategy provides outstanding opportunities to explore the molecular basis of sex-phase synchrony and the rhythmic changes in sex function. A recent preliminary study of the flexistylous ginger *Lanxangia* (=*Amomum*) *tsaoko* reported several genes in this species that may be involved in dichogamy[23].

Here we investigate flexistyly in *Alpinia mutica* (Zingiberaceae) to reveal the molecular mechanisms controlling changes in sexual function in a dichogamous species. Our study addressed three questions: (1) Is flexistyly governed by simple Mendelian inheritance? (2) What gene or genes regulate daily sexual function dynamics? (3) Is there evidence that the causal genetic factors in *Alpinia* play a more widespread role among other dichogamous angiosperms? We demonstrate through controlled crosses that flexistyly is governed by a single Mendelian locus. Our results from genome-wide association studies (GWASs) and investigation of rhythmicity in gene expression indicate that a new gene, *STYLE MOVEMENT and POLLEN EARLY DISPERSAL 1* (*SMPED1*), named because of its synchronized influence on style movement and anther dehiscence, plays a causal role in sex-phase synchrony. Experiments with antisense oligodeoxynucleotides (AS-ODNs) and transgenic plants confirmed that *SMPED1* controls sex-phase synchrony throughout the day. Evolutionary genomics and functional prediction analyses indicate that *SMPED1* is conserved in structure and function across angiosperms. Our findings illuminate the mechanistic basis of dichogamy, a widespread reproductive adaptation in plants.

Results

Determination of the Mendelian inheritance of floral morphs in *A. mutica*

The ratio of PG to PA morphs in wild populations of *Alpinia* species did not deviate significantly from 1:1 (Fig. 1d and Supplementary Table 1), suggesting that flexistyly is governed by a single Mendelian locus. To test this hypothesis, we performed hand-pollinated crosses in *A. mutica* and *A. kwangsiensis*, another flexistylous species. For both species, crosses between morphs yielded F$_1$ offspring exhibiting a PG:PA ratio of 1:1 (Fig. 1e). Self-pollination of PA produced only PA offspring, whereas self-pollination of PG produced both morphs, with a PG:PA ratio of 3:1 (Fig. 1e and Supplementary Table 1). Flexistyly in these *Alpinia* species is thus controlled by a single diallelic Mendelian locus: PG is heterozygous for the dominant allele (*d/D*; *D* for dichogamy) and PA homozygous for a recessive allele (*d/d*). In the wild, PG × PA crosses (*d/D* × *d/d*) produce equal frequencies of each morph and maintain a stable 1:1 morph ratio in populations. Our results are consistent with crossing data from heterodichogamous pecans (*Carya illinoinensis*)[24] and walnuts (*Juglans regia*)[16]. The homozygous PG genotype (*D/D*) is probably rare in natural populations because intermorph mating

maintains heterozygosity in PG, ensuring equal numbers of PG and PA individuals in progenies[15].

Overview of the *A. mutica* genome

We obtained high-quality haplotype-resolved genome assemblies for *A. mutica* to investigate the genetic architecture of dichogamy. The genome size of the PA haplotype was ~2.12 Gb on the basis of 318.96 Gb of sequencing data (Extended Data Fig. 2a and Supplementary Table 2). The initial contig assembly using PacBio HiFi reads (197.25 Gb for PA and 184.94 Gb for PG; Supplementary Table 3) yielded four haplotype-specific-contig-level assemblies for the two morphs with average lengths of 2.18 ± 0.01 Gb. The average contig N50 length of the assemblies was 81.45 ± 3.59 Mb (Supplementary Table 4). Approximately 2.15 Gb of sequences per haplotype-specific-contig-level assembly (98% ± 0.55%) were anchored into 24 pseudochromosomes on the basis of 185.31 Gb of Hi-C reads (Extended Data Fig. 2b,c and Supplementary Tables 4–6; also see Methods), achieving an average completeness of 98.7%, inferred from BUSCO analysis. We predicted an average of 50,444 ± 796 protein-coding genes in each of the four haplotypes, with an average length of 4,268 ± 76 bp (Supplementary Table 4). A high degree of gene synteny among the four haplotypes was evident (Extended Data Fig. 2d). Haplotypes contained a high proportion of transposable element (TE) (85% ± 0.2%), largely due to the expansion of long terminal repeats (LTRs), especially *Copia* elements (Supplementary Tables 4 and 6).

Dichogamy-determining region of *A. mutica* genomes

To locate the dichogamy-determining region (DDR), we obtained whole-genome sequencing short reads with an average depth of 23.55× for 91 individuals from our experimental population (Supplementary Table 7). Using the PA morph H2 genome as a reference, we identified a total of 12,712,117 single-nucleotide polymorphisms (SNPs) (Supplementary Table 8). GWASs of the morphs via univariate linear mixed models in GEMMA revealed a large block of associated SNPs located close to the end of chromosome 8 (Fig. 2a). The top-ranked SNPs (rs100521961 and rs100527109, $P = 2.87 \times 10^{-15}$) were in linkage disequilibrium ($r^2 \geq 0.8$) with other SNPs in this region, which was ~8.36 kb in size (100519807 bp to 100528168 bp; Fig. 2b). Although a highly associated plot was evident for chromosome 13, there was no large linkage disequilibrium block and large TE insertion, as occurred on chromosome 8 (also see details about the evolution of the DDR in 'Evolution of the DDR and conservation of *SMPED1* in angiosperms' section); we therefore focused our attention on the region on chromosome 8.

Dichogamy-associated SNPs should conform to Mendelian inheritance, resulting in more reference alleles in PA individuals and more alternative alleles in PG individuals (Fig. 2c,d). Plotting of alleles in all individuals and sorting by morph revealed many haplotype structures and recombination sites and identified a DDR that was approximately 30 kb in size (rs100517620–rs100547857). The DDR encompassed the region identified by GWASs above (Fig. 2d). Synteny analysis of the 30-kb region revealed a matching counterpart on chromosome 8 in the PG morph assembly spanning ~50 kb in a heterozygous state, with the other copy identical to the 30-kb PA region (Fig. 2b). This heterozygous region is thus a hemizygous locus. We confirmed the heterozygous region by mapping unique short reads from resequenced individuals to the PG morph assembly (Fig. 2c). We thus defined the putative *Alpinia* DDR as the genomic region between 100518121 bp and 100547025 bp on chromosome 8 in the PA morph. Overall genetic diversity around the DDR of the PG morph was significantly lower than in the equivalent region in the PA morph ($P < 0.0001$; Extended Data Fig. 3), but the genetic diversity of Am*SMPED1* in the PG morph was higher than in the PA morph. Because the coding region of Am*SMPED1* between the PG and PA morphs was identical, except for a single synonymous mutation, the higher genetic diversity of Am*SMPED1* in the PG morph is probably associated with an LTR insertion.
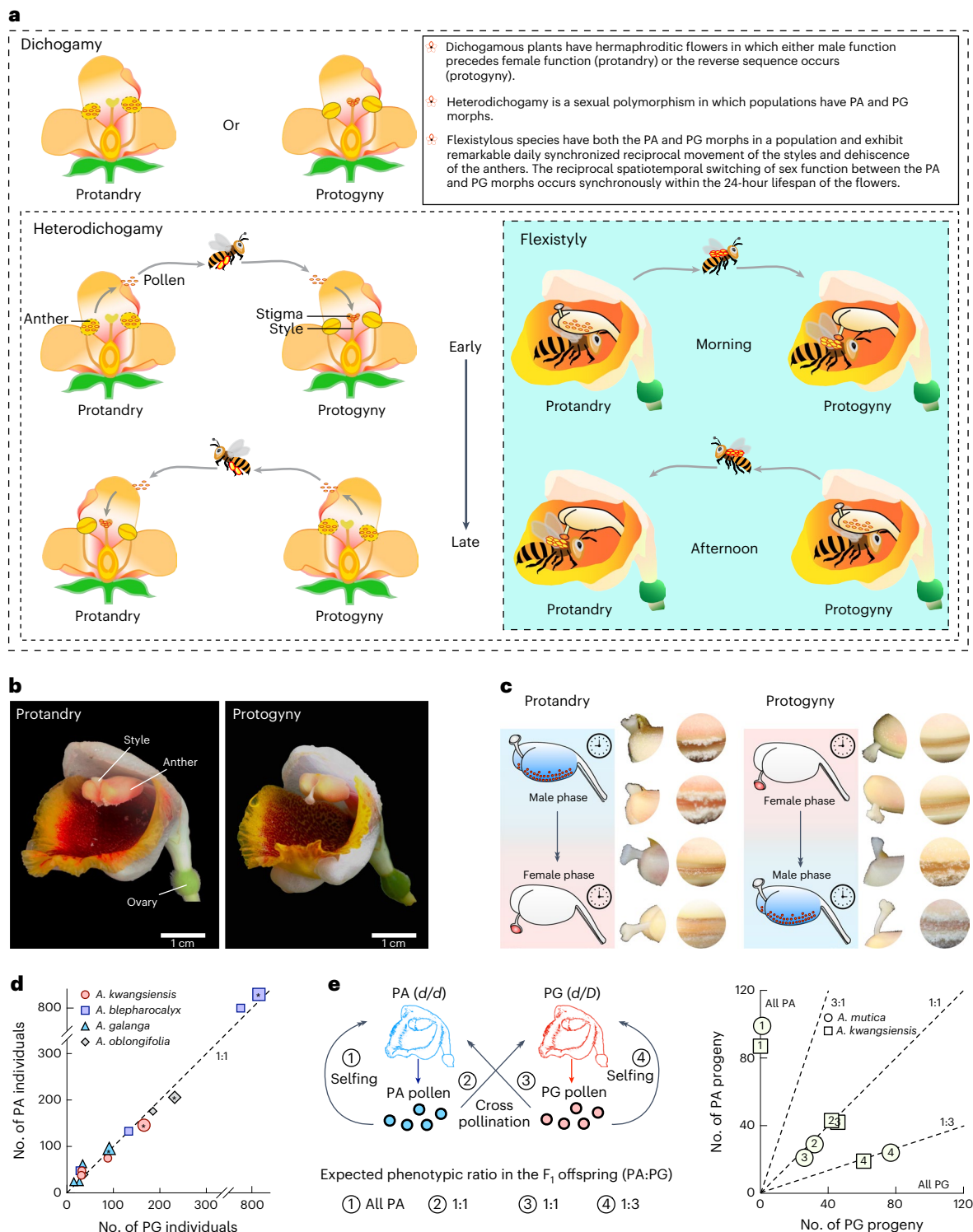
Fig. 1 | Mendelian inheritance of dichogamy in *Alpinia* species. a, Diagram of dichogamy, heterodichogamy and flexistyly. The dashed boxes show the relations among the three systems. **b**, Floral morphology of the PA morph (left) and the PG morph (right) of *A. mutica*. Scale bars, 1 cm. **c**, Daily flower maturation progression of the PA morph and PG morph under natural conditions. Style movement and anther dehiscence are reversed between the two morphs. For each morph, the left column of small images shows the stigma, and the right column shows the stomium, where anthers dehisce. The detailed progression is shown in Extended Data Fig. 1. Under continuous illumination for 24 h, the PA morph converts to the status of PG morph (Extended Data Fig. 1). **d**, Scatter plot of the numbers of individuals of the PA morph and the PG morph in natural populations of four *Alpinia* species, showing a PA:PG ratio of 1:1. The asterisks inside these symbols indicate the ratio of whole samples. **e**, Expected (left) and observed (right) number of F$_1$ offspring from manual crosses between the two morphs in *A. mutica* and *A. kwangsiensis*. The observed morph ratios did not deviate from the expected ratios for Mendelian inheritance for a diallelic locus with dominance (Supplementary Table 1). Illustrations in **a** created with EdrawMax.
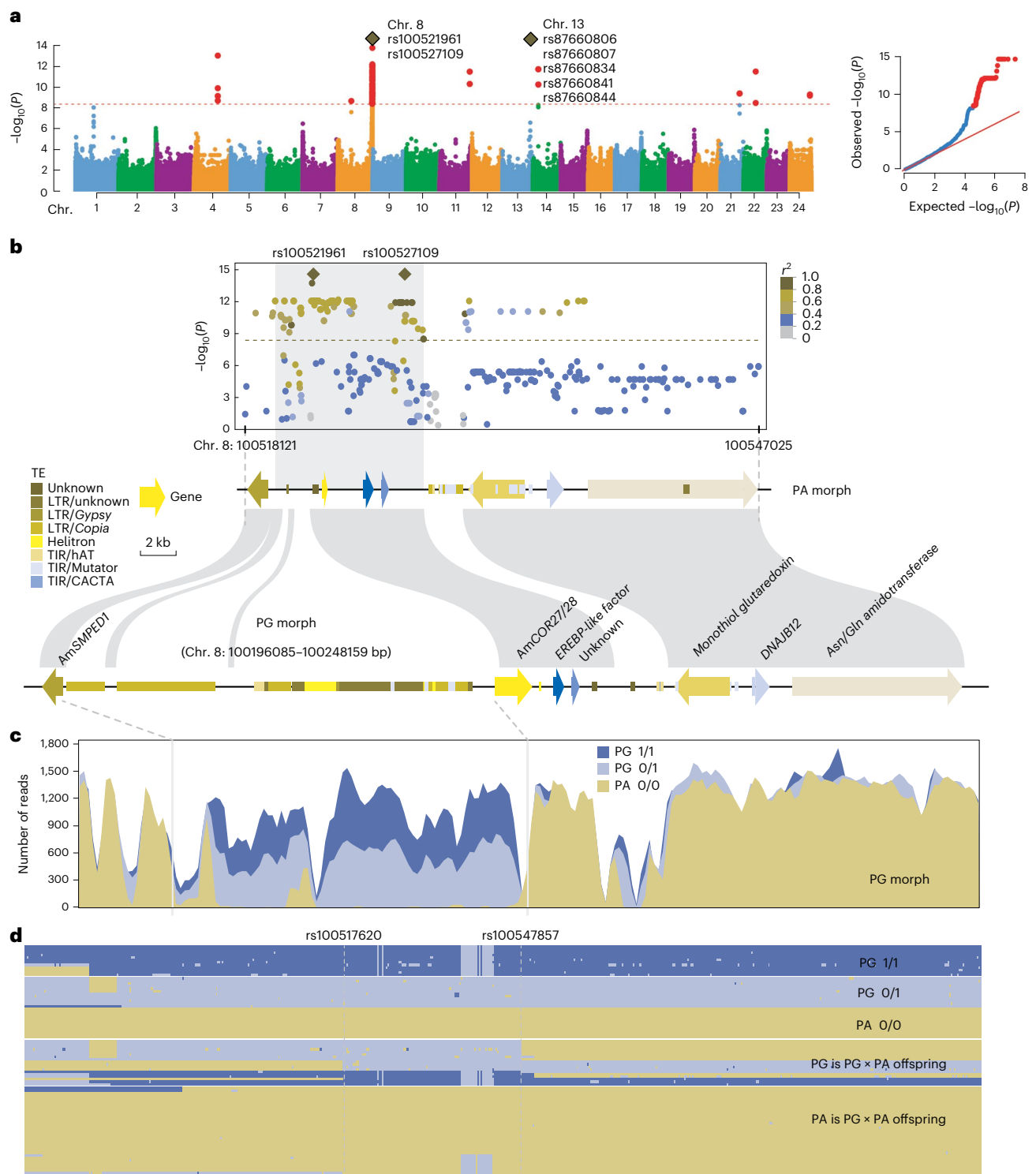
Fig. 2 | Identification of the DDR in the *A. mutica* genome. a, Manhattan plot (left) and QQ plot (right) of a GWAS for dichogamy using genome-wide SNP markers. The GEMMA algorithm was used for the GWAS with a univariate linear mixed model. The red dashed line in the Manhattan plot represents the significance threshold ($-\log_{10}(P) = 8.33$). The brown diamonds show the top-ranked SNPs. The solid red line in the QQ plot is the expected value. **b**, Linkage disequilibrium analysis between the top-ranked SNPs (brown diamonds) and all other SNPs at the end of chromosome 8, calculated as r^2 values. The shaded area indicates the genomic region containing SNPs in a linkage disequilibrium block with the top-ranked SNP ($r^2 > 0.8$). The bottom panel shows the genomic structure of the DDRs in the PA and PG morphs. Syntenic sequences are linked by grey blocks. Genes and TEs are illustrated as coloured arrows and blocks, respectively. The types of TEs are shown in the key on the left. **c**, Sequencing coverage for uniquely mapped reads from individuals of the homozygous PG morph ($1/1$; $n = 12$), the heterozygous PG morph ($1/0$; $n = 12$) and the homozygous PA morph ($0/0$; $n = 12$), plotted against the PG morph genome using a 600-bp sliding window and a 300-bp step. **d**, Manual screening of SNPs associated with dichogamy in an experimental population of *A. mutica*, revealing recombination points and the possible location of the DDR (within the dashed vertical lines). The SNPs at the recombination points (rs100517620 and rs100547857) correspond to 100517620 bp and 100547857 bp on chromosome 8.

The dichogamy-determining gene *SMPED1* and the dichogamy phenotype

We identified six pairs of syntenic genes in the DDR (Fig. 2b). The coding sequences of these genes were highly homologous between morphs, and we assume that dichogamy probably results from differential gene expression. Accordingly, we focused on two candidate genes located within or close to the large hemizygous region. These genes were Am*SMPED1* and *Cold-Regulated Gene 27* (Am*COR27*) (Fig. 2b). Am*SMPED1* was present at the 5' boundary of the hemizygous region, and the coding sequences of Am*SMPED1* in the PA and PG morphs were identical. We collected anther and style tissue from morphs every 3 h during flowering under natural and continuous illumination, starting from the beginning of the day at midnight (0:00) and continuing to 18:00, with at least three replicates at each time point (Fig. 3a and Extended Data Figs. 1 and 4). We performed RNA-seq and reverse-transcription quantitative PCR (RT-qPCR) to evaluate expression levels.

Only Am*SMPED1* exhibited rhythmic expression that was synchronous with anther and style behaviour under natural and continuous illumination (Fig. 3a and Extended Data Figs. 1 and 4a), suggesting that Am*SMPED1* is associated with sex-phase synchrony. Under natural conditions, Am*SMPED1* expression corresponded to that of anther dehiscence in both morphs, showing a clear temporal difference in transcript abundance (Fig. 3a and Extended Data Fig. 4a). Significantly, the expression of Am*SMPED1* changed after sunrise at approximately 9:00. Given our previous observations that pollen in the PG morph is mature in early morning before anther dehiscence[18], we postulate that this gene is involved in the rupture of anther wall layers. Am*SMPED1* expression was also upregulated at 9:00 in the style, corresponding to the time when the stigma moves away from the receptive position in the PG morph.

Am*SMPED1* expression in the PG morph was similar between natural and continuous illumination, whereas Am*SMPED1* expression in the PA morph was completely reversed under continuous illumination (Figs. 1c and 3a and Extended Data Fig. 4a). Am*SMPED1* expression was synchronous with the movement of styles in both morphs under continuous illumination (Figs. 1b,c and 3a and Extended Data Fig. 4a). For the PG morph, anther dehiscence and change in style movement under continuous illumination occurred at 9:00, synchronous with the expression of Am*SMPED1*. The same behaviour in flowers exposed to natural light conditions was evident, and these events occurred at 12:00 (Fig. 3a and Extended Data Figs. 1 and 4a). Our results strongly suggest that the dehiscence of pollen sacs and movement of styles are controlled by Am*SMPED1*.

Allele-specific expression analysis of Am*SMPED1* alleles (*D* and *d*) in the PA (*d/d*) and PG (*d/D*) morphs revealed that the dominant allele *D* had lower expression than the recessive allele *d* in the PG morph (Extended Data Fig. 4b,c). The total expression of *D* plus *d* was consistent with the expression pattern of Am*SMPED1* in the PG morph (Fig. 3a and Extended Data Fig. 4a). Because our reference genome of the PG morph was of the dominant homozygote (*D/D*), the PA and PG promoters in the validation experiment represented the promoters of *d* and *D*, respectively. The validation of promoters indicated that the activity of the *d* promoter was significantly higher than the activity of the *D* promoter (Fig. 3b), possibly causing *d* to have a higher expression level than *D* in the PG morph.

The expression of Am*COR27* and other genes in the DDR was not correlated with the behaviour of anthers or styles (Extended Data Fig. 5a,b). Moreover, patterns of Am*SMPED1* revealed by RNA-seq and RT-qPCR and in situ hybridization analyses of anthers and styles at each time point indicated that Am*SMPED1* was expressed in anther tissue surrounding pollen grains and the inner chamber of the style (Extended Data Fig. 7). We therefore focused on Am*SMPED1* for further functional characterization and evolutionary analysis.

SMPED1 determines sex-phase synchrony

A transgenic system in *A. mutica* was not available; we therefore employed two methods to validate the functions of Am*SMPED1* in style movement and anther dehiscence. First, we applied an AS-ODN treatment to *A. mutica* previously described for styles of Brassicaceae[25,26]. RT-qPCR analysis confirmed that treatment with an AS-ODN specific to Am*SMPED1* (AS-Am*SMPED1*) successfully reduced Am*SMPED1* expression levels in *A. mutica*, compared with mock-treated flowers (no oligonucleotide) and flowers treated with sense oligonucleotides specific to Am*SMPED1* (S-Am*SMPED1*) as a negative control (Fig. 3c). Compared with negative controls, the angles of style curvature in the AS-Am*SMPED1*-treated samples were significantly smaller (Fig. 3d,e). These results confirmed that Am*SMPED1* positively regulates style movement in *A. mutica*.

The style of *A. mutica* is composed of permeable parenchyma tissue and is amenable to AS-ODN treatment. But anther tissue has an impermeable secondary cell wall, and we were unable to apply the AS-ODN treatment to anthers despite many attempts. As an alternative approach, and because the gene and protein structure and function of *SMPED1* are conserved in angiosperms (see below), we used *Arabidopsis thaliana* to assess the function of *SMPED1* in anther dehiscence. In the *A. thaliana* genome, there are two highly conserved genes homologous to Am*SMPED1* with unknown functions so far (At5g67020 and At3g50340), which we named At*SMPED1* and At*SMPED2*, respectively (Fig. 3f and Extended Data Fig. 6a). To explore their functions, we constructed

Fig. 3 | Confirmation that *SMPED1* determines the sex-phase synchrony of dichogamy. a, Expression level of Am*SMPED1* in anthers (left) and styles (right) under natural conditions and continuous illumination (with light exposure during subjective night), as determined by RNA-seq data. The grey shaded areas indicate the night before sunrise at the location where plants were cultivated. The dashed arrows point to the time of anther dehiscence; the transition point of style reciprocal movement is indicated by the solid arrows (see the temporal details in Extended Data Fig. 1). TPM, transcripts per kilobase million. **b**, The activities of Am*SMPED1* promoters of the PA and PG morphs. LUC, luciferase; REN, *Renilla* luciferase. A schematic diagram of the experiment is shown in Extended Data Fig. 6c. **c**, The relative Am*SMPED1* expression level is significantly decreased in *A. mutica* by AS-Am*SMPED1*. S-Am*SMPED1* and mock served as negative controls. 'Mock' means no AS-Am*SMPED1* and no S-Am*SMPED1* in the culture medium. **d**, Only AS-Am*SMPED1* inhibits the style movement of both morphs of *A. mutica*. **e**, Quantification of style curvature for the three treatments shown in **c** and **d**. **f**, GUS staining showing the expression domains of At*SMPED1* (At5g67020) and At*SMPED2* (At3g50340) in *A. thaliana* flowers, demonstrating that these two genes, homologous to Am*SMPED*, are specifically expressed in anthers and styles. At5g67020 and At3g50340 represent unknown proteins so far. **g,h**, Percentage of anther dehiscence in transgenic *A. thaliana* plants overexpressing Am*SMPED1* (*35S::*Am*SMPED1*), At*SMPED1* (*35S::*At*SMPED1*) or At*SMPED2* (*35S::*At*SMPED2*) or knocked down for At*SMPED1* and At*SMPED2* (*ami*At*SMPED1/2*) transcript levels. The expected gene expression levels of these transgenic plants were confirmed via RT-PCR and RT-qPCR (Extended Data Fig. 6b). Plant anthers were collected every 30 min from 6:00 to 10:30, and the percentages of dehiscent anthers at each sampling point were calculated. The data with error bars are presented as the mean ± s.e.m., and the whiskers extend to minima and maxima (**a**–**c**,**e**,**g**,**h**). The data points represent individual biological replicates (**b**,**c**,**e**). The asterisks denote significant differences between samples using two-sided Mann–Whitney *U*-tests (**b**,**c**,**e**). ****P* < 0.001; ***P* < 0.01; **P* < 0.05; NS, *P* > 0.05, not significant. The exact numbers of biological replicates (*n*) are indicated (**a**–**c**,**e**), and each time point has equal *n* (**a**). In **g**,**h**, WT includes 41 individuals, and the sample sizes for each time are 331, 258, 462, 240, 320, 351, 379, 329 and 346, respectively. In **g**, for *35S::*Am*SMPED1*, there are 37 individuals from two transgenic plant lines, and the sample sizes for each time are 314, 307, 285, 330, 287, 300, 311, 318 and 299, respectively. In **h**, all samples are from two transgenic plant lines, and the sample sizes of *35S::*At*SMPED1* (391, 262, 444, 329, 293, 311, 444, 279 and 391 from 38 individuals, respectively), *35S::*At*SMPED2* (400, 339, 525, 318, 359, 322, 485, 336 and 472 from 44 individuals, respectively) and *ami*At*SMPED1/2* (205, 186, 205, 189, 187, 191, 188, 178 and 170 from 28 individuals, respectively) are different. Scale bars, 0.2 cm (**d**), 1 mm (**f**) and 100 μm (**g**).

transgenic plants harbouring a reporter construct consisting of the At*SMPED1* or At*SMPED2* promoter driving the *β-GLUCURONIDASE* (*GUS*) reporter gene (At*SMPED1pro*::*GUS* and At*SMPED2pro*::*GUS*) in the wild-type (WT) Col-0 *Arabidopsis* accession. We detected strong GUS staining in anthers (Fig. 3f), consistent with the results from the *Arabidopsis* eFP browser (https://bar.utoronto.ca/eplant; AtGenExpress

eFP IDs: At5g67020 and At3g50340). In parallel, we generated transgenic lines overexpressing a single *SMPED* gene (35S::Am*SMPED1*, 35S::At*SMPED1* and 35S::At*SMPED2*; Am*SMPED1* represents *SMPED1* of *A. mutica* for clarity) and a knocked-down transgenic line for At*SMPED1* and At*SMPED2* (*ami*At*SMPED1/2*, in which gene expression of At*SMPED1* and that of At*SMPED2* were knocked down simultaneously) to evaluate
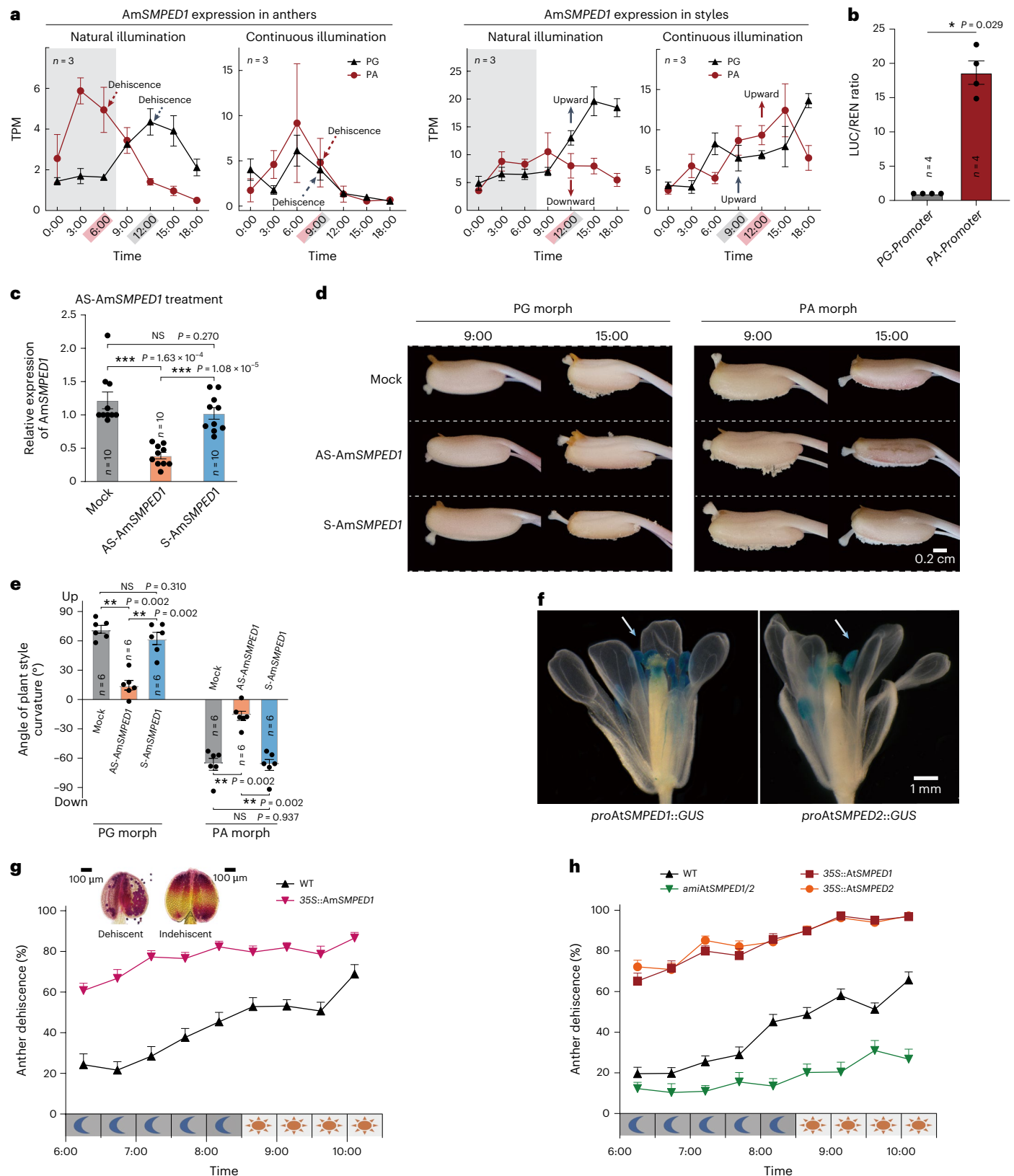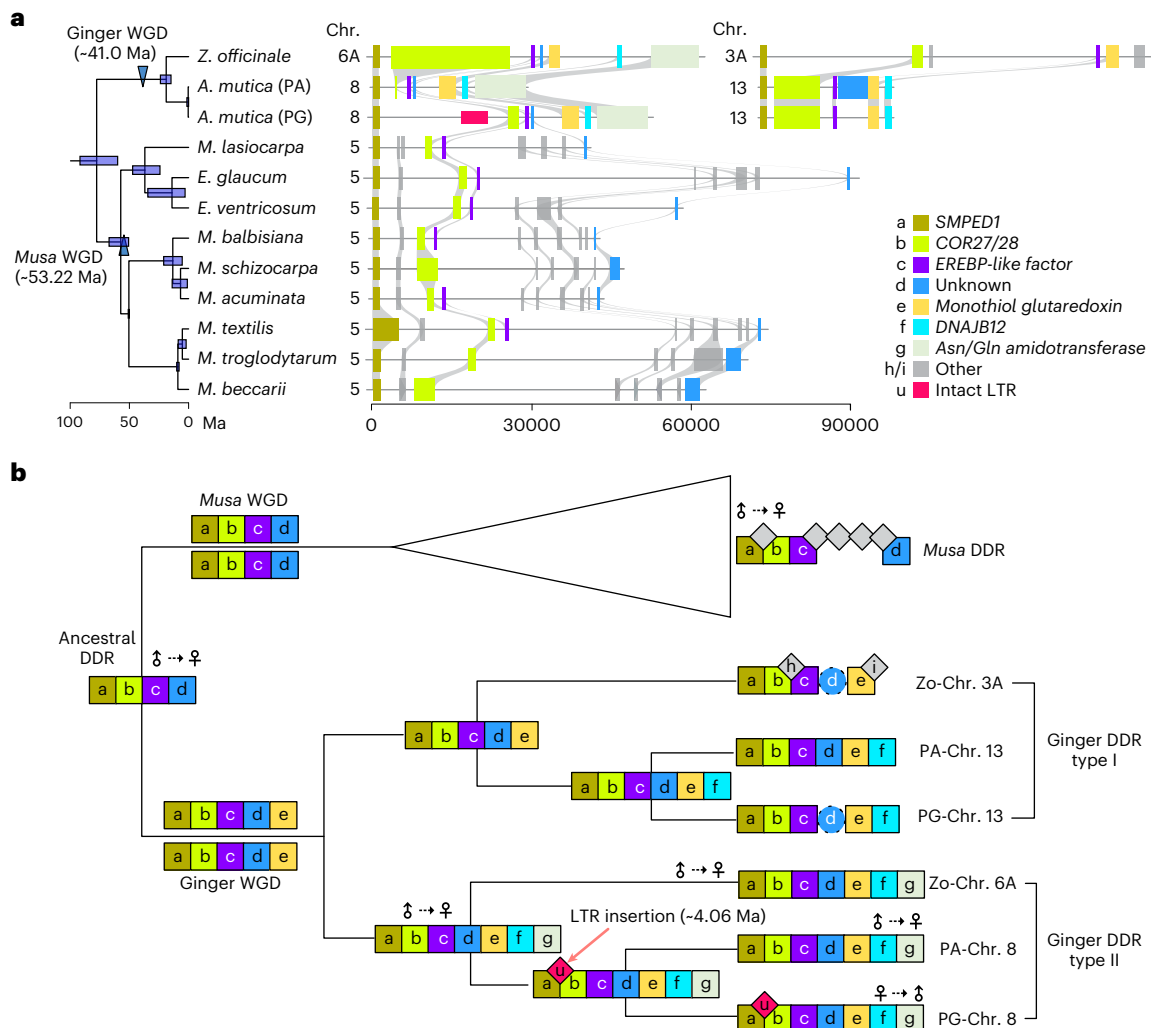
Fig. 4 | Origin of the DDR in ginger and related families. a, Phylogenetic tree of selected Musaceae and Zingiberaceae species (left) and their corresponding DDRs in the genome (right). Support values for all branches are 100%. The light blue boxes show the divergence time frame. WGD events are indicated by the two blue triangles. In each DDR shown to the right, blocks with the same colour represent syntenic genes. The gene models are labelled from a to i for clarity. **b,** Reconstruction of the ancestral DDR via gene models. The squares indicate syntenic genes, the diamonds indicate gene gains and the circles indicate gene losses. The components of the ancestral DDR are named and colour-coded as in **a**.

the effect of *SMPED1* on anther dehiscence. We observed no significant difference in growth status, number of leaves or bolting rate between the WT and our transgenic plants (Extended Data Fig. 8).

At flowering, all transgenic plants overexpressing Am*S-MPED1* (*35*::Am*SMPED1*) or one of its *Arabidopsis* homologues (*35S*::At*SMPED1* and *35S*::At*SMPED2*) showed earlier anther dehiscence by several hours compared with WT plants (Fig. 3g,h). Conversely, At*SMPED1/2*-knockdown lines showed delayed anther dehiscence compared with WT plants (Fig. 3h). These data confirmed that *SMPED1* promotes anther dehiscence. Our results are consistent with the prediction that Am*SMPED1* determines the sex-phase synchrony in *A. mutica*, including reciprocal coordination of style movement and anther dehiscence.

Evolution of the DDR and conservation of *SMPED1* in angiosperms

Phylogenomic analyses indicate that Musaceae is a basal family in Zingiberales[27,28], in which we identified a DDR (Fig. 4a,b). The timing of the ginger-specific whole-genome duplication (WGD) is ~32.5 million years ago (Ma) (Extended Data Fig. 9a,b), confirming a previous report[29]. We identified two DDRs in *A. mutica* and *Zingiber*

officinale. Although the *Musa*-specific WGD occurred ~42.5 Ma (ref. 30), three of the four banana genomes contained only one DDR, suggesting recombination and subsequent loss after WGD. Syntenic comparisons of DDRs across species revealed a high similarity in the arrangement of gene models (Fig. 4a). This ancestral DDR in gingers contains Am*SMPED1*, Am*COR27*, *EREBP-like factor* and an unknown gene (labelled as a, b, c and d in Fig. 4b).

The two copies of ginger DDRs after WGD followed different evolutionary pathways, creating type I and II DDRs with distinct gene models (Fig. 4 and Extended Data Fig. 9c). The type I DDR on chromosome 13 in the two morphs is almost collinear, except for the loss of an unknown gene in the PG morph (Fig. 4a,b). The *Alpinia* DDR described above belongs to the type II group. LTR insertions occurred between Am*SMPED1* and Am*COR27* in the ancestral type II DDR, identifying the origin of the PG morph DDR. Because the LTR-type *Gypsy* Am*LTR1* in *A. mutica* is the only intact LTR structure in this region, we estimate that its insertion occurred ~4.06 Ma (Fig. 4b), offering a lower temporal boundary for the origin of the PG morph DDR. A subsequent LTR deletion that annuls Am*COR27* occurred during the origin of the DDR in the PA morph. Our results are consistent with the hypothesis that the sequence of relevant events commenced with the ancestral DDR
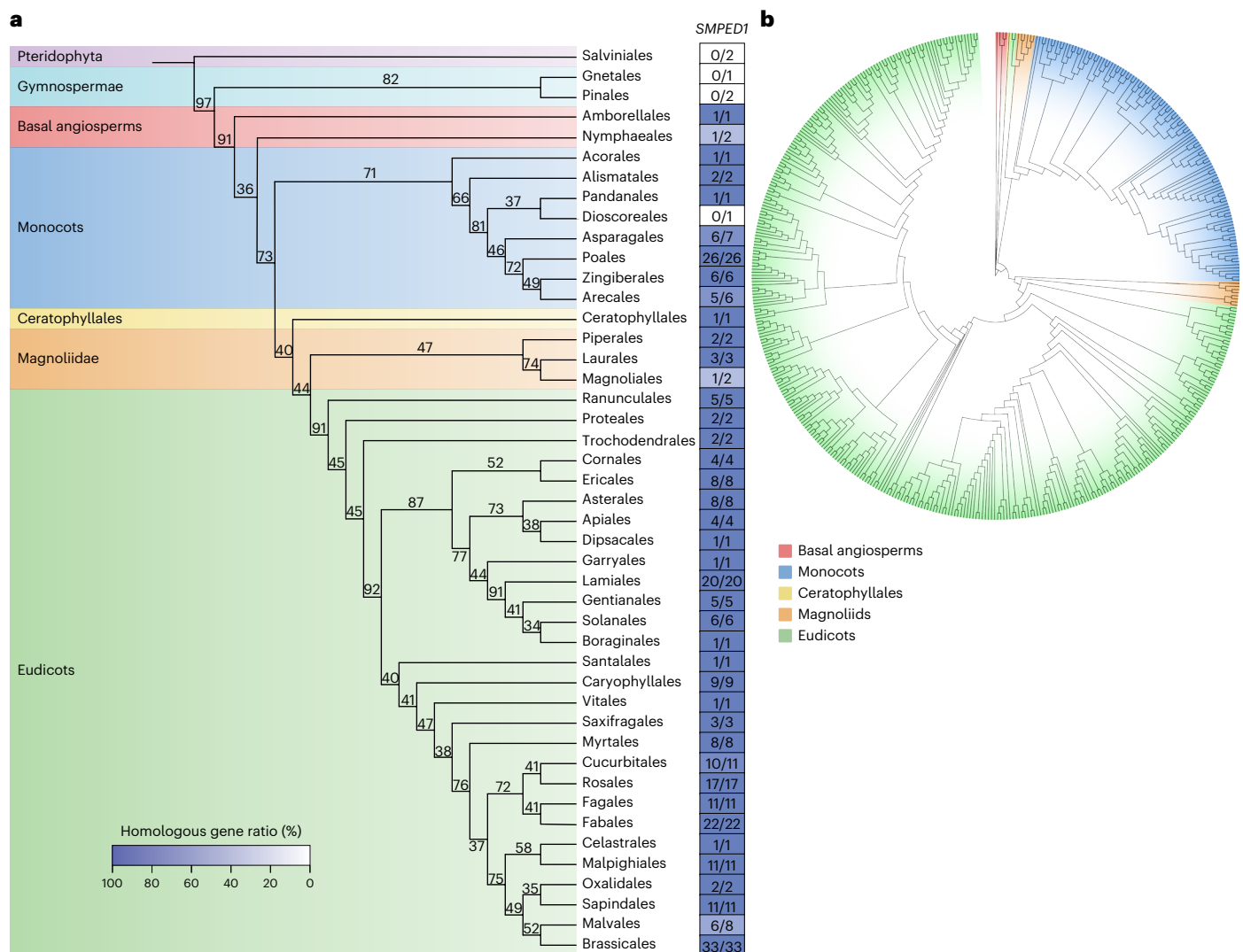
Fig. 5 | Evolution of the putative DDR in angiosperms. a, Species tree at the ordinal level and *SMPED1* in each order. This tree was reconstructed on the basis of the genomes of 281 angiosperms, two pteridophytes and three gymnosperms. The numbers at the nodes are percentage support from 1,000 bootstrap replicates. The right column indicates the distribution of the *SMPED1* homologous genes in angiosperms. The denominator of each fraction is the number of species in the order, and the numerator is the number of species with *SMPED1*. **b**, Gene tree of *SMPED1* in 281 angiosperms, two pteridophytes and three gymnosperms.

conferring the PA phenotype. The PG morph was then subsequently derived from the ancestral PA phenotype via LTR invasion, later giving rise to a PA morph unique to *Alpinia* via LTR deletion.

Unlike the localized distribution of the DDR in Zingiberales, *SMPED1* is widely distributed; it can be identified in almost all angiosperms and probably originated in this group (Fig. 5a,b and Extended Data Fig. 10). The genetic pedigree of *SMPED1*, similar to the angiosperm chloroplast phylogeny[31], suggests that the evolution of *SMPED1* was temporally synchronous with the emergence of basal angiosperms (Fig. 5b). *SMPED1*-like genes in different species groups cluster into one branch. The Magnoliidae *SMPED1* homologues are divided into two branches in the monocots, one of which is Piperales (Magnoliidae *SMPED1*); the other branch contains Laurales and Magnoliales (Magnoliidae *SMPED1*). Despite this dispersed distribution, *SMPED1*-like genes exhibit high similarity (Supplementary Fig. 1a,b) with conserved predicted protein structure and function across angiosperm lineages for their respective encoded proteins (Supplementary Figs. 1c,d and 2). This evolutionary and structural conservation across angiosperms, often with different sexual systems, suggests that *SMPED1* plays a key role in the reproductive biology of diverse lineages.

Discussion

Dichogamy is ubiquitous among angiosperms, with protandry more abundant than protogyny[6,7]. Transitions between these conditions are frequent, but whether protogyny or protandry is ancestral is uncertain[32,33]. The *Juglans G*-locus and its genetic architecture and candidate gene (*TPPD1*) controlling heterodichogamy have been explored in walnut and hickory[17]. However, not all heterodichogamy in Juglandaceae species may be controlled by the *Juglans G*-locus[17], suggesting the formation pathway of heterodichogamy could be diverse. Although the function of *TPPD1* was not further validated in these earlier studies, the findings provide a valuable foundation for subsequent studies on the evolution of heterodichogamy. Our investigations of *Alpinia* validated Mendelian inheritance of heterodichogamy and demonstrated control by a segregating region containing the newly discovered gene *SMPED1*, determining the sex-phase synchrony that characterizes dichogamy in our study system. This finding suggests that a single gene controls the fundamental reproductive behaviours of both male and female organs simultaneously, a form of pleiotropy.

Daily style bending is relatively rare, whereas anther dehiscence occurs in all angiosperm species, suggesting conservation of *SMPED1*

function. Our phylogenetic analysis revealed 42 orders containing *SMPED1*, and the gene probably originated in angiosperms, with its structure and function conserved during the rapid adaptive radiation of this clade, perhaps owing to its vital role in plant reproduction. Our findings support the hypothesis that the ancestral DDR for banana and ginger resulted in a common PA phenotype, suggesting that alterations in *SMPED1* may underlie transitions between protandry and protogyny. If this hypothesis is correct, it could provide insight into how switches between dichogamous states may have occurred. We verified that At*SMPED1*, At*SMPED2* and Am*SMPED1*, which have similar functions, belong to two distant branches of angiosperms: eudicots and monocots. Our results also suggest that dichogamy mediated by the DDR may be unique to Zingiberales, especially the probably newly evolved Am*SMPED1* controlling the daily style movement that characterizes flexistyly. However, considering the widespread presence of *SMPED1* in angiosperms, it is possible that this gene regulates dichogamy in other angiosperm lineages.

Our findings suggest that *SMPED1* is most likely an ancestral gene associated with anther dehiscence. Moreover, an LTR-mediated hemizygous region appears to have created distinct *Alpinia*-specific PG and PA loci from this ancestral locus, resulting in the origin of the dichogamy polymorphism. Identical coding regions of Am*SMPED1* with contrasting gene expressions between the PA and PG morphs suggest a key role of *cis*-regulation elements in the differential expression of Am*SMPED1* controlling dichogamy. The formation of the PA morph was probably caused by the loss of the LTR (Fig. 4b,c), which subsequently induced a change of the promoter (*cis*-regulatory element) of the PA morph (Extended Data Fig. 6c). The promoter difference between the PA and PG morphs revealed that transcription factors or other elements probably regulate the expression of Am*SMPED1* through controlling the activity of promoters (Fig. 3b). Further evidence of the fundamental role of the LTR-mediated hemizygous region involves the strong linkage disequilibrium, reduced genetic diversity and abundant TEs in the DDR of the PG morph (Fig. 2b–d and Extended Data Fig. 3). Our findings mirror sex-specific evolution in *Juglans* and *Carya*[17], the Y-chromosome-like evolution caused by male-promoting factors in sexually dimorphic species[34] and the occurrence of hemizygous regions associated with other plant sexual polymorphisms including dioecy and heterostyly[35,36]. If the association with hemizygosity is more general, it may require revision of existing theories of the genetic mechanisms driving the origin of plant sexual polymorphisms to include chromosomal structural changes.

Non-homologous genetic control of dichogamy in *Juglans/Carya* and *Alpinia* suggests that sexual differentiation involving dichogamy can be achieved through different evolutionary routes and genetic modules. Several important questions remain as to how hemizygosity in *Alpinia* evolved and whether it arose at an early stage of the establishment of the polymorphism via transpositions or other mechanisms, or subsequently evolved in non-recombining regions of the recessive haplotype through gene loss. In *Alpinia*, Haldane's sieve[37], which involves bias against the establishment of recessives[38], may have favoured the spread of the dominant *D* allele governing protogyny, and genomic structural changes were also probably indispensable for the origin and maintenance of the dichogamy polymorphism.

Methods

Morph ratios in natural populations and controlled pollination experiments to determine the inheritance of flexistyly

The number of PG and PA individuals was investigated in 11 natural populations of four *Alpinia* species in southwestern China: *A. kwangsiensis* (*n* = 3 populations), *A. blepharocalyx* (*n* = 3), *A. galanga* (*n* = 3) and *A. oblongifolia* (*n* = 2) (Supplementary Table 1). Our collections of *A. mutica* and *A. kwangsiensis* were maintained in a common garden at Xishuangbanna Tropical Botanical Garden, southwestern China (latitude 21.907226° N, longitude 101.267724° E, altitude 546 m) with

standard care, including cultivation, irrigation, fertilization, pruning and disease control. Both collections were established from a small sample of wild-collected seeds and have been inbred over the past few decades, which should have reduced overall heterozygosity to some extent. For *A. mutica*, 12 individuals from each of the PG and PA morphs were used to conduct manual self-pollination and cross-pollination to generate F_1 seed families. For *A. kwangsiensis*, the number of individuals used for hand-pollination was lower, owing to variation among plants in flower production. Mature seeds were collected and planted, and plants of both species usually bloomed within one year of seed germination. The floral phenotypes of morphs from crosses and self-pollination were scored in the F_1 generation. A *G* test[39] was then used to determine the 'goodness-of-fit' of the observed frequency of morphs against Mendelian expectations for a single locus with two alleles and dominance from controlled pollinations and against a 1:1 morph ratio in natural populations, assuming phenotypic disassortative mating. Details of DNA isolation, sequencing and genome assembly are provided in Supplementary Information.

RNA-seq

To facilitate gene prediction, stigma and anther tissues from both the PA and PG morphs were collected at 0:00 (midnight), 3:00, 6:00, 9:00, 12:00, 15:00 and 18:00. For gene expression validation, styles and anthers were collected for each dichogamy phenotype at 0:00, 3:00, 6:00, 9:00, 12:00, 15:00 and 18:00 under natural illumination and continuous illumination.

We isolated total RNA using an RNAprep Pure Plant Plus Kit (TIANGEN). The integrity, quality and concentration of RNA were detected using a NanoDrop 2000 and an Agilent 2100 Bioanalyzer (Agilent Technologies). Three micrograms of total RNA per sample was used as input material for RNA-seq library preparation. Beads containing oligo (dT) were used to isolate poly(A) mRNA from total RNA. RNA sequencing libraries were then constructed from these mRNAs using a TruSeq RNA Sample Preparation Kit (Illumina), after which the libraries were sequenced as paired-end 150-bp reads on an Illumina HiSeq 4000 instrument.

Details of transcriptome analysis are provided in Supplementary Information.

TE identification, genome annotation and evolution

We identified TEs in the genome via a combination of homology-based and de novo approaches. For homology-based prediction, known TEs were identified using RepeatMasker (version open-4.0.9)[40] against Repbase[41] (Repbase Release 20181026; http://www.girinst.org/repbase/index.html). Extensive de novo TE Annotator (version 2.0.1)[42] was used for de novo annotation of TEs. The lists of TEs identified in the two lines were subsequently combined. The time of insertions of intact LTRs was estimated using a mutation rate of 6.5×10^{-9} per year per nucleotide.

We predicted protein-coding genes from the repeat-masked genome in BRAKER3 (version 3.0.3)[43] using RNA-seq and protein data. The RNA-seq reads were cleaned with Trimmomatic (version 0.39)[44] using the parameters SLIDINGWINDOW 4:15 and MINLEN 70 and then aligned to the genome with HISAT2 (version 2.2.1)[45]. The alignments were then converted to bam format and sorted via SAMtools (version 1.3.1)[46]. The protein data included amino acid sequences of *Z. officinale* var. 'Zhangliang'[29] and *Z. officinale* var. 'Zhugen'[47]. The genes predicted by BRAKER3 were then selected using TSEBRA (version 1.1.2.2)[48] with the parameter intron_support 0.2. The genes were further filtered if the proportion of repeat sequences in their coding region was greater than 50% or if the coding sequence was less than 150 bp.

The Kyoto Encyclopedia of Genes and Genomes-based functional annotations of genes were performed in Kofam (version 1.3.0)[49]. InterProScan functional analysis and Gene Ontology terms were obtained via InterProScan (version 5.53-87.0)[50]. Gene family clustering was performed with the coding sequences of protein-coding genes from

A. thaliana (TAIR10.58)[51], *Oryza sativa* ssp. *japonica* cv. Nipponbare (T2T-NIP)[52], *Ensete glaucum*[53], *E. ventricosum*[54], *Musa acuminata*[55], *M. balbisiana*[56], *M. beccarii*, *M. schizocarpa*[57], *M. textilis*[54], *M. troglodytarum*[58], *Musella lasiocarpa*[54], *Z. officinale* var. 'Zhangliang'[29], PA morph H2 and PG morph H2 in OrthoFinder (version 2.5.4)[59]. Only one transcript remained for a gene. The alignments of the identified single-copy orthologous genes and corresponding rooted species trees generated with OrthoFinder were fed into MCMCTREE in the PAML package (version 4.9j)[60] to infer the species divergence times, and the known divergence times between these species were collected from http://timetree.org.

For the origin of the DDR in Zingiberales, we identified single-copy orthologous genes as defined by OrthoMCL[61], before aligning multiple single-copy genes in MUSCLE. The extracted aligned sequences were used as input for MrBayes (version 3.2)[62] to infer the species phylogeny using a maximum likelihood approach. To estimate the divergence time of each species, information about the known divergence time data for the species was obtained from timetree (http://www.timetree.org/). The divergence time between Musaceae and Zingiberaceae was set to approximately 60 Ma, and the divergence time between *Alpinia* and *Zingiber* was set to 16 Ma. The topology of the maximum likelihood tree was used in MCMCTREE in PAML to reconstruct a divergence time tree and to calculate divergence times.

We performed all-to-all BLASTP[63] (version 2.6.0) analysis of protein sequences for *M. acuminata*, *M. balbisiana*, *Z. officinale* var. 'Zhangliang', *Z. officinale* var. 'Zhugen', PA morph H2 and PG morph H2, and between PA morph H2 and *M. acuminata*, *M. balbisiana* and *Z. officinale* var. 'Zhangliang', using an *e*-value cut-off of 1×10^{-10}. Syntenic regions within each species were identified via MCScanX[64] on the basis of the all-to-all BLASTP results. Protein sequences of homologous/paralogous gene pairs in the identified syntenic regions were first aligned in MUSCLE[65], and the protein alignments were then converted to coding sequence alignments. Finally, we calculated synonymous substitution rates for these coding sequence alignments using KaKs_Calculator (version 1.2)[66] with the NG method.

Whole-genome resequencing and SNP characterization

We selected a total of 91 *A. mutica* plants (42 PG morph and 49 PA morph) from our experimental population in which dominant homozygote PG morph individuals occurred. Young leaves were collected from the plants and snap-frozen in liquid nitrogen. Total DNA was extracted with a DNAsecure Plant Kit (Tiangen), after which 2 μg of genomic DNA from each plant was used to construct a sequencing library using a NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). Paired-end sequencing libraries with an insert size of approximately 400 bp were sequenced on an Illumina HiSeq 4000 instrument at Novogene (Beijing, China).

We filtered paired-end resequencing reads using Fastp (version 0.23.2)[67]. Reads with >40% bases having phred quality ≤20 and shorter than 70 bp were discarded, and the resulting clean reads were mapped to the PA morph H2 with bwa-mem2 (version 2.2.1)[68] using the default parameters. SAMtools was used to convert the mapping results into bam format. Duplicated reads were marked with MarkDuplicates in the Genome Analysis Toolkit (version 4.1.8.1)[69] after the bam file was sorted via SortSam in the Genome Analysis Toolkit.

The detection of sequence variants followed the best-practice workflow recommended by the Genome Analysis Toolkit. In brief, the variants were called for each accession using HaplotypeCaller, and then all gvcf files were combined with CombineGVCFs. Finally, a joint genotyping step was performed on the combined file via GenotypeGVCFs. SNPs were subsequently selected with SelectVariants. In the filtering step, the multi-filter parameters for the SNP filter were set as QD < 5.0, QUAL < 30.0, SOR > 5.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5 and ReadPosRankSum < −8.0. SNPs were removed if not bi-allelic, yielding the basic SNP set. SNP annotation was performed on the basis of the PA morph H2 genome in the package ANNOVAR (version 2015-12-14)[70].

GWAS and manual screening of SNPs

For the GWAS, SNPs with a minor allele frequency <5% and >55% missing calls were removed. The GWAS was performed using the GEMMA (version 0.98.5)[71] algorithm with univariate linear mixed models, and a score test was used. Significant associations were identified via the Bonferroni correction, which was set as $P < 0.05$/total SNPs; therefore, $-\log_{10}(P) = 8.33$. We performed manual screening of SNPs for the detection of signals potentially associated with the dichogamy phenotype. An arbitrary value was assigned to each SNP locus according to how it compared to the reference genome: homozygous reference (Ref, Ref), 0; heterozygous (Ref, Alt), 1; homozygous alternative (Alt, Alt), 2. When we consider an SNP across all individuals, individuals of the PG morph should have more alternative SNPs if this SNP is closely linked to the dichogamy phenotype, thereby producing a high total arbitrary value (theoretical value >42, given the presence of the dominant homozygous PG morph; the screening value was set to 40). Similarly, the PA morph individuals should have more homozygous reference SNPs, thereby producing a total arbitrary value close to 0 (theoretical value of 0; the screening value was set to <10). All SNPs meeting the manual screening criteria above were concatenated to reveal haplotype structures and recombination sites for all individuals that were resequenced.

We applied a 5-kb sliding window with 1-kb step approach to quantify genetic diversity around the DDR on chromosome 8 using VCFtools (version 0.1.16)[72]. A Mann–Whitney test was used to test for a difference between the PG and PA morphs, with $P < 0.05$ indicating a significant difference.

Validation of Am*SMPED1* promoters in the PA and PG morphs

To compare the promoter activities of Am*SMPED1* between the PA (*d/d*) and PG (*d/D*) morphs, upstream DNA fragments of the Am*SMPED1* coding region were cloned into the pGreenII0800-LUC vector. DNA fragments of the PA and PG promoters were 3,016 bp and 3,271 bp, respectively. These constructs were designated as pPG *SMPED1*::LUC and pPA *SMPED1*::LUC, respectively. All plasmids were subsequently transformed into *Agrobacterium tumefaciens* strain EHA105, which was then used to infiltrate *Nicotiana benthamiana* leaves for transient expression assays. After 70 h, the tobacco leaves were collected, and total protein was extracted. The activities of LUC and REN were determined using the Dual-Luciferase Reporter Assay System (Promega). Transcriptional activity was assessed by calculating the ratio of LUC to REN. The REN gene, driven by the 35S promoter in the pGreenII0800-LUC vector, served as an internal control.

RT-qPCR

First-strand cDNA was produced using 500 ng of total RNA extracted from styles and anthers of each dichogamy phenotype collected at 0:00, 3:00, 6:00, 9:00, 12:00, 15:00 and 18:00 under both natural illumination and continuous illumination conditions. RT-qPCR primers were designed for Am*SMPED1*, Am*COR27* and Am*LTR1* using Beacon Designer (version 7; Premier Biosoft). RT-qPCR primers were designed from the target gene coding sequence, selecting amplicons of 80–200 bp. All candidate primers were checked using BLAST to avoid off-target amplification. The sequences of the primers (5′ to 3′) used were as follows: actin-Forward (F), AGATAGCGGTGATTGTCGGG; actin-Reverse (R), AGCTGGTCAAAAGTGAGGCA; Am*SMPED1*-F, TGGACGCGCT-GCTTGTGAAA; Am*SMPED1*-R, CCGGAAGTCGAACACCA; Am*COR27*-F, TTCAACTTCCAACTTCTG; Am*COR27*-R, ATTGCCTAGTCTCATTCT; Am*LTR1*-F, ATAAGTCGTGTAGGTGAT; Am*LTR1*-R, CTTGCTCTGATAAT-GATGA. RT-qPCR was performed using QuantStudio 7 Flex (Applied Biosystems) in a 20-μl reaction volume using a SuperReal PreMix Plus Kit (Tiangen). The solution included 2× SuperReal PreMix Plus 10 μl, Forward primer (10 μM) 0.6 μl, Reverse primer (10 μM) 0.6 μl, 50× ROX Reference Dye 0.4 μl, cDNA 0.25 μl and RNase-free ddH$_2$O 8.15 μl. The PCR procedure was 95 °C for 15 min, followed by 40 cycles of 95 °C for 10 s, 58 °C for 20 s and 72 °C for 32 s. Melting curve analysis was

conducted from 60 °C to 95 °C. Relative expression levels were determined with the $2^{\Delta\Delta CT}$ method[73], and gene expression levels at 0:00 in the PG morph were used as a reference, meaning these levels were set to 1. We used five replicates for anthers under natural illumination and four replicates for other treatments. Three technical replicates of RT-qPCR were used. We used three methods to evaluate whether the primers used were of high efficiency: the product of RT-qPCR was a single band in the gel electrophoresis, the sequence of the RT-qPCR product was homozygous and belonged to Am*SMPED1*, and the melting curve of RT-qPCR was unimodal and the amplification plot smooth.

Locked nucleic acid in situ hybridization
Digoxigenin-labelled locked nucleic acid (LNA) oligonucleotide probes for Am*SMPED1* and *miRCURY* LNA miRNA were synthesized by Qiagen. LNA probes are chemically modified oligonucleotides with enhanced hybridization stability and specificity, enabling the detection of small RNAs with high sensitivity. The Am*SMPED1* complementary LNA-modified DNA probe was used to investigate the expression pattern of Am*SMPED1*. The Scramble-*miRCURY* LNA Detection probe has no complementarity to known plant miRNAs and was used as a negative control[74]. The probe sequences used were 5DiGN/ GTGTAACACGTCTATACGCCCA/3DiG_N (*miRCURY* LNA miRNA detection probe) and 5DiGN/AGTCAGGATCGGGCCACTGCT/3DiG_N (Am*SMPED1* detection probe). Anthers and styles of the PG and PA morphs were fixed in freshly prepared 4% (w/v) paraformaldehyde solution for over 16 h. The fixed samples were dehydrated in an ethanol gradient, cleared with xylene and embedded in paraffin. Sections of 8 μm thickness were cut, rehydrated and treated with proteinase K before hybridization with LNA probes at 45 °C for 16–20 h. After hybridization, the sections were washed with a series of SSC buffers, incubated with anti-digoxigenin-AP antibody (1:1,000 dilution, Roche) and visualized using NBT/BCIP staining as previously described[74,75]. Images were captured with a Leica stereomicroscope (DM3000) under bright-field illumination, using a ×4 objective lens.

Silencing by AS-ODNs
AS-ODNs bind to target nucleic acids to knock down the abundance of specific transcripts. The method was previously described and carried out using the styles of Brassicaceae plants[25,26]. Here this method was used on the styles of *A. mutica* plants with slight modifications. The sequences for the S-ODN and AS-ODN targeting Am*SMPED1* were designed using the RNAfold tool (http://rna.tbi.univie. ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) on the basis of the coding sequence of Am*SMPED1*. The sequences used were as follows: AS-Am*SMPED1*, 5′-AGGGGGCGGAAGGAGAAGA-3′; S-Am*SMPED1*, 5′-TCTTCTCCTTCCGCCCCCT-3′. The ODNs were synthesized by Beijing Qingke Xinye Biotechnology Co., Ltd, with thio-modifications at both ends to enhance stability and resistance to nucleolytic degradation, each covering three bases.

At 23:00 the day before flowering, buds about to bloom were collected from *A. mutica* plants. The petals, labellum and ovary were removed, leaving the remaining floral structures intact. These samples were placed in a cell culture plate containing 1 ml of solution (5 mM $CaCl_2$, 5 mM KCl, 0.01% (w/v) H_3BO_3, 1 mM $MgSO_4$ and 10% (w/v) sucrose, pH 7.5) containing each ODN. Each experimental group included three treatments: S-Am*SMPED1*, AS-Am*SMPED1* and a blank control (water). Two microlitres of S-Am*SMPED1*, AS-Am*SMPED1* at a concentration of 50 μM or water was applied to the style of each flower every 10 min for 2 h. The ODNs were dissolved in a 0.025% (v/v) Tween-20 solution before use. After treatment, the flowers were incubated in the dark at 30 °C and 80% humidity until 6:00 on the day of flowering. They were then transferred to light conditions, and the curvature of the styles was recorded at 9:00 and 15:00 to analyse the differences among the S-Am*SMPED1*, AS-Am*SMPED1* and control samples. Each treatment consisted of six biological replicates for both PG and PA flowers. The

effect of Am*SMPED1* silencing on style movement in *A. mutica* was assessed by observing the changes in style curvature.

Vectors and transgenic plants
The *A. thaliana* Col-0 accession was used in this study. Seeds were surface-sterilized with 0.1% (v/v) sodium hypochlorite for 15 min, followed by four rinses with sterile water. Surface-sterilized seeds were sown on half-strength Murashige and Skoog medium with vitamins (Coolaber, PM1011) supplemented with 1% (w/v) sucrose and solidified with 0.8% (w/v) agar (pH 6.1). The plates were sealed with medical-grade tape, stratified at 4 °C for three days in the dark and then incubated in a growth chamber under long-day conditions (plants were grown under a 16-h light/8-h dark photoperiod using plant growth lights with an intensity of 7,000 lux) at 22 °C with 75% relative humidity. After 15 days, seedlings were transferred to soil composed of a 4:3:1 (v/v/v) mixture of nutrient soil:vermiculite:perlite, supplemented with 0.1% (v/v) liquid fertilizer (Plant-Soul 100% Water Soluble Fertilizer, Wintong). The plants were grown in a controlled environment at 22 °C under the same photoperiod as above.

The promoter regions of At*SMPED1* (At5g67020) and At*SMPED2* (At3g50340) were amplified via PCR using specific primers. The At*SMPED1* promoter spans 813 bp upstream of the translation start site, amplified using the primers F: CCCAAGCTTGCTGCCTCATCG-TAAGCATCCG and R: GTTTCACTCAGGTAAGAGAAAGAGC. The At*SMPED2* promoter spans 2,090 bp upstream of the translation start site, amplified using the primers F: AACTGCAGTAGACAGAG-CAAGTGGAGTGGT and R: GCTCTAGATATAATTTTTTGGGGGGAAA-GAAGAAAAC. These promoter regions were subsequently inserted into the plant transformation vector pCAMBIA1300-221 (Cambia) to create the At*SMPED1pro::GUS* and At*SMPED2pro::GUS* constructs. The tissue localization of At*SMPED1* and At*SMPED2* expression in flowers was determined via GUS staining of At*SMPED1pro::GUS* and At*SMPED2pro::GUS* transgenic plants. Plant tissues were immersed in GUS staining buffer (50 mM sodium phosphate buffer, pH 7.0, 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.1% Triton X-100 and 2 mM X-Gluc), followed by vacuum infiltration for 10 min. The samples were then incubated at 37 °C in the dark for 12–16 h. After staining, the tissues were cleared with 70% ethanol to remove chlorophyll, and the GUS expression patterns were visualized. For the Am*SMPED1* and At*SMPED1/2* overexpression constructs (*35S::*Am*SMPED1*, *35S::*At*SMPED1* and *35S::*At*SMPED2*), the full-length coding sequences of Am*SMPED1* and At*SMPED1/2* were separately amplified via PCR and inserted into the plant transformation vector pCAMBIA1300, which adds a sequence encoding a FLAG tag at the 3′ end of the coding sequence, placed under the control of the CaMV 35S promoter.

The amiRNA against At*SMPED1* and At*SMPED2* (5′-TGATTCACG CTCGAAGAACTC-3′), which targets both At*SMPED1* and At*SMPED2*, was designed with the WMD3 tool (http://wmd3.weigelworld.org/ cgi-bin/webapp.cgi?page=Home;project=stdwmd). The pre-amiRNA sequence was cloned into the modified plant transformation vector pCAMBIA1300 under the control of the *Ubiquitin* promoter, resulting in the *ami*At*SMPED1/2* knockdown construct. All the above constructs were introduced into *A. thaliana* Col-0 via *Agrobacterium tumefaciens* GV3101-mediated transformation via the standard floral-dip method[76]. Transgene expression was confirmed in the transgenic plants using RT-PCR and RT-qPCR. To determine whether the transgenes influenced vegetative growth, all plants were observed carefully, including the WT accession and transgenic plants, from the seedling stage to the adult stage. The number of leaves on the 21st day and bolting rates on the 28th day were compared between WT and transgenic plants.

Anther staining and quantification of anther dehiscence
We grew *A. thaliana* plants in an incubator with a 16-h light/8-h dark photoperiod, where the light cycle started at 8:30. Immediately after collection, the sepals and petals were removed, and the anthers were

placed on a microscope slide. Alexander staining solution (Solarbio, G3050) was applied, and a coverslip was gently placed on top. The samples were then observed under a microscope[77]. The transgenic *Arabidopsis* (*35S*::Am*SMPED1*, *35S*::At*SMPED1*, *35S*::At*SMPED2* and *ami*-At*SMPED1/2*) and WT plants were carefully monitored to characterize anther dehiscence and assess the effect of Am*SMPED1* and its homologous genes on anther dehiscence. The number of anthers collected per sampling point was about 300 (minimum, 240; maximum, more than 360). The percentage of dehiscent anthers was then calculated for each sample. Over all nine sampling points, more than 2,700 anthers in total were observed and counted for each plant line.

Identification and reconstruction of an ancestral DDR
We identified syntenic DDRs in the genomes of banana and ginger species using MCScan (Python version)[65] on the basis of the PA morph H2 DDR on chromosome 8 with the default settings. Syntenic genes were labelled, and each DDR comprised a concatenation of these genes. The ancestral DDR was inferred according to an unrooted phylogenetic tree of DDR loci.

Detailed descriptions of phylogenomic reconstruction and evolution of *SMPED1* in angiosperms are shown in Supplementary Information.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive at the National Genomics Data Center, China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Among them, the HiFi reads, Hi-C reads and full-length transcriptome reads can be found under accession number CRA020599 (reviewer link: https://ngdc.cncb.ac.cn/gsa/s/Su7E6mDl), and the resequencing data are under accession number CRA020857 (reviewer link: https://ngdc.cncb.ac.cn/gsa/s/Z3T7vDQ6). Transcriptome reads of different time points can be shared on request from the corresponding author J.-L.Z. because these data are under further explored. Source data are provided with this paper.

Code availability
All analysis tools used in this study are publicly available, as described in Methods and Reporting Summary.

References
1. Harder, L. D. & Barrett, S. C. H. *Ecology and Evolution of Flowers* (Oxford Univ. Press, 2006).
2. Darwin, C. *The Different Forms of Flowers on Plants of the Same Species* (John Murray, 1877).
3. Barrett, S. C. H. The evolution of plant sexual diversity. *Nat. Rev. Genet.* **3**, 274–284 (2002).
4. Charlesworth, D. & Charlesworth, B. Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Evol. Syst.* **18**, 237–268 (1987).
5. Morran, L. T., Parmenter, M. D. & Phillips, P. C. Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* **462**, 350–352 (2009).
6. Bertin, R. I. & Newman, C. M. Dichogamy in angiosperms. *Bot. Rev.* **59**, 112–152 (1993).
7. Lloyd, D. G. & Webb, C. J. The avoidance of interference between the presentation of pollen and stigmas in angiosperms I. Dichogamy. *N. Z. J. Bot.* **24**, 135–162 (1986).
8. Bertin, R. I. Incidence of monoecy and dichogamy in relation to self-fertilization in angiosperms. *Am. J. Bot.* **80**, 557–560 (1993).
9. Darwin, C. *On the Various Contrivances by Which British and Foreign Orchids Are Fertilized by Insects, and on the Good Effects of Intercrossing* (John Murray, 1862).
10. Darwin, C. *The Effects of Cross- and Self-Fertilization in the Vegetable Kingdom* (Murray, 1876).
11. Renner, S. S. & Muller, N. A. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nat. Plants* **392**, 392–402 (2021).
12. Boualem, A. et al. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* **321**, 836–838 (2008).
13. Huu, C. N. et al. Presence versus absence of *CYP734A50* underlies the style-length dimorphism in primroses. *eLife* **5**, e17956 (2016).
14. Xue, H., Gong, Y., Wright, S. I. & Barrett, S. C. H. The genomic basis of the tristylous floral polymorphism—evidence for a role of gene duplications in a region of restricted recombination. *Mol. Biol. Evol.* **42**, msaf170 (2025).
15. Renner, S. S. How common is heterodichogamy?. *Trends Ecol. Evol.* **16**, 595–597 (2001).
16. Gleeson, S. K. Heterodichogamy in walnuts: inheritance and stable ratios. *Evolution* **36**, 892–902 (1982).
17. Groh, J. S. et al. Ancient structural variants control sex-specific flowering time morphs in walnuts and hickories. *Science* **387**, eado5578 (2025).
18. Li, Q.-J. et al. Flexible style that encourages outcrossing. *Nature* **410**, 432 (2001).
19. Li, Q.-J. et al. Study on the flexistyly pollination mechanism in *Alpinia* plants (Zingiberaceae). *Acta Bot. Sin.* **43**, 364–369 (2001).
20. Cui, X., Wei, R. & Huang, R. A preliminary study on the genetic system of *Amomum tsao-ko*. *J. Yunnan Univ.* **17**, 290–297 (1995).
21. Li, Q.-J. et al. Mating system and stigmatic behaviour during flowering of *Alpinia kwangsiensis* (Zingiberaceae). *Plant Syst. Evol.* **232**, 123–132 (2002).
22. Liu, M., Sun, S. & Li, Q. The relation between stigma position and receptivity in two flexistylous gingers. *Biodivers. Sci.* **15**, 639–644 (2007).
23. Chen, S. et al. Genetic and genomic insights into dichogamy in Zingiberaceae. *Plant Commun.* **6**, 101352 (2025).
24. Thompson, T. E. & Romberg, L. D. Inheritance of heterodichogamy in pecan. *J. Hered.* **76**, 456–458 (1985).
25. Huang, J. et al. Stigma receptors control intraspecies and interspecies barriers in Brassicaceae. *Nature* **614**, 303–308 (2023).
26. Zhang, L. et al. FERONIA receptor kinase-regulated reactive oxygen species mediate self-incompatibility in *Brassica rapa*. *Curr. Biol.* **31**, 3004–3016 (2021).
27. Sass, C., Iles, W. J. D., Barrett, C. F., Smith, S. Y. & Specht, C. D. Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ* **4**, e1584 (2016).
28. Carlsen, M. M. et al. Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: pushing the limits of genomic data. *Mol. Phylogenet. Evol.* **128**, 55–68 (2018).
29. Cheng, S. P. et al. Haplotype-resolved genome assembly and allele-specific gene expression in cultivated ginger. *Hortic. Res.* **8**, 188 (2021).
30. D'Hont, A. et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488**, 213–217 (2012).
31. Li, H.-T. et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* **5**, 461–470 (2019).
32. Willemstein, S. C. *An Evolutionary Basis for Pollination Ecology* Leiden Botanical Series 10 (Brill, 1987).
33. Routley, M. B., Bertin, R. I. & Husband, B. C. Correlated evolution of dichogamy and self-incompatibility: a phylogenetic perspective. *Int. J. Plant Sci.* **165**, 983–993 (2004).

34. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).

35. Harkess, A. et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **8**, 1279 (2017).

36. Kappel, C., Huu, C. N. & Lenhard, M. A short story gets longer: recent insights into the molecular basis of heterostyly. *J. Exp. Bot.* **68**, 5719–5730 (2017).

37. Haldane, J. B. S. A mathematical theory of natural and artificial selection. *Math. Proc. Camb.* **23**, 363–372 (2008).

38. Orr, H. A. & Betancourt, A. J. Haldane's sieve and adaptation from the standing genetic variation. *Genetics* **157**, 875–884 (2001).

39. Maydeu-Olivares, A. & García-Forero, C. in *International Encyclopedia of Education* 3rd edn (eds Peterson, P. et al.) 190–196 (Elsevier, 2010).

40. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **25**, 4.10.11–14.10.14 (2009).

41. Jurka, J. et al. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).

42. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).

43. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).

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

45. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

46. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

47. Li, H. L. et al. Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hortic. Res.* **8**, 189 (2021).

48. Gabriel, L., Hoff, K. J., Bruna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinform.* **22**, 566 (2021).

49. Aramaki, T. et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).

50. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).

51. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).

52. Shang, L. et al. A complete assembly of the rice Nipponbare reference genome. *Mol. Plant* **16**, 1232–1236 (2023).

53. Wang, Z. et al. A chromosome-level reference genome of *Ensete glaucum* gives insight into diversity and chromosomal and repetitive sequence evolution in the Musaceae. *GigaScience* **11**, giac027 (2022).

54. Droc, G. et al. The banana genome hub: a community database for genomics in the Musaceae. *Hortic. Res.* **9**, uhac221 (2022).

55. Martin, G. et al. Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genom.* **17**, 243 (2016).

56. Wang, Z. et al. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nat. Plants* **5**, 810–821 (2019).

57. Belser, C. et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* **4**, 879–887 (2018).

58. Li, Z. et al. The *Musa troglodytarum* L. genome provides insights into the mechanism of non-climacteric behaviour and enrichment of carotenoids. *BMC Biol.* **20**, 186 (2022).

59. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

60. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

61. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

62. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

63. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).

64. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

65. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

66. Zhang, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom. Proteom. Bioinform.* **4**, 259–263 (2006).

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

68. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (IEEE, 2019).

69. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

70. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

71. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

72. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

73. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* **25**, 402–408 (2001).

74. Javelle, M. & Timmermans, M. C. P. In situ localization of small RNAs in plants by using LNA probes. *Nat. Protoc.* **7**, 533–541 (2012).

75. Kramer, E. M. Methods for studying the evolution of plant reproductive structures: comparative gene expression techniques. *Methods Enzymol.* **395**, 617–636 (2005).

76. Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. & Chua, N.-H. *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**, 641–646 (2006).

77. Peterson, R., Slovin, J. P. & Chen, C. A simplified method for differential staining of aborted and non-aborted pollen grains. *Int. J. Plant Biol.* **1**, e13 (2010).

Acknowledgements

Author contributions

J.-L.Z., Y.D., J.-J.H., S.C.H.B., W.C. and Q.-J.L. designed, conceived and supervised the research. A.-D.H., S.-C.D., X.-C.P., J.-L.Z., Q.-J.L., H.L., J.-H.C., Y.D., Y.-L.W., W.-J.W. and Q.-Y.L. collected the samples. J.-L.Z., Q.-J.L. and Y.-L.L. conducted the Mendelian inheritance experiments and the corresponding data analysis. S.-C.D., W.C., Y.D., X.-C.P. and J.-L.Z. performed the genome assembly, annotation, GWAS and evolutionary analysis. X.-C.P. and W.C. performed the SMPED1 protein prediction. A.-D.H., J.-J.H., X.-C.P., X.-M.Z., P.-W.L., X.X. and J.-H.C. validated *SMPED1* and conducted the RT-qPCR experiments in transgenic plants. A.-D.H. and H.L. conducted the RNA in situ hybridizations. A.-D.H., X.-C.P., Y.D., J.-J.H. and J.-L.Z. conducted the experiments using AS-ODNs. J.-L.Z., A.-D.H., S.-C.D. and W.-J.W. analysed the expression data. The paper was drafted by J.-L.Z., Y.D., S.C.H.B., W.C., Q.-J.L., J.-J.H., W.J.K. and B.L. All authors contributed to the review of the paper before submission for publication and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-025-02125-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-025-02125-3.

Correspondence and requests for materials should be addressed to Jian-Li Zhao, Jia-Jia Han, Spencer C. H. Barrett or Wei Chen.

Peer review information *Nature Plants* thanks Takashi Akagi, Roberta Bergero and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

[1]State Key Laboratory of Vegetation Structure, Function and Construction (VegLab), Ministry of Education Key Laboratory for Transboundary Ecosecurity of Southwest China, Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology, and Institute of Biodiversity, School of Ecology and Environmental Science, Yunnan University, Kunming, China. [2]Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, China. [3]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China. [4]Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming, China. [5]Key Laboratory of Subtropical Medicinal Edible Resources Development and Utilization in Yunnan Province, College of Biology and Chemistry, Puer University, Puer, China. [6]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. [7]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada. [8]Deceased: Qing-Jun Li. [9]These authors contributed equally: Jian-Li Zhao, Yang Dong, Ao-Dan Huang, Sheng-Chang Duan, Xiao-Chang Peng. ✉e-mail: jianli.zhao@ynu.edu.cn; hanjiajia@ynu.edu.cn; spencer.barrett@utoronto.ca; wchenntr@gmail.com
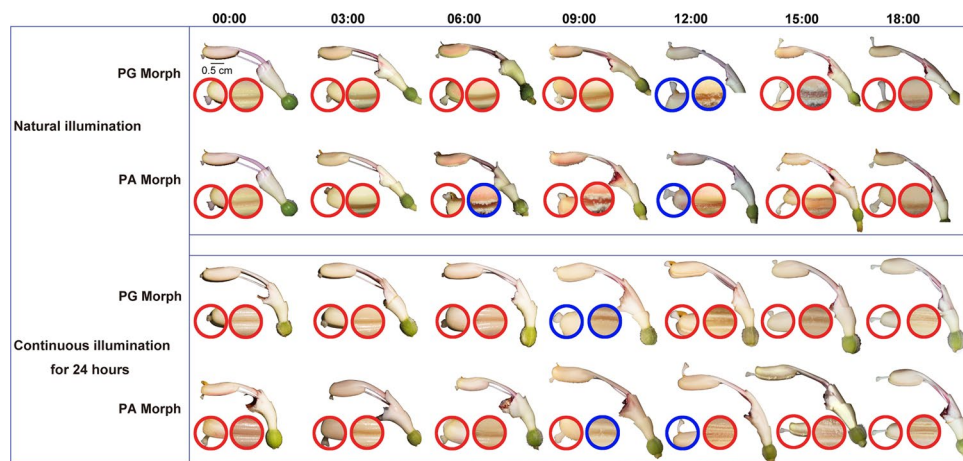
Extended Data Fig. 1 | Style and anther behavior of *Alpinia mutica* flowers under natural light conditions and continuous illumination. Progression of daily flower maturation of the PG and PA morphs under natural light conditions (top) and continuous illumination (bottom). The small circles show close-up images of the stigmas (left circles) and anthers (right circles). Images at the time of anther dehiscence and style rotation are highlighted by blue borders. The scale bar is 1 cm.
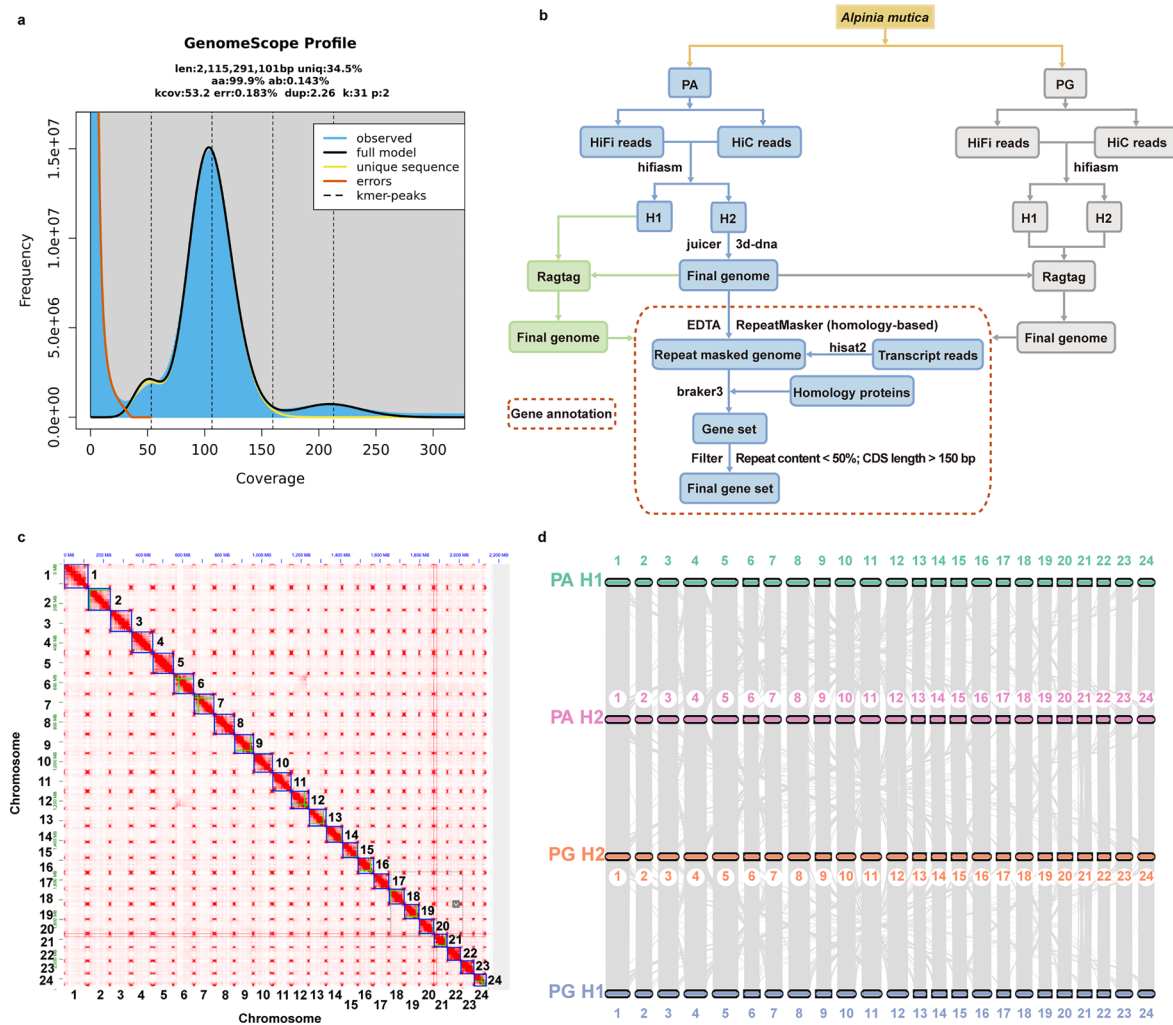
Extended Data Fig. 2 | Genome assembly and annotation of the PA and PG morphs in the *Alpinia mutica* genome. a, Genome survey of the PA morph based on k-mer distribution of Illumina data. The k-mer distributions were constructed on the basis of 31-mers. **b**, Assembly and annotation pipeline of all sequencing data into the four haplotypes of the PA and PG morphs. Red-dashed square indicates the procedure of gene annotation. **c**, Hi-C contact map showing the 24 pseudochromosomes for the PA morph haplotype 2 (H2) genome using Hi-C reads. Numbers along diagonal line are chromosome numbers. **d**, Gene synteny among the four haplotypes of PA and PG morphs.
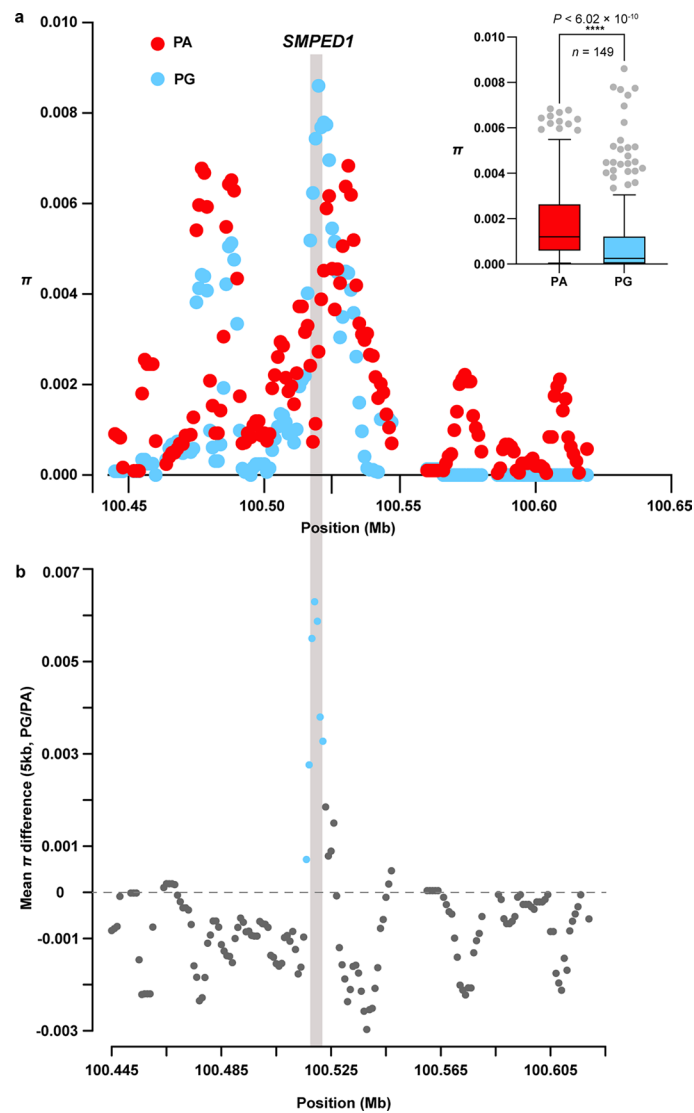
Extended Data Fig. 3 | Comparison of genetic diversity (π) between the two dichogamous morphs in and around the DDR. a, Genetic diversity (π) of PA and PG morphs across the DDR and adjacent region. Box plot in the upper right corner indicates the general π of PG is significantly lower than the π of PA with ****P < 0.0001 estimated by two-tailed Mann-Whitney U test. Number of samples are indicated on bar. The central lines within the box plots represent the medians, the box represents the interquartile range. The data are presented as the mean ± SEM and the whiskers extend to minima and maxima. **b**, Difference of π between PA Vs. PG. The gray vertical line indicates the position of *SMPED1*.
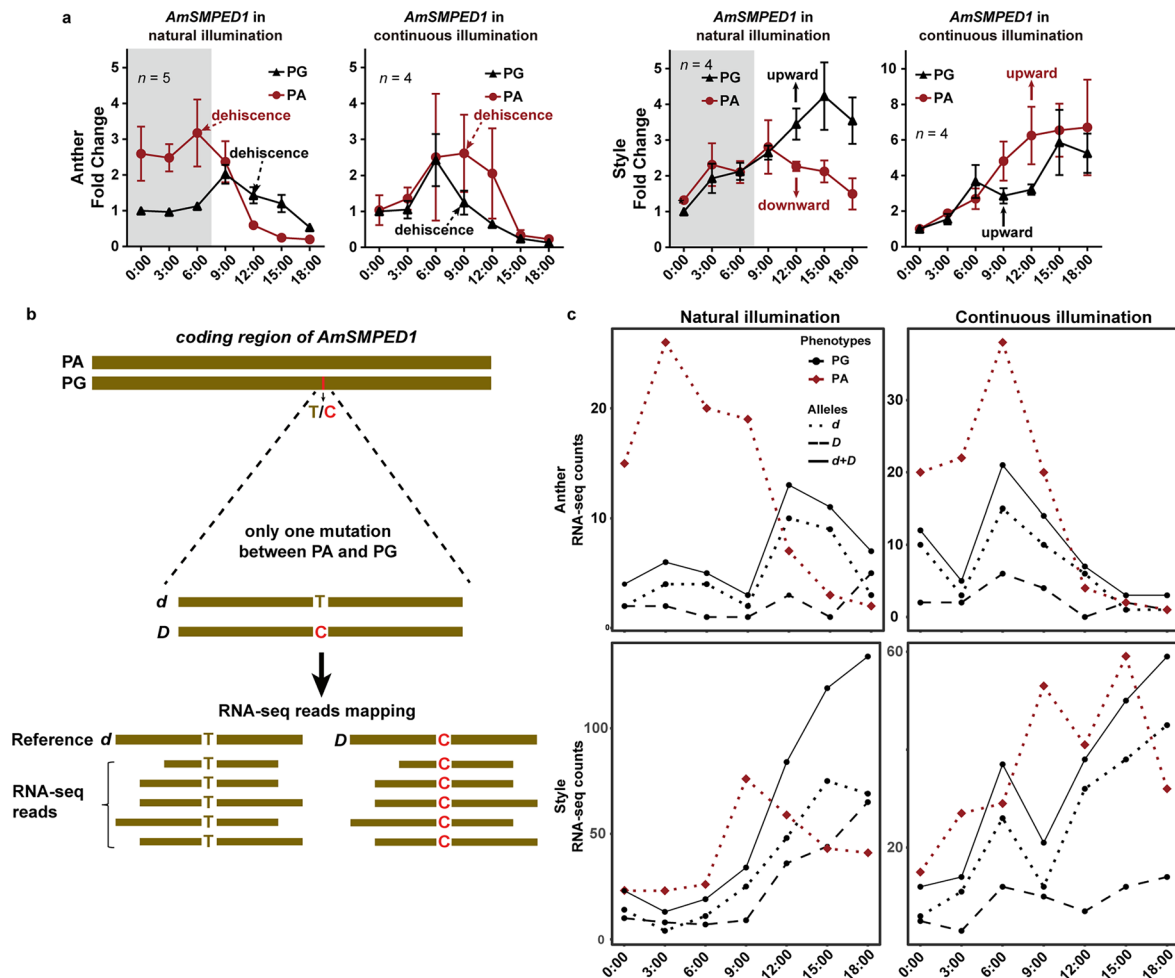
Extended Data Fig. 4 | Relative expression level of *AmSMPED1* and allele-specific expression analysis. a, Relative expression level of *AmSMPED1* in anthers (left) and styles (right) under natural conditions and continuous illumination (with light exposure after sunset), as determined by RT-qPCR data. The transition point of style reciprocal movement is indicated by the directional arrowheads (see temporal details in Extended Data Fig. 1). The data are presented as the mean ± SEM and the whiskers extend to minima and maxima. Biological replicates (*n*) of each time point are indicated. **b**, A synonymous mutation (T Vs. C) is the only difference between recessive allele *d* and dominant allele *D* in the coding region. RNA-seq reads are mapped to different alleles to explore the pattern of allele-specific expression of the recessive allele *d* and dominant allele *D*. Because expressions of the two alleles are extracted from the same RNA sequences (RNA-seq), the amount of gene expression can be calculated by the number of reads (counts) mapped to the reference allele. **c**, Allele-specific expressions of recessive allele *d* and dominant allele *D* at different times in the PA (recessive homozygosity *d/d*) and PG (dominant heterozygosity *d/D*) morphs under different illumination. Allele-specific expression of PA and PG are indicated by different colors. The recessive *d* and dominant allele *D* are represented by different dashed lines. Solid lines are the expression of recessive allele *d* plus dominant allele *D*.
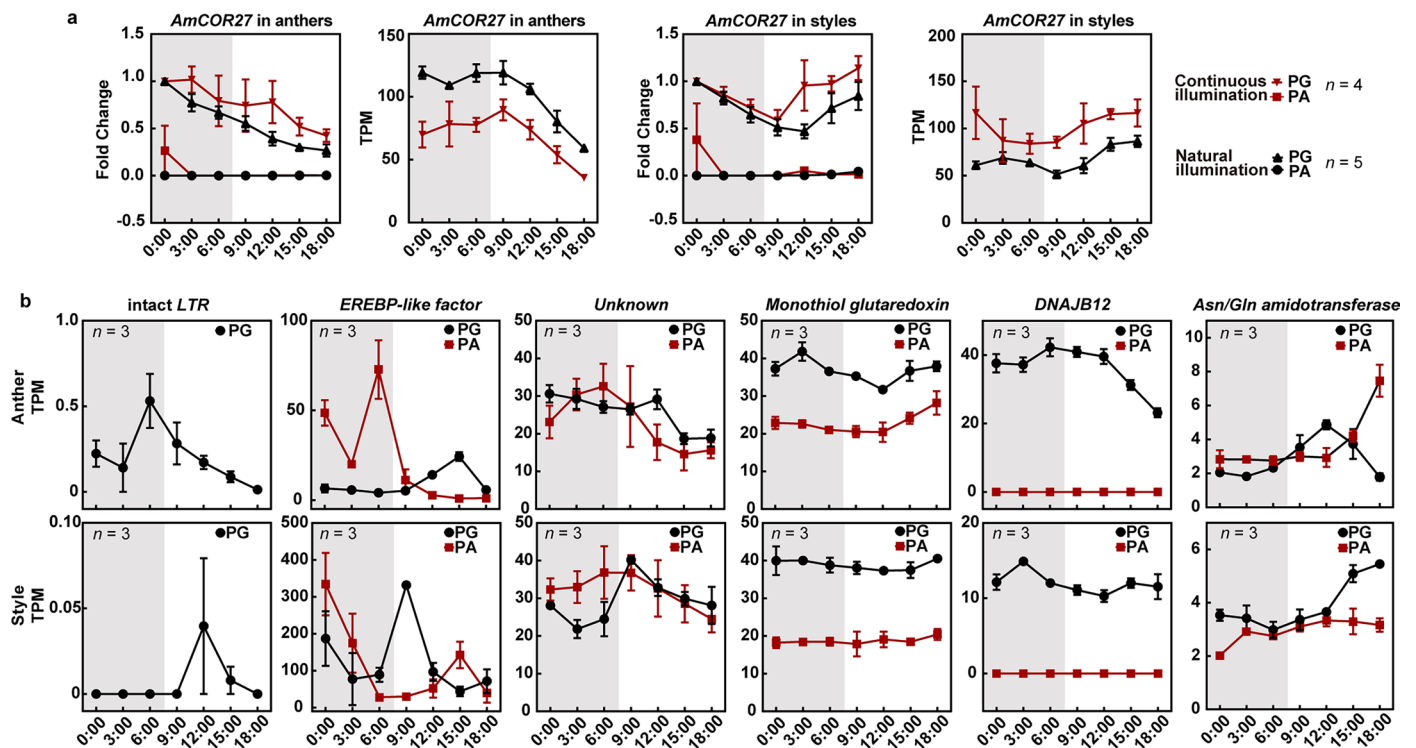
Extended Data Fig. 5 | Expression of other genes related to the DDR.
a, *AmCOR27* expression levels in anthers and styles over the course of one day. The left/right panels are from RT-qPCR/ RNA-seq for anthers and styles, respectively. *AmCOR27* was not expressed in the PA morph. **b**, Expression levels of other genes in the DDR. The gray shading indicates the night before sunrise at the location where the plants were cultivated. The data are presented as the mean ± SEM and the whiskers extend to minima and maxima. Biological replicates (*n*) of each time point are indicated.
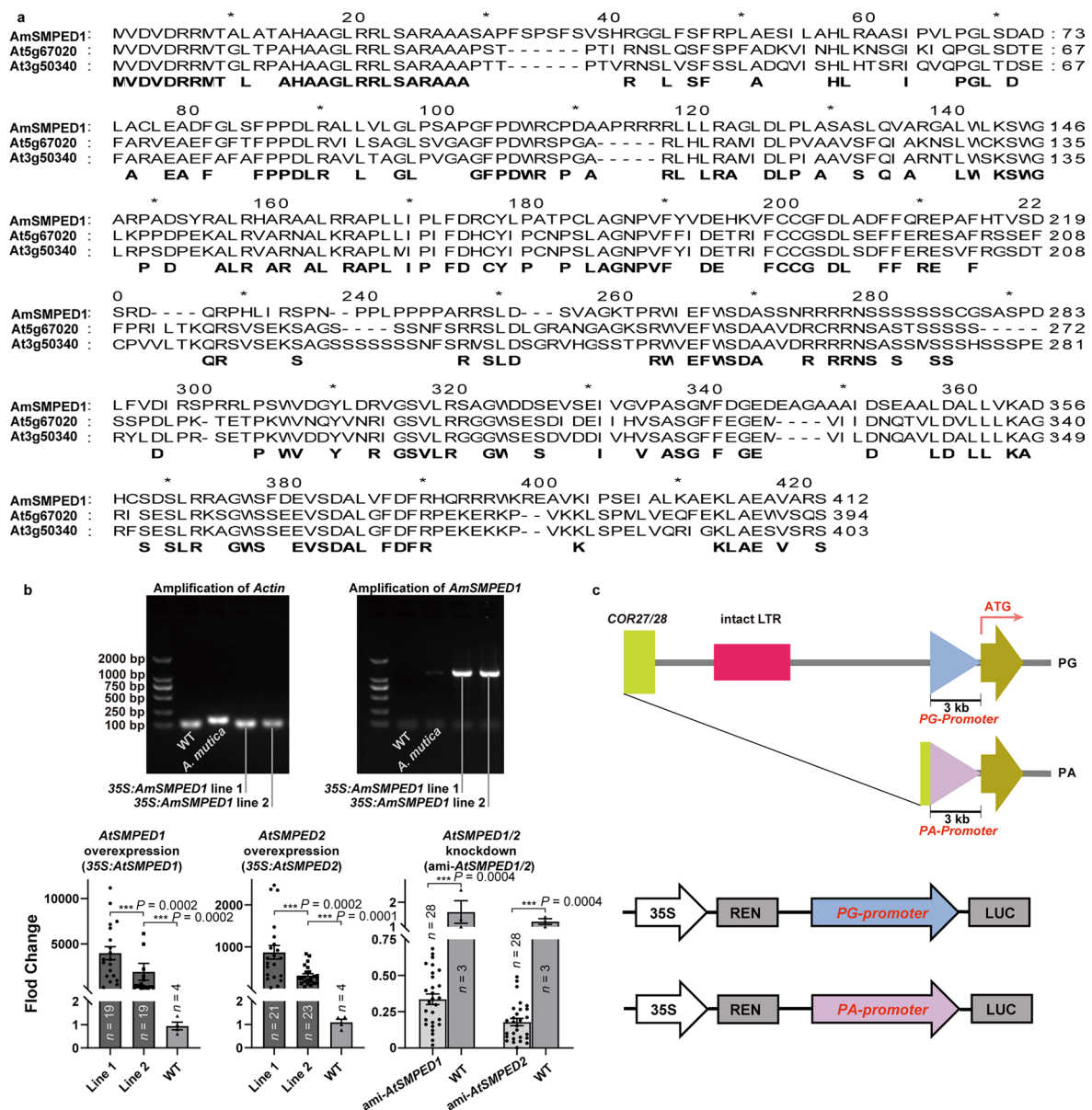
Extended Data Fig. 6 | Sequence of SMPED, overexpression/knockdown of SMPEDs in *Arabidopsis thaliana* and validation of promoters. a, Sequence of AmSMPED1 and its two homologous proteins in *Arabidopsis thaliana* (At5g67020 and At3g50340). **b**, Confirmation of overexpression of *AmSMPED1* (*35S:AmSMPED1*), *AtSMPED1* (*35S:AtSMPED1*), and *AtSMPED2* (*35S:AtSMPED2*), as well as simultaneous knockdown of *AtSMPED1* and *AtSMPED2* (ami-*AtSMPED1/2*) in *A. thaliana*. Full-length amplification of *AmSMPED1* was then performed, and the results showed that, under identical starting RNA concentrations, the amplified bands from overexpression lines were notably brighter compared to those from wild-type and positive control plants, indicating successful overexpression of *AmSMPED1* in the transgenic lines. Overexpression of

AmSMPED1 (*35S:AmSMPED1*) was confirmed by RT-PCR. Total RNA from wild-type plants, positive controls, and transgenic lines was quantified and normalized to the same initial concentration. The data are presented as the mean ± SEM and the whiskers extend to minima and maxima. Biological replicates (*n*) of each treatment are indicated. The blank space on the bar indicates contracted bar due to large values. The asterisks denote significant differences between samples using two-sided Mann–Whitney *U*-test. ***, *P* < 0.0001. **c**, Validation of promoters in the DDR. The upper diagram is the location of promoters in the DDR. "ATG" is the initiation codon of *AmSMPED1*. The lower is a diagram for validation. LUC is luciferase and REN is Renilla luciferase. The results of validation are indicated in Fig. 3b.
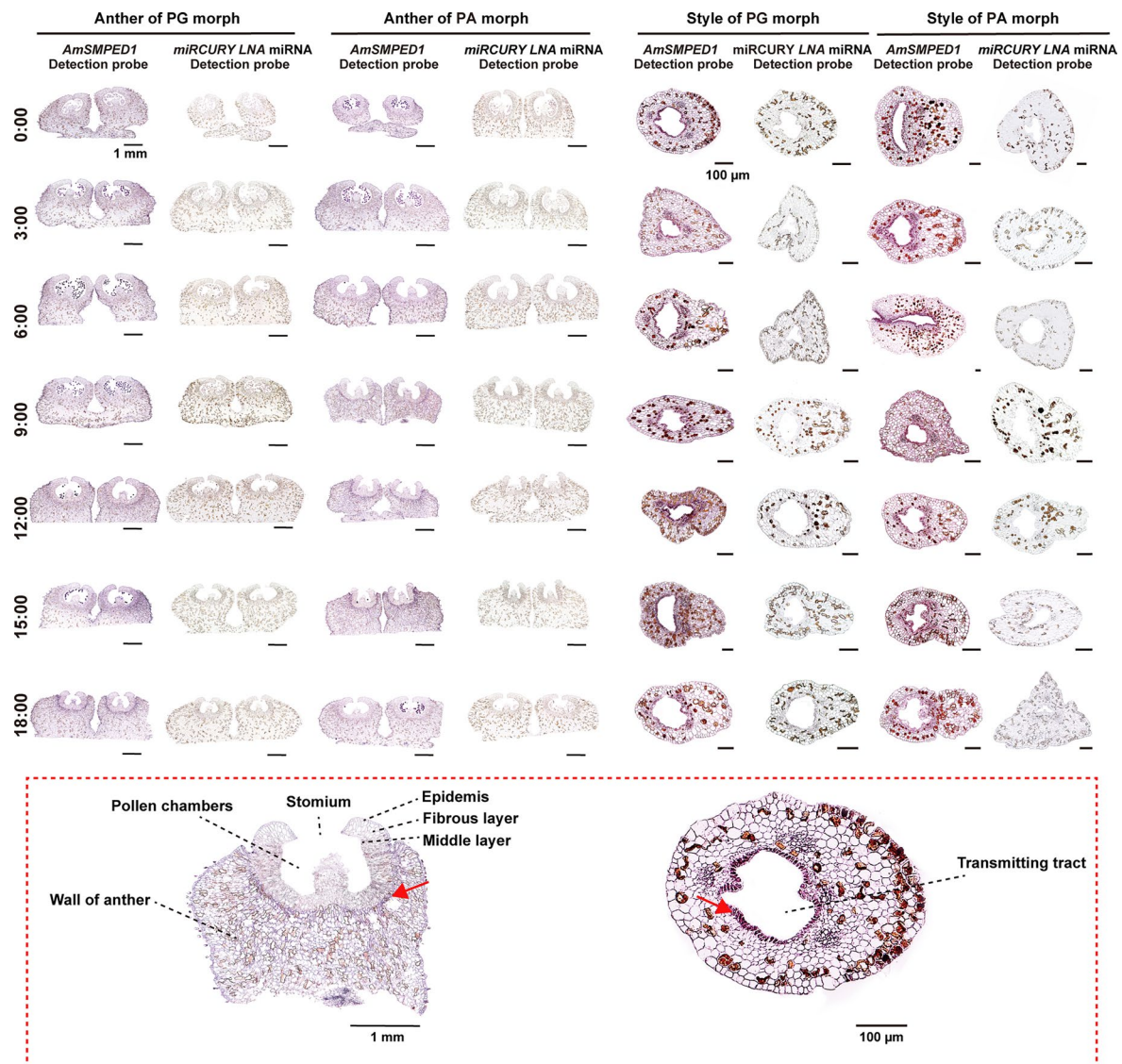
Extended Data Fig. 7 | *In situ* hybridization of *AmSMPED1* in cross sections of anthers and styles of *Alpinia mutica*. The precise localization of *AmSMPED1* in the anthers and styles was detected via digoxigenin (DIG)-labeled locked nucleic acid (LNA) oligonucleotide probes. Positive *in situ* hybridization results are characterized by deep purple signals. The *miRCURY* LNA miRNA detection probe was used as a negative control. Enlarged photographs are in red-dashed square at the bottom to illustrate detail of *in situ* hybridization. Red arrowheads indicated localization of *AmSMPED1* in the anther and style. For each time point, samples were collected three times with two samples each for a total of six biological replicates at each time point for both anthers and styles. All replicates showed similar results. Each scale bar for anthers is 1 mm and for styles is 100 μm.
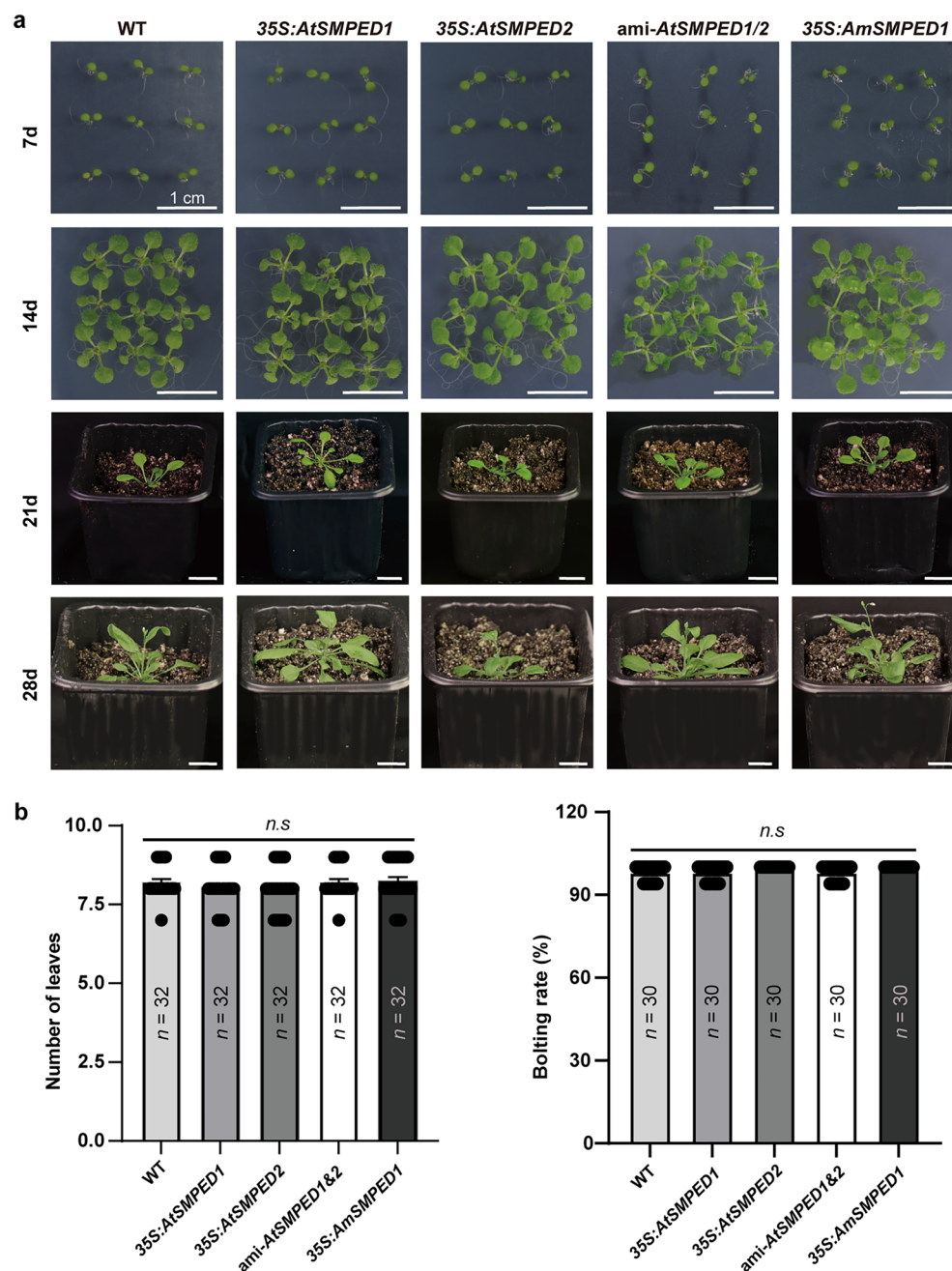
Extended Data Fig. 8 | Growing status of wild-type (WT) and transgenic lines of *Arabidopsis thaliana*. a, Representative photographs of wild-type (WT) *Arabidopsis thaliana* plants and lines overexpressing or knocked down for *AtSMPED1* and/or *AtSMPED2* or overexpressing *AmSMPED1*. WT is the *A. thaliana* Col-0 accession, *35S:AtSMPED1/2* and *35S:AmSMPED1* are the transgenic *A. thaliana* plants that overexpress *AtSMPED1/2* or *AmSMPED1*, respectively, and ami-*AtSMPED1/2* is a transgenic line in which *AtSMPED1* and *AtSMPED2* expression levels are knocked down by artificial microRNAs. The scale bar is 1 cm. **b**, Leaf number on the twenty-first day and bolting rates on the twenty-eighth day. There was no significant difference between WT and transgenic plants in their vegetative traits. The data are presented as the mean ± SEM and the whiskers extend to minima and maxima. Biological replicates (*n*) are indicated.
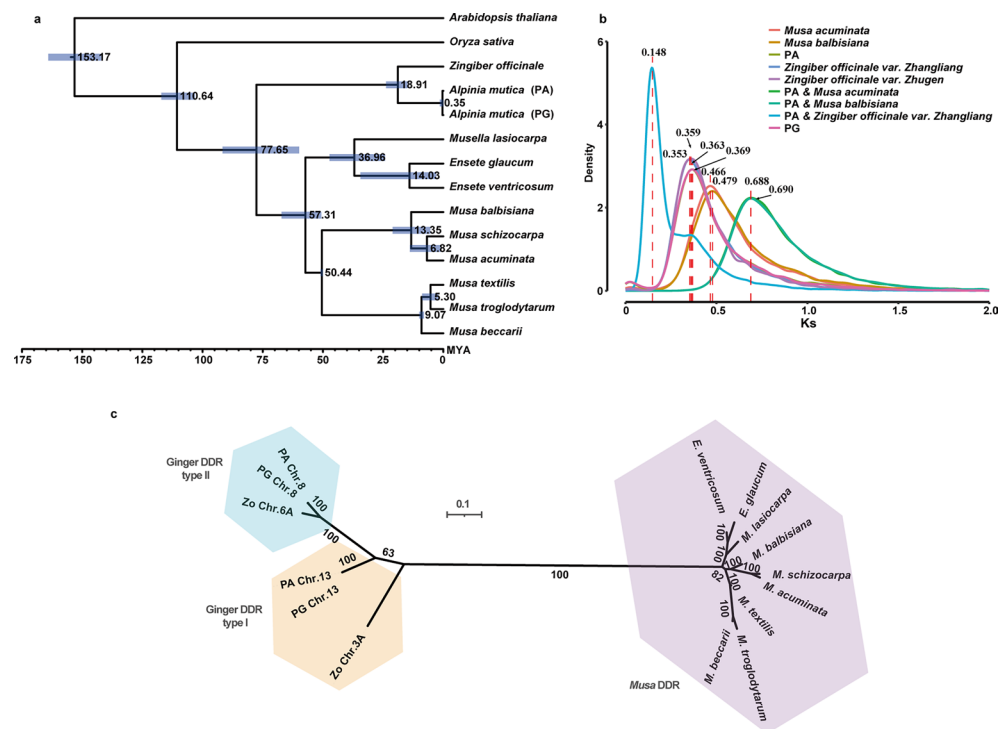
Extended Data Fig. 9 | Whole-genome duplication events and the phylogenetic tree of dichogamy-determining regions in ginger and related families. **a**, Phylogenetic tree of Musaceae and Zingiberaceae species with *Oryza sativa* and *A. thaliana* as outgroups. The blue boxes show the divergence time frames. **b**, Whole-genome duplication and divergence events revealed by *Ks* distribution. **c**, Unrooted phylogenetic tree of DDRs in Musaceae and Zingiberaceae species revealing three distinctive types of DDRs.
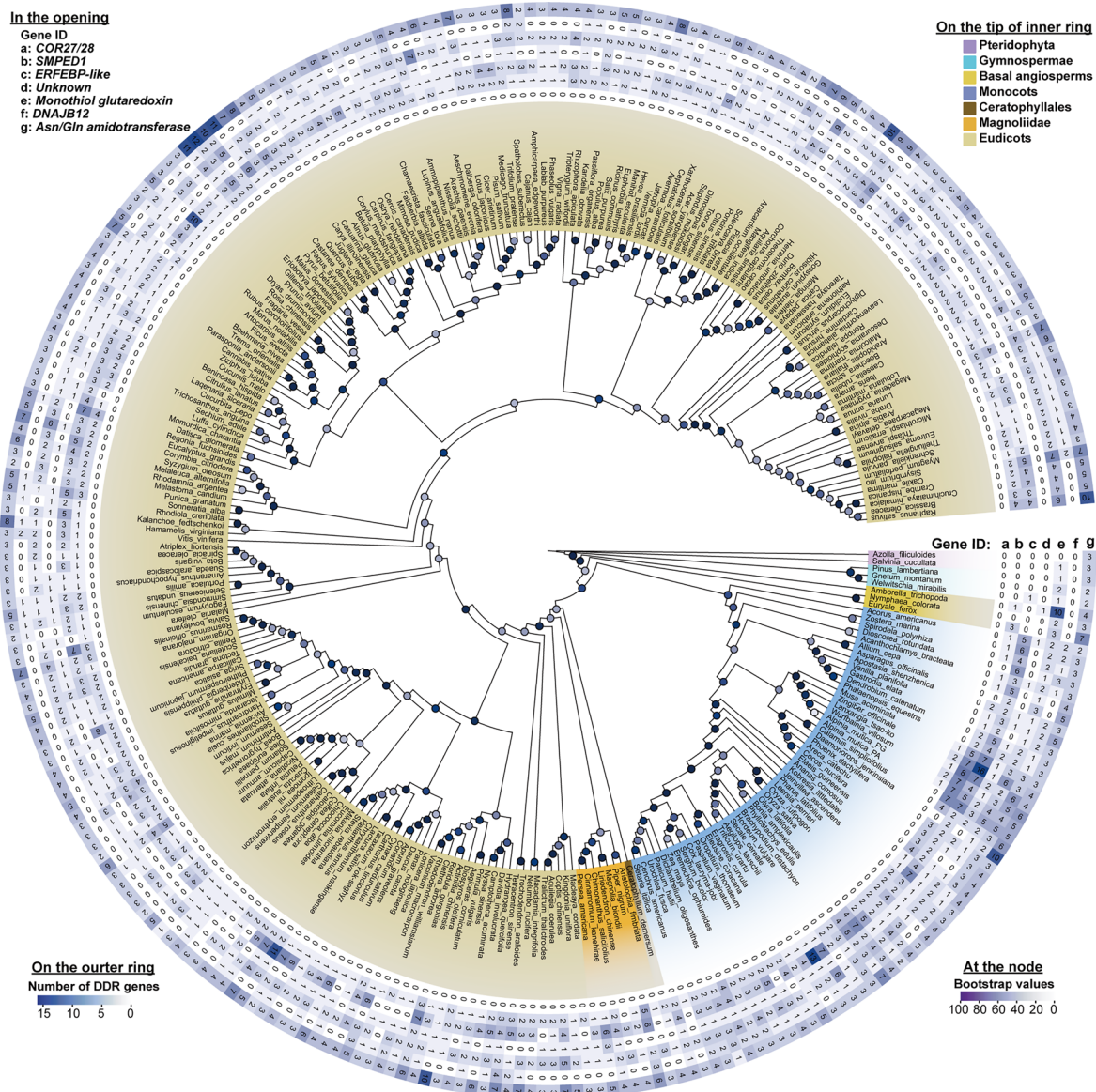
Extended Data Fig. 10 | Numbers of homologous genes in the dichogamy-determining regions (DDR) and their distributions in angiosperms. Numbers on the outer ring represent the numbers of homologous genes in the DDR indicated in Fig. 2b. Gene symbols (*COR27/28, SMPED1, EREBP-like, Unknown, Monothiol glutaredoxin, DNAJB12, Asn/Gln amidotransferase*) are represented by lowercases a-h (legend on the upper left). Darker color indicates more numbers

of genes (legend on the lower left). The different color on the middle ring represents different plant groups (legend on the upper right). Dots on the node represent bootstrap values and the darker color higher bootstrap values (legend on the lower right). The bolds with underline above the legends indicate where the legend are.

nature portfolio

Corresponding author(s): Jian-Li Zhao, Jia-Jia Han, Spencer C. H. Barrett, Wei Chen

Last updated by author(s): Aug 24, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing platforms used to generate the raw data are listed as followed: PacBio Sequel Revio, Illumina HiSeq, Illumina NovaSeq.
Data analysis	The softwares used in this manuscript include: Jellyfish v2.2.10, GenomeScope2 v2.0, Hifiasm v0.18.5-r499, 3D-DNA v180922, juicer v1.7.6, Juicebox Assembly Tools v1.11.08, Ragtag v2.1.0, TSEBRA v1.1.2.2, Kofam v1.3.0, InterProScan v5.53--87.0, BUSCO v5.5.0, RepeatMasker v4.0.9, Repbase, EDTA v2.0.1, BRAKER3 v3.0.3, Trimmomatic v0.39, HISAT2 v2.2.1, SAMtools v1.3.1, OrthoFinder v2.5.4, PAML v4.9j, MCScanX , MUSCLE, KaKs_Calculator v1.2, Fastp v0.23.2, bwa-mem2 v2.2.1, GATK v4.1.8.1, ANNOVAR v2015-12-14, GEMMA v0. 98.5, VCFtools v0.1.16, featureCounts software, GraphPad Prism v9.2.0, Beacon Designer v7, OrthoMCL, MrBayes v3.2, IQ-TREE, ASTRAL, BLASTP v2.6.0, TBtools v0.6669, SPSS v26.0, Alphafold2 based on DNN, Alphafold2 based on LLM, Colabfold, MSTA mode of US-Align, VEGAN, AutoSite, and DeepFRI.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences (GSA) and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Among them, the HiFi reads, Hi-C reads and full-length transcriptome reads can be found under accession number CRA020599 (reviewer link: https://ngdc.cncb.ac.cn/gsa/s/Su7E6mDl), and the resequencing data are under accession number CRA020857 (reviewer link: https://ngdc.cncb.ac.cn/gsa/s/Z3T7vDQ6). Transcriptome reads of different time point can be shared on request from the corresponding author Jian-Li Zhao because these data are under further explored.

Research involving human participants, their data, or biological material

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. For any experiments in this study, we have sufficient samples for analysis. For experiments to determine inheritance of flexistyly, wo selected 4 different species (A. kwangsiensi, A. blepharocalyx, A. galanga, and A. oblongifolia) (Table S1). Our collections of A. mutica and A. kwangsiensis) with 2-3 populations to investigate the ratio of PG and PA in nature. Total number of individuals for the 4 species are 312, 1824, 187, 438 respectively (details in Supplementary Table 1). We also used 12 individuals of A. mutica and 8 individuals of A. kwangsiensis as parents to do selfing and outcrossing. One individual of PA morph and One individual of the dominant homozygote PG morph (identified by progeny testing) were used for genome assembly. A. mutica is a random outcrossing species, we randomly selected 49 PA individuals and 42 PA individuals for GWAS. Samples from the population of A. mutica in the common garden which was derived from few seeds that can exclude impact of other genetic factors not related to flexistyly. To test gene expression model in a day, we collected RNA samples at 7 different time points and at 3 replicates for RNAseq. There were 4-5 biological replicates of RT-qPCR for gene expression validation with three technical replicates. In the experiments of transgenic plants, the number of anthers collected per sampling point (9 time points for every treatment) was 240-360. Vegetative traits were observed by three times. 32 and 30 plant individuals were used to measure the number of leaves and bolting rates, respectively. For the phylogenomic evolutionary analysis, genomes of 281 angiosperms were selected, covering most of orders with genomes were sequenced in angiosperms.
Data exclusions	No data were excluded from the analysis.
Replication	Three biological replicates were used for RNAseq. Four to five biological replicates were used for RT-qPCR with three technical replicates. For in situ hybridization, for each time point, samples were collected three times with two samples each for a total of six biological replicates at each time point for both anthers and styles. All replications were successful and were used.
Randomization	The sampling processes were conducted randomly.

Blinding	Blinding was not necessary during sampling collection, since we knew the morphotypes of Alpinia individuals and also knew the treatment of transgenic Arabidopsis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-DIG–AP, Fab fragments (supplier: Roche; catalog number: 11093274910). Working dilution: 1:1,000 in blocking buffer for RNA in situ hybridization.
Validation	The commercial anti-Digoxigenin–AP (Fab fragments) antibody has been widely validated for RNA in situ hybridization in multiple published studies and in our study (Extended Date Fig. 7). Product profile: https://www.sigmaaldrich.com/catalog/product/roche/11093274910?lang=en®ion=AU

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	Wild type Arabidopsis thaliana Col-0 accession were obtain from our laboratory. Transgenic plants were built in our laboratory. Alpinia mutica was collected from Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences.
Novel plant genotypes	Overexpression of AtSMPED1, AtSMPED2, AmSMPED1 in Arabidopsis thaliana and knockdown of AtSMPED1, AtSMPED2 in Arabidopsis thaliana, producing corresponding genotypes. To obtain genotypes of overexpression, promoter regions were inserted into the plant transformation vector pCAMBIA1300-221 (Cambia, Australia). To obtain genotypes of knockdown, the pre-amiRNA sequence was cloned into a modified plant transformation vector pCAMBIA1300 under the control of the Ubiquitin (Ubi) promoter. All above constructs were introduced into A. thaliana Col-0 via Agrobacterium (Agrobacterium tumefaciens) GV3101-mediated transformation via the standard floral-dip method. There were two lines and 30-45 individuals for each treatment.
Authentication	Gene expressions of all individuals of Arabidopsis thaliana for anther observations were characterized by RT-qPCR. The tissue localization of AtSMPED1 and AtSMPED2 in flowers was determined by GUS staining. Validation of AmSMPED1 promoters was conducted in Agrobacterium tumefaciens strain EHA105 using pGreenII0800-LUC vector.