nature plants

Article https://doi.org/10.1038/s41477-026-02223-w

A distorter–restorer system drives quantitative reproductive isolation in rice

Received: 26 June 2025

Accepted: 8 January 2026

Yu Zhang[1,2,7] ✉, **Ying Yang**[1,7], **Chuanlin Shi**[3,4,7], **Qiuhong Pu**[1], **Jiawu Zhou**[1], **Wenchuang He**[3], **Peng Xu**[5], **Qiang Xu**[3], **Fengyi Hu**[6], **Xuanchen Song**[1,6], **Xiaohan Jiang**[1], **Yonggang Lv**[1], **Liu He**[1,6], **Xianneng Deng**[1], **Lianguang Shang**[3] ✉ & **Dayun Tao**[1] ✉

Hybrid sterility and segregation distortion are the major forms of postzygotic reproductive isolation in rice, yet the molecular basis of their quantitative variation remains unclear. Here we identify *S44*, a natural distorter–restorer system in *Oryza longistaminata*/Asian cultivated rice hybrids, comprising four tightly linked elements—*Reproductive Isolation Distorter* (*RID*), *Reproductive Isolation Restorer* (*RIR*), *Reproductive Isolation Activator* (*RIA*) and *Reproductive Isolation Suppressor* (*RIS*)—which collectively regulate hybrid male sterility and segregation distortion. The distorter *RID* triggers the elimination of *O. sativa* cultivar RD23 pollen, whereas the restorer *RIR* selectively safeguards *O. longistaminata* gametes, thereby preferentially transmitting its allele into the progeny. *RIS* and *RIA* fine-tune segregation distortion. We further demonstrate that the allelic conflicts at the *S44* locus drive quantitative reproductive isolation between *O. longistaminata* and other rice lineages, and CRISPR-engineered *RID* knockout can universally overcome *S44*-mediated reproductive barriers in the AA genome, enabling revolutionary cross-species breeding. This distorter–restorer system provides a unique genetic module for deciphering speciation mechanisms and advancing crop breeding strategies.

Reproductive isolation is a common biological phenomenon that establishes barriers to genetic exchange between divergent populations, serving dual roles in maintaining species identity and driving speciation processes[1–4]. Hybrid sterility emerges as a predominant postzygotic mechanism that reduces hybrid fitness through deleterious genetic interactions between differentiated lineages[5,6]. In rice, this phenomenon not only restricts genetic introgression of advantageous traits but also impedes the exploitation of heterotic potential between wild and cultivated species. To date, 15 hybrid sterility loci have been characterized, including *S5*, *S7*, *HSA1* and *ESA* for hybrid female sterility[7–10]; *Sa*, *Sc*, *qHMS1*, *qHMS7*, *S27*/*S28*, *DGS1*/*DGS2*, *S22B*, *S22A* and *pf12*/*RHS12*/*Se* for hybrid male sterility[11–20]; *S1*, responsible for both male and female sterility[21–23]; and *DPL1*/*DPL2*, associated with hybrid pollen germination[24].

The genetic models of hybrid sterility can be categorized into two principal paradigms in rice. The first involves a poison–antidote

---

[1] Yunnan Seed Laboratory, Yunnan Key Laboratory for Rice Genetic Improvement, Food Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, China. [2] College of Agriculture and Life Sciences, Kunming University, Kunming, China. [3] State Key Laboratory of Genome and Multi-omics Technologies, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [4] Ministry of Education Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Jiangxi Agricultural University, Nanchang, China. [5] State Key Laboratory of Plant Diversity and Specialty Crops, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, China. [6] Innovation Center of Perennial Rice Technology in Yunnan, New Cornerstone Science Laboratory, School of Agriculture, Yunnan University, Kunming, China. [7] These authors contributed equally: Yu Zhang, Ying Yang, Chuanlin Shi. ✉e-mail: zhangyu_rice@163.com; shanglianguang@caas.cn; taody12@aliyun.com

or killer–protector system requiring coordinated action of two or three components: a sporophytically expressed toxin (killer) inducing non-selective gamete elimination, and a gametophytic antidote (protector) conferring allele-specific protection. This model is exemplified by the *S1*, *S5*, *qHMS7*, *pf12*/*RHS12*/*Se* and *qHMS1* loci, where tightly linked genes orchestrate selective gamete survival[7,9,14,16–19,22,23]. The second paradigm features a killer–target system comprising *trans*-acting lethal factors that interact with specific targets to eliminate gametes[25], as demonstrated by the *Sc* locus in intraspecific hybrids[20].

Segregation distortion (SD), characterized by non-Mendelian allele transmission, frequently accompanies hybrid sterility through mechanisms involving meiotic drivers or post-meiotic gamete elimination[25]. Meiotic drivers such as *Ab10* in maize and *CDC42* in *Mus musculus* mediate allele-specific chromosome segregation[26,27], while gamete killers, such as *wtf* genes in fission yeast[28], *Spok* genes in *Podospora anserina*[29], the SD system in *Drosophila melanogaster*[30,31] and the sup-35/pha-1 element in *Caenorhabditis elegans*[32], execute post-meiotic elimination of specific haplotypes. In rice, the cumulative effects of multiple sterility loci establish varied degrees of reproductive isolation[18], yet the capacity of allelic conflicts at single loci to generate graduated reproductive isolation remains underexplored. The *S44* locus, previously mapped to chromosome 6 in *O. longistaminata*/RD23 hybrids[33], presents an exceptional model to investigate this phenomenon.

Here we demonstrate that the *S44* locus constitutes a unique four-element distorter–restorer system governing reproductive isolation. *Reproductive Isolation Distorter* (*RID*), encoding a sporophytic distorter, induces RD23 gamete elimination, dependent on the activation effect of *Reproductive Isolation Activator* (*RIA*). *Reproductive Isolation Restorer* (*RIR*) produces a gametophytic restorer for self-protection. *Reproductive Isolation Suppressor* (*RIS*) as a suppressor and *RIA* as an activator fine-tune SD in the offspring. *RIR* targets *RID* for degradation and neutralizes its cytotoxic effect. Furthermore, our results reveal that allelic conflicts of distinct *S44* haplotypes generate quantitative variation in reproductive isolation between *Oryza longistaminata* and other lineages in rice. Compared with the classical killer–protector or toxin–antidote system[7,14], this quadripartite distorter–restorer system expands the conventional genetic model by incorporating regulatory modifiers that enable tunable reproductive barriers, providing unprecedented resolution for studying speciation. Our findings establish a genetic framework for understanding how single-locus allelic conflicts can drive speciation processes while maintaining adaptive genetic exchange under specific haplotype combinations.

Results

S44 governs hybrid male sterility and SD

To clone the *S44* locus responsible for hybrid pollen semi-sterility in *O. longistaminata*/RD23 hybrids, we developed a near-isogenic line (NIL-*S44*) by introgressing a segment spanning *S44* from *O. longistaminata* into the RD23 genetic background (Extended Data Fig. 1a,b). The two parents RD23 and NIL-*S44* and their F_1 hybrids showed indistinguishable vegetative and reproductive morphologies (Fig. 1a); however, the hybrids displayed pollen semi-sterility with normal spikelet fertility (Fig. 1b,c). Semi-thin sections revealed that approximately half of the pollen grains in the hybrids were malformed at the trinucleate stage, ultimately leading to pollen abortion, manifested by sparse starch content, irregularly thicker pollen walls and disorganized mitochondrial matrices (Extended Data Fig. 1c,d).

Transmission ratio distortion analysis revealed that female gametes transmitted both parental alleles equally, whereas male gametes carrying the *O. longistaminata S44* haplotype exhibited a marked transmission advantage in the progeny (Fig. 1d). This suggests that RD23-type pollen was partially aborted and failed to transmit to the progeny. Consistent with this pattern, the population segregating for semi-sterile and fertile pollen exhibited a bimodal distribution and fitted a 1:1 ratio in the segregation population (Extended Data Fig. 1e). Further segregation analysis showed severe SD at the *S44* locus, with a significantly higher prevalence of the NIL-*S44* genotype than the RD23 genotype (Extended Data Fig. 1f). These results confirm that *S44* is a single-locus determinant of hybrid male sterility, selectively eliminating RD23-derived pollen.

The *S44* locus was initially mapped to the region between RM5814 and RM20695 on chromosome 6 (ref. 33). To clone *S44*, we further delimited it to a 29.4-kb region flanked by the markers OL3019-2 and OL3025-1. Comparative genomic analysis revealed three genes ($ORF1^L$, *ORF2* and *ORF3*) in the *S44* region in *O. longistaminata*, which retains only an $ORF1^R$ allele in RD23 (Fig. 1e). $ORF1^R$ and $ORF1^L$ were predicted to encode homologous coiled-coil-domain-containing protein 90A, with $ORF1^R$ distinguished by a 9-bp insertion in the first exon and four synonymous single-nucleotide mutations, resulting in a three-amino-acid deletion in $ORF1^L$ (Fig. 1f). Expression profiling revealed ubiquitous transcription of $ORF1^L$/$ORF1^R$ across tissues, including mature pollen (Extended Data Fig. 2a,d,e). In contrast, *ORF2* and *ORF3*, which encode an unknown protein and an F-box protein, respectively, were highly expressed in the anther, particularly in the pollen grains (Extended Data Fig. 2b–e).

A distorter *RID* functions as a sporophytic pollen killer

To elucidate the genetic mechanism of *S44*, we first employed the CRISPR–Cas9 editing system to knock out *ORF2* in the F_1 hybrids. Sequencing confirmed a frameshift mutation caused by a G deletion, leading to the loss of function of *ORF2* in all edited lines (Extended Data Fig. 3a). All independent *ORF2* knockout lines completely rescued pollen fertility and restored the Mendelian segregation ratio (1:2:1) for *RR*:*RL*:*LL* genotypes in the T_1 progeny (Fig. 2a–c and Extended Data Fig. 3b). These results suggest that *ORF2* is indispensable for hybrid pollen abortion and SD.

To validate the function of *ORF2*, we introduced the *ORF2* genomic fragment, including its native promoter and coding region, into RD23 and F_1 hybrids (Extended Data Fig. 3c). There was no significant difference in pollen fertility between RD23 and RD23-t*ORF2* in the T_0 and T_1 generations (Extended Data Fig. 3d,e), whereas all transgenic lines displayed complete or high pollen sterility in F_1 hybrids, with T_1 segregants showing severe transmission ratio distortion and preferential elimination of RD23-derived gametes (Fig. 2d–f and Extended Data Fig. 3f,g). Progeny lines expressing *ORF2* in the NIL-*S44* background exhibited pollen semi-sterility in the T_1 segregation population (Extended Data Fig. 3g). To confirm this, we introduced the *ORF2* genomic fragment into NIL-*S44*, and the pollen of transgenic plants was semi-sterile, indicating that *ORF2* kills the pollen independent of heterozygosity (Extended Data Fig. 3h). We also observed a significant negative correlation between the elevated expression level of *RID* and the decreased pollen fertility in the F_1-t*RID* and NIL-*S44*-t*RID* transgenic plants, compared with the controls (Extended Data Fig. 3i,j). These results demonstrate that *ORF2*, acting as a distorter, could kill RD23 gametes in the hybrids in a parent-genotype-dependent fashion, which confers transmission advantages to *O. longistaminata* gametes in heterozygotes. We thus designated *ORF2* as *RID*.

A restorer *RIR* acts as a gametophytic pollen protector

The selective abortion of RD23-derived pollen in the F_1 hybrids implied the existence of a protective factor in NIL-*S44*. To test this, we designed two target sites in the first and third exons to edit the *ORF3* gene (Extended Data Fig. 4a). However, no homozygous *orf3* mutant was recoverable in the NIL-*S44* and F_1 hybrids, suggesting that *ORF3* is indispensable for pollen fertility. To further validate the function of *ORF3*, we introduced its native genomic sequence into the F_1 hybrids (Extended Data Fig. 4b). Transgenic T_0 plants (*RL*; t*ORF3*/−) showed approximately 75% pollen fertility (Fig. 2g,h), while T_1 homozygotes

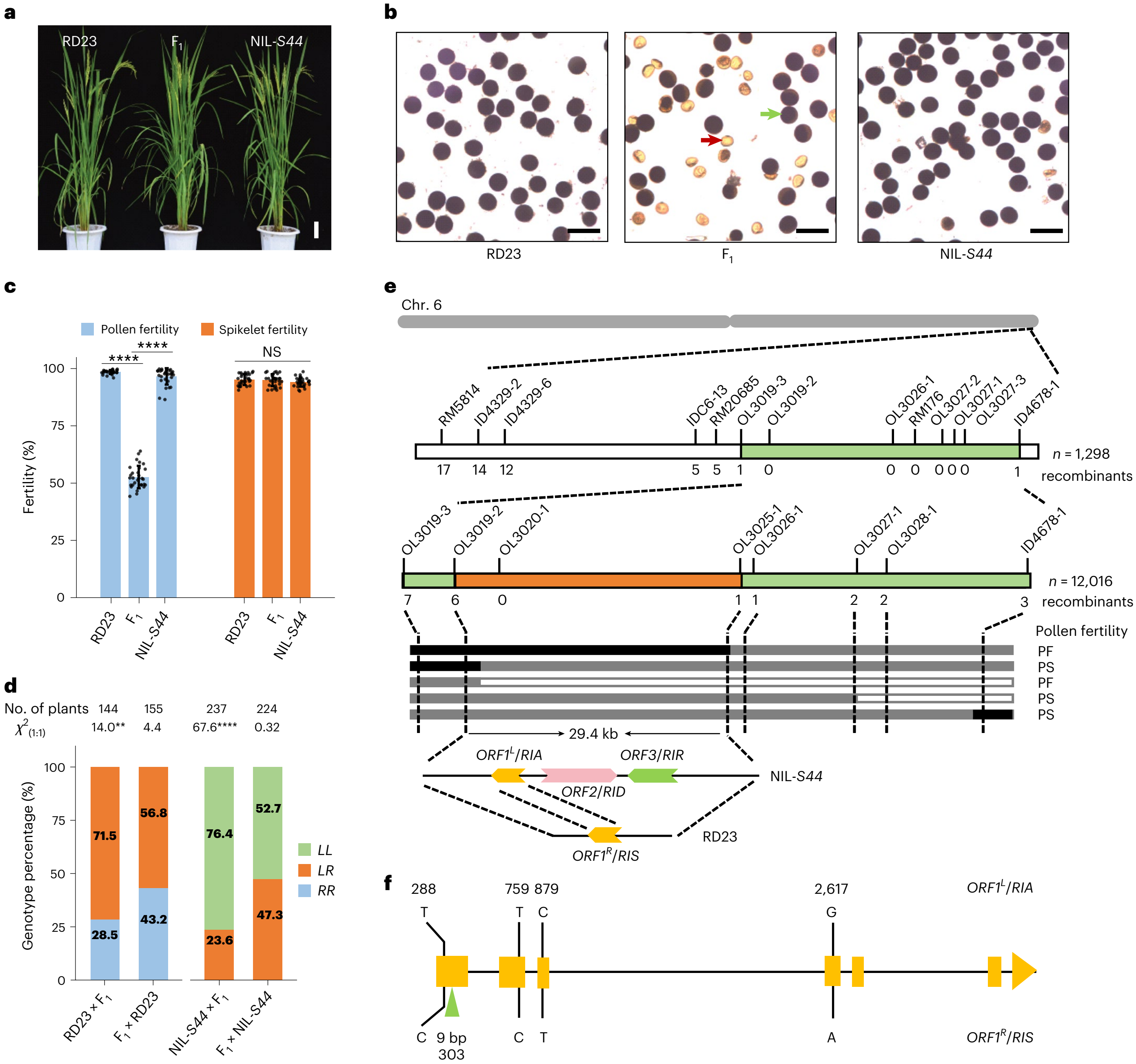


Fig. 1 | Identification and cloning of *S44*. **a**, Morphological comparison of RD23, hybrid F_1 and NIL-*S44*. Scale bar, 10 cm. **b**, Pollen fertility in RD23, NIL-*S44* and their F_1 offspring is shown via I_2-KI staining, with the red arrow indicating aborted pollen and the green arrow showing fertile pollen. Scale bars, 100 μm. **c**, Quantification of pollen fertility and spikelet fertility in RD23, NIL-*S44* and their F_1 hybrid offspring. The data are presented as mean ± s.d. (n = 20 plants). ****P < 0.0001 according to one-way analysis of variance (ANOVA) and Tukey's honest significant difference (HSD) test. NS, not significant. **d**, Gamete transmission in the F_1 hybrids was tested in reciprocal crosses between the F_1 and RD23 or NIL-*S44*. The individuals derived from those crosses were genotyped using the SSR markers OL3019-2 and OL3027-1 to distinguish among three types of alleles. *RR*, RD23 genotype; *LL*, *O. longistaminata* genotype; *RL*, heterozygous. **P<0.01; ****P<0.0001 according to χ^2 tests. **e**, *S44* was first delimited to a 108.94-kb region between OL3019-3 and ID4687-1 using 1,298 BC_8F_2 individuals and then further narrowed down to the genomic region between markers OL3019-2 and OL3025-1 using 12,016 BC_8F_3 individuals from BC_8F_2 heterozygous plants at the *S44* locus. Sequence analysis revealed three genes (*ORF1^L/RIA*, *ORF2/RID* and *ORF3/RIR*) within a 29.4-kb region in NIL-*S44*, and only a 10.4-kb region in the corresponding RD23 genome containing one gene (*ORF1^R/RIS*). PF, pollen fertility; PS, pollen semi-sterility. The black boxes, grey boxes and white boxes denote the *O. longistaminata* genotype, the heterozygous genotype and the RD23 genotype, respectively. The green and orange segments indicate the primary and secondary mapping intervals, respectively. The white segments represent Chr. 6 segments. **f**, Genomic structure of the candidate genes *ORF1^R/RIS* and *ORF1^L/RIA*. The orange vertical bars represent exons, with four single nucleotide polymorphism variations and a 9-bp insertion. The positions of variations are displayed in the coding regions. The green triangle denotes the insertion sequence. *ORF1^L/RIA*, *ORF1* allele from *O. longistaminata*; *ORF1^R/RIS*, *ORF1* allele from RD23. For all P values in **c** and **d**, see Source Data.

(*RL*; t*ORF3*/*ORF3*) achieved full fertility restoration with restored Mendelian segregation (Fig. 2i and Extended Data Fig. 4c). These results suggest that *ORF3* can rescue pollen fertility in a gametophytic manner and recover SD.

The lethality of homozygous *orf3* mutants in both NIL-*S44* and F_1 hybrids suggested a lethal interaction between *ORF3* and *RID*. To test whether *ORF3* is indispensable for viability under *RID* activity, we generated CRISPR–Cas9 knockouts targeting both loci in the F_1 hybrids

(Extended Data Fig. 4d). The double mutant restored full pollen fertility (Fig. 2j,k and Extended Data Fig. 4e), with selfed progenies fitting the 1:2:1 segregation ratio for the homozygous *RR* type, the heterozygous type and the homozygous *LL* type (Fig. 2l). This genetic epistasis demonstrates that *ORF3* specifically antagonizes *RID*-mediated distortion; we therefore renamed it *RIR*.

Both $ORF1^{R}$ and $ORF1^{L}$ are essential to male gametogenesis

To elucidate the molecular mechanism underlying *S44*-mediated hybrid sterility, we systematically investigated the function of $ORF1^{R}$ and $ORF1^{L}$ through genetic complementation and loss-of-function analyses. Transgenic lines expressing the $ORF1^{R}$ genomic fragment in the hybrid F$_1$ and NIL-*S44* backgrounds exhibited pollen fertility comparable to that of the controls (Extended Data Fig. 5a,b), indicating that $ORF1^{R}$ alone does not affect the hybrid sterility phenotype. To further dissect its biological role, we generated an $orf1^{r}$ knockout mutant via CRISPR–Cas9 editing in the RD23 background (Extended Data Fig. 5c). Two independent mutant lines carrying either a single-nucleotide (C) or dinucleotide (CT) deletion exhibited significantly shorter, curved filaments and semi-sterile pollen, accompanied by severely reduced spikelet fertility (Extended Data Fig. 5d,e), indicating that $orf1^{r}$ is a critical regulator of male gametogenesis.

Similarly, ectopic expression of $ORF1^{L}$ in the F$_1$ hybrids and RD23 failed to alter pollen fertility (Extended Data Fig. 6a,b). However, CRISPR-mediated knockout of $ORF1^{L}$ in the NIL-*S44* background resulted in underdeveloped stamens, reduced pollen fertility and diminished spikelet fertility (Extended Data Fig. 6c–f). In vitro germination assays further confirmed that pollen germination rates of both mutants were significantly lower than those of the controls (Extended Data Fig. 6g). These parallel phenotypes in $orf1^{r}$ and $orf1^{l}$ mutants indicated that their roles in pollen development are the same. Further analysis revealed that compared with fertile pollen, aborted pollen exhibited two distinct pollen subtypes: type I, spherical grains with wrinkled surfaces, and type II, irregularly shaped grains. Both types of pollen grains in both mutants were filled with cytoplasm and a few starch granules and were characterized by irregular thicker exines, unclear matrix structure and intermembrane space in mitochondria (Extended Data Figs. 7a and 8a). To assess female gamete development, we performed saturation pollination. Using both RD23 pollen and self-pollen on the $orf1^{r}$ and $orf1^{l}$ mutants resulted in significantly higher seed set rates than open pollination and bagged pollination (Extended Data Figs. 7b,c and 8b,c). In addition, mature embryo sacs observed via confocal microscopy revealed high fertility rates in $orf1^{r}$ (91.5%) and $orf1^{l}$ (93.7%) (Extended Data Figs. 7d and 8d). These results indicate that female gamete development is normal in both mutants. The low seed set is therefore primarily attributable to their inadequate fertile pollen and lower germination rate. Collectively, these findings demonstrate that $ORF1^{R}$ and $ORF1^{L}$ are essential for pollen development.

RIS and *RIA* fine-tune SD

To investigate whether $ORF1^{R}$ and $ORF1^{L}$ mutations affect hybrid sterility, we observed pollen fertility in the F$_1$ background. The results indicated comparable pollen fertility in the F$_1$-$orf1^{r}$ mutant and the F$_1$-$orf1^{r}$ $orf1^{l}$ double mutant, while the F$_1$-$orf1^{l}$ mutant exhibited a modest increase in pollen fertility relative to the F$_1$ hybrids (Fig. 2m). No impact of $ORF1^{R}$ and $ORF1^{L}$ mutations on spikelet fertility was observed in the F$_1$ hybrids, whereas spikelet fertility in F$_1$-$orf1^{r}$ $orf1^{l}$ plummeted to levels akin to those of homozygous $orf1^{r}$ and $orf1^{l}$ mutants (Fig. 2m), demonstrating that either $ORF1^{R}$ or $ORF1^{L}$ is sufficient and necessary for normal male gamete development in rice, and neither allele contributes to hybrid pollen sterility.

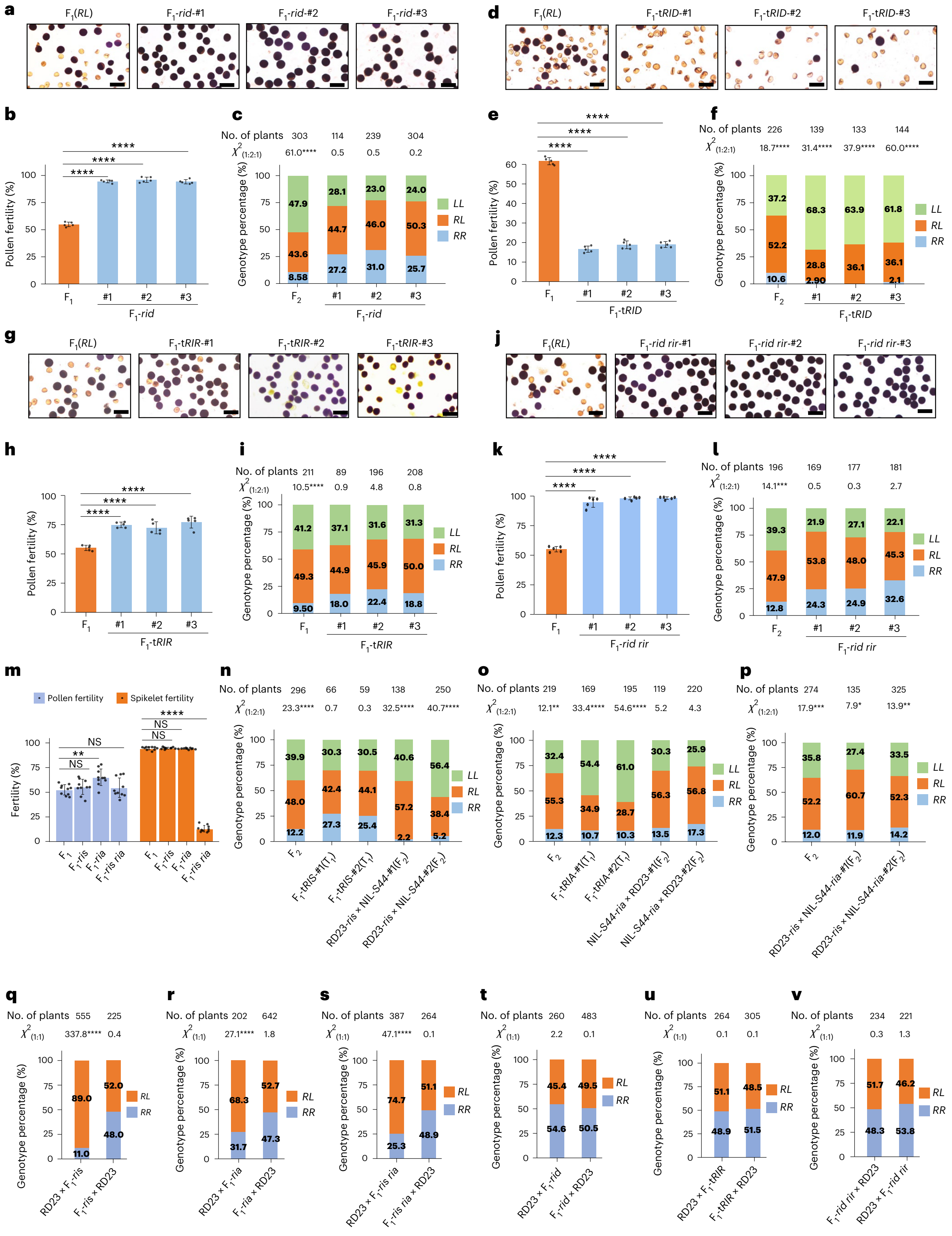
To investigate whether $ORF1^{R}$ and $ORF1^{L}$ are responsible for SD, we genotyped the transgenic and knockout progenies. The T$_1$ transgenic lines with the introduction of $ORF1^{R}$ into the F$_1$ hybrids restored normal Mendelian segregation ratios, whereas the loss of function of $ORF1^{R}$ aggravated SD in hybrid progeny (Fig. 2n). Conversely, the introduction of $ORF1^{L}$ into F$_1$ exacerbated SD compared with non-transgenic controls, while its knockout lines alleviated the degree of SD in T$_1$ progeny (Fig. 2o). Intriguingly, F$_2$ plants from $orf1^{r}$ and $orf1^{l}$ crosses still deviated from the Mendelian expectations (Fig. 2p), revealing a regulatory equilibrium where $ORF1^{R}$ suppresses and $ORF1^{L}$ activates SD. These results demonstrate that $ORF1^{R}$ and $ORF1^{L}$ fine-tune SD as a suppressor and an activator, respectively. We therefore designated $ORF1^{R}$ as *RIS* and $ORF1^{L}$ as *RIA*.

To examine the roles of male and female gametes in SD, we analysed gamete-specific transmission via reciprocal tests between corresponding F$_1$ mutants and RD23. When we used F$_1$-*ris*, F$_1$-*ria* and F$_1$-*ris ria* as the pollen donors, severe SD occurred, with *RIS* knockout showing the strongest effect (Fig. 2q–s). Female gamete transmission remained unaffected, pinpointing these genes as male-specific regulators (Fig. 2q–s). Reciprocal tests further confirmed that *RID* knockout, *RIR* introduction or dual *RID RIR* knockout in the F$_1$ hybrids resulted in an equal transmission ratio of male gametes, confirming their complementary roles in regulating transmission ratios to Mendelian expectations. Similarly, female gamete transmission was normal (Fig. 2t–v). These results further establish *RIS*, *RIA*, *RID* and *RIR* as male-gamete-specific regulators of SD, functioning as a suppressor, an activator, a distorter and a restorer, respectively.

RIA activates the distorter effect of *RID*

Previous investigations revealed that introducing the *RID* genomic fragment into RD23 failed to induce pollen sterility, whereas *RID* significantly reduced pollen fertility in both F$_1$ hybrids and NIL-*S44* (Fig. 2a–f

Fig. 2 | The RID–RIR–RIS/RIA element establishes a distorter–restorer system. a, Knocking out *RID* rescues pollen fertility in F$_1$ hybrids in the T$_0$ generation. Scale bars, 100 μm. **b**, Quantification of the pollen fertility of *rid* mutants in the T$_0$ generation. The data are presented as mean ± s.d. ($n = 5$ plants). **c**, Numbers and percentages of *RR*:*RL*:*LL* individuals from F$_1$-*rid* plants in the T$_1$ generation. **d**, Transformation of the exogenous *RID* gene (t*RID*) into F$_1$ hybrids results in high pollen sterility in the T$_0$ generation. Scale bars, 100 μm. **e**, Quantification of pollen fertility in transgenic F$_1$ plants as shown in **d**. The data are presented as mean ± s.d. ($n = 5$ plants). **f**, Segregation ratio of T$_1$ progeny from selfed T$_0$ plants carrying the t*RID* transgene in the F$_1$ background. **g**, Pollen fertility in F$_1$ plants with exogenous heterozygous *RIR* (t*RIR*/−) in the T$_0$ generation. Scale bars, 100 μm. **h**, Quantification of pollen fertility in F$_1$ hybrids with exogenous heterozygous *RIR* in the T$_0$ generation. The data are presented as mean ± s.d. ($n = 5$ plants). **i**, Complementation (t*RIR*) lines display the normal Mendelian ratio in T$_1$ progeny. **j**, I$_2$-KI-stained pollen of *rid rir* double mutants in the T$_0$ generation. Scale bars, 100 μm. **k**, Pollen fertility of three independent *rid rir* double mutants in the F$_1$ background. The data are presented as mean ± s.d. ($n = 5$ plants). **l**, Segregation ratio among progeny of F$_1$-*rid rir* mutants. **m**, Quantification of pollen and spikelet fertility in F$_1$ and knockout lines. The data are presented as mean ± s.d. ($n = 10$ plants). **n**, The impact of exogenous *RIS* (t*RIS*) in the T$_1$ generation or the knockout of *RIS* in F$_2$ populations on the ratio of *LL*:*RL*:*RR*. **o**, The percentages of the three genotypes with transgenic *RIA* (t*RIA*) or knocked out *RIA* in the F$_1$ background. **p**, Knocking out both *RIA* and *RIS* results in SD. **q**, Segregation ratios from reciprocal crosses between the F$_1$-*ris* mutant and RD23. **r**, The effect of the *RIA* mutation on gamete transmission assessed via reciprocal crosses between RD23 and the F$_1$-*ria* mutant. **s**, Segregation ratio of F$_1$-*ris ria* double mutants was detected via reciprocal crosses. **t**, *RID* mutation in the F$_1$ background produces equal male gametes transmission ratio. **u**, Introduction of the *RIR* gene into the F1 background restores the Mendelian segregation ratio of the progeny. **v**, The effect of both *RID* and *RIR* mutations on gamete transmission. In **b**, **e**, **h**, **k** and **m**, $^{*}P < 0.05$, $^{**}P < 0.01$ or $^{****}P < 0.0001$ according to one-way ANOVA and Tukey's HSD test. In **c**, **f**, **i**, **l** and **n**–**v**, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ or $^{****}P < 0.0001$ according to χ^2 tests. For all *P* values, see Source Data.


a
F1(RL)
F1-rid-#1
F1-rid-#2
F1-rid-#3
d
F1(RL)
F1-tRID-#1
F1-tRID-#2
F1-tRID-#3
b
Pollen fertility (%)
****
F1
#1
#2
#3
F1-rid
c
No. of plants 303 114 239 304
χ2 (1:2:1) 61.0**** 0.5 0.5 0.2
Genotype percentage (%)
47.9 28.1 23.0 24.0
43.6 44.7 46.0 50.3
8.58 27.2 31.0 25.7
LL
RL
RR
F2
F1-rid
e
Pollen fertility (%)
F1-tRID
f
No. of plants 226 139 133 144
χ2 (1:2:1) 18.7**** 31.4**** 37.9**** 60.0****
37.2 68.3 63.9 61.8
52.2 28.8 36.1 36.1
10.6 2.90 2.1
F1-tRID
g
F1(RL)
F1-tRIR-#1
F1-tRIR-#2
F1-tRIR-#3
j
F1(RL)
F1-rid rir-#1
F1-rid rir-#2
F1-rid rir-#3
h
F1-tRIR
i
No. of plants 211 89 196 208
χ2 (1:2:1) 10.5**** 0.9 4.8 0.8
41.2 37.1 31.6 31.3
49.3 44.9 45.9 50.0
9.50 18.0 22.4 18.8
F1-tRIR
k
F1-rid rir
l
No. of plants 196 169 177 181
χ2 (1:2:1) 14.1*** 0.5 0.3 2.7
39.3 21.9 27.1 22.1
47.9 53.8 48.0 45.3
12.8 24.3 24.9 32.6
F1-rid rir
m
Pollen fertility
Spikelet fertility
Fertility (%)
NS
**
****
F1
F1-ris
F1-ria
F1-ris ria
n
No. of plants 296 66 59 138 250
χ2 (1:2:1) 23.3**** 0.7 0.3 32.5**** 40.7****
39.9 30.3 30.5 40.6 56.4
48.0 42.4 44.1 57.2 38.4
12.2 27.3 25.4 2.2 5.2
F2
F1-tRIS-#1(T1)
F1-tRIS-#2(T1)
RD23-ris × NIL-S44-#1(F2)
RD23-ris × NIL-S44-#2(F2)
o
No. of plants 219 169 195 119 220
χ2 (1:2:1) 12.1** 33.4**** 54.6**** 5.2 4.3
32.4 54.4 61.0 30.3 25.9
55.3 34.9 28.7 56.3 56.8
12.3 10.7 10.3 13.5 17.3
F2
F1-tRIA-#1(T1)
F1-tRIA-#2(T1)
NIL-S44-ria × RD23-#1(F2)
NIL-S44-ria × RD23-#2(F2)
p
No. of plants 274 135 325
χ2 (1:2:1) 17.9*** 7.9* 13.9**
35.8 27.4 33.5
52.2 60.7 52.3
12.0 11.9 14.2
F2
RD23-ris × NIL-S44-ria-#1(F2)
RD23-ris × NIL-S44-ria-#2(F2)
q
No. of plants 555 225
χ2 (1:1) 337.8**** 0.4
89.0 52.0
11.0 48.0
RD23 × F1-ris
F1-ris × RD23
r
No. of plants 202 642
χ2 (1:1) 27.1**** 1.8
68.3 52.7
31.7 47.3
RD23 × F1-ria
F1-ria × RD23
s
No. of plants 387 264
χ2 (1:1) 47.1**** 0.1
74.7 51.1
25.3 48.9
RD23 × F1-ris ria
F1-ris ria × RD23
t
No. of plants 260 483
χ2 (1:1) 2.2 0.1
45.4 49.5
54.6 50.5
RD23 × F1-rid
F1-rid × RD23
u
No. of plants 264 305
χ2 (1:1) 0.1 0.1
51.1 48.5
48.9 51.5
RD23 × F1-tRIR
F1-tRIR × RD23
v
No. of plants 234 221
χ2 (1:1) 0.3 1.3
51.7 46.2
48.3 53.8
F1-rid rir × RD23
RD23 × F1-rid rir

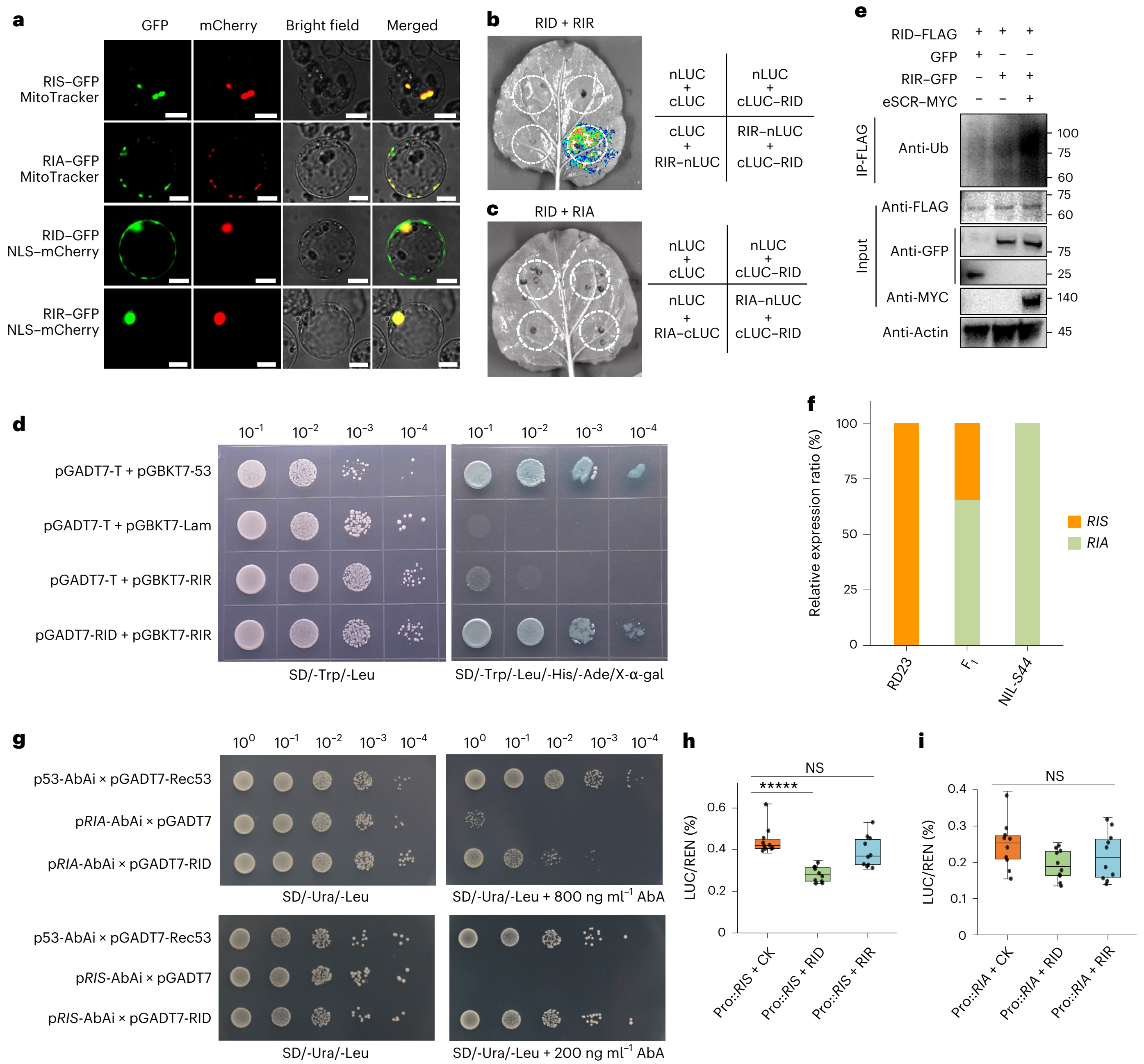


Fig. 3 | Interactions of RID with RIR and of RID with the *RIS* promoter. **a**, Subcellular localization of RIS/RIA, RID and RIR. RIS and RIA are localized to the mitochondria, while RID is localized to both the nucleus and the membrane. RIR is exclusively localized to the nucleus. $n = 3$ independent biological replicates. Scale bars, 10 μm. **b**, Split-luciferase complementation assay revealed the interaction of RID and RIR in *Nicotiana benthamiana* leaves. **c**, Split-luciferase complementation assay showed no interaction between RID and RIA in *N. benthamiana* leaves. In **b** and **c**, $n = 3$ independent biological replicates; nLUC and cLUC represent the amino and carboxy termini of luciferase, respectively. **d**, Yeast two-hybrid analysis indicated that RID interacts with RIR. SD/-Leu/-Trp denotes synthetic dropout medium; SD/-Trp/-Leu/-His/-Ade/X-α-gal is the selective medium. $n = 3$ independent biological replicates. **e**, RIR ubiquitinates RID in vivo. The polyubiquitinated products were detected with anti-ubiquitin antibody. $n = 2$ independent biological replicates. **f**, Relative expression ratios of *RIS* and *RIA* in RD23, NIL-*S44* and their F_1 hybrids. **g**, Yeast one-hybrid showed that RID could bind to the promoters of *RIS* and *RIA*. SD/-Ura/-Leu is the synthetic dropout medium, and SD/-Ura/-Leu with 200 ng ml^{-1} AbA (Aureobasidin A) or 800 ng ml^{-1} AbA is the selective medium. $n = 3$ independent biological replicates. **h**, Dual-luciferase reporter assay showed that RID can inhibit the expression of *RIS*, whereas RIR cannot. REN, *Renilla* luciferase. CK, empty vector. **i**, Dual-luciferase reporter assay showed that RID and RIR have no impact on the expression of *RIA*. CK, empty vector. In **h** and **i**, the data are presented as mean ± s.d. ($n = 10$ independent biological replicates). From bottom to top, the horizontal lines in the box plots represent the minima, 25th percentiles, medians, 75th percentiles and maxima. $^{****}P < 0.0001$ according to one-way ANOVA and Tukey's HSD test.

and Extended Data Fig. 3d,e). This suggests that *RID* requires cofactors to induce gamete lethality. Given that *RIR* serves as a restorer counteracting *RID*-mediated distortion, we hypothesized that *RIA* cooperates with *RID* to form a pollen-lethal module. To test this hypothesis, we crossed NIL-*S44* carrying the exogenous *RID* as the female parent with RD23 harbouring the transformed *RIA* as the male parent. F_1 hybrids inheriting both *RID* and *RIA* exhibited significantly lower pollen fertility than controls lacking both transgenes or containing *RID* alone (Supplementary Table 1). Notably, when F_1 plants harbouring *RID* were crossed with *RIA*-expressing RD23, theoretically, the *R*/*tRIA*-type male gamete could be produced, and the following female gamete genotypes could be generated: *R*/*tRID*, *R*/−, *L*/*tRID* and *L*/−. Since *RIR* is absent in RD23, zygotes with the *RR-tRID*/*tRIA* genotype cannot be detected in the hybrid population (Supplementary Table 1). These findings indicate that *RID* functions as a core distorter whose gametocidal activity requires *RIA*-mediated activation, forming an interdependent molecular module that targets the RD23-derived gametes.

RIR ubiquinates RID, and RID inhibits the expression of *RIS*

To elucidate the molecular mechanism of *S44*-mediated hybrid sterility, we first analysed the spatiotemporal expression dynamics of the target genes. RNA in situ hybridization revealed similar expression patterns: *RIS* is highly expressed in tapetal cells and developing microspores in RD23 anthers, whereas *RIA*, *RID* and *RIR* are highly expressed in these cell types in NIL-*S44* and F_1 hybrids (Extended Data Fig. 9a–c), suggesting their conserved roles in pollen development. Subcellular localization assays demonstrated a compartmentalized distribution: RIS and RIA localized to mitochondria, RID partitioned in the nucleus and plasma membrane, and RIR exclusively accumulated in the nucleus (Fig. 3a). This spatiotemporal distribution implies their distinct functions in hybrid sterility and SD.

Given the antagonistic relationship between RID and RIR, coupled with RIA's role as an activator of the distorter, we investigated their physical interactions. Split-luciferase complementation assays revealed a physical interaction between RIR and RID, establishing a molecular interface for their function (Fig. 3b). No interaction was detected between RID and RIA, suggesting that RIA activates RID activity through indirect mechanisms (Fig. 3c). Yeast two-hybrid assays further confirmed the RID–RIR interaction, reinforcing their roles as core antagonistic elements (Fig. 3d). Additionally, an in vivo ubiquitination assay in tobacco demonstrated that RIR, acting as an F-box protein, can assemble into a functional SCF E3 ubiquitin ligase complex that ubiquitinates RID, suggesting that RID undergoes protein degradation via the ubiquitin-proteasome pathway (Fig. 3e).

A remarkable sequence divergence in the RIS/RIA promoter regions (57.14% identity) correlated with their differential expression, with *RIA* expression significantly higher than that of *RIS* in the F_1 hybrids (Fig. 3f). Bioinformatic analysis predicted that RID harbours DNA-binding domains of transcription factors (https://smart.embl.de/). We thus hypothesized that RID might bind the promoters of *RIS* and *RIA*, thereby regulating their expression. To test this, we examined the interactions between the *RIS*/*RIA* promoters and the RID protein. Yeast one-hybrid assays demonstrated that RID directly binds both the *RIS* and *RIA* promoters, with preferential affinity for the *RIS* promoter (Fig. 3g). Functional validation via dual-luciferase reporters revealed that RID selectively represses *RIS* transcription but does not affect *RIA* expression (Fig. 3h,i). This evidence suggests that RID could inhibit the expression level of *RIS* by binding to its promoter.

A subcellular localization assay revealed that both RIS and RIA are localized in the mitochondria, leading us to speculate that *S44* may induce mitochondrial dysfunction. NBT staining showed a temporal shift in superoxide anion accumulation. RD23, NIL-*S44* and the F_1 hybrid first exhibited detectable superoxide anion at the S8 stage (Extended Data Fig. 10a–c), whereas reactive oxygen species (ROS) burst occurred earlier in NIL-*S44*-t*RID* (S7 stage) (Extended Data Fig. 10d) and was markedly premature in RD23-*ris* and NIL-*S44*-*ria* (S6 stage) (Extended Data Fig. 10e,f), suggesting disruption of the mitochondrial electron transport chain.

Analysis of the cellular NAD^+/NADH ratio in leaves demonstrated that compared with the parental lines, the F_1 hybrid and *ris*/*ria* mutants had reduced ratios, indicating impaired mitochondrial function (Extended Data Fig. 10g,i,j). In contrast, introducing *RID* significantly increased the NAD^+/NADH ratio, reflecting the presence of oxidative stress (Extended Data Fig. 10h).

RD23 exhibits higher ATP levels than NIL-*S44* and the F_1 hybrid (Extended Data Fig. 10k). The introduction of *RID* into NIL-*S44* did not result in a significant difference in ATP content compared to the control (Extended Data Fig. 10l), whereas mutations in *RIA* or *RIS* led to a marked decrease in ATP (Extended Data Fig. 10m,n). Collectively, these findings indicate that the loss of *RIS*/*RIA* function impairs the mitochondrial electron transport chain and disrupts redox homeostasis, ultimately resulting in an inadequate energy supply. Furthermore, the introduction of *RID* exacerbates oxidative reactions mediated by accumulated ROS, underscoring the redox imbalance caused by mitochondrial damage.

Evolutionary trajectory of the *S44* locus

To elucidate the evolutionary dynamics of the *S44* locus, we systematically investigated the gene architecture of its four elements (*RIS*/*RIA*, *RID* and *RIR*) across 416 accessions representing AA-genome species through PCR amplification and pan-genome sequencing[34] (Supplementary Table 2). While *RIS*/*RIA* exhibited universal conservation across all accessions (100% detection rate), *RID* and *RIR* displayed lineage-specific divergence, with *RID*–*RIR* pairs identified in 50.5% (210/416) of accessions (Supplementary Tables 2 and 3). Additionally, *RIR* probably originated earlier than *RID*, as evidenced by its presence in the outgroups *Leersia perrieri*, *O. granulata* and *O. punctata* (Fig. 4a). This evolutionary precedence aligns with *RIR*'s role as a restorer gene.

The 15 identified haplotypes of RIS/RIA were divided into four evolutionary classes according to the amino acid variations in the first exon: ancestral Q1-type, derived Q3-type, structural variant Q4-type and recombinant R2Q-type. The derived Q3-type was fixed in Asian cultivated rice, probably correlated with SD (Fig. 4a and Supplementary Table 3). The 34 haplotypes of RID resolved into Omer-type (an Australia-specific pseudogene via a 28-bp deletion resulting in a truncated protein), del-type (a 2-bp deletion in the coding region leading to premature stop codons), mix-type and full-type. African wild species, including *O. longistaminata*, *O. barthii* and *O. glaberrima*, retained full-type RID, showing significant contrast to del-type dominance in Asian *japonica*, which indicates the different selection pressures during domestication (Fig. 4a and Supplementary Table 4). The 25 haplotypes of RIR are categorized into T-type, M-type and shift-type according to whether the 446th amino acid is Met or Thr and whether the RIR protein exhibits frameshift (Fig. 4a and Supplementary Table 5). Full-type RID exhibited absolute linkage with T-type RIR, mirroring co-evolution of the two proteins (Fig. 4a).

These haplotypes of RIS/RIA–RID–RIR formed a total of 67 haplotype combinations across all samples (Supplementary Table 2), which formed four phylogenetically distinct groups (Fig. 4b and Supplementary Table 6). Group I (RID/RIR-deficient) dominated Asian and American taxa, including *O. rufipogon* (20/35 accessions), *indica* (157/159 accessions) and *O. glumaepatula* (7/9 accessions), facilitating the shaping of interspecific gene flow. Group II (Omer-type RID) was exclusive to Australian *O. meridionalis* (10/10 accessions), characterized by RID pseudogenization and RIR diversification. Group III (del-type RID) was prevalent in Asian wild (*O. nivara*, 8/9) and cultivated rice (*japonica*, 49/66). Group IV (functional RID/RIR) consisted of African wild species (*O. longistaminata* (10/10), *O. barthii* (28/29) and *O. glaberrima* (86/86)), which formed a monophyletic group (100% full-type or mix-type RID/T-type RIR) (Fig. 4c). Taken together, the haplotypes of *S44* exhibit significant geographic structuring and species-specific differentiation. This pattern is consistent with a role in reproductive isolation and may have contributed to the divergence of AA-genome rice species.

A widely compatible haplotype confers hybrid pollen fertility

To investigate the genetic mechanism underlying hybrid fertility mediated by the *S44* locus, we identified a critical 2-bp deletion in

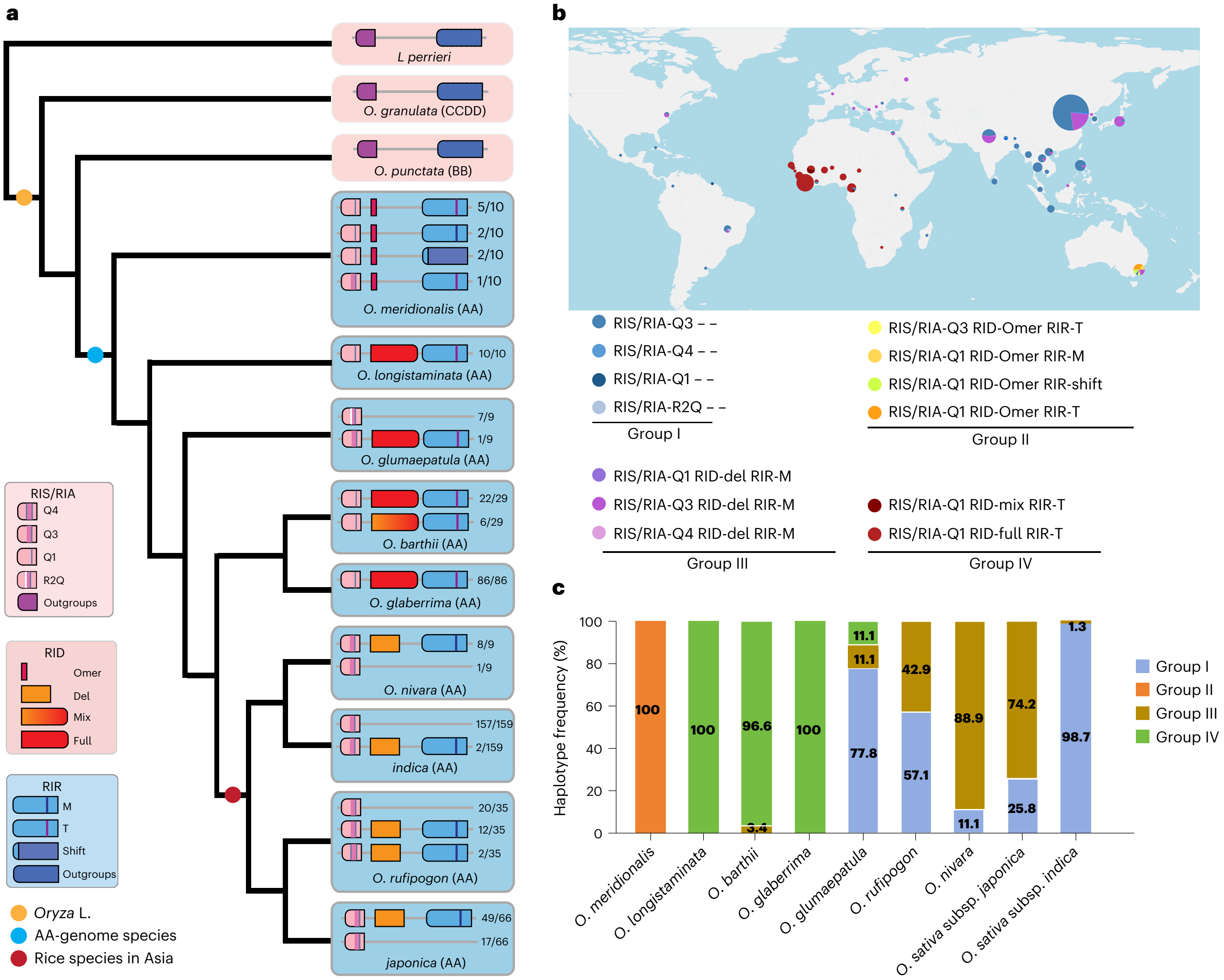


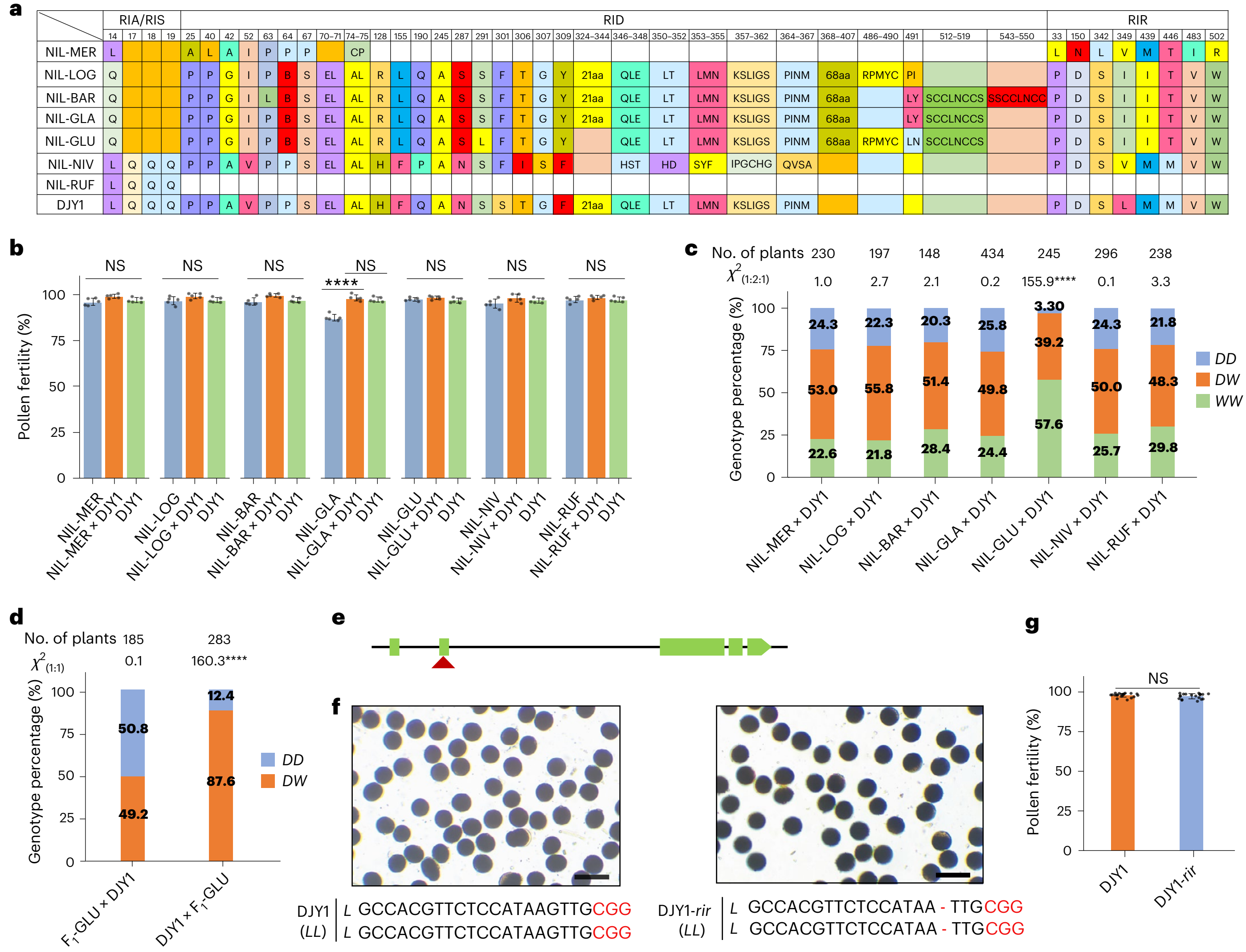
Fig. 4 | Model of the evolutionary history and distribution of the *S44* locus. **a**, The *S44* locus emerged after the divergence of AA-genome species, followed by parallel evolution in independent Asian and African lineages within the *Oryza* genus. The number of haplotypes and the total number of accessions analysed are shown on the right. *L. perrieri*, *O. granulata* and *O. punctata* were used as the outgroups. **b**, Geographic distribution of the accessions used in this study. Each circle represents the respective frequencies of these haplotypes at each locality. All haplotypes are categorized into four groups on the basis of the *S44* sequence. Group I is characterized by the absence of RID and RIR elements. Group II contains the Omer-type RID and diverse RIR types. Group III exhibits the del-type RID and the M-type RIR. Group IV comprises the functional RID and RIR. **c**, The distribution frequencies of *S44* haplotypes among AA-genome species.

RID that generates a premature stop codon (producing a truncated 348-amino-acid protein) in *O. rufipogon*, *O. nivara* and Asian cultivated rice. To functionally validate the role of truncated *RID*, we engineered a series of BC_6F_2 NILs in the DJY1 background (a *japonica* variety carrying Q3-type RIS, del-type RID and M-type RIR haplotypes) by introgressing divergent haplotypes from *O. meridionalis*, *O. longistaminata*, *O. barthii*, *O. glaberrima*, *O. glumaepatula*, *O. rufipogon* and *O. nivara*, designated as NIL-MER, NIL-LOG, NIL-BAR, NIL-GLA, NIL-GLU, NIL-RUF and NIL-NIV, respectively (Fig. 5a). All F_1 hybrids derived from crosses between DJY1 and NILs exhibited normal pollen fertility (Fig. 5b), with SD exclusively observed in the DJY1/NIL-GLU cross (Fig. 5c). Reciprocal test analyses revealed preferential transmission of the *O. glumaepatula* male allele (Fig. 5d), suggesting that NIL-GLU confers hybrid SD in the *S44* region through other unknown mechanisms independent of the *S44* locus. Collectively, the defective RID variant (del-type) in DJY1, which fails to initiate gamete elimination, contributes to its compatibility.

If the truncated *RID* in DJY1 is a non-functional distorter, pollen development should proceed normally in the absence of the restorer gene *RIR*. To validate this, we knocked out *RIR* in the DJY1 background using CRISPR–Cas9 editing technology (Fig. 5e). The homozygous *rir* mutant maintained full pollen fertility without affecting vegetative growth or floral development (Fig. 5f,g). These results confirm that the truncated RID in DJY1 loses distorter activity, thereby effectively neutralizing its gamete elimination capacity. Crucially, the preserved *RIR* function in this haplotype provides restoration activity, establishing the haplotype as a naturally occurring wide-compatibility system.

Allelic conflicts at the *S44* locus drive quantitative reproductive isolation

The *S44* locus functions as a speciation hotspot governing male hybrid sterility between African perennial *O. longistaminata* and Asian *indica* cultivars. Phylogenomic analyses reveal striking haplotype fixation patterns: African accessions predominantly carry Hap58/59, while Asian



	RIA/RIS				RID																														RIR							
	14	17	18	19	25	40	42	52	63	64	67	70–71	74–75	128	155	190	245	287	291	301	306	307	309	324–344	346–348	350–352	353–355	357–362	364–367	368–407	486–490	491	512–519	543–550	33	150	342	349	439	446	483	502
NIL-MER	L				A	L	A	I	P	P	P		CP																						L	N	L	V	M	T	I	R
NIL-LOG	Q				P	P	G	I	P	B	S	EL	AL	R	L	Q	A	S	S	F	T	G	Y	21aa	QLE	LT	LMN	KSLIGS	PINM	68aa	RPMYC	PI			P	D	S	I	I	T	V	W
NIL-BAR	Q				P	P	G	I	L	B	S	EL	AL	R	L	Q	A	S	S	F	T	G	Y	21aa	QLE	LT	LMN	KSLIGS	PINM	68aa		LY	SCCLNCCS	SSCCLNCC	P	D	S	I	I	T	V	W
NIL-GLA	Q				P	P	G	I	P	B	S	EL	AL	R	L	Q	A	S	S	F	T	G	Y	21aa	QLE	LT	LMN	KSLIGS	PINM	68aa		LY	SCCLNCCS		P	D	S	I	I	T	V	W
NIL-GLU	Q				P	P	G	I	P	B	S	EL	AL	R	L	Q	A	S	L	F	T	G	Y		QLE	LT	LMN	KSLIGS	PINM	68aa	RPMYC	LN	SCCLNCCS		P	D	S	I	I	T	V	W
NIL-NIV	L	Q	Q	Q	P	P	A	V	P	P	S	EL	AL	H	F	P	A	N	S	F	I	S	F		HST	HD	SYF	IPGCHG	QVSA						P	D	S	V	M	M	V	W
NIL-RUF	L	Q	Q	Q																																						
DJY1	L	Q	Q	Q	P	P	A	V	P	P	S	EL	AL	H	F	Q	A	N	S	S	T	G	F	21aa	QLE	LT	LMN	KSLIGS	PINM						P	D	S	L	M	M	V	W

DJY1 (LL) | L GCCACGTTCTCCATAAGTTGCGG
L GCCACGTTCTCCATAAGTTGCGG

DJY1-*rir* (LL) | L GCCACGTTCTCCATAA - TTGCGG
L GCCACGTTCTCCATAA - TTGCGG

Fig. 5 | Identification of a widely compatible haplotype. a, The *S44* haplotypes of NILs derived from the crosses between DJY1 and AA-genome species in rice. NIL-MER, NIL-LOG, NIL-BAR, NIL-GLA, NIL-GLU, NIL-NIV, NIL-RUF and DJY1 denote the introgression fragments originating from *O. meridionalis*, *O. longistaminata*, *O. barthii*, *O. glaberrima*, *O. nivara*, *O. rufipogon* and the *japonica* variety DJY1 at the *S44* locus, respectively. **b**, Pollen fertility in crosses between DJY1 and DJY1-NILs carrying the *S44* locus from different AA species. The data are presented as mean ± s.d. ($n = 5$ plants). **c**, Segregation ratios observed in progeny populations from crosses between DJY1 and other NILs. **d**, Gamete transmission was tested in reciprocal crosses between F$_1$-GLU and DJY1. In **c** and **d**, *DD*, *WW* and *DW* represent the homozygous DJY1 genotype, the homozygous NIL genotype and the heterozygous genotype, respectively. **e**, Target site in *RIR* for CRISPR–Cas9-mediated mutagenesis in the DJY1 background. **f**, The DJY1-*rir* mutant exhibited normal pollen fertility. Scale bars, 100 µm. **g**, Quantification of pollen fertility in DJY1 and the DJY1-*rir* mutant. The data are presented as mean ± s.d. ($n = 20$ plants). In **b** and **g**, $^{****}P < 0.0001$ according to one-way ANOVA and Tukey's HSD test. In **c** and **d**, $^{****}P < 0.0001$ according to χ^2 tests. For all *P* values, see Source Data.

lineages (including *O. rufipogon*, *O. nivara* and cultivated rice) retain Hap6/11, mirroring their continental divergence and reproductive trajectories (Fig. 4a–c and Supplementary Table 2). To investigate the role of the *S44* locus in reproductive isolation between Asian and African rice species, we first detected the expression level of the *S44* locus. The results revealed that *RIA*/*RIS* transcripts were present in all NILs. *RID* was expressed in DJY1, NIL-LOG and NIL-NIV, but not in NIL-BAR, NIL-GLA or NIL-RUF. *RIR* transcripts were detected in all NILs except for NIL-RUF (Fig. 6a–c). To dissect how *S44* allelic conflicts quantitatively modulate reproductive barriers, we developed three indexes: the Hybrid Fertility Index (HFI) quantifies the ratio of reproductive fitness between hybrid progeny and parental species, reflecting hybrid fertility barriers (range, 0 (complete hybrid sterility) to 1 (equivalent fertility to parents)); the Gametic Selection Index (GSI) assesses the deviation of gamete genotype frequencies from Mendelian inheritance ratios, revealing selection pressure at the gamete level (range, 0 (perfect Mendelian segregation) to 1 (complete deviation from Mendelian segregation)); and the Composite Reproductive Isolation Index (CRII) integrates fertility barriers (HFI) and gametic selection (GSI) to comprehensively evaluate the strength of reproductive isolation between species (range, 0 (no reproductive isolation) to 1 (complete reproductive isolation)).

Among the tested hybrid combinations, NIL-LOG/NIL-GLA, NIL-LOG/NIL-RUF, NIL-LOG/DJY1, NIL-LOG/NIL-NIV and NIL-LOG/NIL-BAR, the HFI values increased progressively. Specifically, for NIL-LOG/DJY1, NIL-LOG/NIL-NIV and NIL-LOG/NIL-BAR, their HFIs approached or equalled 1, suggesting that there are no significant hybrid fertility barriers between NIL-LOG and these lines (Fig. 6d). In contrast, the HFI values for the NIL-LOG/NIL-GLA and NIL-LOG/NIL-RUF combinations were 0.53 and 0.71, respectively (Fig. 6d). This indicates that the hybrid fertility barrier is stronger in the NIL-LOG/NIL-GLA combination than in the NIL-LOG/NIL-RUF hybrids.

Progeny from NIL-LOG crossed with DJY1, NIL-NIV, NIL-GLA, NIL-RUF and NIL-BAR exhibited increasingly severe SD (Fig. 6e). The NIL-LOG/DJY1 hybrids showed the lowest GSI, indicating negligible

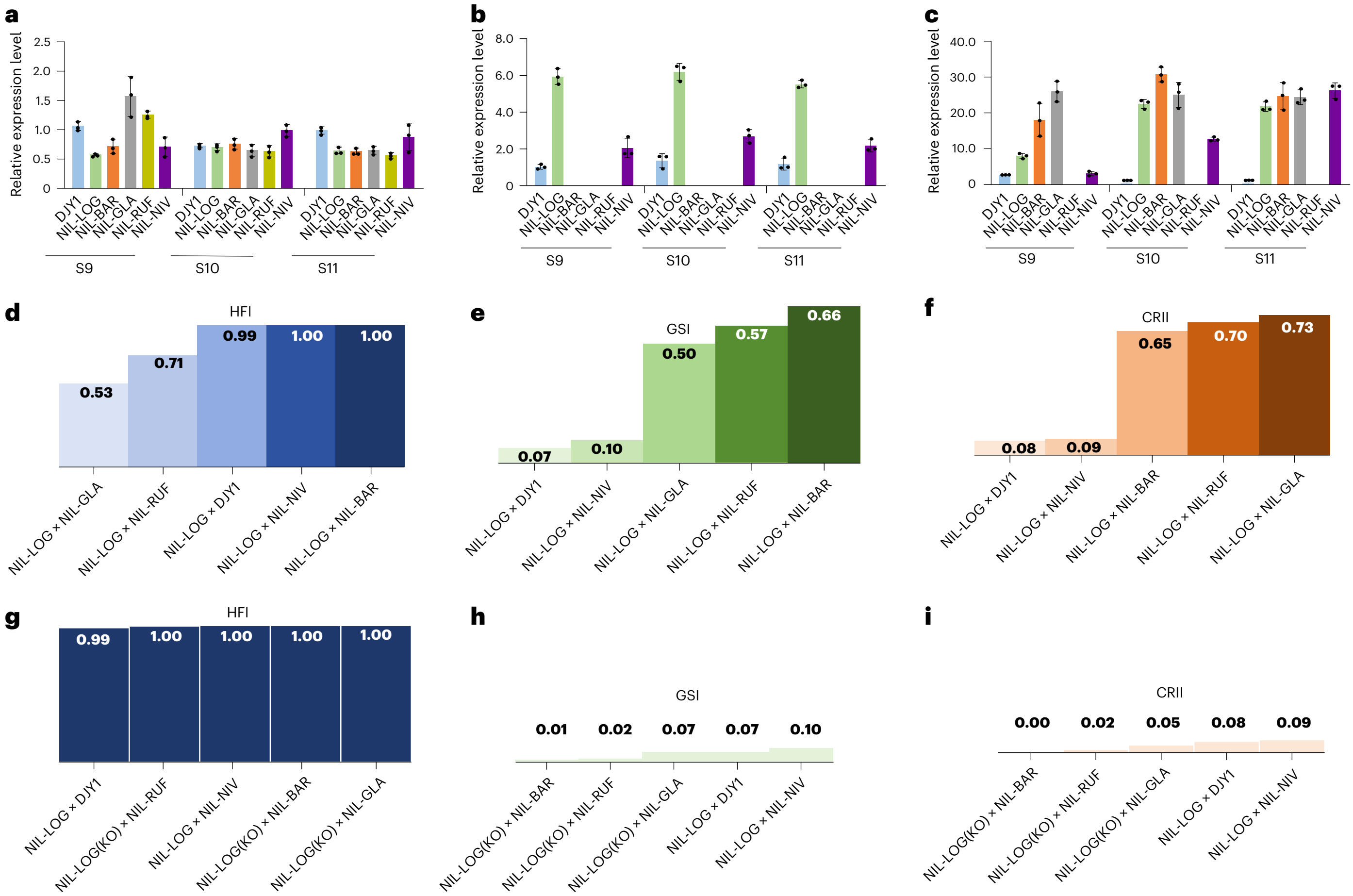


Fig. 6 | The *S44* locus contributes to reproductive isolation between *O. longistaminata* and other lineages. a–**c**, Relative expression levels of *RIS*/*RIA* (**a**), *RID* (**b**) and *RIR* (**c**) in the NILs across stages S9 to S11. The data are presented as mean ± s.d. (*n* = 3 independent biological replicates). **d**, HFI of hybrids between NIL-LOG and other NILs. The data are presented as mean ± s.d. (*n* = 5 plants). **e**, GSI of hybrids between NIL-LOG and other NILs. **f**, CRII of hybrids between NIL-LOG and other NILs. **g**, Knocking out *RID* in NIL-LOG could recover the HFI. **h**, Knocking out *RID* in NIL-LOG could restore the GSI. **i**, Knocking out *RID* in NIL-LOG could recover the CRII.

SD, whereas the NIL-LOG/NIL-BAR cross displayed the highest GSI, reflecting significant SD (Fig. 6e).

To quantify reproductive isolation, we integrated the HFI and GSI into the CRII. The CRII values for NIL-LOG/DJY1 and NIL-LOG/NIL-NIV were 0.08 and 0.09, respectively, both close to 0, indicating no reproductive isolation between NIL-LOG and DJY1 or NIL-NIV (Fig. 6f). In contrast, the CRII values for NIL-LOG/NIL-BAR, NIL-LOG/NIL-RUF and NIL-LOG/NIL-GLA were 0.65, 0.70 and 0.73, respectively, indicating significant reproductive isolation between these crosses (Fig. 6f). These results establish that *S44*-mediated reproductive isolation follows a hierarchical pattern: strongest in the NIL-LOG/NIL-GLA cross, followed by moderate isolation in NIL-LOG/NIL-RUF and weak in NIL-LOG/NIL-BAR, with undetectable barriers in NIL-LOG/DJY1 or NIL-LOG/NIL-NIV hybrids. These results implicate *S44*-mediated allelic conflicts as the primary driver of reproductive barriers.

To validate whether the distorter gene *RID* from *O. longistaminata* drives the reproductive isolation in these combinations, we crossed the *RID* knockout line with NIL-GLA, NIL-RUF and NIL-BAR. All HFI values approached or equalled 1, indicating complete restoration of fertility comparable to control crosses (Fig. 6g). Furthermore, the GSI values were lower than 0.1, suggesting no SD in hybrid progeny (Fig. 6h). Consistently, the CRII values of all the crosses were close to zero, indicating no reproductive isolation in the hybrids (Fig. 6i). These findings demonstrate that knockout of the distorter gene *RID* effectively neutralizes *S44*-driven reproductive barriers. On the basis of these results, we propose a haplotype-driven selection model: (1) the *O. longistaminata* haplotype (Hap58) maintains strong *S44*-mediated hybrid sterility and SD as a gene flow barrier; (2) the *O. barthii* haplotype (Hap59) selectively confers SD without a fertility barrier, enabling targeted allele propagation; and (3) the Asian haplotype (Hap6) acquires *RID* non-functionality coupled with RIR restoration capacity, achieving wide interspecific compatibility (Fig. 7a–c). Our results revealed a quadripartite distorter–restorer system consisting of RIA/RIS, RID and RIR. The interactions within this system are as follows: RIA activates RID, while RIR targets RID for degradation through ubiquitination. Meanwhile, RID inhibits the expression of RIS, and RIA/RIS induces mitochondrial dysfunction by regulating ROS burst (Fig. 7d). By strategically selecting parents with target haplotypes of the *S44* locus, this framework enables precision engineering of wide-compatibility lines to overcome speciation barriers—a paradigm shift in designing interspecific hybrid breeding programmes.

Discussion

In this study, we dissected the *S44* locus as a major regulator of reproductive isolation between *O. longistaminata* and *O. sativa* cultivar RD23, uncovering an unprecedented four-element distorter–restorer system (*RIS*/*RIA*–*RID*–*RIR*) that quantitatively controls both pollen sterility and SD. Unlike the classic two/three-component hybrid incompatibility systems in plants[7,11,14,16–19,23], this homeostatic distorter–restorer system operates through synergistic interactions:

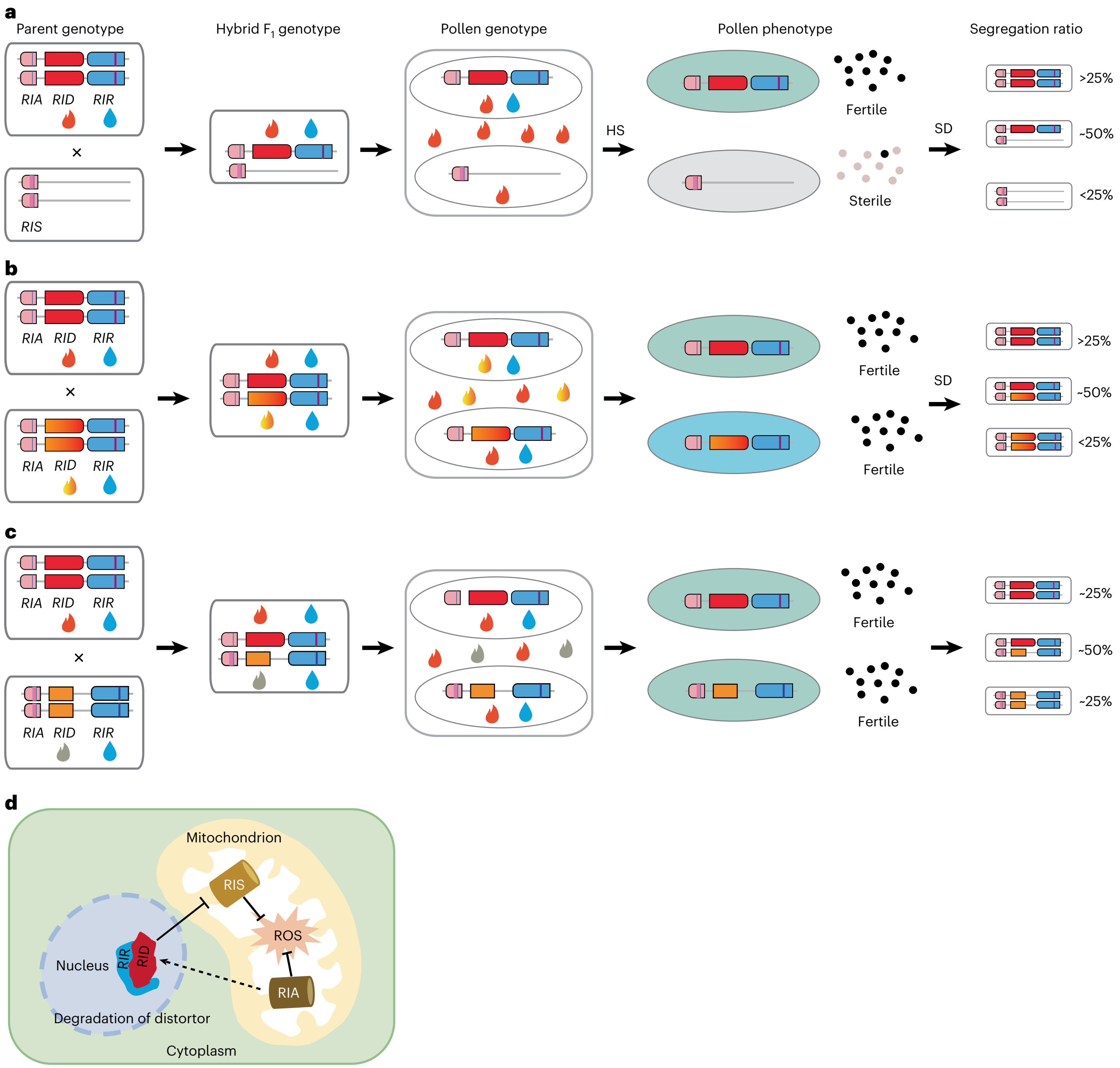


Fig. 7 | Hybrid pollen sterility and SD in combinations between two *S44* alleles as a guide in breeding programmes. **a**, The interaction of haplotypes with functional *RID* and *RIR* (Hap58) and haplotypes lacking *RID* and *RIR* confers hybrid sterility and SD. **b**, Hybrids derived from crosses between varieties with functional *RID* and *RIR* (Hap58) and varieties with mutated *RID* and functional *RIR* (Hap59) show SD in their progeny. **c**, Varieties with non-functional *RID* and functional *RIR* (Hap6) are widely compatible at the *S44* locus. **d**, A working model for the distorter–restorer system. RIA activates the distorter effect of RID, and RID physically interacts with RIR and inhibits the expression of RIS. RID is degraded by RIR. HS, hybrid sterility. The red, orange and grey flame icons indicate the strong, weak and non-functional distorter. The blue water icon denotes the restorer.

RID (distorter) requires *RIA* (activator) to induce gamete elimination, while *RIR* (restorer) provides gametophytic protection (Fig. 2 and Supplementary Table 1). This genetic behaviour of the *S44* locus is consistent with the "one-locus sporogametophytic interaction model"[35–37], which has been reported for most hybrid sterility loci, such as *S1* (ref. 23), *S5* (ref. 7), *Sa*[11], *Sc*[20], *qHMS1* (ref. 18), *qHMS7* (ref. 14) and *pf12*/*RHS12*/*Se*[16,17,19]. Cytological experiments, such as the TUNEL assay, will help us understand the molecular mechanism underlying pollen abortion mediated by sporophyte–gametophyte interaction. The *RIS* and *RIA* paralogues in rice regulate distortion strength through competitive inhibition (Fig. 2). Pollen sterility was observed in the F$_1$ hybrids and the F$_1$-*ris ria* double mutant. *S44* regulates gamete elimination in the F$_1$ hybrids due to allelic conflict, whereas pollen abortion in F$_1$-*ris ria* may result from developmental defects during gametogenesis. Cytological observations reveal that pollen abortion in F$_1$ hybrids is mainly characterized by spherical abortion or typical abortion (Fig. 1b), whereas the *ria ris* mutants primarily display the staining abortion phenotype. These findings suggest that the underlying cytological and genetic mechanisms responsible for pollen sterility differ between F$_1$ hybrids and *ria ris* mutants.

Phylogenomic dating shows *RIA*/*RIS* diverged before 2.1 million years ago, preceding *RID*/*RIR* emergence, suggesting the stepwise assembly of this reproductive barrier (Fig. 4). This study has identified the *RIS* and *RIA* genes as modified factors that can fine-tune SD in rice, like the function of *Tiam2s* and *Tiam2l* in *t*-haplotype transmission in mouse[38]. The *S44* system is a good model for elucidating the genetic complexity of reproductive isolation in rice. Interestingly, we observed that transgenic plants expressing *RID* in the NIL-*S44* genetic background exhibit semi-sterility (Extended Data Fig. 3h). This phenotype could arise for two reasons. One is that *RID* indiscriminately kills pollen in a sporophytic manner. Another is that *RID* and *RIR* regulate hybrid sterility and SD in a dosage-dependent manner. Consequently, the introduction of an exogenous *RID* copy disrupts the *RID*–*RIR* dosage balance, leading to non-selective killing of self-pollen by *RID*. This is in accordance with the finding that a single copy of *JIEYAO* can counteract a single copy of *DUYAO*[16].

The SD system controlled by the *S44* locus shares conceptual parallels with the SD system in *D. melanogaster*, which includes a multi-locus genetic system, comprising the segregation distorter (*Sd*), responder (*Rsp*), enhancer (*E*(*SD*)), suppressor (*Su*(*SD*)) and stabilizer (*St*(*SD*))[31,39]. However, there are three key differences between these two systems. First, *E*(*SD*) and *Su*(*SD*) in *Drosophila* operate distortion independently of *Sd*. In contrast, *RIA* requires physical linkage and co-expression with *RID* for functionality. Second, modifiers of the SD system in *D. melanogaster* could be either linked or unlinked with *Sd*; that is, *E*(*SD*) evolved in *cis* with the distorter, affecting distortion strength in a dose-dependent manner, whereas *Su*(*SD*) can evolve in *trans* anywhere in the genome. In rice, the *S44* system evolved a stricter *cis*-regulatory system, favouring tight genetic linkage (Fig. 1e). Finally, the SD system in *Drosophila* conforms to the killer–target model. By contrast, the SD system controlled by the *S44* locus fits the distorter–restorer model. These divergences highlight distinct evolutionary trajectories for SD systems in plants and animals.

SD controlled by the mouse *t*-haplotype, a ~10–15-Mb recombination-suppressed region on chromosome 17, is a well-characterized model of non-Mendelian inheritance. The *t*-haplotype contains multiple genes with synergistic interactions. Distorter genes (*Tcd1*–*Tcd4*), a responder gene (*Tcr*) and modifier genes (*Tiam2l* and *Tiam2s*) govern the transmission ratio of gametes by regulating sperm motility[38,40,41]. *S44* in rice and the *t*-haplotype in mouse both confer SD by impairing one type of gamete. However, the former relies on single-locus interactions, and three tightly linked genes maintain the integrity of the locus (Fig. 1e), whereas the latter employs multigenic collaboration and large-scale genomic regulation to achieve complex drive mechanisms, and the integrity of the loci are maintained through the suppression of chromosomal recombination[38,40,41]. The SD system is of significant evolutionary importance. The SD system in *D. melanogaster* is susceptible to being weakened by unlinked suppressors *Su*(*SD*)[31,39]. The transmission advantage of the mouse *t*-haplotype comes at the cost of homozygous embryo lethality[38,40,41]. *S44* is a simple and efficient system for gamete SD without homozygous fertility penalties or suppressor vulnerability. Therefore, *S44* can stably maintain genetic diversity in rice and provides a unique model for understanding species evolution. These studies can provide insights into the evolutionary conservation and divergence of mechanisms that drive non-Mendelian inheritance across different species.

Hybrid sterility is a complex biological phenomenon controlled by multiple loci[5,6]. The regulatory mode and signalling pathways associated with each locus are unique, contributing to the diversity and complexity of hybrid sterility mechanisms[42]. Domains of unknown function (DUFs) are generalized protein domains characterized by two distinct features: relatively conserved amino acid sequences and uncharacterized functions. They play key roles in plant growth, development and responses to biotic and abiotic stress[43,44]. In recent years, significant progress has been made in characterizing hybrid sterility loci, drawing attention to the role of DUF proteins in hybrid sterility. For example, two tightly linked genes at the *hsa1* locus—*HSA1a*, encoding a DUF1618 protein, and *HSA1b*, encoding an unknown protein—contribute to hybrid female sterility between two rice subspecies[8]. Similarly, *Sc*, which also encodes a DUF1618-domain-containing protein, is essential for hybrid sterility regulated by allelic suppression between *japonica* and *indica* rice subspecies[20]. Additionally, *S22A* and *S22B* encode a protein containing DUF1668 that is responsible for hybrid pollen semi-sterility between *O. glumaepatula* and *O. sativa*[15]. In this study, the allelic genes *RIS* and *RIA*, encoding mitochondrial precursor proteins with DUF1640, were found to act as a suppressor and an activator in SD, respectively (Fig. 2n,o). Unlike *HSA1* and *S22B*, *RIA* and *RIS* do not directly regulate hybrid fertility but are indispensable for male gamete development (Extended Data Figs. 5 and 6). Their roles are similar to that of *Sc*, which is also essential for pollen development[20]. Further investigations revealed that mutations in *RIA*/*RIS* lead to thicker pollen walls and insufficient starch grain accumulation, potentially resulting in reduced pollen germination ability and fertilization success (Extended Data Figs. 7 and 8). A previous study indicated that *STS1*, a protein containing DUF726, plays an important role in tapetum degradation and pollen wall formation by regulating lipid biosynthesis in rice[45]. These findings collectively suggest an association between DUF proteins and pollen cell wall formation. Numerous studies have also documented that male sterility genes often function in mitochondria by triggering ROS bursts, aberrant programmed cell death or cytotoxicity, leading to pollen abortion[46–49]. In this study, both *RIA* and *RIS* localize to mitochondria (Fig. 3a). *RIA*/*RIS* mutation may cause mitochondrial dysfunction by advancing ROS burst, disrupting redox homeostasis and resulting in an insufficient supply of energy (Extended Data Fig. 10). Interestingly, RIS and RIA differ by three amino acids, which turn RIS from a repressor to an activator (Figs. 1f and 2n,o). Similar functional switches have been reported in previous studies. A two-residue substitution (T304M and D306A) in CERK6's SD1 region was sufficient to confer gain of function for nodulation symbiosis, thereby reprogramming this chitin receptor to activate symbiotic signalling in *Lotus japonicus*[50]. In this study, the specific mechanism by which a 9-bp insertion could change the function of the original activator (*RIA*) into a repressor (*RIS*) remains to be further elucidated. Notably, sequence divergence exists between the *RIA* and *RIS* promoters. Analysis using the PlantCARE database revealed that 21 *cis*-regulatory elements are shared by both promoters, and several are unique to each. The *RIS* promoter specifically contains ABRE3a, ABRE4, ACE and ARE elements. In contrast, the *RIA* promoter exclusively harbours the GATA-motif, GC-motif, MYB-like sequence, RY-element, TCT-motif, W-box and WUN-motif. These distinct *cis*-element profiles may underlie the functional differences between *RIA* and *RIS*. A key objective for future research will be to identify which of these elements are directly bound by RID.

Previous studies have shown that JIEYAO detoxifies DUYAO by redirecting it to autophagosomes for degradation, thereby neutralizing its toxic effect on pollen[16]. The degradation of the CMS protein WA352 is mediated by the ubiquitination pathway in vegetable tissue, allowing WA352 to accumulate in anthers due to the downregulation of an F-box protein[51]. In this study, RID is predicted to encode a protein containing a DNA-binding domain, a transmembrane region and a ubiquitin-3-binding domain. RIR encodes an F-box protein with leucine-rich repeats localized in the nucleus (Fig. 3a). In addition, RIR degrades RID through direct interaction to eliminate its toxicity to gametes via the ubiquitination pathway (Fig. 3e). These findings imply that the activation of toxic proteins leading to pollen abortion results from the failure of their degradation through ubiquitination or autophagy pathways, which may represent a universal mechanism regulating male sterility.

In this study, systematic expression analysis of *RID* and *RIR* in a series of NILs in the DJY1 background revealed a consistent genotype–expression–phenotype relationship. Functional *RID* and *RIR* were both expressed in NIL-LOG. DJY1 and NIL-NIV both carried wide-compatibility haplotypes with non-functional *RID* and functional *RIR* expression. Accordingly, hybrids between NIL-LOG and DJY1 or NIL-NIV were fully fertile and showed no SD. NIL-BAR encoded a functional *RID* allele but showed no *RID* expression, while *RIR* was highly expressed. As a result, their hybrids with NIL-LOG exhibited SD with normal pollen fertility, underscoring the dosage-dependent effect of *RID* expression. NIL-RUF lacked the *RID* and *RIR* genes. Its hybrids with NIL-LOG displayed semi-sterility and marked SD (Fig. 6). This phenotypic variation supports a quantitative model of reproductive isolation established by the *S44* locus.

The mechanism by which reproductive isolation is established in the AA genome of the *Oryza* genus remains unclear. In this study, we observed that allelic conflicts of *S44* could induce varying degrees of reproductive isolation. Simple classifications such as fertility/sterility or distorted/non-distorted segregation fail to reflect the phenotypic complexity of these interactions. We therefore pioneered the CRII metric to quantitatively assess interspecific reproductive isolation strength. Reproductive isolation mediated by *S44* was strongest between *O. longistaminata* and *O. glaberrima*, followed by that between *O. longistaminata* and *O. rufipogon* and that between *O. longistaminata* and *O. barthii*, whereas no reproductive isolation was observed between *O. longistaminata* and the cultivated rice variety DJY1 or between *O. longistaminata* and *O. nivara* (Fig. 6d–f). Previous studies demonstrated that additive and epistatic effects across loci collectively determine reproductive isolation strength[18,52]. This study elucidated that a single locus in rice controls the strength of reproductive isolation through allelic conflict. This differential reproductive isolation leads to varying levels of gene flow between populations, thereby driving the emergence of genetic diversity across both interspecific and intraspecific groups. Regarding the four elements, further investigation is required to identify specific functional domains, critical amino acid substitutions or regulatory sequence differences underlying their functional divergence.

The African wild rice species *O. longistaminata* is a valuable genetic resource for improving the Asian domesticated rice, *O. sativa*[53–55]. Our findings indicate that neutral loci, knocking out *RID* or introducing *RIR* can break the reproductive isolation mediated by the *S44* locus, enabling the transfer of favourable genes into Asian cultivated rice. A recent report indicated that artificial gene drive systems use the poison–antidote mechanism to override Mendelian laws of inheritance in *Arabidopsis thaliana*, aiming to spread traits that are beneficial to humans in wild populations through synthetic gene drive technology[56]. In this study, the distorter–restorer system holds promising applications for enhancing breeding efficiency through targeted genetic manipulation. For example, we could enhance the transfer frequency of genes responsible for favourable traits by incorporating this system involving the *O. longistaminata* haplotype into the genome, such as perennial habit and resistance to biotic and abiotic stress[53,55,57–59]. Indeed, the distorter–restorer system controlled by *S44* could facilitate the spread of favourable genes without decreasing fertility, making it a potentially valuable tool for rice breeding programmes.

Methods

Plant material and growth conditions

RD23 (an *indica* variety from Thailand), as the female parent, was crossed with an *O. longistaminata* accession derived from the International Rice Research Institute serving as the donor parent. The F_1 hybrids were obtained through embryo rescue, and BC_6F_{10} individuals were raised via successive backcrosses, phenotype selection for pollen sterility and self-crosses[33]. Individuals in the BC_7F_2 population derived from a cross between a BC_6F_{10} plant and RD23 were genotyped using rice 6k Chips (Illumina). Individuals harbouring the homozygous genome fragment on the long arm of chromosome 6 from *O. longistaminata* were selected as NIL-*S44*. NIL-*S44* was crossed with the recurrent parent RD23 to produce BC_8F_1, which was then self-fertilized to generate the BC_8F_2 population. BC_8F_3 populations were derived from heterozygous individuals in BC_8F_2. Both populations were used for fine mapping.

A *japonica* cultivar, Dianjingyou 1 (DJY1), as the female parent and AA-genome accessions with different *S44* haplotypes as the donor and male parents were used to produce the F_1 hybrids. The SSR and KASP markers were employed to detect the *S44* locus, and individuals with a heterozygous *S44* locus were backcrossed six times using DJY1 as the recurrent parent from the F_1 generation. The genetic background of NIL-*S44* was scanned using a rice 6k SNP Chip (Illumina) in the BC_6F_2 generation.

All accessions of wild and cultivated rice for haplotype and evolutionary analysis were sourced from our laboratory, the International Rice Research Institute (Los Baños, Philippines) and the National BioResource Project.

All plant materials were planted at a spacing of 20 cm × 25 cm in the paddy field at the Experimental Station in Jinghong (22.00° N, 100.77° E), Yunnan Province, China.

Evaluation of pollen and spikelet fertility

Spikelets at the pre-flowering stage were fixed in 75% ethanol. Pollen grains from three to five mature flowers were stained with 1% I_2-KI solution. Spikelet fertility was examined as the seed-setting rate in the three main panicles of each individual plant.

In vitro pollen germination

The pollen germination medium contained 15% sucrose, 0.01% H_3BO_3 and 0.01% $Ca(NO_3)_2$ and was stored at 4 °C. A single drop of the medium was placed onto a glass slide. Florets with just-opened glumes were selected, and pollen grains were deposited onto the medium droplet via gentle shaking. Pollen was incubated in the dark at 30 °C for 15–20 min. Germination was assessed under a microscope. A pollen grain was considered germinated if the length of its pollen tube exceeded the grain's diameter. The pollen germination rate was calculated as the percentage of germinated pollen grains relative to the total number of pollen grains observed.

Cytological analyses

To observe the development of microspores, anthers at different developmental stages were first fixed in FAA solution and then embedded in Spur resin (ERL4221:DER736:NSA, 10:6:29). Semi-thin sections of 2 μm were prepared using a microtome (RM2235, Leica) and stained with 0.25% toluidine blue. The stained sections were then observed using a light microscope (BX51, Olympus).

For scanning electron microscopy, mature anthers from RD23, NIL-*S44* and the F_1 hybrids were pre-fixed in 2.5% glutaraldehyde for 24 h. The samples were rinsed three times with 0.1 M phosphate buffer (pH 7.0) and post-fixed in 1.5% osmium tetroxide (OsO_4) for 2 h in PBS (pH 7.0), followed by a series of ethanol dehydration steps, and then dried using a critical point drier with CO_2. Subsequently, the pollen grains were coated with gold and observed using a scanning electron microscope (GeminiSEM 300, ZEISS).

For transmission electron microscopy, anther samples were fixed in 2.5% glutaraldehyde for 24 h and 1.5% osmium tetroxide (OsO_4) for 2 h. The samples were then dehydrated through a series of ethanol treatments and embedded in resin. Ultrathin sections (60 nm) were stained with uranyl acetate and citrate aqueous solution. Images were obtained with a transmission electron microscope (JEM1400, JEOL) operated at 80 kV.

For embryo sac observation, spikelets were fixed in FAA solution for two days at 4 °C, followed by dehydration in the ethanol and methyl salicylate series. The samples were then treated with pure methyl salicylate twice until visually transparent and stored at 4 °C. Embryo sac fertility was imaged using a confocal microscope (Fluoview, FV1000, Olympus).

Fine mapping of the *S44* locus

The *S44* locus was initially mapped to a region on the long arm of chromosome 6 between markers OL3019-3 and ID4678-1, using the RD23 × NIL-*S44* BC_8F_2 population consisting of 1,298 individuals. The mapping region was subsequently narrowed to a 29.4-kb region between markers OL3019-2 and OL3025-1 using the BC_8F_3 population, which included 12,016 individuals derived from the progeny of the BC_8F_2 population. The genomic sequence of the *S44* region was amplified from RD23 and NIL-*S44* using KOD-FX DNA polymerase (Toyobo). The sequences were then blasted and assembled using DNAStar Lasergene v.11. All primers are listed in Supplementary Table 7.

RNA extraction and quantitative PCR with reverse transcription analysis

Samples of roots, stem bases, leaves and anthers at different developmental stages were collected from RD23, NIL-*S44* and the F_1 hybrids. Total RNA was extracted using the MiniBEST Universal RNA Extraction Kit (TaKaRa). First-strand cDNA was synthesized using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific). Quantitative PCR with reverse transcription was performed using iTaq Universal SYBR Green Supermix (Biorad) on a CFX96 system (Biorad) with three biological replicates. PCR amplification was conducted according to the manufacturer's instructions. The data were normalized using the *ubiquitin1* gene (Os03g0234350) as the endogenous control. All primers are listed in Supplementary Table 7.

Construct for genetic transformation

For the complementation test, 7,675 bp of the *RIS* genomic fragment from RD23 and 8,676 bp of the *RIA* genomic fragment, 4,904 bp of the *RID* genomic fragment and 9,512 bp of the *RIR* genomic fragment from NIL-*S44* were amplified. The genomic fragments of *RIS*, *RID* and *RIR* were fused into the binary vector pCAMBIA1305.1 to obtain the t*RIS*, t*RID* and t*RIR* constructs, respectively. The genomic sequence of *RIA* was inserted into pCAMBIA1300 to generate the t*RIA* construct. These vectors were transformed into the callus of RD23, NIL-*S44* and the F_1 hybrids via an *Agrobacterium*-mediated transformation system. Positive transgenic individuals were identified via PCR using HPT-specific primers and target-gene-specific primers.

For CRISPR–Cas9 vector construction, the 24-bp target-specific sequences were fused with the pBWA (V)H vector to generate knockout constructs, which were then introduced into the callus of RD23, NIL-*S44* or the F_1 hybrids. Positive edited transformants were identified via sequencing. The target sequences and sequencing primers are listed in Supplementary Table 7.

Rice transformations were performed by Biorun BIO-TECH CO, Ltd, Wuhan, China.

Subcellular localization

For subcellular localization, the coding regions of *RIS*, *RIA*, *RID* and *RIR* (lacking stop codons) were cloned into the pCAMBIA 1390 vector carrying a GFP tag. The *RIS-eGFP* and *RIA-eGFP* constructs were transformed into rice protoplasts and incubated in the dark at 30 °C for 15 h. Transformed protoplasts were stained with the mitochondrion-specific dye MitoTracker Deep Red (Invitrogen) to visualize mitochondria signals. *RID-eGFP* and *RIS-eGFP*, along with the nuclear marker *NLS-RFP*, were co-transformed into rice protoplasts to observe their nuclear signals. Fluorescence imaging was performed using confocal laser scanning microscopy (LSM780, ZEISS). All primers are listed in Supplementary Table 7.

β-Glucuronidase histochemical staining

The upstream promoter sequences of 2,253 bp, 3,129 bp, 2,648 bp and 2,374 bp from the start codons of *RIA*, *RIS*, *RID* and *RIR*, respectively, were amplified from the RD23 or NIL-*S44* genome and cloned into the binary vector pCAMBIA 1305 to drive GUS reporter gene expression. These vectors were introduced into the callus of RD23 via an *Agrobacterium tumefaciens*-mediated transformation system. Roots, stem bases, leaves and young panicles of T_0 plants were stained with GUS dye and fixed in 75% ethanol. Images were visualized using a stereoscopic microscope and a light microscope (BX51 and SZX16, Olympus). All primers are listed in Supplementary Table 7.

RNA in situ hybridization

Young spikelets at different developmental stages were fixed in FAA solution at 4 °C overnight. The samples were embedded in paraffin following a series of dehydration steps. Gene-specific probes were amplified from RD23 or NIL-*S44* cDNA and cloned into the pGEMT Easy vector (Promega). The hybridization probes were synthesized in vitro using a DIG Northern Starker Kit (Roche). Transverse sections of anthers were hybridized with gene-specific probes. Sections hybridized by sense probes and anti-sense probes were observed using a light microscope (BX51, Olympus). All primers are listed in Supplementary Table 7.

Yeast two-hybrid assay

The coding sequences of *RID* and *RIR* were amplified and inserted into the yeast two-hybrid bait and prey vectors, pGADT7 and pGBKT7 (Clontech), respectively. The bait and prey vectors were co-transformed in various combinations into *Saccharomyces cerevisiae* strain AH109 according to the manufacturer's instructions (Clontech). The pGADT7-T and pGBKT7-53 combination served as a positive control, while the pGADT7-T and pGBKT7-Lam combination was used as a negative control. Yeast cells were grown on non-selective SD/-Trp/-Leu synthetic dropout medium without tryptophan and leucine, and selective SD/-Trp/-Leu/-His/-Ade medium containing X-α-Gal without tryptophan, leucine, histidine and adenine at 28 °C for three to five days. All primers are listed in Supplementary Table 7.

Yeast one-hybrid assay

The promoter sequences of *RIA* and *RIS* were amplified and inserted into the yeast one-hybrid vector pABAi. The coding sequence of *RID* was cloned into pGADT7. These vectors were co-transformed into *S. cerevisiae* strain AH109 according to the manufacturer's instructions. Yeast cells were grown on synthetic dropout medium (SD/-Ura/-Leu) and selective medium (SD/-Ura/-Leu 200 ng ml^{-1} AbA or SD/-Ura/-Leu 800 ng ml^{-1} AbA) at 28 °C for three to five days. All primers are listed in Supplementary Table 7.

Split-luciferase complementary assay

The coding regions of target genes were cloned into split-luciferase vectors. RID was fused to the C-terminal half (cLUC–RID) and RIR and RIA to the N-terminal half (RIR–nLUC and RIA–nLUC) of the split luciferase; cLUC and nLUC served as controls. The plasmids were transformed into *Agrobacterium* strain GV3101, cultured overnight and resuspended in activation buffer (10 mM $MgCl_2$, 10 mM MES, 200 μM acetosyringone, pH 5.7) to $OD_{600} = 1$. The two mixed suspensions were co-infiltrated into leaves of four-week-old *N. benthamiana*. Luciferase activity was measured using luciferin as a substrate. Images were captured using the Tanon-5200 Chemiluminescent imaging system (Tanon). All primers are listed in Supplementary Table 7.

Ubiquitination assay

The coding sequence of *RID* was amplified and cloned into the *35S*::*FLAG* vector to generate the *35S*::*RID-FLAG* construct. Similarly, the *RIR* coding sequence was inserted into the *35S*::*GFP* vector to create the *35S*::*RIR-GFP* construct. The backbone sequence of the SCF E3 ubiquitin ligase complex eSCR was sourced from a previous study[60] and inserted into the *35S*::*MYC* vector to create the *35S*::*eSCR-MYC* construct. All constructs were transformed into *Agrobacterium tumefaciens* strain GV3101, followed by infiltration of leaves from four-week-old tobacco plants. To inhibit the ubiquitin-proteasome pathway, 50 μM MG132 (a proteasome inhibitor) was infiltrated 12 h before sample collection. Subsequently, the samples were pulverized in liquid nitrogen, and the resulting powder was added to protein extraction buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, 0.1% (v/v) NP-40, 0.5 mM DTT, 1 mM PMSF and 1× protease inhibitor cocktail), followed by incubation at 4 °C for 30 min with gentle agitation. The homogenate was centrifuged at 4 °C to obtain a clarified tissue lysate supernatant. For immunoprecipitation analysis, anti-FLAG magnetic beads were incubated with the supernatant for 3 h at 4 °C with end-over-end rotation. Bead-bound proteins were then collected via magnetic separation, washed three times with ice-cold extraction buffer and subsequently subjected to western blot using antibodies against ubiquitin, GFP, FLAG and MYC epitopes.

Dual-luciferase reporter assay in rice protoplasts

Approximately 2.2-kb and 3.1-kb promoter sequences of *RIS* and *RIA* were amplified and inserted into the pGreenII 0800-LUC vector to generate the pro*RIS*-LUC and pro*RIA*-LUC constructs, respectively. The coding sequences of *RID* and *RIR* were also inserted into the pGreenII-62SK vector. The plasmids were co-transformed into rice protoplasts. The ratio of firefly luciferase activity to *Renilla* luciferase activity (LUC/REN) was measured with the Luciferase Assay System (Promega) and calculated on the basis of ten replicates. All primers are listed in Supplementary Table 7.

NBT staining

Anthers at different developmental stages were extracted from RD23, NIL-*S44*, F_1, NIL-t*RID*, RD23-*ris* and NIL-*S44*-*ria* and immersed immediately in NBT staining solution (Servicebio) for 6 h in the dark at room temperature. After staining, the anthers were rinsed with sterile water, treated with 95% ethanol for 12 h and then visualized through a stereomicroscope (SZX16, Olympus).

Measurement of NAD^+/NADH and ATP content

Fresh leaf samples from RD23, NIL-*S44*, F_1, NIL-t*RID*, RD23-*ris* and NIL-*S44*-*ria* were collected and thoroughly homogenized in lysis buffer. The homogenate was centrifuged at 12,000 *g* and 4 °C for 10 min, and the supernatant was collected for NAD^+/NADH content measurement using the manufacturer's protocol (Beyotime).

For ATP content analysis, fresh leaf samples were homogenized in lysis buffer. After lysis, the samples were centrifuged at 12,000 *g* and 4 °C for 5 min. The supernatant was collected, and ATP content was determined according to the manufacturer's instructions (Beyotime).

Evolution and haplotype analysis of the *S44* locus

Diverse wild and cultivated rice accessions were used to isolate the *S44* locus via PCR, and the obtained genomic fragments were sequenced. Additionally, the *S44* regions from both rice pan-genome sequencing data and outgroup genomes were located and extracted using MUMmer (v.4.0.0)[61] followed by SAMtools (v.1.19.2)[62]. The coding sequences of the *RIS*/*RIA*, *RID* and *RIR* genes were translated into protein sequences using Biopython (1.8.3)[63]. These coding sequences and protein sequences were subsequently subjected to multiple sequence alignment using Mafft (v.7.526)[64], facilitating the analysis of their evolutionary history and haplotypes. Haplotype analysis was conducted with DNAsp (v.6.12.03)[65] to identify and define distinct haplotypes. The R package rworldmap (v.1.3-8)[66] was used to generate colour-coded maps illustrating the geographical distribution of haplotypes. All quantitative analyses were performed using custom Python scripts (3.9.12) (https://www.python.org/). All primers are listed in Supplementary Table 7.

The CRII

The HFI is calculated by comparing the mean reproductive fitness of hybrid progeny to that of the parental species. The formula for the HFI is HFI = $F_{\mathrm{hybrid}}/F_{\mathrm{parent}}$. The range of the HFI values extends from 0 to 1. When the HFI equals 1, it signifies that the reproductive fitness of hybrid offspring is equivalent to that of the parental species, indicating the absence of reproductive isolation. When the HFI equals 0, it indicates complete sterility in hybrid offspring, implying the presence of complete reproductive isolation.

The GSI is a measure used to assess the degree of deviation of gamete segregation from the expected Mendelian ratios. It is calculated on the basis of the observed proportion of a particular genotype in hybrid offspring (denoted as p_{obs}) and the expected proportion of that genotype according to Mendelian inheritance (denoted as p_{exp}). The formula for the GSI is GSI = $(\sum|(p_{\mathrm{obs}} - p_{\mathrm{exp}})/p_{\mathrm{exp}}|)/n$. The range of GSI values extends from 0 to 1. When the GSI equals 0, it indicates that the observed genotype proportion is identical to the expected Mendelian proportion, suggesting no gametic selection or deviation from Mendelian segregation. When the GSI equals 1, it denotes that the observed genotype proportion deviates completely from the expected Mendelian proportion, suggesting complete gametic selection or bias in gamete segregation.

The CRII is a metric that integrates multiple factors to assess the overall degree of reproductive isolation between species. It combines the HFI and the GSI to provide a comprehensive measure. The formula for the CRII is CRII = 1 − (HFI × (1 − GSI)). The range of CRII values extends from 0 to 1. When the CRII equals 0, it indicates the absence of reproductive isolation, meaning that there are no barriers to successful reproduction between the species. When the CRII equals 1, it signifies complete reproductive isolation, meaning that there are strong barriers preventing successful reproduction between the species.

Reporting summary

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

Data availability

The genotypes of part varieties were collected from RiceSuperPIRdb (http://www.ricesuperpir.com). The chromosomal-level genomes of outgroup species *L. perrieri* (GCA_000325765.3), *O. granulata* (GCA_005223365.2) and *O. punctata* (GCA_000573905.2) were retrieved from NCBI GenBank. The nucleotide sequences of *RIS* (PQ723328), *RIA* (PQ723325), *RID* (PQ723326) and *RIR* (PQ723327) in this study have been uploaded to NCBI. Source data are provided with this paper. All other data are available within this article and its Supplementary Information.

References

1. Wu, C. I. & Ting, C. T. Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122 (2004).
2. Fishman, L. & Sweigart, A. L. When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annu. Rev. Plant Biol.* **69**, 707–731 (2018).
3. Westram, A. M., Stankowski, S., Surendranadh, P. & Barton, N. What is reproductive isolation? *J. Evol. Biol.* **35**, 1143–1164 (2022).
4. Zhao, Z., Shen, R. & Liu, Y. G. Hybrid sterility genes with driving force for speciation in rice. *Sci. Bull. (Beijing)* **68**, 1845–1848 (2023).
5. Ouyang, Y., Liu, Y. G. & Zhang, Q. Hybrid sterility in plant: stories from rice. *Curr. Opin. Plant Biol.* **13**, 186–192 (2010).
6. Ouyang, Y. & Zhang, Q. Understanding reproductive isolation based on the rice model. *Annu. Rev. Plant Biol.* **64**, 111–135 (2013).
7. Yang, J. et al. A killer–protector system regulates both hybrid sterility and segregation distortion in rice. *Science* **337**, 1336–1340 (2012).
8. Kubo, T., Takashi, T., Ashikari, M., Yoshimura, A. & Kurata, N. Two tightly linked genes at the *hsa1* locus cause both F_1 and F_2 hybrid sterility in rice. *Mol. Plant* **9**, 221–232 (2016).
9. Yu, Y. et al. Hybrid sterility in rice (*Oryza sativa* L.) involves the tetratricopeptide repeat domain containing protein. *Genetics* **203**, 1439–1451 (2016).

10. Hou, J. et al. ESA1 is involved in embryo sac abortion in interspecific hybrid progeny of rice. *Plant Physiol*. **180**, 356–366 (2019).
11. Long, Y. et al. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proc. Natl Acad. Sci. USA* **105**, 18871–18876 (2008).
12. Yamagata, Y. et al. Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *Proc. Natl Acad. Sci. USA* **107**, 1494–1499 (2010).
13. Nguyen, G. N. et al. Duplication and loss of function of genes encoding RNA polymerase III subunit C4 causes hybrid incompatibility in rice. *G3 (Bethesda)* **7**, 2565–2575 (2017).
14. Yu, X. et al. A selfish genetic element confers non-Mendelian inheritance in rice. *Science* **360**, 1130–1132 (2018).
15. Sakata, M. et al. Domain unknown function DUF1668-containing genes in multiple lineages are responsible for F_1 pollen sterility in rice. *Front. Plant Sci*. **11**, 632420 (2020).
16. Wang, C. et al. A natural gene drive system confers reproductive isolation in rice. *Cell* **186**, 3577–3592 e3518 (2023).
17. Wang, D. et al. Two complementary genes in a presence–absence variation contribute to indica–japonica reproductive isolation in rice. *Nat. Commun*. **14**, 4531 (2023).
18. You, S. et al. A toxin–antidote system contributes to interspecific reproductive isolation in rice. *Nat. Commun*. **14**, 7528 (2023).
19. Zhou, P. et al. A minimal genome design to maximally guarantee fertile inter-subspecific hybrid rice. *Mol. Plant* **16**, 726–738 (2023).
20. Shen, R. et al. Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. *Nat. Commun*. **8**, 1310 (2017).
21. Xie, Y. et al. Interspecific hybrid sterility in rice is mediated by Og*TPR1* at the *S1* locus encoding a peptidase-like protein. *Mol. Plant* **10**, 1137–1140 (2017).
22. Koide, Y. et al. Lineage-specific gene acquisition or loss is involved in interspecific hybrid sterility in rice. *Proc. Natl Acad. Sci. USA* **115**, E1955–E1962 (2018).
23. Xie, Y. et al. An asymmetric allelic interaction drives allele transmission bias in interspecific rice hybrids. *Nat. Commun*. **10**, 2501 (2019).
24. Mizuta, Y., Harushima, Y. & Kurata, N. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl Acad. Sci. USA* **107**, 20417–20422 (2010).
25. Fishman, L. & McIntosh, M. Standard deviations: the biological bases of transmission ratio distortion. *Annu. Rev. Genet*. **53**, 347–372 (2019).
26. Akera, T. et al. Spindle asymmetry drives non-Mendelian chromosome segregation. *Science* **358**, 668–672 (2017).
27. Dawe, R. K. et al. A Kinesin-14 Motor activates neocentromeres to promote meiotic drive in maize. *Cell* **173**, 839–850 e818 (2018).
28. Bravo Nunez, M. A., Lange, J. J. & Zanders, S. E. A suppressor of a wtf poison-antidote meiotic driver acts via mimicry of the driver's antidote. *PLoS Genet*. **14**, e1007836 (2018).
29. Grognet, P., Lalucque, H., Malagnac, F. & Silar, P. Genes that bias Mendelian segregation. *PLoS Genet*. **10**, e1004387 (2014).
30. Merrill, C., Bayraktaroglu, L., Kusano, A. & Ganetzky, B. Truncated RanGAP encoded by the *Segregation Distorter* locus of *Drosophila*. *Science* **283**, 1742–1745 (1999).
31. Larracuente, A. M. & Presgraves, D. C. The selfish *Segregation Distorter* gene complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).
32. Ben-David, E., Burga, A. & Kruglyak, L. A maternal-effect selfish genetic element in *Caenorhabditis elegans*. *Science* **356**, 1051–1055 (2017).
33. Zhao, J. et al. A new gene controlling hybrid sterility between *Oryza sativa* and *Oryza longistaminata*. *Euphytica* **187**, 339–344 (2012).
34. Shang, L. et al. A super pan-genomic landscape of rice. *Cell Res*. **32**, 878–896 (2022).
35. Oka, H. Analysis of genes controlling f(1) sterility in rice by the use of isogenic lines. *Genetics* **77**, 521–534 (1974).
36. Ikehashi, H. & Araki, H. Genetics of F_1 sterility in remote crosses of rice. *Rice Genet*. https://doi.org/10.1142/9789812814265_0011 (2008).
37. Sano, Y. The genic nature of gamete eliminator in rice. *Genetics* **125**, 183–191 (1990).
38. Charron, Y. et al. Two isoforms of the RAC-specific guanine nucleotide exchange factor TIAM2 act oppositely on transmission ratio distortion by the mouse t-haplotype. *PLoS Genet*. **15**, e1007964 (2019).
39. Kusano, A., Staber, C. & Ganetzky, B. Segregation distortion induced by wild-type RanGAP in *Drosophila*. *Proc. Natl Acad. Sci. USA* **99**, 6866–6870 (2002).
40. Bauer, H., Willert, J., Koschorz, B. & Herrmann, B. G. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nat. Genet*. **37**, 969–973 (2005).
41. Bauer, H., Veron, N., Willert, J. & Herrmann, B. G. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes Dev*. **21**, 143–147 (2007).
42. Xie, Y., Shen, R., Chen, L. & Liu, Y. G. Molecular mechanisms of hybrid sterility in rice. *Sci. China Life Sci*. **62**, 737–743 (2019).
43. Luo, C. et al. Domain of unknown function (DUF) proteins in plants: function and perspective. *Protoplasma* https://doi.org/10.1007/s00709-023-01917-8 (2023).
44. Lv, P. et al. Unraveling the diverse roles of neglected genes containing domains of unknown function (DUFs): progress and perspective. *Int. J. Mol. Sci*. https://doi.org/10.3390/ijms24044187 (2023).
45. Yuan, G. et al. SWOLLEN TAPETUM AND STERILITY 1 is required for tapetum degeneration and pollen wall formation in rice. *Plant Physiol*. **190**, 352–370 (2022).
46. Komori, T. et al. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant J*. **37**, 315–325 (2004).
47. Wang, Z. et al. Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* **18**, 676–687 (2006).
48. Wang, K. et al. ORFH79 impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. *New Phytol*. **198**, 408–418 (2013).
49. Huang, W. et al. Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. *Proc. Natl Acad. Sci. USA* **112**, 14984–14989 (2015).
50. Tsitsikli, M. et al. Two residues reprogram immunity receptors for nitrogen-fixing symbiosis. *Nature* https://doi.org/10.1038/s41586-025-09696-3 (2025).
51. Zhang, Z. et al. Ubiquitin-mediated degradation restricts spatiotemporal accumulation of the cytoplasmic male sterility protein WA352 to anthers in rice. *Proc. Natl Acad. Sci. USA* **122**, e2504381122 (2025).
52. Kubo, T., Yoshimura, A. & Kurata, N. Pollen killer gene *S35* function requires interaction with an activator that maps close to *S24*, another pollen killer gene in rice. *G3 (Bethesda)* **6**, 1459–1468 (2016).
53. Khush, G., Bacalangco, E. & Ogawa, T. A new gene for resistance to bacterial blight from *O. longistaminata*. *Rice Genet. Newslett*. **7**, 121–122 (1991).

54. Labroo, M. R. et al. Solving the mystery of Obake rice in Africa: population structure analyses of *Oryza longistaminata* reveal three genetic groups and evidence of both recent and ancient introgression with *O. sativa*. *Front. Plant Sci.* **14**, 1278196 (2023).
55. Zhang, S. et al. Sustained productivity and agronomic potential of perennial rice. *Nat. Sustain.* **6**, 28–38 (2023).
56. Liu, Y., Jiao, B., Champer, J. & Qian, W. Overriding Mendelian inheritance in *Arabidopsis* with a CRISPR toxin-antidote gene drive that impairs pollen germination. *Nat. Plants* **10**, 910–922 (2024).
57. Song, W. Y. et al. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* **270**, 1804–1806 (1995).
58. Giuliani, R. et al. Coordination of leaf photosynthesis, transpiration, and structural traits in rice and wild relatives (genus *Oryza*). *Plant Physiol* **162**, 1632–1651 (2013).
59. Lian, X. et al. Spatiotemporal transcriptomic atlas of rhizome formation in *Oryza longistaminata*. *Plant Biotechnol. J.* https://doi.org/10.1111/pbi.14294 (2024).
60. Liu, H. et al. An engineered platform for reconstituting functional multisubunit SCF E3 ligase in vitro. *Mol. Plant* **15**, 1285–1299 (2022).
61. Marcais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
62. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* https://doi.org/10.1093/gigascience/giab008 (2021).
63. Cock, P. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
64. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
65. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
66. South, A. rworldmap: A new R package for mapping global data. *R J.* **3**, 35–43 (2011).

Acknowledgements

We thank F. Chen, F. Zhang and J. Jiang from the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, for their invaluable contributions to the molecular and cytological experiments and for their insightful revisions to the manuscript. We also appreciate the discussions on evolutionary analysis provided by S. Ge from the Institute of Botany, Chinese Academy of Sciences, and W. Qian from the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences. This work was supported by grants from the National Natural Science Foundation of China (grant nos. 32460503, 31660380 and 31201196 to Y.Z. and 32522075, 32494773 to L.S.), the State Key Laboratory of Molecular Developmental Biology (grant no. 054 to Y.Z.), Yunnan Fundamental Research Projects (grant no. 202001AS070003 to Y.Z.) and the Yunnan Provincial Government (grant nos YNWR-QNBJ-2018-359 and 5300002100 00000013809 to Y.Z.).

Author contributions

Y.Z., L.S. and D.T. conceived and designed the experiments. Y.Z., Y.Y. and C.S. performed most experiments. Q.P., J.Z., P.X., X.S., X.D. and Y.L. developed some NILs and performed the genetic analysis. W.H., Q.X., F.H., X.J. and L.H. performed the evolutionary analysis and biochemical assays. Y.Z. analysed the data and wrote the manuscript. D.T., L.S., P.X. and C.S. revised the manuscript.

Competing interests

The authors declare no competing interests.

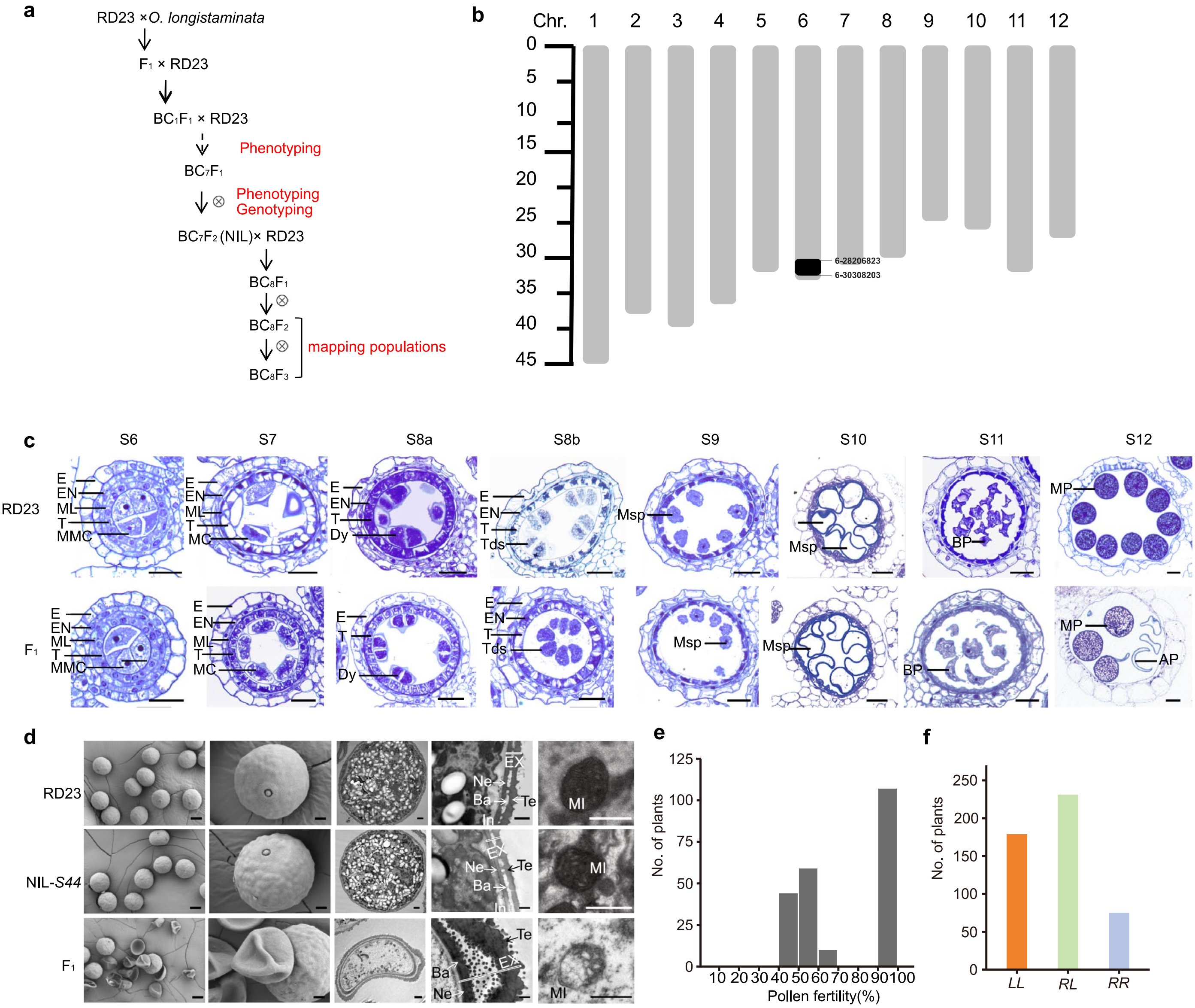
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-026-02223-w.

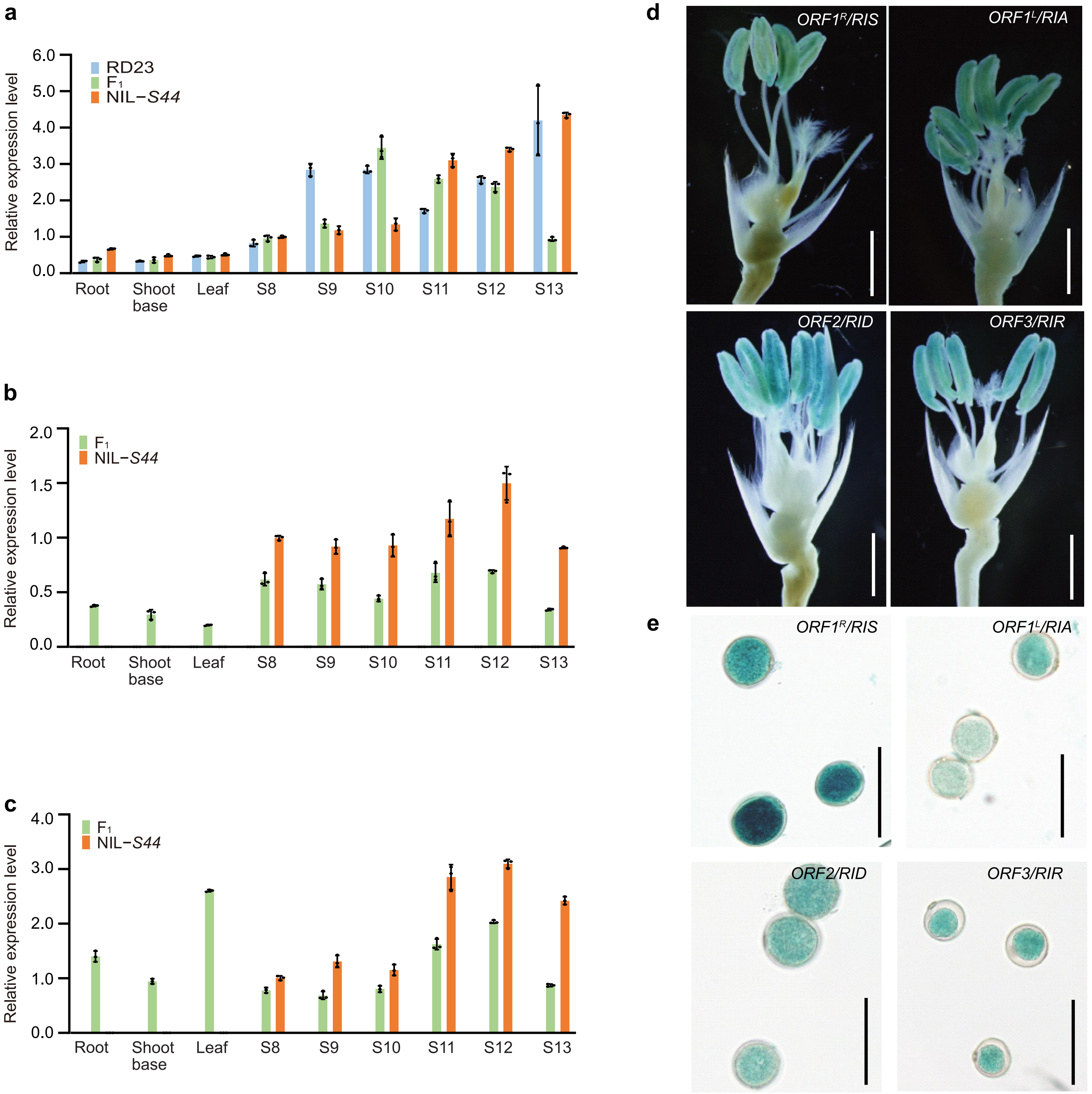
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-026-02223-w.

Correspondence and requests for materials should be addressed to Yu Zhang, Lianguang Shang or Dayun Tao.

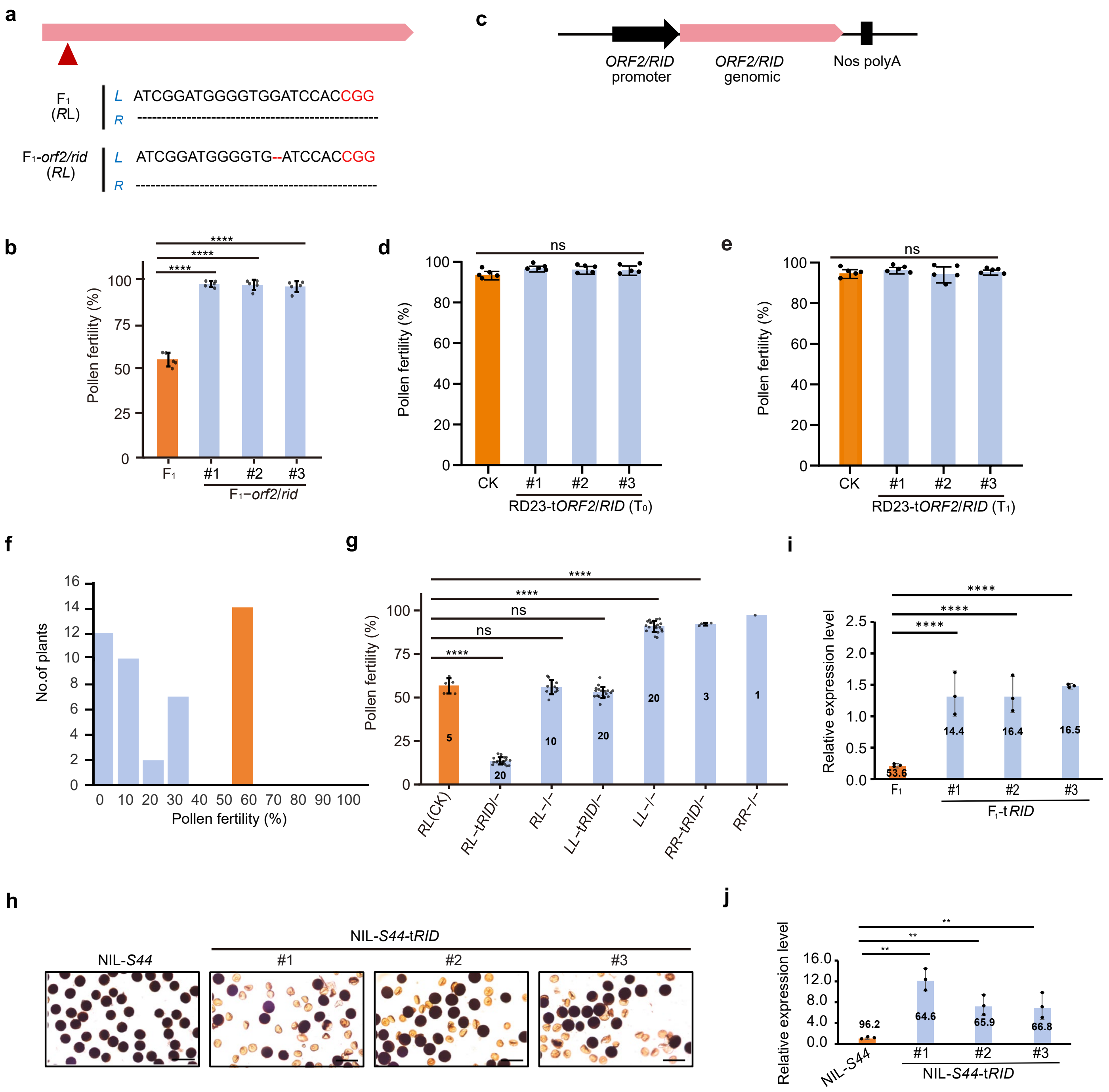
Peer review information *Nature Plants* thanks Bao Liu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



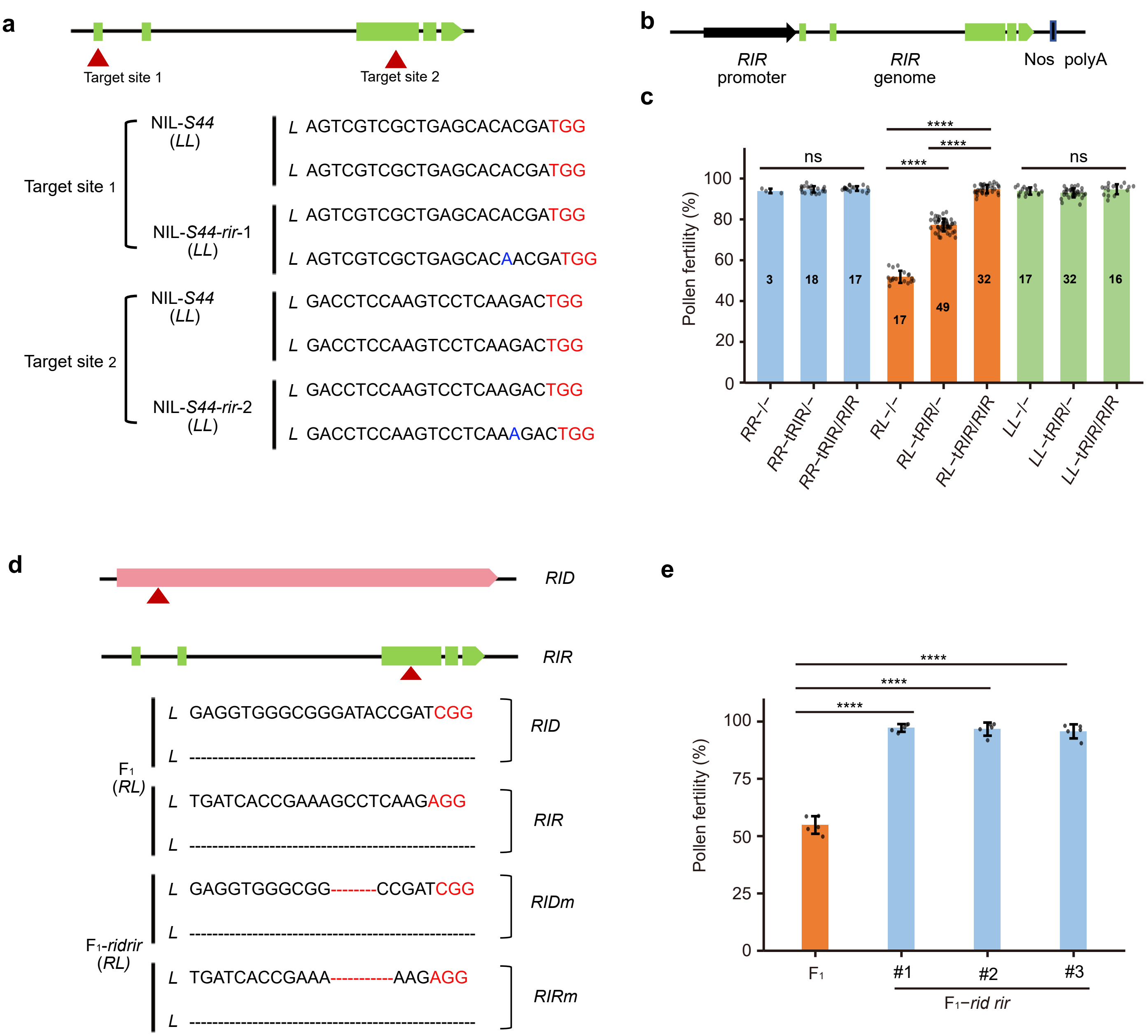
Extended Data Fig. 1 | Phenotype of RD23, NIL-*S44* and their F_1, genetic analysis of *S44*. **a**, Introgression of *S44* from *O. longistaminata* to RD23 by backcrossing to create NIL-*S44*. **b**, Genomic fragment from *O. longistaminata* identified by whole-genome scanning with Illumina 6k SNP Chip. The *O. longistaminata* fragment harboring *S44* associated with pollen sterility is shown in a black box and grey box indicates RD23 genome. **c**, Cytological observation of anther development at S6-S12 stages in RD23 and the F_1 hybrids. E, epidermis; En, endothecium; ML, middle layer; T, tapetum; Dy, dyad cell; Tds, tetrads; Msp, microspore; BP, bicellular pollen; MP, mature pollen; AP, abortive pollen. S6, meiocyte mother cell (MMC) stage; S7, meiosis stage; S8a, dyad stage; S8b, tetrad stage; S9, early microspore stage; S10, vacuolated microspore stage; S11, bicellular pollen stage; S12, tricellular pollen stage. Scale bar, 50 μm. n = 3 independent biological samples. **d**, Observation of RD23, NIL-*S44* and F_1 at mature pollen stages. The first column indicates the pollen shape imaged by scanning electron microscopy (SEM), Scale bar, 20 μm. The second column magnified normal or aborted pollen grains from the first column. Scale bar, 5 μm. The third column shows normal pollen grain with abundant starch granules in RD23 and NIL-*S44*, whereas abnormal pollen grain with few starch granules in F_1. Scale bar, 2 μm. The fourth column displays thin pollen walls in two parents and abnormal thick pollen wall in F_1. Scale bar, 200 nm. The last column depicts normal mitochondria in the parents and abnormal mitochondria in F_1. Scale bar, 500 nm. n = 5 independent biological replicates. Ex, exine; In, intine; Te, tectum; Ne, nexine; Ba, bacula; Ne, nexine; MI, mitochondria. **e**, Frequency distribution of pollen fertility in RD23×NIL-*S44* F_2 population. $\chi^2_{(1:1)} = 0.16$ $P = 0.69$ in χ^2 test. **f**, Genotype frequency in the F_2 population derived from the cross between RD23 and NIL-*S44*. *LL*, $\chi^2_{(1:1)} = 45.69$, $P = 1.20E^{-10}$ in χ^2 test. *LL*, *RL*, *RR* stand for homozygous *O.longistaminata* genotype, heterozygous genotype and homozygous RD23 genotype.



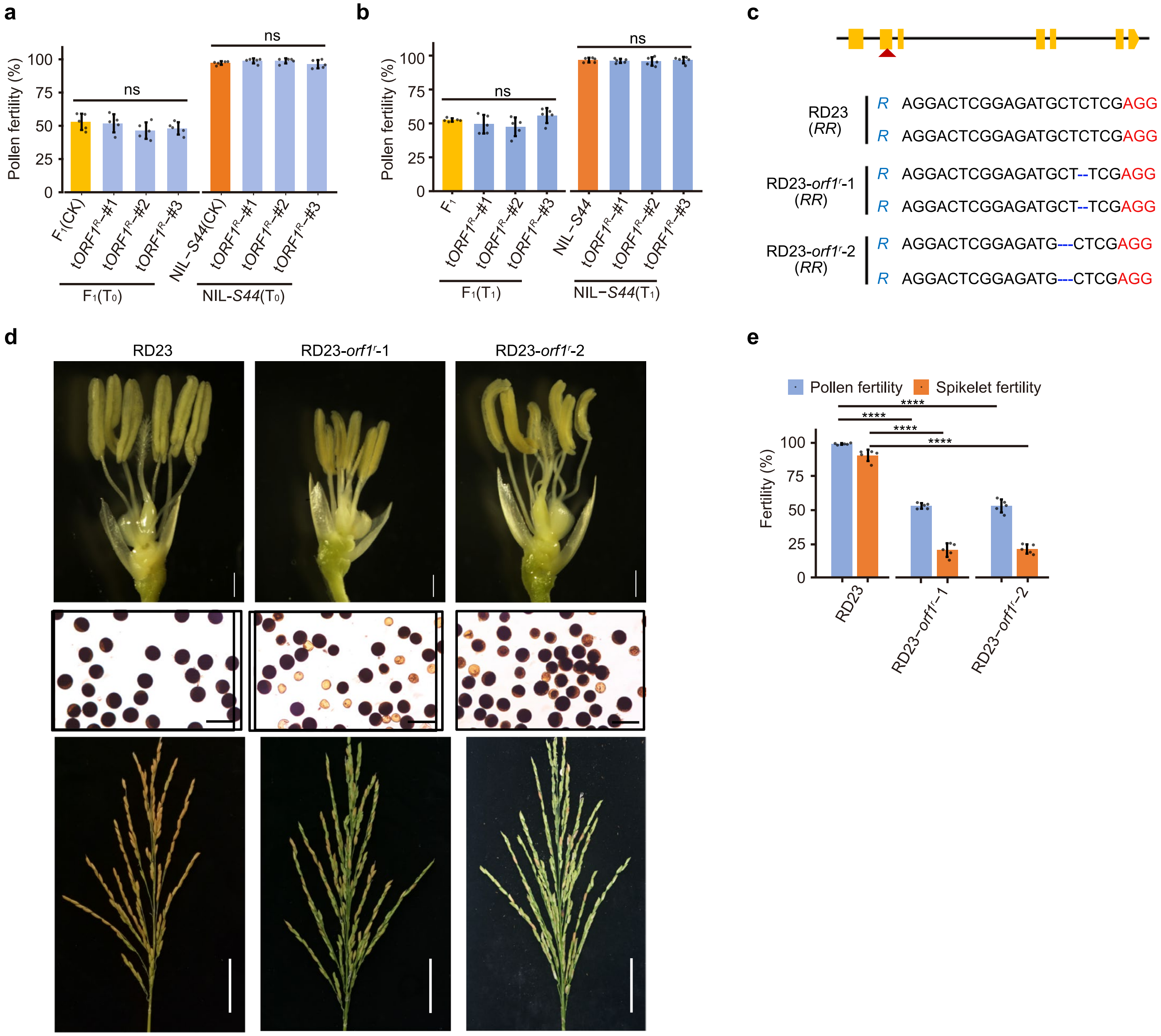
Extended Data Fig. 2 | Expression patterns of genes at the *S44* locus. **a**, Expression pattern of *ORF1*(*RIS/RIA*) in RD23, NIL-*S44*, and F$_1$ at the vegetative and reproductive stages. Transcripts of *ORF1* are detected in all vegetative and reproductive organs. The expression level of *ORF1* in both RD23 and NIL-*S44* is significant higher at stage S13 compared to the other stages, while its expression in F$_1$ is lower than that in two parents at stage S13. **b**, Expression level of *ORF2/RID* in RD23, NIL-*S44* and F$_1$ at the vegetative and reproductive stage. The expression level of *ORF2/RID* peaks at stage S12. **c**, Comparison of *ORF3/RIR* expression in RD23, NIL-*S44* and F$_1$ at the vegetative and reproductive stage. A high expression level of *ORF3/RIR* is detected at stage S12. **d**, GUS staining of transgenic plants harboring the promoter fragments of *ORF1*^{R}*/RIS*, *ORF1*^{L}*/RIA*, *ORF2/RID*, *ORF3/RIR* and GUS gene. Expression of all four genes is detected in the stamens. n = 5 biological samples. Scale bar, 200 µm. **e**, Expression of *ORF1*^{R}*/RIS*, *ORF1*^{L}*/RIA*, *ORF2/RID*, *ORF3/RIR* is observed in pollen grains. n = 5 biological samples. Scale bar, 10 µm. In **a**, **b** and **c**, Data are presented as mean ± SD, n = 3 independent biological replications.



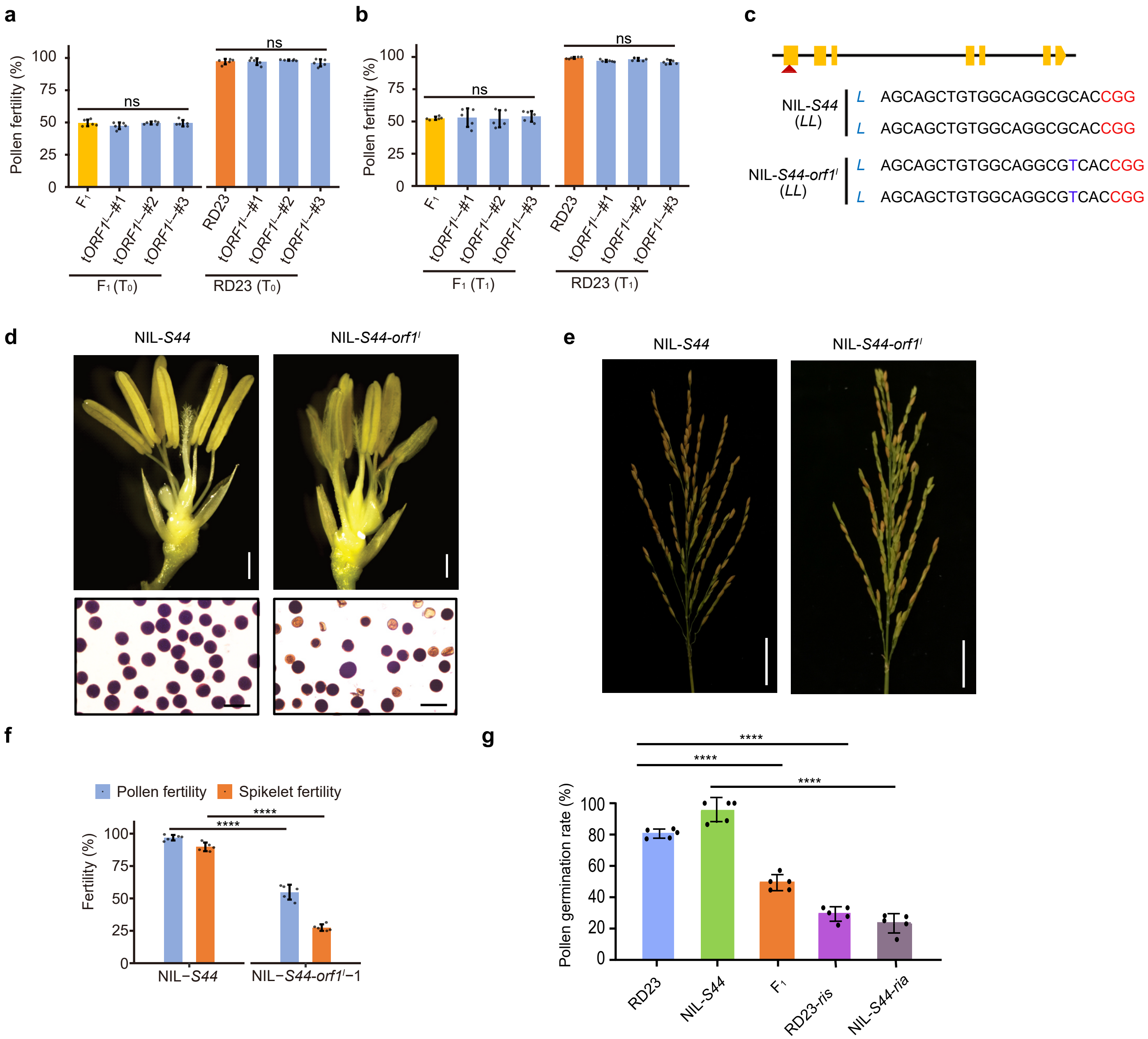
Extended Data Fig. 3 | *RID* kills pollen in a sporophytic manner. a, Target site in *RID* for CRISPR/Cas9-mediated mutagenesis. The PAM sequence is highlighted in red letters, and deletion is marked by a red dotted line. **b**, Knocking out *RID* rescues normal pollen fertility of F$_1$ hybrids in the T$_1$ generation, n = 5 plants. **c**, A *RID* genomic fragment (including 2.6 kb promoter and 1.2 kb 3′ UTR) from NIL-*S44* was used for genetic complementation test. **d**, Pollen fertility of RD23 harboring exogenous *RID* in the T$_0$ generation, n = 5 plants. **e**, Pollen fertility of RD23 harboring exogenous *RID* in the T$_1$ generation, n = 5 plants. **f**, Fertility frequency distribution of complementation (F$_1$-t*RID*) lines in the T$_0$ generation. **g**, Pollen fertility of complementation (F$_1$-t*RID*) lines in the T$_1$ generation, with the number of different genotype plants noted in the column. **h**, Introduction of the *RID* gene into NIL-*S44* results in pollen semi-sterility. Scale bars, 100 µm. **i**, Relative expression level of *RID* in the F$_1$ and F$_1$-t*RID*. Pollen fertility (%) is presented in the columns. n = 3 independent biological replicates. **j**, Transcript level of *RID* in the NIL-*S44* and NIL-*S44*-t*RID*. Pollen fertility (%) is presented in the columns. n = 3 independent biological replicates. In **b**, **d**, **e**, **g**, **i** and **j**, Data are presented as mean ± SD, ****$P<0.0001$ by one-way ANOVA, Tukey's HSD test. ns, not significant. For all *P* values, see Source Data.



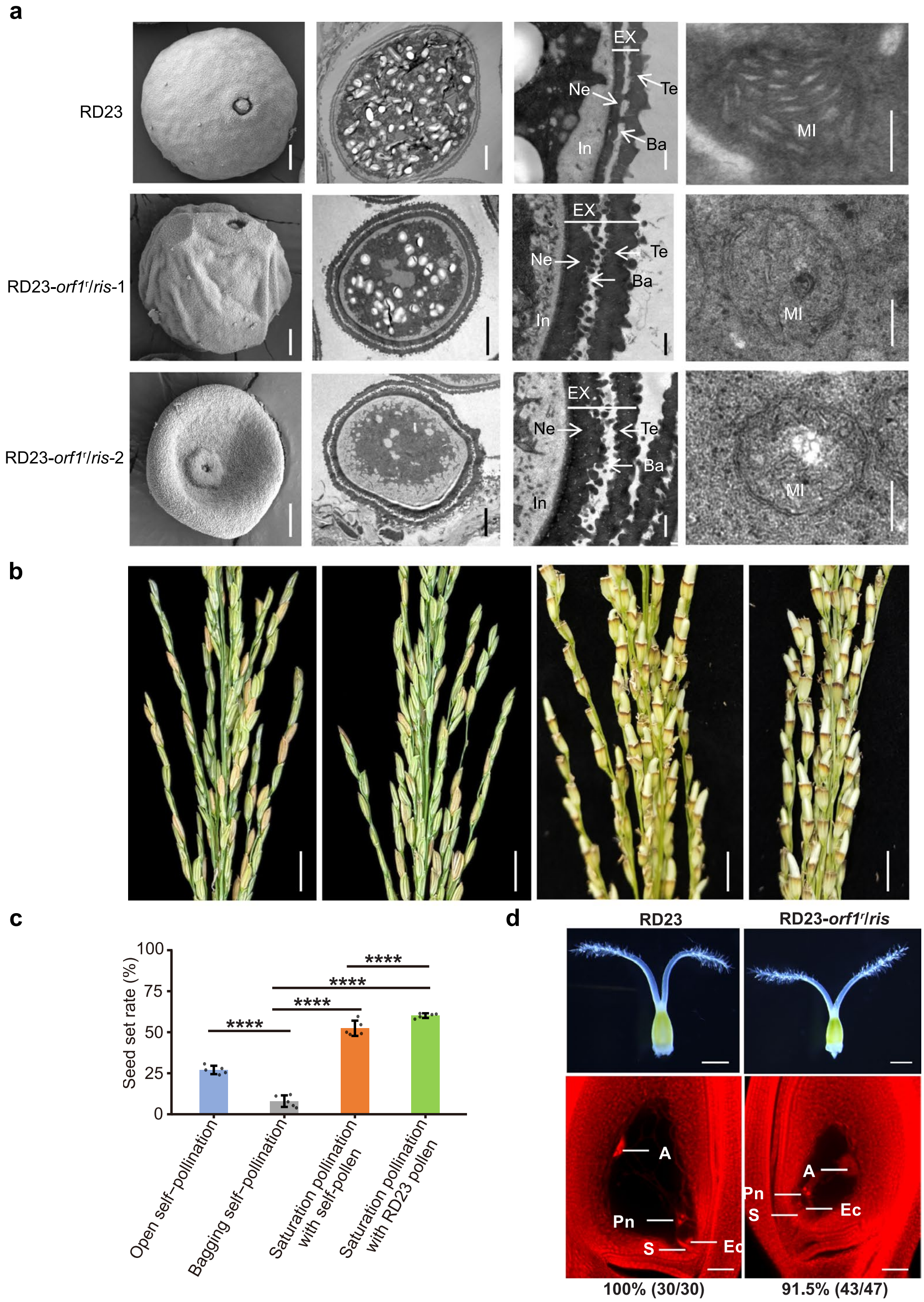
Extended Data Fig. 4 | *RIR* as a restorer rescues RD23-type pollen in a gametophytic manner. a, Two target sites in *RIR* are indicated by red arrowheads. The PAM sequences are highlighted in red letters, and inserted bases are marked by blue letters. **b**, A *RIR* genomic fragment (including a 2.1 kb promoter and 1.6 kb 3′ UTR) from NIL-*S44* was used for genetic complementation test. **c**, Pollen fertility of T_1 transgenic *RIR* plants in the F_1 background. Data are presented as mean ± SD. The number of different genotype plants notes in the column. **d**, Position of two editing targets in *RID* and *RIR* are shown by red arrowhead. Deletion bases are marked by red dotted lines. **e**, Pollen fertility of *rid rir* double mutant in the T_1 generation, Data are mean ± SD (n = 5 plants). In **c** and **e**, $^{****}P<0.0001$ by one-way ANOVA, Tukey's HSD test. ns, not significant. For all *P* values, see Source Data.



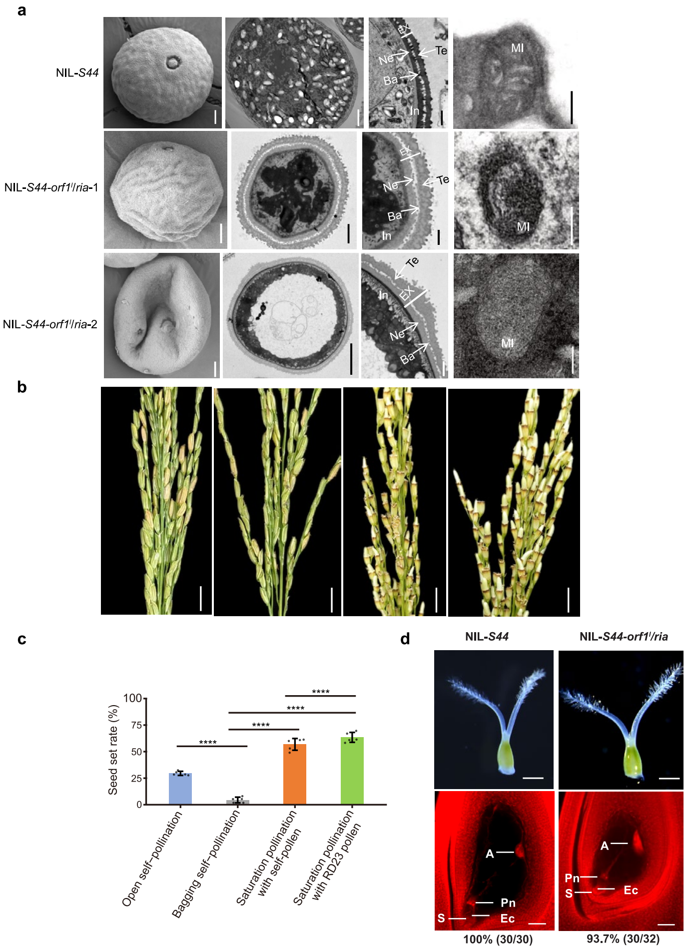
Extended Data Fig. 5 | *ORF1*^{R}/*RIS* is a pollen-essential gene. a and **b**, Pollen fertility in transgenic lines with *ORF1*^{R}/*RIS* genomic fragment in both F$_1$ and NIL-*S44* backgrounds. Introduction of *ORF1*^{R}/*RIS* does not affect hybrid pollen fertility, compared with non-transgenic plants. **c**, The target site in *ORF1*^{R}/*RIS* is indicated by red arrowhead. The PAM sequence is highlighted in red letters, and deletion bases are marked by blue dotted lines. **d**, Mutations in *ORF1*^{R}/*RIS* result in smaller and crushed stamens, partial pollen abortion and high spikelet sterility. Scale bar, 500 µm for stamen, 100 µm for pollen fertility and 5 cm for spikelet fertility. **e**, Quantitative analysis of pollen fertility and spikelet fertility in the two knockout *orf1*^{r}/*ris* mutants. In **a**, **b** and **e**, the data represent the mean ± SD (n = 5 plants). ****$P<0.0001$ by one-way ANOVA, Tukey's HSD test. ns, not significant. For all *P* values, see Source Data.



Extended Data Fig. 6 | *ORF1^L/RIA* functions in the pollen development. **a** and **b**, Pollen fertility in F$_1$ and RD23 carrying *ORF1^L/RIA* transgene in T$_0$ and T$_1$ generations, respectively. No significant difference in pollen fertility is observed between the transgenic lines harboring *ORF1^L/RIA* and non-transgenic lines. **c**, The target site in *ORF1^L/RIA* is indicated by a red arrowhead. The PAM sequence is highlighted in red, and insertion base is marked by a blue letter. **d**, Mutation in *ORF1^L/RIA* leads to shorter and underdeveloped stamens, and reduced pollen fertility. Scale bar, 500 µm for stamens, 100 µm for pollen fertility. **e**, The *orf1^l/ria* mutant shows high spikelet sterility. Scale bar, 5 cm. **f**, Quantification of pollen and spikelet fertility in NIL-*S44* and NIL-*S44-orf1^l*. **g**, Pollen germination rate of two parents, F$_1$, *orf1^r/ris* and *orf1^l/ria*. In **a**, **b**, **f** and **g**, data are mean ± SD (n = 5 plants). **** $P<0.0001$. ns, not significant (one-way ANOVA, Tukey's HSD test). For all *P* values, see Source Data.

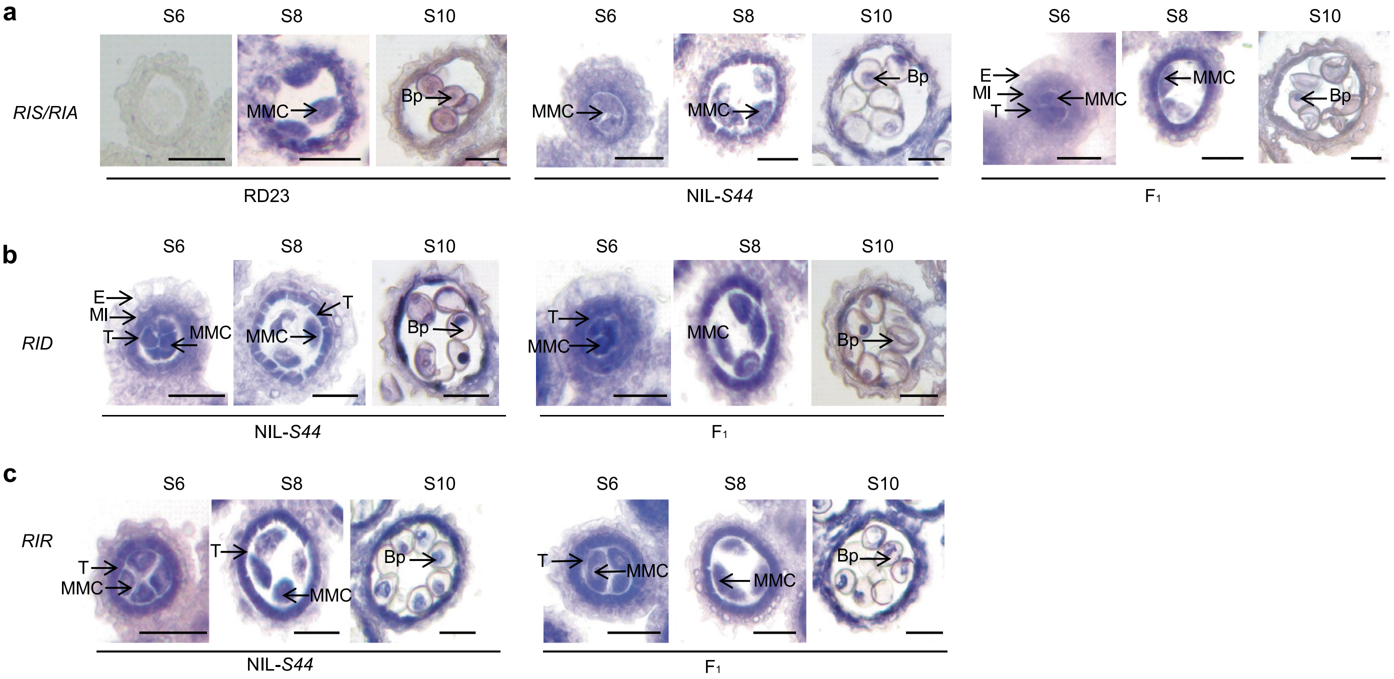


Extended Data Fig. 7 | Phenotype characterization of the *orf1*^{r}/*ris* mutant. **a**, Observation of pollen grains from RD23 and the *orf1*^{r}/*ris* mutant. The first column presents the morphology of mature pollen grains as analyzed by SEM, with representative samples from the *orf1*^{r}/*ris* mutant including round and wrinkled (middle), and irregular shaped pollen grains (bottom). Scale bar, 5 µm. The second column exhibits the transverse section of mature pollen grains from RD23 and the *orf1*^{r}/*ris* mutant as observed by TEM. Pollen grains from RD23 are filled with starch granules (top), while few starch granules are observed in the *orf1*^{r}/*ris* mutant (middle and bottom). Scale bar, 5 µm. The third column illustrates the mature pollen wall from RD23 and the *orf1*^{r}/*ris* mutant by TEM. The pollen wall is thin and uniform in RD23 (top), whereas the pollen wall exhibits uneven thickening in *orf1*^{r}/*ris* mutant (middle and bottom). Scale bar, 500 nm. The final column depicts normal mitochondria in RD23 and abnormal mitochondria in the RD23-*orf1*^{r}/*ris* mutant, Scale bar, 200 nm. n = 5 biological samples. Ex, exine; In, intine; Te, tectum; Ne, nexine; Ba, bacula; Ne, nexine; MI, mitochondria. **b**, Saturation pollination in the *orf1*^{r}/*ris* mutant. A significant difference in seed set rate is observed among open self-pollination, bagged self-pollination, saturation pollination with *orf1*^{r}/*ris* pollen, and saturation pollination with RD23 pollen (from left to right). Scale bar, 5 cm. **c**, Quantitative analysis of seed set rate following saturation pollination assay in the *orf1*^{r}/*ris* mutant. Data are mean ± SD (n = 5 plants), *****P*<0.0001 by one-way ANOVA, Tukey's HSD test. For all *P* values, see Source Data. **d**, Ovary and mature embryo sac phenotype. The number indicates fertility of the embryo sac. n = 5 spikelets for ovary observation, n = 30 embryo sacs for RD23 and n = 47 embryo sacs for RD23-*orf1*^{r}/*ris*. Scale bar, 500 µm for ovary and 50 µm for embryo sacs. A, antipode; Pn, polar nuclei; Ec, egg cell; S, synergic.

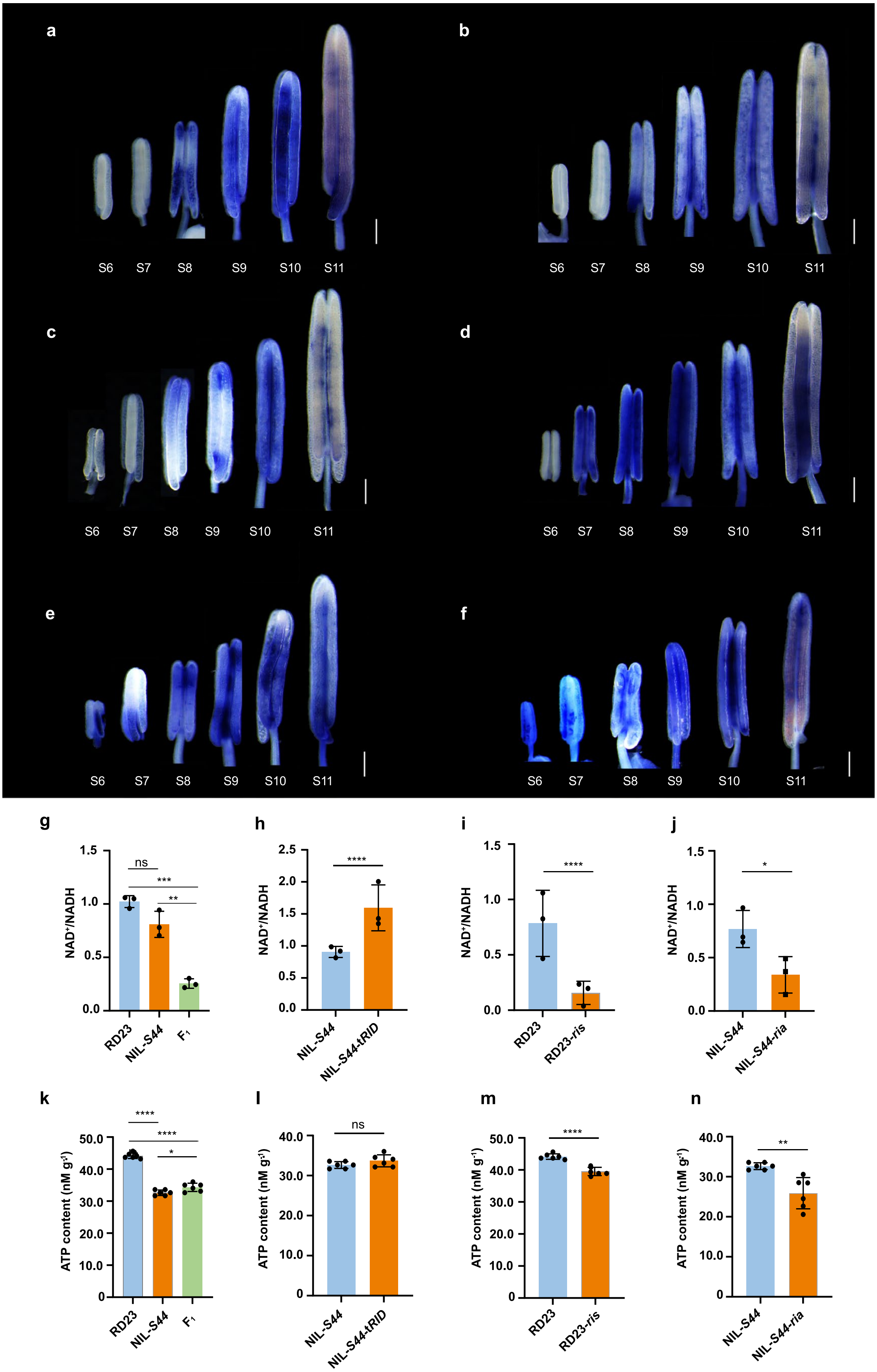


Extended Data Fig. 8 | Phenotype characterization of the *orf1*^{l}*/ria* mutant. **a**, Observation of pollen grains from NIL-*S44* and the *orf1*^{l}*/ria* mutant. The first column presents normal and round mature pollen in NIL-*S44* (top), and small wrinkle (middle) and mis-shaped grains (bottom) in the *orf1*^{l}*/ria* mutant as analyzed by SEM. Scale bar, 5 µm. The second column shows transverse section of mature pollen grains from NIL-*S44* and the *orf1*^{l}*/ria* mutant by TEM. Pollen grains from NIL-*S44* are filled with abundant starch granules (top), while few or no starch granules are observed in the *orf1*^{l}*/ria* mutant (middle and bottom). Scale bar, 5 µm. The third column illustrates the mature pollen wall from NIL-*S44* and the *orf1*^{l}*/ria* mutant by TEM. The pollen wall in NIL-*S44* is thin and uniform, whereas an irregularly thickened pollen wall is observed in the *orf1*^{l}*/ria* mutant. Scale bar, 500 nm. The final column depicts normal mitochondria in NIL-*S44* and abnormal mitochondria in the *orf1*^{l}*/ria* mutant. n = 5 biological samples. Scale bar, 200 nm. Ex, exine; In, intine; Te, tectum; Ne, nexine; Ba, bacula; Ne, nexine; MI, mitochondria. **b**, Saturation pollination in the *orf1*^{l}*/ria* mutant. The seed rates of open self-pollination, bagged self-pollination, saturation pollination with *orf1*^{l}*/ria* pollen, and saturation pollination with RD23 pollen (from left to right) are showed. Scale bar, 5 cm. **c**, Quantitative analysis of the seed set rate following saturation pollination assay in the *orf1*^{l}*/ria* mutant. Data are mean ± SD (n = 5 plants). **** $P<0.0001$ by one-way ANOVA, Tukey's HSD test. For all *P* values, see Source Data. **d**, Ovary and mature embryo sac phenotype. The number indicates fertility of the embryo sac. n = 5 spikelets for ovary observation, n = 30 embryo sacs for NIL-*S44* and n = 32 embryo sacs for NIL-*S44*-*orf1*^{l}*/ria*. Scale bar, 500 µm for ovary and 50 µm for embryo sacs. A, antipode; Pn, polar nuclei; Ec, egg cell; S, synergic.

https://doi.org/10.1038/s41477-026-02223-w



Extended Data Fig. 9 | RNA *in situ* hybridization analysis of *RIS*/*RIA*, *RID and RIR*. **a**, The expression patterns of *RIS* and *RIA* at S6, S8 and S10 stages during the pollen development in RD23, NIL-*S44* and their F_1 hybrids are depicted. S6 stage in RD23 is hybridized with a sense probe as a negative control; other anther sections are hybridized with antisense probes. Scale bar, 50 µm. **b**, The expression pattern of *RID* at S6, S8 and S10 stages in NIL-*S44* and F_1 plants. Scale bar, 50 µm. **c**, The expression pattern of *RIR* at S6, S8 and S10 stages in NIL-*S44* and F_1 plants. Scale bar, 50 µm. S6, microspore mother cell stages; S8, meiotic stages; S10, bicellular stages. MMC, microspore mother cell; ML, middle layer; T, tapetum; Bp, bicellular pollen. In **a**, **b**, and **c**, n = 5 biological samples.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | *S44* regulates mitochondrial activity. (a-f), Detecting superoxide anion production by NBT staining from stages 6 to 11. **a**, RD23; **b**, NIL-*S44*; **c**, F_1; **d**, NIL-*S44*-t*RID*; **e**, RD23-*ris*; **f**, NIL-*S44*-*ria*. n = 5 spikelets. Scale bar, 200 µm. **(g-j)**, NAD^+/NADH ratio in RD23, NIL-*S44*, and their F_1 hybrids, NIL-*S44* with exogenious t*RID* as well as *ris/ria* mutants. Data are presented as mean ± SD (n = 3 independent biological replicates). **(k-n)**, ATP content in RD23, NIL-*S44*, and their F_1 hybrids, NIL-*S44* with exogenious t*RID* and *ris/ria* mutants. Data are presented as mean ± SD (n = 6 independent biological replicates). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ by one-way ANOVA, Tukey's HSD test, respectively; ns, not significant. For all *P* values, see Source Data.

nature portfolio

Corresponding author(s): Dayun Tao, Yu Zhang, Lianguang Shang

Last updated by author(s): 18/11/25

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	Semi-thin sections were examined using a Leica RM2235. Scanning electron microscopy (SEM) was performed on a ZEISS GeminiSEM 300. Transmission electron microscopy (TEM) was conducted with a JEOL JEM1400. Quantitative PCR (qPCR) was carried out using a Biorad CFX96. Subcellular localization was assessed with a ZEISS LSM780 confocal microscope. GUS and NBT staining was visualized under an Olympus BX51 and SZX16 microscope. RNA in situ hybridization was performed on an Olympus BX51 microscope. Luciferase (LUC) assays were imaged using a Tanon-5200 Chemiluminescent imaging system. NAD+/NADH was detected by BioTek SYNERGY HTX. ATP content was measured using PerkinElmer Enspire. Embryo sac was observed by Confocal microscope (Fluoview, FV1000, Olympus, Japan).
Data analysis	Data were analyzed using WPS ,GraphPad Prism 9, R (version 4.4.0; R Core Team, 2024) for statistical analysis, DNASTARLasergene 11 for sequence blast, Primer Premier 5 for primer design, Snapgene 6.0.2, Mafft and DNAsp for sequence blast and haplotype analysis, Biopython for translating CDS sequences, MUMmer for locating the target genes positions, samtools for extracting gene sequence. R package rwordmap for drawing map, custom Python scripts for quantitative analyses.

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

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data are available within the article and its supplementary information. Gene sequences have been deposited in NCBI with accession numbers PQ723328, PQ723325, PQ723326, and PQ723327.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

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

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	Sample sizes were determined based on research in the field and previous publications on similar experiments (https://doi.org:10.1016/j.cell.2023.06.023; https://doi.org:10.1126/science.aar4279)
Data exclusions	No data was excluded from the analysis.
Replication	All experiments were successfully repeated at least three times. The number of replicates is indicated for each experiment.
Randomization	All samples were grown in a random order in the field.
Blinding	Assessments of pollen fertility and spikelet fertility were performed blindly, without knowledge of genotype information.

Behavioural & social sciences study design

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

Study description	NA
Research sample	NA
Sampling strategy	NA
Data collection	NA

Timing	NA
Data exclusions	NA
Non-participation	NA
Randomization	NA

Ecological, evolutionary & environmental sciences study design

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

Study description	NA
Research sample	NA
Sampling strategy	NA
Data collection	NA
Timing and spatial scale	NA
Data exclusions	NA
Reproducibility	NA
Randomization	NA
Blinding	NA

Did the study involve field work? ☐ Yes ☒ No

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	NA
Validation	NA

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	NA
Authentication	NA

Mycoplasma contamination	NA
Commonly misidentified lines (See ICLAC register)	NA

Palaeontology and Archaeology

Specimen provenance	NA
Specimen deposition	NA
Dating methods	NA

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	NA

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	NA
Wild animals	NA
Reporting on sex	NA
Field-collected samples	NA
Ethics oversight	NA

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NA
Study protocol	NA
Data collection	NA
Outcomes	NA

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	All transgenic lines, mutants, and near-isogenic lines used in this study were constructed in our laboratory.
Novel plant genotypes	Mutants and transgenic lines were generated via Agrobacterium-mediated callus transformation. For complementation tests, genomic fragments of JIN, HUO, and SHUI were cloned into the binary vector pCAMBIA1305.1 to create JIN-COM, HUO-COM, and SHUI-COM constructs, which were then introduced into RD23, NIL-S44, and F1 plants to generate transgenic plants. The genomic sequence of MU was inserted into pCAMBIA1300 to create the MU-COM construct, which was transformed into F1 and NIL-S44 plants. The promoter sequences of JIN, MU, HUO, and SHUI were fused to pCAMBIA1305.1 and transformed into RD23 callus to generate JINpro::GUS, MUpro::GUS, HUOpro::GUS, and SHUIpro::GUS plants. For mutant lines, the target sequences of JIN, MU, HUO, and SHUI were individually fused into the pBWA (V)H vector and introduced into RD23, NIL-S44, or F1 plants, resulting in jin, mu, huo homozygous mutants, heterozygous shui mutant, and homozygous huoshui mutant. The Jin/mu mutant in the F1 background was generated by crossing RD23-jin and NIL-S44-mu. For mapping S44, NIL was developed from crosses derived from O. longistaminata as the donor parent and RD23 as the recurrent parent. Near-isogenic lines were derived from crosses between AA genome species as donor parents and DJY1 as the recurrent parent.
Authentication	Transgenic lines were authenticated by PCR using HPT gene and target gene-specific primers. CRISPR-Cas9 editing mutants were confirmed by PCR and sequencing. Near-isogenic lines were validated by PCR and SNP 6k chip analysis.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	NA
Files in database submission	NA
Genome browser session (e.g. UCSC)	NA

Methodology

Replicates	NA
Sequencing depth	NA
Antibodies	NA
Peak calling parameters	NA
Data quality	NA
Software	NA

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	NA
Instrument	NA
Software	NA
Cell population abundance	NA
Gating strategy	NA

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	NA
Design specifications	NA
Behavioral performance measures	NA

Acquisition

Imaging type(s)	NA
Field strength	NA
Sequence & imaging parameters	NA
Area of acquisition	NA

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	NA
Normalization	NA
Normalization template	NA
Noise and artifact removal	NA
Volume censoring	NA

Statistical modeling & inference

Model type and settings	NA
Effect(s) tested	NA

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	NA
Correction	NA

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis