An asymmetric allelic interaction drives allele transmission bias in interspecific rice hybrids

Yongyao Xie1,2,3,4,7, Jintao Tang1,2,4,7, Xianrong Xie1,3,4, Xiaojuan Li1,2,4, Jianle Huang1,4, Yue Fei1,2,4, Jingluan Han1,2,4, Shuifu Chen1,3,4, Huiwu Tang1,3,4, Xiucai Zhao1,3,4, Dayun Tao5, Peng Xu5,6, Yao-Guang Liu1,3,4 & Letian Chen1,2,3,4

Hybrid sterility (HS) between Oryza sativa (Asian rice) and O. glaberrima (African rice) is mainly controlled by the S1 locus. However, our limited understanding of the HS mechanism hampers utilization of the strong interspecific heterosis. Here, we show that three closely linked genes (S1A4, S1TPR, and S1A6) in the African S1 allele (S1-g) constitute a killer-protector system that eliminates gametes carrying the Asian allele (S1-s). In Asian-African rice hybrids (S1-gS1-s), the S1TPR-S1A4-S1A6 interaction in sporophytic tissues generates an abortion signal to male and female gametes. However, S1TPR can rescue S1-g gametes, while the S1-s gametes selectively abort for lacking S1TPR. Knockout of any of the S1-g genes eliminates the HS. Evolutionary analysis suggests that S1 may have arisen from newly evolved genes, multi-step recombination, and nucleotide variations. Our findings will help to overcome the interspecific reproductive barrier and use Asian-African hybrids for increasing rice production.
Hybrid sterility (HS) is a major mechanism of postzygotic reproductive isolation, which can limit reproduction among divergent populations during speciation and thus contribute to the maintenance of species identity.1,2 The classic Bateson–Dobzhansky–Muller (BDM) model3 attributes postzygotic reproductive isolation to an incompatible genetic interaction between divergent alleles of at least two loci,4,5 although mounting evidence suggests that HS caused by the interaction of alleles comprising multiple genes at a single locus also fits the BDM model.6–12 The HS loci act as selfish genetic elements, which can eliminate the gametes carrying competing alleles, thus gaining a transmission advantage over other alleles, similar to meiotic drive. This can result in distorted segregation and non-Mendelian transmission of the alleles in the progeny.6–13 The killer–protector system and killer meiotic driver models have been proposed to explain the biased allele transmission of HS in eukaryotes.14–16 In this model, the killer gene (usually functioning in the sporophyte) produces a detrimental sterility signal that indiscriminately kills all meiotic cells16. The protector gene (usually tightly linked to the killer gene) functions in the gamete to eliminate this detrimental effect and rescue the gamete that contains the correct allele of the protector gene. Therefore, gametes carrying the protector gene survive; gametes lacking the protector gene die.16

The Oryza genus comprises 21 wild species and 2 cultivated species, Asian rice (O. sativa L., including the subspecies japonica and indica) and African rice (O. glaberrima Steud.). The broad genetic diversity within Oryza can contribute to heterosis, also known as hybrid vigor, in which the hybrid performs better than its parental inbred lines.17,18 However, HS of different forms (e.g., male sterility, female sterility, and both male and female sterilities) is common between different species and subspecies in the Oryza genus and this HS hinders the ability of plant breeders to use the strong heterosis in the production of high yield, robust hybrid rice. Therefore, understanding the mechanisms of the HS could enable breeding of improved hybrid rice. A number of HS loci have been genetically studied in Oryza19,20 and several HS genes have been cloned from various loci, including the single-locus HS loci S1, S3, HSA1, S7, S1, Sc, qSHMS7, ESA1, and the two-locus DPL1/DPL2, S27/S28, and DGS1/DGS2 pairs.8,12,14,15,21–29 Despite extensive studies of these loci, our understanding of the molecular mechanisms governing HS and its effect on genome evolution remains limited.

S1 is a typical single-locus-type HS locus that affects hybrids produced by crossing African and Asian rice.23,27,30–32 In these Asian–African rice hybrids, male and female gametes carrying the Asian rice S1 allele (S1-s) are selectively aborted, leading to very strong preferential transmission of the African rice S1 allele (S1-g), which reaches an allele frequency of ca. 0.95 in F2 progeny.27 Recently, we cloned OgTPRI, the peptidase-encoding causal gene of the HS effect of the S1 locus; another S1-related gene, SSP, also encoding a peptidase, was later identified by Koide and colleagues.33 Despite these advances, the molecular mechanism governing the selective abortion and survival of gametes with different alleles in interspecific and intraspecific hybrids in plants remains poorly understood, and the evolutionary origins of HS loci and their relationship to speciation remain largely unknown. Here, we identify additional HS-related genes at S1, describe the S1 tripartite gamete killer–protector system, and explore the evolutionary relationship of this complex locus with the allopatric speciation of the related Oryza species.

Results
The S1A4 gene is required for S1 HS. Our previous study revealed the existence of structural variation between the African rice S1 allele (S1-g) and the Asian rice S1 allele (S1-s).27 We found that OgTPRI (hereafter named SITPR) at S1-g is required for SI-mediated HS. Moreover, S1-s contains an SITPR allele, named SITP, which is a truncated form of SITPR due to a premature stop codon caused by a single-nucleotide mutation (Fig. 1a).27 According to our sequence analysis, besides SITPR, the S1-g region contains six African rice-specific putative genes, SIA1–SIA6 (Fig. 1a). We determined the expression profiles of these genes in a near-isogenic line carrying S1-g (NIL-g) and its recurrent parental line RP-s (japonica rice carrying S1-s), and their F1 plants. This showed that SIA2–SIA6 were transcriptionally active in anthers and young panicles (Supplementary Figs. 1a and 2). Moreover, SITPR was expressed at high levels in the microspores (Supplementary Fig. 1b).

To test whether SITPR is sufficient to cause gamete abortion, we generated transgenic plants that contained an SITPR transgene (SITPR+) in the RP-s background. The T0 plants hemizygous for SITPR+ (SITPR+), where dash indicates absence of the T-DNA/transgene in the chromosome site) had fully fertile pollen and spikelets (Fig. 1b), indicating that SITPR is not sufficient to induce HS. These results suggest that S1 HS requires other O. glaberrima-specific S1 component(s) in addition to SITPR.

To investigate whether any of the anther- and panicle-expressed S1 genes (SIA2–SIA6) are involved in S1 HS, we used CRISPR/Cas9 to individually knock out their functions in NIL-g. All the obtained knockout mutants (sia2–sia6) showed normal male and female fertility, which indicates that these genes are not essential for gamete development (Supplementary Fig. 3a and Supplementary Table 1). If a specific gene is required for S1 HS, knocking it out in NIL-g should create a neutral allele that does not show HS in crosses with RP-s. When these mutants were crossed with RP-s, the mutant F1 plants containing the sia2, sia3, or sia5 mutant alleles were semi-sterile (ca. 50% sterile pollen and spikelets) (Fig. 1b), indicating that SITPR is not sufficient to induce HS. These results suggest that S1 HS requires other O. glaberrima-specific S1 component(s) in addition to SITPR.

SIA4-SITPR-SIA6 constitutes a killer–protector system. To understand whether SIA4, SITPR, and SIA6 are sufficient to kill gametes, we transformed RP-s with transgenes containing different combinations of these three genes, then looked for sterility and distortion of segregation of the transgenes. If the killing process requires all three genes, gamete abortion would not occur in plants carrying only one or two of these components. Indeed, all the plants transformed with one (SITPR+, SIA4+, or SIA6+), or two transgenes (the SIA4–SIA6 and SITPR-SIA6 transgenes carrying two genes, or the pyramided transgenes in the F1 of SIA4×SITPR), all in a hemizygous state, did not show the typical semi-sterile phenotypes of pollen and spikelets (Fig. 1c and Supplementary Fig. 5). In contrast, when the three transgenes were pyramided together by a cross between the homozygous SIA4–SIA6 and SITPR lines, the pollen and spikelets of the
transgenic F1 plants were semi-sterile (Fig. 1d). Furthermore, abnormal embryo sacs were observed in the F1 plants of S1A4–S1A6\textsubscript{t} × S1TPR\textsubscript{t} (Supplementary Fig. 6), consistent with previous reports in Asian–African rice hybrids\textsuperscript{23,31}. These results suggested that these three genes are necessary and sufficient to constitute a gamete-killer system.

We further analyzed the segregation of the SIA4–SIA6\textsubscript{t} and SITPR\textsubscript{t} transgenes. The segregation of SITPR\textsubscript{t} and SIA4–SIA6\textsubscript{t} in their T\textsubscript{1} generations fit the 1:2:1 ratio, consistent with the hypothesis that the HS requires all three components (Fig. 1e, Supplementary Tables 3 and 4). As expected, the F1 of the SIA4–SIA6\textsubscript{t} × SITPR\textsubscript{t} cross (sporophytic genotype SIA4–SIA6\textsubscript{t}–/A1\textsuperscript{T0} of TPR\textsubscript{t} (T\textsuperscript{–}))

| Segregation rate | 25.4 (T\textsuperscript{T}) | 23.7 (T\textsuperscript{T}) | 23.8 (T\textsuperscript{T}) |
|------------------|-----------------------------|-----------------------------|-----------------------------|
| TPR\textsubscript{t} (T\textsuperscript{T}) | 48.4 (T\textsuperscript{T}) | 51.0 (T\textsuperscript{T}) | 50.9 (T\textsuperscript{T}) |
| A4–A6\textsubscript{t} × TPR\textsubscript{t} (F\textsubscript{2}) | 26.2 (T\textsuperscript{T}) | 25.3 (T\textsuperscript{T}) | 25.3 (T\textsuperscript{T}) |
| A4–A6\textsubscript{t} (T\textsuperscript{1}) | 24.2 (T\textsuperscript{T}) | 24.2 (T\textsuperscript{T}) | 24.2 (T\textsuperscript{T}) |

| **No. of plants** | 248 | 190 | 281 | 281 |
|-------------------|-----|-----|-----|-----|
| χ\textsuperscript{2} | 0.29 | 0.18 | 0.20 | 397.7*** |

| Segregation rate | 25.4 (T\textsuperscript{T}) | 23.7 (T\textsuperscript{T}) | 23.8 (T\textsuperscript{T}) |
|------------------|-----------------------------|-----------------------------|-----------------------------|
| TPR\textsubscript{t} (T\textsuperscript{T}) | 48.4 (T\textsuperscript{T}) | 51.0 (T\textsuperscript{T}) | 50.9 (T\textsuperscript{T}) |
| A4–A6\textsubscript{t} × TPR\textsubscript{t} (F\textsubscript{2}) | 26.2 (T\textsuperscript{T}) | 25.3 (T\textsuperscript{T}) | 25.3 (T\textsuperscript{T}) |
| A4–A6\textsubscript{t} (T\textsuperscript{1}) | 24.2 (T\textsuperscript{T}) | 24.2 (T\textsuperscript{T}) | 24.2 (T\textsuperscript{T}) |

| **No. of plants** | 248 | 190 | 281 | 281 |
|-------------------|-----|-----|-----|-----|
| χ\textsuperscript{2} | 0.29 | 0.18 | 0.20 | 397.7*** |

| Segregation rate | 25.4 (T\textsuperscript{T}) | 23.7 (T\textsuperscript{T}) | 23.8 (T\textsuperscript{T}) |
|------------------|-----------------------------|-----------------------------|-----------------------------|
| TPR\textsubscript{t} (T\textsuperscript{T}) | 48.4 (T\textsuperscript{T}) | 51.0 (T\textsuperscript{T}) | 50.9 (T\textsuperscript{T}) |
| A4–A6\textsubscript{t} × TPR\textsubscript{t} (F\textsubscript{2}) | 26.2 (T\textsuperscript{T}) | 25.3 (T\textsuperscript{T}) | 25.3 (T\textsuperscript{T}) |
| A4–A6\textsubscript{t} (T\textsuperscript{1}) | 24.2 (T\textsuperscript{T}) | 24.2 (T\textsuperscript{T}) | 24.2 (T\textsuperscript{T}) |

| **No. of plants** | 248 | 190 | 281 | 281 |
|-------------------|-----|-----|-----|-----|
| χ\textsuperscript{2} | 0.29 | 0.18 | 0.20 | 397.7*** |
SITPR\textsuperscript{1,2} was semi-sterile. Moreover, the segregation of SITPR\textsuperscript{4} was severely distorted, with most (75.8\%) of the F\textsubscript{2} individuals carrying homozygous SITPR\textsuperscript{4} and the rest containing hemizygous SITPR\textsuperscript{4} (Fig. 1e and Supplementary Table 5). This indicated that the gametophytic genotype is what matters for the transmission advantage of SITPR\textsuperscript{4}. However, the segregation ratio of SIA4–SIA6\textsuperscript{d} in the F\textsubscript{2} progeny of SIA4–SIA6\textsuperscript{d} × SITPR\textsuperscript{4} fit the 1:2:1 ratio, similar to the segregation of SITPR\textsuperscript{4} and SIA4–SIA6\textsuperscript{d} in their T\textsubscript{1} generations (Fig. 1e, Supplementary Tables 3–5). Since SIA4–SIA6\textsuperscript{d} and SITPR\textsuperscript{4} are not linked, this is consistent with SITPR\textsuperscript{4} in the gametophyte providing protection from the sterile effect of SIA4–SIA6\textsuperscript{d} and SITPR\textsuperscript{4} in the sporophyte. In contrast to the situation in the native S\textsubscript{1}-g allele, here the transgenes are unlinked and the distortion of SITPR\textsuperscript{4} segregation does not affect the normal SIA4–SIA6\textsuperscript{d} segregation. Moreover, it is consistent with the effect of SIA4–SIA6\textsuperscript{d} and SITPR\textsuperscript{4} acting in the sporophyte.

These observations prompted us to propose a model using a Punnett square to explain the results of this genetic analysis (Fig. 1f). We reasoned that if SITPR\textsuperscript{4} and SIA6 might act together in the sporophyte to produce a detrimental sterility signal that kills the male and female gametes. However, the SITPR\textsuperscript{4} transgene alone is capable of protecting the corresponding gametes (with or without SIA4–SIA6\textsuperscript{d}) from the sterility, thereby rescuing the gametes harboring SITPR\textsuperscript{4} and leading to a severely distorted segregation favoring SITPR\textsuperscript{4} in the F\textsubscript{2} progeny but not affecting segregation of the unlinked SIA4–SIA6\textsuperscript{d}.

SITPR\textsuperscript{4} is required for killer and protector function. If SITPR\textsuperscript{4} in the gamete is sufficient to provide the protector function at the SI locus, gametes containing the S\textsubscript{1}-s allele should be partially rescued by the SITPR\textsuperscript{4} transgene. To test this, we crossed the hemizygous SITPR\textsuperscript{4} line with NIL-g to produce F\textsubscript{1} hybrids (SI-gS1-s/SITPR\textsuperscript{4}–) that contain the S\textsubscript{1}-g allele in the sporophytic tissue (and thus should activate HS), but will segregate the SITPR\textsuperscript{4} transgene in the gametes. Indeed, the F\textsubscript{1}, and F\textsubscript{2} plants that are heterozygous for S\textsubscript{1} but lack the SITPR\textsuperscript{4} transgene were semi-sterile (~50\%). By contrast, the fertilities of the pollen and spikels of the F\textsubscript{1} and F\textsubscript{2} plants that are heterozygous for S\textsubscript{1} and carry the SITPR\textsuperscript{4} transgene in a hemizygous condition increased to ~75\%. Moreover, the S\textsubscript{1}-heterozygotes with homozygous SITPR\textsuperscript{4} were fully fertile (Fig. 2a and Supplementary Table 6), indicating that the SITPR\textsuperscript{4} allows transmission of the S\textsubscript{1}-s allele carrying this transgene. The F\textsubscript{2} plants homozygous for S\textsubscript{1}(S\textsubscript{1}-S\textsubscript{1}-s or S\textsubscript{1}-gS1-g) were fully fertile regardless of whether they contained SITPR\textsuperscript{4} or not (Fig. 2a and Supplementary Table 6). This suggests that SITPR\textsuperscript{4} indeed rescued the gametes containing S\textsubscript{1}-s in a gametophytic manner.

Based on these findings and our hypothesis, we proposed that in the F\textsubscript{1} plants (SI-gS1-s/SITPR\textsuperscript{4}–) of SITPR\textsuperscript{4} × NIL-g, the S\textsubscript{1}-g allele produced a sterility signal in the sporophytic cells. However, all the S\textsubscript{1}-g-containing gametes survive due to the presence of the endogenous SITPR\textsuperscript{4}, and the S\textsubscript{1}-s-containing gametes are aborted unless they carried the SITPR\textsuperscript{4} transgene. The segregation ratios for the genotypes of S\textsubscript{1}(S\textsubscript{1}-gS\textsubscript{1}-gS1-sS1-s) and SITPR\textsuperscript{4} (SITPR\textsuperscript{4}SITPR\textsuperscript{4}SITPR\textsuperscript{4}–––) were therefore predicted to be 4:8:4 in the F\textsubscript{2} population (Fig. 2b). Indeed, this segregation ratio perfectly fit the ratio from genetic analysis of the SITPR\textsuperscript{4} × NIL-g F\textsubscript{2} population (Fig. 2c and Supplementary Table 6).

To further verify that SITPR\textsuperscript{4} also participates in the killing process, we crossed the homozygous SITPR\textsuperscript{4} plants with the s1tpr mutant (the function of SITPR\textsuperscript{4} was completely knocked out, resulting in homozygous mutated S\textsubscript{1}-g\textsuperscript{m}\textsuperscript{27}). We reasoned that if the SITPR\textsuperscript{4} transgene can function together with the endogenous SIA4 and SIA6 (still present at the S\textsubscript{1}-g\textsuperscript{m} allele) genes to kill the gametes, gamete abortion would be observed in the resultant F\textsubscript{1} hybrids. As expected, the F\textsubscript{1} plants exhibited semi-sterile pollen and spiklets, like those of RP-s × NIL-g (Fig. 3a–c). Since SITPR\textsuperscript{4} makes up part of the killer system, and has another role in the protection of SITPR\textsuperscript{4}-containing gametes, we therefore hypothesized a segregation model, in which the hybrids (S\textsubscript{1}-g\textsuperscript{m}S\textsubscript{1}-s/SITPR\textsuperscript{4}–) produce a detrimental sterility signal (in a sporophytic manner) to kill the gametes carrying S\textsubscript{1}-s or S\textsubscript{1}-g\textsuperscript{m}, but the SITPR\textsuperscript{4}-containing gametes survive (Fig. 3d). As expected, in the F\textsubscript{2} population of the SITPR\textsuperscript{4} × s1tpr the segregation ratio of the S\textsubscript{1} genotypes (S\textsubscript{1}-g\textsuperscript{m}S\textsubscript{1}-g\textsuperscript{m}S\textsubscript{1}-sS1-s) fit the 1:2:1 ratio, and the SITPR\textsuperscript{4} genotypes had a significantly distorted segregation ratio (Fig. 3e and Supplementary Table 7). These results also suggested that the abortion of the S\textsubscript{1}-s or S\textsubscript{1}-g\textsuperscript{m}-containing gametes in the hybrids is independent of the SITP\textsuperscript{4} gene at S\textsubscript{1}-s or other gametophytic responder genes.

SIA6-S1A4-S1TIPR\textsuperscript{4} is a tripartite complex in the nucleus. In this gamete killer–protector system, SIA6 encodes a smaller peptide with similar features to the SITPR peptide\textsuperscript{23,27}, and SIA4 encodes an uncharacterized protein of 261 amino acids without any putative conserved domains (Supplementary Fig. 7). To study the subcellular localization of these proteins, SIA4, SIA6, and SITPR were fused with the GFP gene sequence, respectively, and expressed in rice protoplasts. All these proteins localized in the nucleus (Fig. 4a), suggesting that they may interact with each other in the nucleus. Consistent with this, bimolecular fluorescence complementation (BiFC) and pull-down assays indicated that SIA4 interacted with SIA6 and SITPR respectively in the nucleus, but SITPR did not interact with SIA6 directly (Fig. 4b–d).

Possible biochemical effect of the SI-HS system. Given that the SI HS factors interact in the nucleus, we next examined their
The STIPPR rescues the gametes carrying S1s in hybrids. a Pollen and spikelet fertilities of the F1 plants derived from the cross between the hemizygous STIPPR (TPR, ss/T-) and NIL-g (gg/--) and various genotypes of the F2 segregants. Error bars indicate S.D. NA, not available. b A proposed model for the segregation behavior of the endogenous S1 alleles and TPR in the F2 plants derived from TPR x NIL-g. The F1 male and female gametes containing St-g and/or T are considered fertile, and those with S1s (ss/--) but lacking T are generally sterile. Thus, the expected segregation ratios for gg:ss:ss and TT:T--:-- are 4:4:1. The color codes are consistent with the genotype of the individuals in the F2 population. Black represents homozygous TPR; gray represents hemizygous TPR; white represents lacking TPR; green represents homozygous St-s; red represents heterozygous St; blue represents free homozygous St-g. The segregation rates of the S1 alleles and the transgene (T) in the analyzed F2 population fit the expected ratio. Source data of (a, c) are provided as a Source Data file

Evolutionary origin of the SI gamete killer–protector system. According to the BDM model, hybrid incompatibility is caused by a detrimental interaction between the divergent alleles of two independent lineages, which may have been derived from a recent common ancestor. To trace the evolution of the SI locus, we first used the STIPPR, STIP, SIA4, and SIA6 nucleotide coding sequences to perform a BLAST search for putative orthologs in the Poaceae sequences using the GenBank database (https://www.ncbi.nlm.nih.gov/), and constructed a phylogenetic tree of the
candidates. We found that the nucleotide coding sequences of SITPR, but not S1A4 and S1A6, were significantly similar to rice in the genomes of Zea mays, Sorghum bicolor, Brachypodium distachyon, Hordeum vulgare, Aegilops tauschii, Triticum aestivum, and Setaria italic (Supplementary Fig. 11). Notably, the orthologous gene (LOC101754700) in S. italic has the highest nucleotide identity (ca. 82%) with SITPR and SITP. Thus, the SITPR ortholog in S. italic is used as one of the appropriate outgroup references for analyzing the divergence of SITPR and SITP.

To further confirm the point of divergence in the SI locus of Oryza, we identified the SITPR, SITP, S1A4, and S1A6 sequences in the AA genome species of Oryza (O. meridionalis, O. longistaminata, O. barthii, O. rufipogon, O. glaberrima, and O. sativa) and several non-AA genome species, including O. officinalis (CC genome), O. rhizomatis (CC genome), O. eichingeri (CC genome), and O. minuta (BBC genome), which are all closely related to the AA genome Oryza species based on molecular evidence.33 These non-AA genome species were also used as outgroups in the divergent analysis of the gene. We found that only SITPR and/or SITP, but not S1A4 and S1A6, were present in these outgroup species, indicating that S1A4 and S1A6 likely newly evolved in the Oryza species with AA genomes (Supplementary Fig. 12 and Supplementary Data 1).

Since seven single-nucleotide polymorphisms (SNPs, sites 1–7) are present between the SITPR and SITP coding sequences (Supplementary Fig. 12)27, we further analyzed the patterns of these seven SNPs in 443 accessions of the AA-genome Oryza species and some other outgroup species (including S. italic) to trace the sequence divergence of the SITPR and SITP genes in these species. An SITPR-type allele (Allele 1) was detected in S. italic, O. officinalis, O. minuta, O. longistaminata, and O. meridionalis (Supplementary Fig. 12 and Supplementary Data 1), thus likely representing the primitive form of SITPR. Notably, the nucleotides at SNP site 7 (C to A variation causing the premature stop codon in SITP) were polymorphic not only among the AA genome species, but also among the analyzed CC genome species. In contrast, the polymorphisms of the SNP sites 1–6 were present only in AA genome species (Supplementary Fig. 12 and Supplementary Data 1), suggesting that SNP 7 in SITPR and SITP arose early in the evolution of the Oryza genus and that various alleles co-existed in the primitive Oryza gene pool.

On the basis of these detected SNP patterns, we identified at least fifteen SITP alleles (Alleles 1–1 to 1–15) that contained variations at the SNP sites 1, 3, 4, 6, and 7 in AA-genome Oryza species, including O. longistaminata, O. meridionalis, O. rufipogon, and O. sativa. All O. rufipogon accessions (68) and all O. sativa accessions (116) carried the one-gene SI haplotype containing SITP (haplotype). Furthermore, we found four SITPR alleles (Alleles 2–1 to 2–4) that carried variations at the SNP sites 2 and 5 in O. barthii and O. glaberrima (Supplementary Fig. 12).

The non-overlapping natural variations in the SITPR and SITP alleles between O. glaberrima and O. sativa suggested that at least two independent lineages evolved from the common ancestral lineage carrying the ancestral SITPR gene (Fig. 5a, Supplementary
Figs. 12 and 13). In one lineage, the resultant SITP variants (Alleles 1-1 to 1-16) passed a bottleneck and some (Alleles 1-3 to 1-8) were transmitted into O. rufipogon, resulting in the eventual fixation of two (Alleles 1-3 and 1-8) as the current S1-s in O. sativa; this allele is present in all populations of O. sativa (Fig. 5a, Supplementary Figs. 12 and 13). By contrast, S1-g likely evolved in another lineage (Fig. 5a, Supplementary Figs. 12 and 13). Given that SIA4 and SIA6 are absent in the CC and BBCC genomes of other Oryza species, we speculated that the SIA4-SITPR/SITP-SIA6 alleles might newly evolve in the AA-genome Oryza species. Consistent with this hypothesis, three types of SITP-containing structures with the genes SIA4 and/or SIA6, inserted upstream and downstream of SITP, respectively, were identified in O. meridionalis (Fig. 5a, Supplementary Figs. 12 and 13). Seven accessions had the SIA4-SITP structure (Alleles 1–9 to 1–11), five accessions had the SITP-SIA6 structure (Alleles 1–12 and 1–13), and two accessions had the SIA4-SITP-SIA6 structure (Alleles 1–14 and 1–15) (Fig. 5a, Supplementary Figs. 12 and 13). Similarly, three types of SITPR-containing structures with SIA4 and/or SIA6 (Alleles 2–2 to 2–4) were identified in O. barthii, the wild progenitor of O. glaberrima (Fig. 5a, Supplementary Figs. 12 and 13): 10 accessions had the SIA4-SITPR structure, one accession had the SITPR-SIA6 structure, and 52 accessions had the three-gene structure SIA4-SITPR-SIA6 (Fig. 5a, Supplementary Figs. 12 and 13).

Despite the incomplete lineage sorting in the AA-genome Oryza species, O. meridionalis is considered to be the sister group of the other AA species33,34. Intermediate variants containing SITP and the flanking genes SIA4 and/or SIA6 (SIA4-SITP, SITP-SIA6, SIA4-SITP-SIA6) may therefore have appeared first in O. meridionalis, followed by the two-gene intermediate structures carrying SITPR (SIA4-SITPR and SITPR-SIA6) and the final functional SIA4-SITPR-SIA6 complex (Fig. 5a and Supplementary Fig. 13). These complexes probably arose during allelic recombination events in natural hybrids between ancestral species possessing the intermediate structures (SITPR/SIA4-SITP, SITPR/SITP-SIA6, and SIA4-SITPR/SITPR-SIA6) (Fig. 5a and Supplementary Fig. 13), finally generating SIA4-SITPR-SIA6 in the O. barthii lineage.

All 133 analyzed accessions of O. glaberrima were found to possess the complete SIA4-SITPR-SIA6 structure, the same allele (Allele 2–4) was detected in O. barthii, indicating that the S1-g allele had been fixed in African rice (Fig. 5a, Supplementary Figs. 12 and 13). In addition, 13 accessions of O. meridionalis possessed another structure of SIA4-SITPR-SIA6, where SITPR is the primitive form of Allele 1 (Supplementary Figs. 12 and 13).

A working model for the S1-mediated killer–protector system.

Our results revealed that the three closely linked genes SIA4, SITPR, and SIA6 at the S1-g allele constitute a tripartite gamete killer–protector complex that acts as an ultra-selfish genetic complex generating a sterility signal via physical interaction of the three encoded proteins in the sporophytic cells. SITPR also serves as a protector in S1-g gametophytic cells (Fig. 5b). In O. glaberrima plants with a homozygous S1-g allele, all gametes escape abortion because SITPR eliminates the detrimental effect induced by S1-g. The S1-s allele in O. sativa harbors only the defective SITP gene and lacks the functional gamete killer and protector; therefore, all gametes are viable. In the interspecific hybrids, the tripartite SIA4-SITPR-SIA6 complex of S1-g causes the selective abortion of gametes containing the S1-s allele, resulting in a transmission advantage for S1-g (Fig. 5b).
Discussion

To date, there are two genetic models for HS in rice\textsuperscript{15}. The HS loci S5, S7, Sa, Sc, and qHMS7 fit the one-locus model\textsuperscript{12,14,15,29}, while S271/S28, DPL1/DPL2, and DGS1/DGS2 fit the two-locus model\textsuperscript{12,26,28}. There is emerging molecular evidence to support the conclusion that the loci involved in one-locus HS systems are usually complex loci, comprising multiple adjacent and functionally related genes, as we previously reported for the Sa locus containing the adjacent genes SaF and SaM\textsuperscript{2}. In the Sa and S5 systems, the symmetric allelic HS interactions are represented by the molecular interactions of the proteins from both alleles of the parental lines\textsuperscript{3,4}; for example, the interacting proteins in the three-component Sa complex, SaF\textsuperscript{2} and SaM\textsuperscript{3}, are contributed by the indica allele, while the SaM\textsuperscript{3} protein is encoded by the japonica allele\textsuperscript{4}. Similarly, in the S5 gamete-killer system, the ORF5\textsuperscript{+} protein is encoded by the indica allele, but ORF4\textsuperscript{+} is contributed by the japonica allele\textsuperscript{14}.

SI is considered the predominant HS locus found in Oryza, because of its strong genetic effect on HS of male and female gametes in interspecific hybrid progenies. In this study, we showed that the functional African rice allele, SI-g, consists of three closely linked actively expressed genes SIA4, SITPR, and SIA6 (SSP) (Fig. 5b). The gamete-killer function of the SI HS system requires all the three components (SIA4, SITPR and SIA6) only from SI-g, while the protector function depends on solely on SI-g-derived SITPR; no component from the Asian rice allele is required for this gamete killer–protector system. The SI gamete killer–protector system is therefore determined by the African rice allele, representing an asymmetric allelic interaction. This characteristic is distinct from the S5 and Sa systems, which are symmetric allelic interactions involving components from both divergent alleles. The SITPR has dual roles in killer and protector function, which is distinguished from the reported killer–protector systems\textsuperscript{14,15}, in which killer and protector are conferred by different factors. Thus, this SI gamete killer–protector system expands our understanding of the single-locus BDM model.

According to our transcriptome sequencing analysis, six genes in the BCAA degradation pathway were expressed at significantly lower levels in the RP-s plants than in the F1 and NIL-g plants (Supplementary Figs. 8–10). Dysfunctional BCAA biosynthesis is known to cause the abortion of both male and female gametophyte development\textsuperscript{35}. Therefore, we propose that the complex of the three SI-g gene products may induce excessive BCAA degradation in the F1, and NIL-g plants, resulting in the sterility effect that affects gamete development. The SITPR produced in SI-g-type gametes may result in adequate levels of BCAA via its peptidase function, as it resembles trypsin peptidase in animals\textsuperscript{36}. The gametes carrying SI-s in the F1 hybrids lacking functional SITPR would fail to restore fertility due to their BCAA deficiency. In Asian rice cultivars (SI-s1/Sl-s1), the genes involved in BCAA degradation are expressed at low levels, meaning that the gametes have enough BCAA to continue their development. In addition, we reasoned that the upregulation of photosynthesis-related genes in the F1 and NIL-g plants may be the result of a regulatory feedback loop monitoring nutrient or energy deficiency (Supplementary Figs. 8 and 10), which is an interesting topic for future study.

The hypothesized Gondwanaland origin of the Oryza genus explains the extensive geographic distribution of the Oryza in modern species\textsuperscript{27,18,35,37}. However, the current divergence times estimated from the molecular evolution between species in the Oryza genus are not in accordance with this theory\textsuperscript{34,38}. These divergence times were based on sequencing analysis of the currently available species, which may cause artifacts or bias in the algorithm used due to the inability to sequence extinct ancestral species. If the Oryza genus did indeed originate in Gondwanaland, we can hypothesize that the SI alleles in Asian rice and African rice might have originated from a common progenitor in Gondwanaland, and later evolved in parallel but independent lineages after their geographical separation (Fig. 5a). The ancient Oryza populations located on the ancient Australian continent might have been separated from the majority of the Oryza taxa due to the breakup of Gondwanaland and subsequent continental drift, causing the SIT allele variants (without SIA4 or SIA6) to pass through a genetic bottleneck before entering the O. rufipogon and O. sativa lineages in Asia. Nevertheless, other intermediate structures of SITPR and/or SITP variants carrying SIA4 and/or SIA6 in the ancestral Oryza species might have continued to evolve on the ancient supercontinent. The functional SIA4-SITPR-SIA6 structure was eventually generated in the O. barthii lineage and was further transferred into O. glaberrima and fixed on the African continent. The origin and evolution of the SI alleles, from the intermediate (neutral) haplotypes (SIA4-SITPR and SITPR-SIA6) to the functional ultra-selfish genetic complex, might therefore be associated with or contribute to the speciation of the related Oryza species.

Polymorphism has been shown to be inevitable in the progress from allele origination to fixation at BDM loci\textsuperscript{4}, consistent with the formation of the SI HS allele (Supplementary Figs. 12 and 13 and Supplementary Data 1). Extensive sequence variations were observed in SITPR and SITP at the SI locus in wild rice species (Supplementary Figs. 12 and 13 and Supplementary Data 1), allowing us to propose that their common ancestral sequences may be polymorphic in the Oryza species. Although the outline of the Oryza phylogeny is clear, the exact relationships among the Oryza species are elusive due to discordance between the phylogenetic trees for different genes. These issues are caused by factors such as incomplete lineage sorting\textsuperscript{34,39,40}, meaning that the SI alleles in Asian rice and African rice might have originated from an unknown common progenitor. The geographical distribution of polymorphisms at the SI locus might have arisen from the long-distance dispersal of these species, which was followed by selection and fixation.

Besides SI described in this study, several other complex HS loci that are composed of two or three closely linked HS genes (such as Sa, S5 and qHMS7) have been identified in Oryza species\textsuperscript{8,14,15}. These findings suggest that during speciation, generation of such complex HS loci may have advantages such as simple inheritance (as single functional genetic units having minimum recombination between the closely linked genes/alleles) and maximum genetic effect for postzygotic reproductive isolation.

The functional SI-g system is predominantly fixed in O. glaberrima populations, which explains why researchers failed to identify natural hybrid-compatible (neutral) SI alleles that could be used to break down the reproductive barrier in interspecific crosses between O. glaberrima and O. sativa. Our findings suggest that artificial hybrid-compatible SI alleles could be created for the utilization of distant heterosis by disrupting any one of the three genes in SI-g by CRISPR/Cas9 knockout\textsuperscript{12,27,41}. As an alternative strategy, replacing the premature stop codon in O. sativa cultivars, using base editing\textsuperscript{12} would rescue SITP to functional SITPR and thereby allow gamete normal development in interspecific hybrids.

Methods

Plant materials. A near-isogenic line NIL-g containing SI-g was developed using an African rice (O. glaberrima) line IRGCO2203 as the SI donor and an Asian rice (O. sativa) line IRAT216 containing SI-s as the recurrent parent (RP-s). The genomic sequences of SITPR, SIA4, and SIA6 were amplified by specific primers from BAC OG-BBa0049I08, which was kindly provided by Dr. Rod A.

"ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10488-3
8 NATURE COMMUNICATIONS | (2019) 10:2501 | https://doi.org/10.1038/s41467-019-10488-3 | www.nature.com/naturecommunications"
Wing (the University of Arizona). The genomic sequences were sub-cloned into the binary vector pCAMBIA1300 using a Gibson assembly assay to generate the functional-complementation constructs. 26 The target genes were inserted into the wRNA expression cassettes and sub-cloned into the CRISPR/Cas9 constructs 27 for knocking out genes S1A2–S1A6. All functional complementation constructs and knockout plasmids were transformed into RP-s and NIL-g using the Agrobacterium tumefaciens-mediated method. The primers used for vector construction are listed in Supplementary Table 8.

Phenotyping of pollen and spikelet fertility. The plant materials were grown in Guangzhou, China during the normal growing season and in Sanya, China during the winter season. The anthers of 3–5 mature flowers from each independent flowering individual were stained with I–KI solution to enable the observation and imaging of pollen fertility using a light microscope (Axio Observer D1, Carl Zeiss, Oberkochen, Germany). When the seeds ripened, the spikelet fertility was examined as the seed-setting rate in the main panicle of each individual.

Historical analysis. The historical analysis was performed using paraphin sectioning method 22. Briefly, spikiets of RP-s, NIL-g, the F2 plants from the RP-s × NIL-g and the F2 plants from the S1A4–S1A6 × STTPR crosses were fixed in FAA (1:1:18, formalin/glacial acetic acid/95% ethanol) for at least 24 h. Their ovules were then dehydrated, embedded in paraffin and cut into 8-μm longitudinal sections. The sections were stained with toluidine blue before being observed using a light microscope.

Genotyping of transgenes and S1 alleles in F2 populations. The T-DNA flanking sequences of STTPR and S1A4–S1A6 were determined using hiTAIL-PCR 23. PCR products were cloned into the F2-T-DNA tag using PCR with the specific primer sets (Supplementary Table 8). In addition, the genotypes of S1-alleles were determined by the S1-linked InDel marker 217907 (Supplementary Table 8).

Reverse transcription, expression and transcriptome analyses. The anthers, panicles and microspores were collected from PR-s and NIL-g at different stages and their total RNA was extracted. For each sample, 2 μg of total RNA was reverse transcribed to synthesize the first-strand cDNA according to the manufacturer’s instructions (Toyobo, Osaka, Japan). The quantitative RT-PCR was conducted using gene-specific primers (Supplementary Table 8) with three biological replicates. The data were normalized using OaActin as the endogenous control.

The transcriptome sequencing of the anther samples was performed by GENE DENovo Co. Ltd (Guangzhou, China). Genes with a significantly different expression level (upregulated by at least 2.82-fold or downregulated by at least 0.35 fold) in the F2 and NIL-g plants relative to RP-s were selected for further analysis. The genes that were commonly upregulated or downregulated in the F2 and NIL-g plants were identified and analyzed using the KEgg pathway database (http://www.genome.jp/kegg/). The expression patterns of the differentially expressed genes involved in high-sampling pathways were further validated using quantitative RT-PCR.

Subcellular localization and BiFC assays. The coding sequences of STTPR, S1A4, and S1A6 were cloned into the pYdiGF vector carrying the GFP tag to assay their subcellular localization. They were also cloned into pVN and pVC and were fused with the N-terminal or C-erminal sequence of their subcellular localization. They were also cloned into pVN and pVC and were fused with the N-terminal or C-erminal sequence of their subcellular localization. They were also cloned into pVN and pVC and were fused with the N-terminal or C-erminal sequence of their subcellular localization.

In vitro protein pull-down assay. The pull-down assays were conducted according to the manufacturer’s instructions (New England BioLabs, MA, USA). Briefly, the S1A4 protein was fused to an MBP-tag as the bait, while the S1A6 or STTPR proteins were fused with a GST-tag as the prey. The MBP-S1A4, GST-S1A6, or GST-STTPR proteins were expressed in Escherichia coli Rosetta (DE3). The cells containing these recombinant proteins were harvested in phosphate-buffered saline (pH 7.4) and then were ruptured by sonication. The lysates containing MBP-S1A4 were pulled down using amylose resin and mixed with lysates containing GST-S1A6 or GST-STTPR. The proteins were pulled down and detected using western blotting with anti-MBP (TransGen Biotech, Beijing, China; #HT701, 1:5000 dilution) and anti-GST antibodies (TransGen Biotech, Beijing, China; #HT601, 1:5000 dilution), respectively. Uncropped blots are presented in the Source Data file.

Phylogenetic and evolutionary analysis of the S1 locus. To determine the orthology of S1A4, S1A6, and STTPR, their nucleotide sequences were used as templates for a BLAST search of the Poaceae family in the GenBank database (https://www.ncbi.nlm.nih.gov/). The sequences of the putative orthologs were downloaded and used for the phylogenetic study. The phylogenetic tree was constructed using the maximum likelihood method via MEGA7 (www.megasoftware.net), with 10,000 bootstrap replications. To trace the ancestral STTPR and STTP sequences, seven non-synonymous SNPs were analyzed in two accessions of S. italica and 458 rice accessions. For the bioinformatics analysis, the publicly available sequences from the OMA project 48,49, the wild rice genome project 50, the 3000 rice genomes project 51 and the African wild rice genome 52 were downloaded from NCBI. The short reads were aligned to the genomic sequences of the STTPR and S1-alleles using Burrows-Wheeler Aligner 53. The depth of coverage and the SNPs were detected using the SAMtools package 54. To determine the presence of S1A4 and S1A6, coverage was calculated based on their depth at each nucleotide position with a Python program. Primer sets covering the genomic region of STTPR, STTP, S1A4, and S1A6 were designed to amplify the target region for the validation of the sequences in several collected accessions of wild and cultivated rice species to trace the divergence of the S1 locus.

Data availability
Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The source data underlying Figs. 1e, 2a, 2c, 3e, and 4c, 4d and Supplementary Figs. 1b, 2, 4d, 4e, 4f, 8a, 8b, 10 and Supplementary Tables 2–7 are provided as Source Data file. The genomic DNA and messenger RNA sequences of gene S1A4 have been deposited in NCBI GenBank database under accession number MK105813 and MK105814, respectively. The raw data of transcriptome sequencing have been deposited in NCBI Sequence Read Archive under accession number PRJNA540398. All data are available from the corresponding author upon request.

Received: 15 November 2018 Accepted: 15 May 2019
Published online: 07 June 2019

References
1. Bombolies, K. Doodmed: mechanisms of isolation and incompatibility in plants. Annu. Rev. Plant Biol. 61, 109–124 (2010).
2. Nosil, P. & Schluter, D. The genes underlying the process of speciation. Trends Ecol. Evol. 26, 161–167 (2011).
3. Muller, H. Isolating mechanisms, evolution, and temperature. Biol. Symp. 6, 7–125 (1942).
4. Cutter, A. D. The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. Trends Ecol. Evol. 27, 209–218 (2012).
5. Presgraves, D. C. The molecular evolutionary basis of species formation. Nat. Rev. Genet. 11, 175–180 (2010).
6. Hammond, T. M., Rehard, D. G., Xia, H. & Shiu, P. K. Molecular dissection of Neurospora Spore killer meiotic drive elements. Proc. Natl Acad. Sci. USA 109, 12093–12098 (2012).
7. Herrmann, B. G., Koschor, B., Wertz, K., McLaughlin, K. J. & Kispert, A. A protein kinase encoded by the 1 complex responder gene causes non-mendelian inheritance. Nature 402, 141–146 (1999).
8. 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).
9. Merrill, C., Bayraktaroglu, L., Kusano, A. & Ganetzky, B. Truncated RanGAP encoded by the segregation distorter locus of Drosophila. Science 283, 1742–1745 (1999).
10. Nuckolls, N. L. et al. wtf genes are fulid microwave antiseptic meiotic drivers, elife 6, e26057 (2017).
11. Phadnis, N. & Orr, H. A. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science 323, 376–379 (2009).
12. Shen, R. et al. Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. Nat. Commun. 8, 1310 (2017).
13. Sweigart, A. L. & Willis, J. H. Molecular evolution and genetics of postzygotic reproductive isolation in plants. Fl1000 Biol. Rep. 4, 23 (2012).
14. Yang, J. et al. A killer–protector system regulates both hybrid sterility and segregation distortion in rice. Science 337, 1336–1340 (2012).
15. Yu, X. et al. A selfish genetic element confers non-Mendelian inheritance in rice. Science 360, 1130–1132 (2018).
16. Bravo, N. M., Nuckolls, N. L. & Zanders, S. E. Genetic villains: killer meiotic drivers. Trends Genet. 34, 424–433 (2018).
17. Chang, T. T. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rice. Euphytica 25, 425–441 (1976).
18. Kusum, G. S. Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. 35, 23–34 (1997).
19. Ouyang, Y. & Zhang, Q. Understanding reproductive isolation based on the rice model. Annu. Rev. Plant Biol. 64, 111–135 (2013).
20. Ouyang, Y. D. Progress of indica-japonica hybrid sterility and wide-compatibility in rice. Chin. Sci. Bull. 61, 3833–3841 (2016).
21. Chen, J. et al. A triallelic system of SS is a major regulator of the reproductive barrier and compatibility of japonica-indica hybrids in rice. Proc. Natl Acad. Sci. USA 105, 11436–11441 (2008).
22. Hou, J. et al. E51 is involved in embryo sac abortion in interspecific hybrid progeny of rice. Plant Physiol. https://doi.org/10.1104/pp.18.01374 (2019).
23. 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).

24. Kubo, T., Takashi, T., Ashikari, M., Yoshimura, A. & Kurata, N. Two tightly linked genes at the S1 locus cause both F1 and F2 hybrid sterility in rice. Mol. Plant 9, 221–232 (2016).

25. 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).

26. Nguyen, G. N. et al. Duplication and loss of function of genes encoding RNA polymerase III subunit C4 causes hybrid incompatibility in rice. G3 7, 2565–2575 (2017).

27. Xie, Y. et al. Interspecific Hybrid sterility in rice is mediated by OgTPRI at the S1 locus encoding a peptidase-like protein. Mol. Plant 10, 1137–1140 (2017).

28. Yamagata, Y. et al. Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. Proc. Natl Acad. Sci. USA 107, 1494–1499 (2010).

29. Yu, Y. et al. Hybrid sterility in rice (Oryza sativa L.) involves the tetratricopeptide repeat domain containing protein. Genetics 203, 1439–1451 (2016).

30. Garavito, A. et al. A genetic model for the female sterility barrier between Asian and African cultivated rice species. Genetics 185, 1425–1440 (2010).

31. Koide, Y. et al. Sex-independent transmission ratio distortion system responsible for reproductive barriers between Asian and African rice species. New Phytol. 179, 888–900 (2008).

32. Sano, Y. The generic nature of gamete eliminator in rice. Genetics 125, 183–191 (1990).

33. Ge, S., Sang, T., Lu, B. R. & Hong, D. Y. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl Acad. Sci. USA 96, 14400–14405 (1999).

34. Stein, J. C. et al. Genomes of 13 domesticated and wild rice relatives highlight conservation and utilization of African wild rice genetic resources. Proc. Natl Acad. Sci. USA 115, 6588–6593 (2018).

35. Zhang, C. et al. Dihydroxyacid dehydratase is important for gametophyte development and disruption causes increased susceptibility to salinity stress in Arabidopsis. J. Exp. Bot. 66, 879–888 (2015).

36. Szabo, R. & Bugge, T. H. Membrane-anchored serine proteases in vertebrate cell and developmental biology. Annu. Rev. Cell Dev. Biol. 27, 213–235 (2011).

37. Wambu, P. W., Brooyanka, M., Furtado, A., Waters, D. L. & Henry, R. J. Relationships of wild and domesticated rices (Oryza AA genome species) based upon whole chloroplast genome sequences. Sci. Rep. 5, 13957 (2015).

38. Tang, L. et al. Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. Mol. Phylogenet. Evol. 54, 266–277 (2010).

39. Degnan, J. H. & Rosenberg, N. A. Tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340 (2009).

40. Wambu, P. W., Furtado, A., Waters, D. L., Nyamongo, D. O. & Henry, R. J. Conservation and utilization of African Oryza genetic resources. Rice 6, 29 (2013).

41. Xie, Y. et al. Suppression or knockdown of SaF/SaM overcomes the Sa-mediated hybrid male sterility in rice. J. Integr. Plant Biol. 59, 669–679 (2017).

42. Ren, B. et al. Cas9-NG greatly expands the targeting scope of genome-editing toolkit by recognizing NG and other atypical PAMs in rice. Mol. Plant https://doi.org/10.1016/j.molp.2019.03.010 (2019).

43. Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol. 498, 349–363 (2011).

44. Ma, X. et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol. Plant 8, 1274–1284 (2015).

45. Liu, Y. G. & Chen, Y. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. Biotechniques 43, 649–656 (2007).

46. Xie, Y. et al. The intrinsic cis element SEI recruits trans-acting repressor complexes to repress the expression of ELONGATED UPPERMOST INTERNODE1 in rice. Mol. Plant 11, 720–735 (2018).

47. Chen, S. et al. A highly efficient transient protoplast system for analyzing defence gene expression and protein–protein interactions in rice. Mol. Plant Biol. 7, 417–427 (2006).

48. Wang, M. et al. The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nat. Genet. 46, 982–988 (2014).

49. Wang, Z. et al. The Oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol. Biol. 59, 53–62 (2005).

50. Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. Nature 490, 497–501 (2012).

51. Li, Z. et al. The 3,000 rice genomes project. Gigasciencce 3, 7 (2014).

52. Zhang, Y. et al. Genome and comparative transcriptomics of African wild rice Oryza longistaminata provide insights into molecular mechanism of rhizomatousness and self-incompatibility. Mol. Plant 8, 1683–1686 (2015).

53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1734–1760 (2009).

54. Li, H. et al. The sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

Acknowledgements
We thank Dr. Chen Li from Rice Research Institute, Guangdong Academy of Agricultural Sciences for providing wild rice materials. This work was supported by grants from the National Key Research and Development Program of China (2016YFD0100804), the National Natural Science Foundation of China (31471564, 31701499), the Transgenic Research Program of the Ministry of Agriculture of China (2009ZX08009-1078), the Key Research Program of Guangzhou Science Technology and Innovation Commission (201707020016), and the China Postdoctoral Science Foundation (2018M630955).

Author contributions
L.C. and Y.X. conceived this work and wrote the manuscript; Y.X. and J.T. performed most of the experiments; X.X., L.L., J. Huang, Y.F., J. Han., X.Z., H.T. and S.C. performed some of the experiments; D.T., P.X. and Y.-G.L. provided the genetic materials and gave constructive advice on the experiments and manuscript preparation.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10488-3.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/.

Journal peer review information: Nature Communications thanks Kinya Toriyama, Jianmin Wan and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019