nature plants

LETTERS

PUBLISHED: 8 MAY 2017 | VOLUME: 3 | ARTICLE NUMBER: 17064

A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication

Wenguang Wu[1†], Xiaoyun Liu[1†], Muhua Wang[2†‡], Rachel S. Meyer[3], Xiaojin Luo[4], Marie-Noelle Ndjiondjop[5], Lubin Tan[1], Jianwei Zhang[2], Jianzhong Wu[6], Hongwei Cai[1], Chuanqing Sun[1,7], Xiangkun Wang[1], Rod A. Wing[2,8]* and Zuofeng Zhu[1]*

Grain size is one of the most important components of grain yield and selecting large seeds has been a main target during plant domestication. Surprisingly, the grain of African cultivated rice (*Oryza glaberrima* Steud.) typically is smaller than that of its progenitor, *Oryza barthii*. Here we report the cloning and characterization of a quantitative trait locus, *GL4*, controlling the grain length on chromosome 4 in African rice, which regulates longitudinal cell elongation of the outer and inner glumes. Interestingly, *GL4* also controls the seed shattering phenotype like its orthologue *SH4* gene in Asian rice. Our data show that a single-nucleotide polymorphism (SNP) mutation in the *GL4* gene resulted in a premature stop codon and led to small seeds and loss of seed shattering during African rice domestication. These results provide new insights into diverse domestication practices in African rice, and also pave the way for enhancing crop yield to meeting the challenge of cereal demand in West Africa.

The population in sub-Saharan Africa is estimated to exceed 1.5 billion by 2050[1]. It is critical to increase crop productivity to meet the food demands of a rapidly growing population[2]. Native African crops are valuable species to meet this demand: many are fast growing, nutritious, low input and have great historical and cultural importance. In West Africa, a locally cultivated rice (*O. glaberrima* Steud.) was domesticated thousands of years ago from its wild progenitor *O. barthii*[3–6]. Although development of Asian rice (*Oryza sativa* L.) for commercial agriculture in West Africa has greatly displaced native crops, smallholder farmers have maintained African rice landraces and their potential as climate-ready crops is gaining increasing attention[7,8], for example as an improvement target and for creating resilient interspecific hybrids such as New Rice for Africa (NERICA). In areas with the most adverse ecological conditions, *O. glaberrima* is favoured by farmers for its adaptability and resistance to multiple constraints[9,10].

Grain size is one of the most important factors determining grain yield and has been proposed to be among the earliest traits selected in the cereal crop domestication process[11,12]. Wheat, sorghum and Asian rice have larger seeds than their wild relatives[13–15]. Thus far, several genes for grain size, such as *GW2*, *qSW5*, *GS5*, *GW8*, *GS2*/*GL2*, *GS3* and *GL7*/*GW7*, have been identified in Asian rice[16–26]. However, unlike the tendency for increased grain size in other crops, the seed of *O. glaberrima* typically is shorter and smaller than that of its progenitor *O. barthii*[27] (Supplementary Fig. 1). Therefore, revealing the molecular genetic basis of selection small seeds during the domestication of *O. glaberrima* will contribute to improving grain yield of *O. glaberrima*.

In an attempt to identify the genes controlling grain size in African rice, we constructed a set of introgression lines using African wild rice accession W1411 (IRGC101248, *O. barthii*) as a donor, and a cultivar of African cultivated rice (IRGC102305, *O. glaberrima*) as a recipient. We obtained one introgression line (GIL25) displaying longer and larger grain size than the recipient parent IRGC102305 (Fig. 1a), which harboured two chromosomal segments on chromosomes 4 and 10 from *O. barthii* (W1411) (Fig. 1b and Supplementary Table 1). Compared with IRGC102305, the grain length and 1,000 grain weight of GIL25 increased by 13.6 and 16.8%, respectively, ultimately resulting in a dramatic increase of grain yield per plant (+14.3%) (Fig. 1c–f and Supplementary Fig. 2). Further observation of the outer and inner glumes using scanning electron microscopy found that GIL25 has longer epidermal cells than that of IRGC102305 (Fig. 1g–i), causing longer grain.

To identify genetic factors for grain size in African rice, we developed an F_2 population containing 186 individuals derived from a cross between GIL25 and IRGC102305, and detected a locus controlling grain length (referred to as *GL4*) between the markers RM3335 and RM5608 on chromosome 4, with the *O. barthii*-derived allele contributing to increased grain length (Fig. 2a). To map the *GL4* locus precisely, we assayed another 6,500 plants using newly developed markers between RM3335 and RM5608. A progeny test of homozygous segregates further narrowed down the *GL4* locus to a 5.9 kb region between markers M3 and M4 (Fig. 2b). In this region, there was only one annotated gene *ORGLA04G0254300* according to the annotation information of *O. glaberrima* IRGC96717 (var. CG14) genome (http://ensembl.gramene.org/Oryza_glaberrima/). Using quantitative polymerase chain reaction with reverse transcription (RT–PCR), we found that *ORGLA04G0254300* mainly expressed in the young panicle and weakly expressed in the stem, but did not express in the

[1]National Centre for Evaluation of Agricultural Wild Plants (Rice), MOE Key Laboratory of Crop Heterosis and Utilization, Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China. [2]Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA. [3]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California 90095, USA. [4]State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China. [5]Africa Rice Centre, Cotonou 2031, Benin. [6]Advanced Genomics Breeding Section, Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba 305-8602, Japan. [7]State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, Beijing 100193, China. [8]International Rice Research Institute, T.T. Chang Genetic Resources Centre, Los Baños, Laguna, Philippines. ‡Present address: Friedrich Miescher Laboratory of the Max Planck Society, Tübingen 72076, Germany. †These authors contributed equally to this work. *e-mail: zhuzf@cau.edu.cn; rwing@mail.arizona.edu

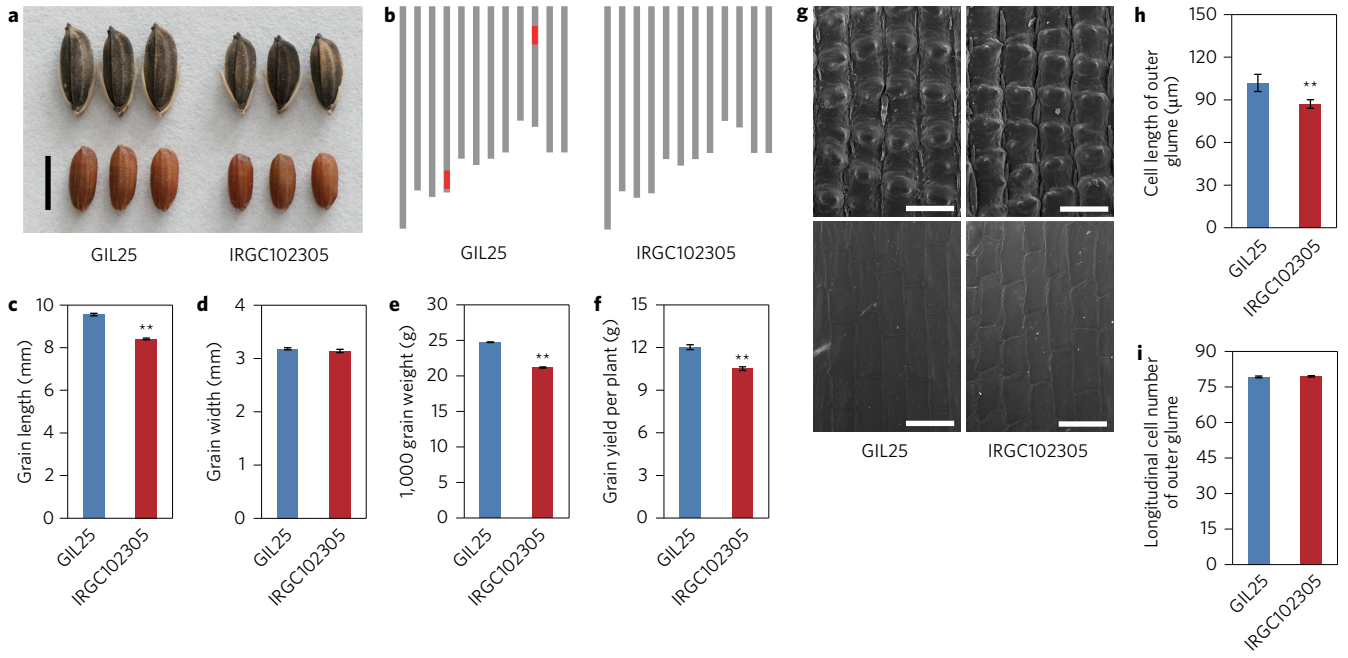

Figure 1 | Characterization of the grains traits between African cultivated rice IRGC102305 and introgression line GIL25. a, Grains of the African cultivated rice IRGC102305 and introgression line GIL25. Scale bar, 5 mm. **b**, Graphical genotypes of GIL25 and IRGC102305. **c**–**f**, Comparison of the grain traits between GIL25 and IRGC102305: grain length (**c**); grain width (**d**); 1,000 grain weight (**e**); grain yield per plant (**f**). **g**, Scanning electron microscope images of the outer glume (top) and inner epidermal cells of the lemma (bottom). Scale bars, 100 μm. **h**, Comparison of average cell number along the longitudinal axis between IRGC102305 and GIL25 grains. **i**, Cell length. $n = 10$ in **c**,**d**,**h**; $n = 3$ in **e**,**f**,**i**. All data are mean ± s.d. $^{**}P < 0.01$. The Student's *t*-test was used.

leaf and root (Supplementary Fig. 3). Thus, we proposed that *ORGLA04G0254300* was a candidate gene for *GL4*.

To verify this hypothesis, we generated a construct (pGL4) by placing a 4.8 kb genomic fragment from GIL25, covering the entire *ORGLA04G0254300* gene, into the vector pCAMBIA1300, and introduced this construct into IRGC102305. All 12 independent transgenic lines showed longer and larger grains than the controls (Fig. 2c–h), demonstrating that *ORGLA04G0254300* was the *GL4* gene, and that mutation in the *GL4* gene caused the small grain size in African cultivated rice IRGC102305.

To identify the causative mutations, we uncovered five mutations in the *GL4* coding sequences between the mapping parents GIL25 and IRGC102305 (Fig. 3a and Supplementary Fig. 4). These included three mutations in the first exon, including a 1 bp substitution (SNP1, G294A), a 6 bp insertion/deletion (indel) and a 1 bp substitution (SNP2, C760T) resulting in a premature stop codon of the IRGC102305 allele, and two SNP mutations (SNP3 and SNP4) in the intron. We sequenced the *GL4* genes of both 16 accessions of *O. barthii* and 67 diverse *O. glaberrima* cultivated rice from 13 West African countries (Supplementary Table 2) and analysed their haplotypes. We found that the 16 accessions of wild rice clustered in five haplotypes (H1 to H4) and the 67 *O. glaberrima* accessions only clustered in two haplotypes (H4 and H5) (Fig. 3b). An association test with grain length and nucleotide polymorphism revealed that the strongest signal was present at the SNP2 site (Fig. 3b,c). A medium signal was detected at the SNP3 site, which was in high linkage disequilibrium with the SNP2 site ($r^2 > 0.8$) (Fig. 3d). Further sequence comparisons demonstrated that the nucleotide at the SNP2 site of all 16 *O. barthii* accessions was C, and that of 62 *O. glaberrima* accessions was T. In addition, a point mutation was made at the SNP2 site in the *GL4* allele of IRGC102305 to develop another genomic construct (pGL4-2) by substituting the nucleotide T with C, which correcting the mutation and make an intact protein. We introduced this construct into IRGC102305. Observation of grain size found that all 15 independent transgenic lines showed complementation of grain length and grain weight (Fig. 3e–g). These results reflected that the SNP2 (C760T) mutation in *GL4* might account for the difference in the grain length between W1411 and IRGC102305.

The *GL4* allele of GIL25 was predicted to encode a 387 residue polypeptide. Analysis using the PredictNLS program found that the GL4 protein had a nuclear localization signal (NLS) at 302–306 amino acids. The premature stop codon caused by the SNP2 (C760T) mutation in *GL4* led to truncation of 253 amino acid residues without the NLS (Supplementary Fig. 5). To verify the impact of the deletion of NLS in the GL4 protein, we fused a green fluorescent protein (GFP) with GL4 alleles of GIL25 and IRGC102305 to make $GL4^{Ob}$–GFP and $GL4^{Og}$–GFP transient expression vectors, respectively, under the control of the CaMV35S promoter. The two constructs were introduced into onion epidermal cells. Observation of fluorescence signals found that the $GL4^{Ob}$–GFP fusion protein localized to the nucleus, but the $GL4^{Og}$–GFP fusion protein localized to both the cytoplasm and nucleus (Supplementary Fig. 6). These results suggested that the premature stop of the GL4 protein translation in *O. glaberrima* might cause dysfunction of its nuclear localization.

Notably, BLASTP analysis (http://www.ncbi.nlm.nih.gov/BLAST) showed that the GL4 protein has a high identity with *SH4/SHA1*, its orthologous gene in Asian wild rice. A domain search of GL4 in the Pfam and PROSITE databases found a Myb-like domain in its sequence, indicating GL4 is a Myb-like protein like SH4/SHA1 in *O. sativa*. The Ensembl gene tree (http://ensembl.gramene.org/Oryza_glaberrima/Gene/Compara_Tree?db=core;g=ORGLA04G0254300;r=4:25150788-25152622;t=ORGLA04G0254300.1) shows this gene is conserved in the *Oryza* genus and other major crop species, indicating it has functional constraint during evolution. SH4 was previously reported to control the transition from shattering to non-shattering during the domestication of Asian cultivated rice *O. sativa* by selecting another SNP mutation (G237T)[28,29]. Therefore, we further evaluated grain shattering among GIL25, IRGC102305 and transgenic plants. The results showed that GIL25 displayed a grain-shattering characteristic with

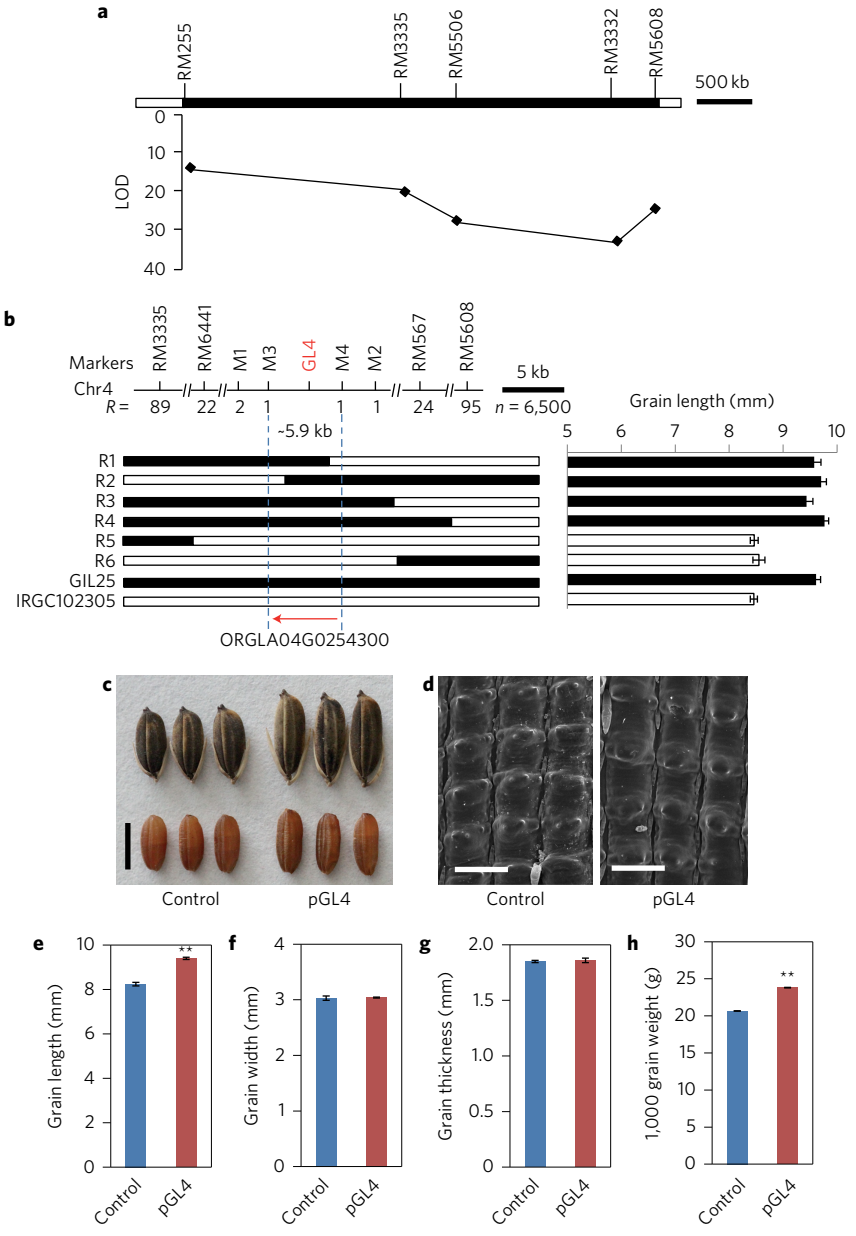

Figure 2 | Map-based cloning of *GL4*. **a**, The *GL4* locus was localized between the marker RM3335 and RM5608 on chromosome 4. **b**, The *GL4* locus was narrowed down to a 5.9 kb interval between markers M3 and M4 using 6,500 plants. The number of recombinants between *GL4* and markers is shown on the chromosome. R1–R6, six recombinant lines. **c**, The grain phenotype of the control and the transgenic plant of IRGC102305 expressing *pGL4*. The control plants were a transgenic line carrying an empty vector. **d**, Scanning electron microscope images of the inner epidermal cells of the lemma. Scale bars, 100 μm. **e**–**h**, Comparison of the grain-related traits between the control and transgenic plant of IRGC102305 expressing pGL4: grain length (**e**); grain width (**f**); grain thickness (**g**); 1,000 grain weight (**h**). $n = 10$ in **e**–**g** and $n = 3$ in **h**. All data are mean ± s.d. $^{**}P < 0.01$. The Student's *t*-test was used.

a complete abscission zone between the grain and the pedicel (Fig. 4a). However, the seed of IRGC102305 stayed on the plants at maturity, and displayed a deficiency in abscission zone development near the vascular bundle (Fig. 4b). The transgenic plant expressing the *GL4* allele from GIL25 recovered a complete abscission zone and exhibited seed-shattering characteristics (Fig. 4c and Supplementary Figs 7 and 8). These results indicated that *GL4/ObSH4* also controlled seed shattering in African wild rice; the SNP2 mutation in the *GL4* resulted in the loss of seed shattering in IRGC102305.

To compare the genetic effect in both grain size and shattering between the two mutant alleles, we developed two near-isogenic lines (NIL-GL4^{Or} and NIL-GL4^{Og}) under the Asian cultivated rice *O. sativa* var. Teqing (NIL-GL4^{Os}) genetic background, which contained a very small GL4 region from Asian wild rice (*Oryza rufipogon* Griff.) and African cultivated rice *O. glaberrima*, respectively (Supplementary Figs 9 and 10). The NIL-GL4^{Or} showed a grain-shattering phenotype with a complete abscission zone between the grain and the pedicel (Fig. 4d). Both NIL-GL4^{Os} and NIL-GL4^{Og} lost the shattering trait. The seed of the NIL-GL4^{Og} was much harder to shed than that of NIL-GL4^{Os}, which is consistent with the dramatic deficiency in abscission zone development in NIL-GL4^{Og} compared with that in NIL-GL4^{Os} (Fig. 4e,f and Supplementary Fig. 11). At the same time, both grain length and weight were not significantly different between NIL-GL4^{Os} and NIL-GL4^{Or}, implying that the selection of the *Ossh4* mutation allele did not change grain size during Asian rice domestication. However, the seed of NIL-GL4^{Og} was smaller than that of NIL-GL4^{Os}, suggesting the *Oggl4* mutant allele led to smaller grain (Fig. 4g–k).

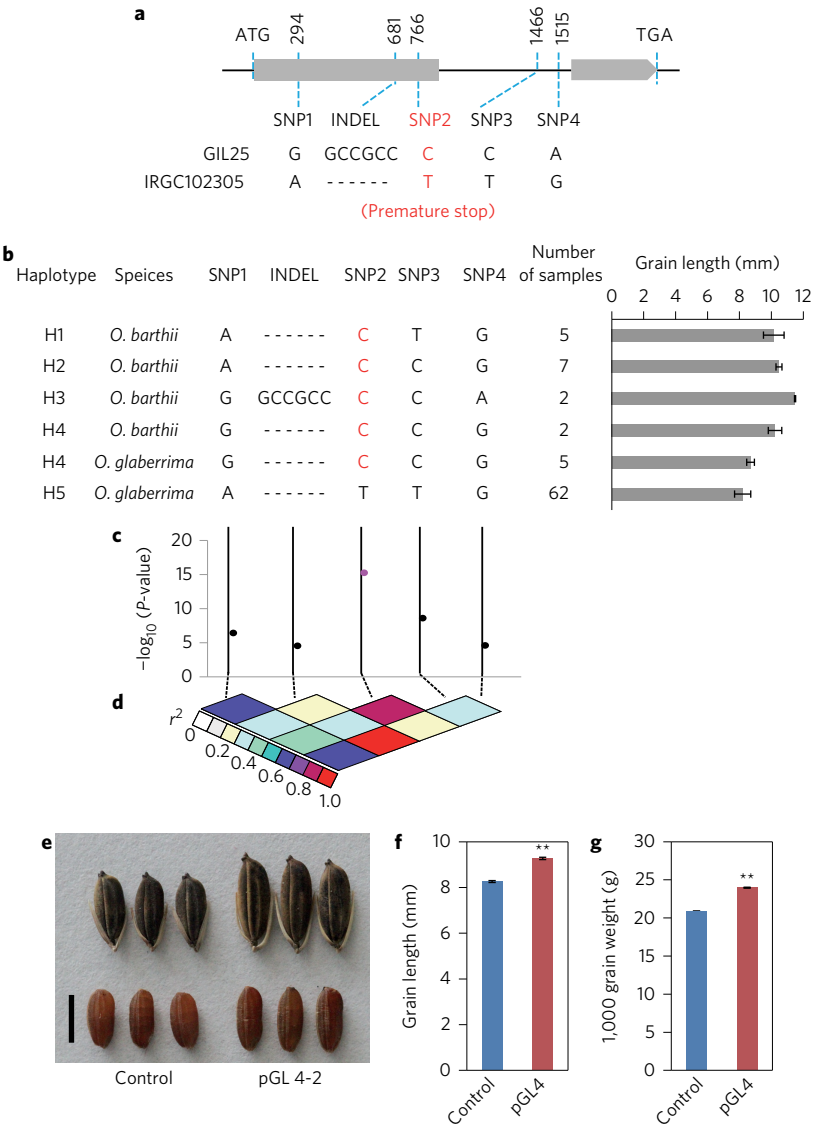

Figure 3 | Identification of the causative mutation for the *GL4* gene. a, The variations in the *GL4* coding region between GIL25 and IRGC102305. **b**, Haplotype analysis of the *GL4* gene on 16 accessions of wild rice and 67 accessions of domesticated African rice using the five mutation sites between GIL25 and IRGC102305. **c**, Association testing of five variants in the *GL4*. Black dots, four variations; red dot, proposed causative mutation (SNP2). **d**, Triangle matrix of pairwise linkage disequilibrium. **e**, The grain phenotype of the control and the transgenic plant of IRGC102305 expressing pGL4-2. Scale bar, 5 mm. **f**,**g**, Comparison of the grain length and the grain weight between the control and the transgenic plant of IRGC102305 expressing pGL4-2: grain length (**f**); 1,000 grain weight (**g**). The control plants were a transgenic line carrying an empty vector. $n = 10$ in **f** and $n = 3$ in **g**. All data are mean ± s.d. $^{**}P < 0.01$. The Student's *t*-test was used.

To elucidate the evolutionary trajectory of *GL4* during African rice domestication, we reanalysed 93 *O. glaberrima* and 94 *O. barthii* accessions using previously published resequencing data that span the species ranges[5,6]. Principal component analysis (PCA) of genetic variations 5 kb upstream and downstream of *GL4* separates 11 *O. glaberrima* accessions carrying the ancestral allele (C) at SNP2 (chromosome 4, 25152034) from *O. glaberrima* accessions carrying the derived allele (T), indicating two haplotypes of GL4 segregate in *O. glaberrima* (Supplementary Fig. 12 and Supplementary Table 3). The extended haplotype homozygosity (EHH) was calculated for both ancestral and derived alleles upstream and downstream of SNP2 in *O. glaberrima* (Fig. 5a). The decay of EHH for derived alleles is much slower than for ancestral alleles, suggesting that the haplotype carrying T but not the one carrying C was selected during domestication. In addition, we detected a reduction of nucleotide diversity (π) in *O. glaberrima* accessions carrying T at SNP2 compared with *O. barthii* (Supplementary Fig. 13). Together, these analyses show *GL4* was under selection during *O. glaberrima* domestication.

Interestingly, the variation in *GL4* is spatially restricted to an area in the southwestern (SW) forest range, which adds evidence supporting deep and separate roots of domestication practices in the west versus the eastern cultivation range[5]. The accessions carrying C at SNP2 were collected from Guinea, Liberia or Sierra Leone (Fig. 5b and Supplementary Table 4), and clustered in PCA (Supplementary Fig. 14). The lack of fixation on *GL4* in the SW region leads to two possible explanations: either other loci might have been involved in the selection of small grain size during the gradual domestication process in that geographically distinct region, or the selection for small non-shattering grains began at that location, and the variation observed is due to the persistence of ancestral haplotypes in landraces. Only those with the non-shattering and small grain phenotype would have been brought out of the region after the onset of domestication.

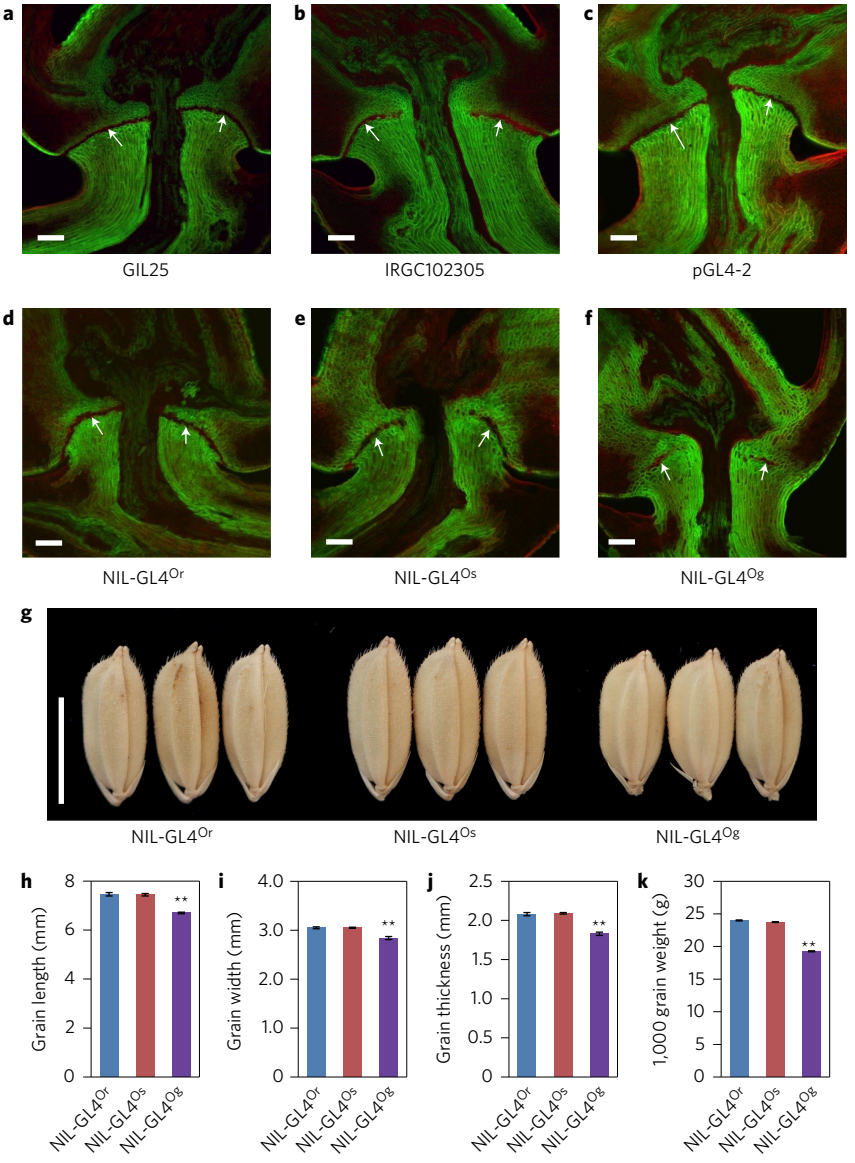

Figure 4 | Comparison of the effects of different *GL4* alleles. a–f, Fluorescence images of a longitudinal section of the flower and pedicel junction. White arrows, abscission layer. Scale bars, 50 μm. **a**, GIL25 parental line of the mapping population, with a complete abscission layer (indicated by an arrow). **b**, IRGC102305 parental line of the mapping population, with an incomplete abscission layer. **c**, Transgenic plants of IRGC102305 expressing *pGL4-2*, with a complete abscission layer. **d**, GL4 near-isogenic lines NIL-GL4^{Or} (which contains a very small GL4 region from Asian wild rice in a Teqing genetic background), with a complete abscission layer. **e**, Asian cultivated rice Teqing (NIL-GL4^{Os}), with an incomplete abscission layer. **f**, GL4 near-isogenic lines NIL-GL4^{Og} (containing a very small GL4 region from African cultivated rice in a Teqing genetic background), with a severely incomplete abscission layer. **g**, The grain phenotype of GL4 near-isogenic lines NIL-GL4^{Or}, NIL-GL4^{Og} and Teqing (NIL-GL4^{Os}). Scale bar, 5 mm. **h–k**, Comparison of the grain shape among NIL-GL4^{Or}, NIL-GL4^{Os} and NIL-GL4^{Og}: grain length (**h**); grain width (**i**); grain thickness (**j**); 1,000-grain weight (**k**). $n = 10$ in **h–j**; $n = 3$ in **k**. All data are mean ± s.d. $^{**}P < 0.01$. The Student's *t*-test was used.

Taken together, in the present study we demonstrated that *GL4*/*SH4* is a key domestication gene with pleiotropic effects controlling both grain size and shattering in rice. During crop domestication, reducing seed shattering and increasing seed size were two main selection targets in most grains, but it appears that in West Africa, selection proceeded in a different form leading to a decrease in seed size. Although different mutant alleles (*Ossh4* and *Ogsh4*) resulting in the loss of the grain-shattering characteristic were selected in parallel during the domestication of both African and Asian rice, either intentionally or unconsciously, selecting the *Ossh4* mutation did not change the seed size in Asian cultivated rice, but selecting *Ogsh4* led to small seeds in African cultivated rice. Why was this *Ogsh4* allele, which reduces seed size and inevitably results in lower yield, selected during the domestication of African rice? There might be two possibilities: first, it is possible that the high iron toxicity in much of West Africa, which can increase lodging in the already lodging-prone *O. glaberrima* crop, may have led farmers to favour smaller grains; second, there might be no other functional mutations available for selection that affect the SH4 gene function in the limited genetic pool in Africa rice. Further, the genetic architecture of the seed size trait is complex, involving over 400 genes[30], and therefore should be quite canalized, suggesting genetic mutations in other genes and regulators such as microRNAs will only have miniscule effect[31]. Further, it appears that some genes important to Asian rice seed size may be absent or unimportant in other *Oryza* (for example GS3[32]), which limits the ability to leverage findings in one species to catalyse discovery in another. Nonetheless, given the

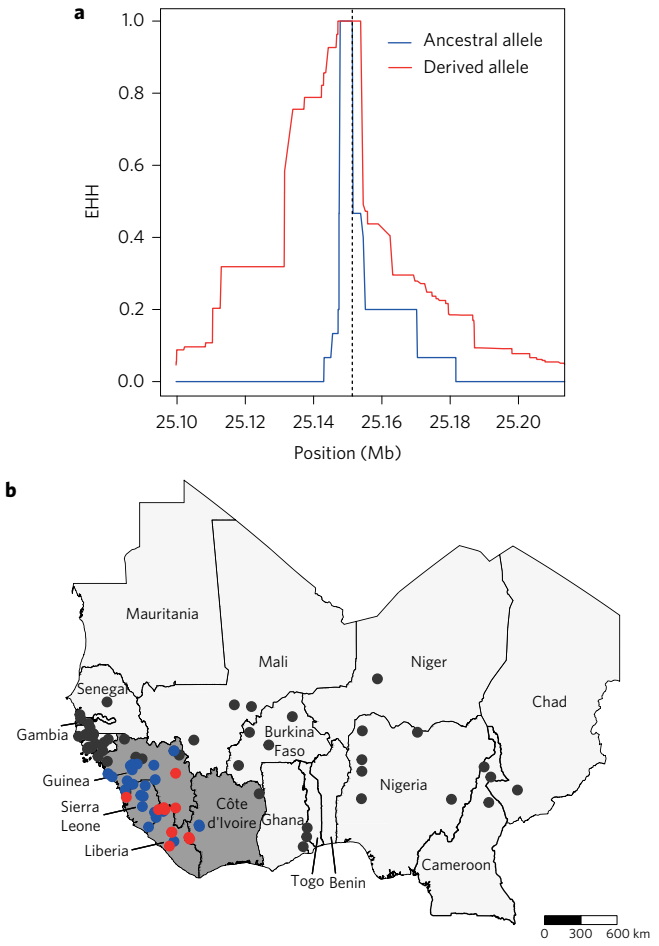

Figure 5 | Evolutionary analysis of *GL4/SH4*. **a**, Extended haplotype homozygosity (EHH) analysis of GL4. The focal SNP (SNP2) of the analysis is indicated by a dotted line. **b**, Biogeographic analysis of genetic variation of SNP2. Dots on the map represent collection sites of *O. glaberrima* accessions used in the analysis. *O. glaberrima* accessions carrying C at SNP2 are red; accessions carrying T at SNP2 from the southwestern population are blue; accessions carrying T at SNP2 from other populations are black. The shaded area on the map indicates countries where *O. glaberrima* accessions carrying C at the SNP2 position were collected.

contemporary preference for larger grain size and the need to increase crop yield, we argue that replacing the *Ogsh4* allele with the *Ossh4* allele would enhance the grain yield of *O. glaberrima* in breed practice and would make it a more valuable crop for future food security.

With improvement, the yield of *O. glaberrima* can match or even exceed Asian cultivated rice, and breeders can take advantage of local trait combinations adapted to the myriad cultivation settings, capacities and practices in West Africa[33–35]. To meet the food demands of a rapidly growing population in Africa, the scaling up of resilient native crop cultivation is one of the most important targets for a food-secure future. We suggest selection to increase the grain size in *O. glaberrima* and exploiting its low-input requirements and adaptations to many West African marginal landscapes. This will improve the food security of sub-Saharan Africa, preserve cultural heritage, and provide a more nutritious and taste-preferred rice for the growing population.

Methods

Plant materials. The introgression line GIL25 was generated by repetitive backcross progeny derived from a cross between W1411 (*O. barthii*) × IRGC102305 (*O. glaberrima*) with IRGC102305 as the recurrent parent using marker-assisted selection. It harbours a segment at locus GL4 and another segment at chromosome 10. The NIL-GL4^{Og} and the NIL-GL4^{Or} plants have different *GL4* alleles from *O. glaberrima* and YJWCR, respectively, under the same background of Teqing (*O. sativa* L. ssp. *indica* variety). The primer sequences for the genotyping assays are provided in Supplementary Table 1. Other cultivars and wild-rice accessions used for haplotype tests in this study are listed in Supplementary Table 2.

Growing conditions and trait measurement. Rice plants were cultivated in Sanya, Hainan province, China, under field conditions with an interplant spacing of 15 × 15 cm^2 for transplanting. For grain yield per plant analysis, a total of 50 plants for each line were grown in a paddy field with three completely randomized blocks. The 1,000 grain weight, tiller numbers per plant and grain yield per plant were measured using 30 plants from three completely randomized blocks. Grain number per panicle was investigated using the main stem panicle. Thirty fully filled grains were chose and measured by electronic digital-display Vernier calliper to obtain grain length, width and thickness. The entire spikelet hull, including the lemma and palea, was included in these measurements. Three panicles per line were measured for rice grain-breaking tensile strength at 35 days after heading. Longitude cell numbers of grain were observed under a stereoscope (Zeiss) when grains were fully matured.

Histological analysis. Panicles were obtained at the flowering stage; a longitudinal section was achieved by hand cutting through the grain and pedicel junction. Sections were stained with acridine orange and observed under an Olympus FV1000 laser scanning microscope. A 488 nm and a 543 nm laser line were used.

Scanning electron microscopy. The outer surface and inner epidermal cells of lemma from mature seeds were gold plated and observed using a Hitachi S-2460 scanning electron microscope at 15 K. The cell number was calculated along the longitudinal axis. The cell length was measured using Image J software. The spikelets at ~35 days after heading were collected for scanning the separation section between grain and pedicel as above.

RNA extraction, quantitative RT–PCR. Total RNA was extracted from various rice tissues and reverse transcribed using the M-MLV reverse transcriptase kit (Promega). Quantitative RT–PCR was carried out using SYBR Green QPCR mix (Bio-Rad), and the *actin1* gene was used as an endogenous control to normalize detected gene expression. The cycling conditions included incubation for 30 s at 95 °C followed by 40 cycles of amplification (95 °C for 5 s and 60 °C for 30 s). All samples were repeated at least three times. The primers are given in Supplementary Table 1.

Vector construction and complementation test. To generate the complementary vector pGL4, a total length of 4.8 kb DNA fragments that contained 2.2 kb upstream of the *GL4* transcription start site and 0.5 kb downstream of its termination site were amplified from W1411 to place into a pCAMBIA1300 vector. The 3.1 kb region upstream of the SNP2 and 1.7 kb downstream of SNP was amplified from IRGC102305, respectively. The SNP2 was changed in the primer and was placed into pCAMBIA1300 by homologous recombination to construct the pGL4-2 vector. We obtained 12 for pGL4 and 15 for pGL4-2 independent transgenic plants, respectively. Primers used are listed in Supplementary Table 1.

Subcellular localization. The coding sequences of *GL4* gene were amplified from W1411 and IRGC102305 to generate p35S::*GL4–GFP* and p35S::*Oggl4–GFP* vectors. The two vectors were bombarded into onion epidermal cells using a helium biolistic device (Bio-Rad PDS-1000). 4′-6-Diamidino-2-phenylindole (DAPI) was used to indicate the nuclei under an Olympus FV1000 fluorescence microscope.

Evolutionary analysis of GL4. Resequencing reads of 93 *O. glaberrima* and 94 *O. barthii* were downloaded from the NCBI Short Read Archive (SRA) database. The Burrows–Wheeler Aligner (BWA) V0.7.10 (ref. 36). was used to map raw pair-end reads onto chromosome 4 of *O. glaberrima* CG14 genome assembly[5]. SNP calling and filtering were performed using the Genome Analysis Toolkit (GATK) V3.4 (ref. 37). In total, 680,362 SNPs were identified and used in the following analysis. PCA of genetic variations 5 kb upstream and downstream of GL4 was performed using the smartpca program implemented in EIGENSOFT package V6.0.1 (ref. 38). Genotypes of *O. glaberrima* accessions were phased and imputed using Beagle 4.1 (refs 39,40). The EHH was calculated using R package REHH 2.0 (ref. 41) with SNP2 (chromosome 4, 25152034) as the focal SNP. As group V *O. barthii* accessions described previously[5] are genetically closed to *O. glaberrima*, the ancestral state was inferred as the major allele in group I to IV *O. barthii* accessions at each SNP position. The nucleotide diversity (π) of *O. glaberrima* accessions carrying T at SNPs and all *O. barthii* accessions was calculated using VCFtools V0.1.14 (ref. 42). The result of PCA using genome-wide genetic variations of 93 *O. glaberrima* accessions was downloaded from Meyer *et al.*[6] and plotted with *O. glaberrima* accessions carrying C at the SNP2 position highlighted in red (Supplementary Fig. 14). Two *O. glaberrima* accessions (IRRI_103993 and IRRI_104573) with missing calls at the SNP2 position were excluded from biogeographic analysis of *GL4*. The GPS coordinates of 91 *O. glaberrima*

accessions were downloaded from Meyer *et al.*[6]. The map of biogeographic analysis was drawn using R.

Data availability. The authors declare that the data supporting the findings of this study are available within the paper and the Supplementary Information.

Gene sequences of *GL4* have been deposited in the GenBank with the following accession codes: KY427262 (*GL4* in W1411), KY427263 (*GL4* in IRGC102305).

Received 21 December 2016; accepted 4 April 2017; published 8 May 2017

References

1. United Nations-Department of Economic and Social Affairs. *World Population Prospects, The 2015 Revision* (United Nations, New York, 2015).
2. Martin, K. I. *et al.* Can sub-Saharan Africa feed itself? *Proc. Natl Acad. Sci. USA* **113,** 14964–14969 (2016).
3. Sweeney, M. & McCouch, S. The complex history of the domestication of rice. *Ann. Bot.* **100,** 951–957 (2007).
4. Harlan, J., De Wet, J. & Stemler, A. in *Origins of African Plant Domestication* (ed. Harlan, J. R.) 3–20 (Mouton Publishers, 1976).
5. Wang, M. *et al.* The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* **46,** 982–988 (2014).
6. Meyer, R. S. *et al.* Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* **48,** 1083–1088 (2016).
7. Linares, O. F. African rice (*Oryza glaberrima*): history and future potential. *Proc. Natl Acad. Sci. USA* **99,** 16360–16365 (2002).
8. Rhodes, E. R., Jalloh, A. & Diouf, A. *Review of Research and Policy for Climate Change Adaptation in the Agriculture Sector of West Africa* (AfricaInteract/FAC, 2014).
9. Jones, M. P., Dingkuhn, M., Aluko, G. K. & Semon, M. Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. *Euphytica* **92,** 237–246 (1997).
10. Sarla, N. & Swamy, B. P. *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. *Curr. Sci.* **89,** 955–963 (2005).
11. Dorian, F. & Robin, A. Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation. *Annu. Plant Rev.* **38,** 238–295 (2009).
12. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14,** 840–852 (2013).
13. Gegas, V. C. *et al.* A genetic framework for grain size and shape variation in wheat. *Plant Cell* **22,** 1046–1056 (2010).
14. Han, L. *et al.* Fine mapping of *qGW1*, a major QTL for grain weight in sorghum. *Theor. Appl. Genet.* **128,** 1813–1825 (2015).
15. Zuo, J. R. & Li, J. Y. Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* **48,** 99–118 (2014).
16. Song, X., Huang, W., Shi, M., Zhu, M. & Lin, H. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39,** 623–630 (2007).
17. Shomura, A. *et al.* Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **40,** 1023–1028 (2008).
18. Li, Y. *et al.* Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* **43,** 1266–1269 (2011).
19. Ishimaru, K. *et al.* Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat. Genet.* **45,** 707–711 (2013).
20. Wang, S. *et al.* Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* **44,** 950–954 (2012).
21. Che, R. *et al.* Control of grain size and rice yield by GL2-mediated brassinosteroid responses. *Nat. Plants* **2,** 15195 (2015).
22. Duan, P. *et al.* Regulation of *OsGRF4* by OsmiR396 controls grain size and yield in rice. *Nat. Plants* **2,** 15203 (2015).
23. Wang, Y. *et al.* Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat. Genet.* **47,** 944–948 (2015).
24. Wang, S. *et al.* The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* **47,** 949–954 (2015).
25. Mao, H. *et al.* Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl Acad. Sci. USA* **107,** 19579–19584 (2010).
26. Si, L. *et al. OsSPL13* control grain size in cultivated rice. *Nat. Genet.* **48,** 447–456 (2016).
27. Katayama, T. C. & Sumi, A. Studies on agronomic traits of Africa rice (*Oryza glaberrima* steud.), 3: some grain morphological aspects of domestication and decrement. *Jpn. J. Crop. Sci.* **64,** 807–814 (1995).
28. Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311,** 1936–1939 (2006).
29. Lin, Z. *et al.* Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* **226,** 11–20 (2007).
30. Huang, R. Y. *et al.* Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci.* **18,** 218–226 (2013).
31. Tang, W. *et al.* SNP-based analysis of genetic diversity reveals important alleles associated with seed size in rice. *BMC Plant Biol.* **16,** 1 (2016).
32. Takano-Kai, N. *et al.* Evolutionary history of *GS3*, a gene conferring grain length in rice. *Genetics* **182,** 1323–1334 (2009).
33. Rodenburg, J. *et al.* Weed competitiveness of the lowland rice varieties of NERICA in the southern Guinea savanna. *Field Crop Res.* **114,** 411–418 (2009).
34. Agnoun, Y. *et al.* The African rice *Oryza glaberrima* steud: knowledge distribution and prospects. *Int. J. Biol.* **4,** 158–180 (2012).
35. Balasubramanian, V., Sie, M., Hijmans, R. J. & Otsuka, K. Increasing rice production in Sub-Saharan Africa: challenges and opportunities. *Adv. Agron.* **94,** 79–80 (2007).
36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler *Transform. Bioinformatics* **25,** 1754–1760 (2009).
37. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
38. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** 2074–2093 (2006).
39. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81,** 1084–1097 (2007).
40. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98,** 116–126 (2016).
41. Gautier, M., Klassmann, A. & Vitali, R. REHH 2.0: a reimplementation of the R package REHH to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17,** 78–90 (2017).
42. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).

Acknowledgements

We thank the International Rice Research Institute and the United States Department of Agriculture Research Service for providing the wild rice and cultivated rice samples. This research was supported by the National Key R&D Program for Crop Breeding (2016YFD0100901), the Ministry of Agriculture of China (2016ZX08009-003) and the National Natural Science Foundation of China (Grant 31471457). We acknowledge the NSF Plant Genome Program for support to Meyer (1202803). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Z.Z. and R.A.W. designed and supervised this study. W.W. conducted characterization of introgression line, map-based cloning, genetic transformation, gene expression analysis. W.W. and X. Liu constructed the introgression lines. M.W., R.S.M. and J.Z. performed evolutionary analysis of GL4. X. Luo bred the NIL-GL4^{Og}. M.-N.N., L.T., J.W., H.C., C.S. and X.W. conducted the collection of rice germplasm and phenotypic data. Z.Z., R.S.M. and R.A.W. wrote the manuscript.

Additional information

Supplementary information is available for this paper.

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

Correspondence and requests for materials should be addressed to R.A.W. and Z.Z.

How to cite this article: Wu, W. *et al.* A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. *Nat. Plants* **3,** 17064 (2017).

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

Competing interests

The authors declare no competing financial interests.