The genetic control of glabrous glume during African rice domestication

Leqin Chang[a], Min Hu[a], Jing Ning[a], Wei He[a], Jiayu Gao[a], Marie-Noelle Ndjiondjop[b], Yongcai Fu[a], Fengxia Liu[c], Hongying Sun[a], Ping Gu[a], Chuanqing Sun[c], Zuofeng Zhu[a, *]

[a] State Key Laboratory for Agrobiotechnology, National Center for Evaluation of Agricultural Wild Plants (Rice), Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China.

[b] Africa Rice Center, Cotonou, Benin.

[c] State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, Beijing 100193, China.

[*] Corresponding author.Email address: zhuzf@cau.edu.cn (Z. Zhu)

Abstract

African cultivated rice, *Oryza glaberrima*, is characterized by its glabrous glumes. During domestication, the pubescent glumes of its wild ancestor, *Oryza barthii*, lost their trichomes, and in this study, we show that *GLABROUS GLUME 5* (*GLAG5*), a *WUSCHEL*-like homeobox transcription factor on chromosome 5, is required for trichome development. DNA methylation associated with an *hAT* transposable element inserted in the promoter region of *GLAG5* was found to reduce its expression, leading to the formation of glabrous glumes and leaves in African cultivated rice. Among 82 African cultivated rice varieties investigated in this study, 59 (approximately 71%) lines exhibited glabrous glumes and harbored the *hAT* transposon; however, the other 23 varieties, which exhibit pubescent glumes (approximately 29%), lacked the *hAT* transposon, indicating that *glag5* had undergone strong artificial selection. The π_w/π_c ratios also showed the *hAT* transposon insertions influenced the genetic diversity of an approximately 150-kb interval encompassing the *GLAG5* locus. The identification of the *GLAG5* gene provides new insights into the domestication of cultivated rice in Africa. We speculate that the selection of varieties with mutations in their promotor regions is an important aspect of crop domestication.

Introduction

African cultivated rice (*Oryza glaberrima*) was domesticated from its wild progenitor *Oryza barthii* in Africa more than 3000 years ago (Purugganan, 2014). During domestication, farmers selected for morphological traits that facilitated cultivation and harvesting, which included reduced seed shattering, erect plant architecture and short-awn or awnless seeds. African rice is a valuable genetic resource containing resistance genes to biotic and abiotic stresses, such as Rice Yellow Mottle Virus, high temperature, soil acidity, iron toxicity and drought (Linares, 2002; Thiemele et al 2010; Jones et al., 2013; Wambugu et al., 2013; Li et al., 2015). Owing to its genetic importance, the origin and evolution of African rice has been studied extensively over several decades (Ammiraju et al., 2008; Van Andel, 2010; Molina et al., 2011; Sakai et al., 2011; Chen et al., 2013; Stein et al., 2018). Recent studies have identified the genes controlling the reduction of seed shattering and erect plant architecture characteristics in African cultivated rice (Wu et al., 2017; Hu et al., 2018; Lv et al., 2018), and this information has provided new insights into its domestication.

Oryza glaberrima is known as glabrous rice, and it is characterized by glabrous glumes and leaves. However, African wild rice is typically densely covered by hard (thorn-like) trichomes on the seeds and soft trichomes on the leaves, which aid seed dispersal and provide protection against seed predation by birds and other animals. The glabrous glumes of *O. glaberrima* are valued for hand-harvesting, because they decrease skin irritation during threshing, drying and milling, and reduce the amount of storage space required (Lim et al., 1984). The loss of trichomes is considered to be an important event during the domestication of wild rice. However, the molecular genetic mechanisms involved in the loss of trichomes remain largely unknown.

Transposable elements (TEs), or fragments of DNA that can move from one location in the genome to another, are major components of eukaryotic genomes (Feschotte et al., 2002). In rice, over 35% of the genomic content is composed of TEs, and generic sequences are embedded in these vast expanses of TEs (Ma et al., 2020). TEs can influence nearby gene expression either through *cis*-acting elements residing

in their own sequences or by changing the DNA or chromatin methylation states of adjacent genes (Studer et al., 2011; Friedli and Trono, 2015; Li et al., 2017; Peng et al., 2019). To maintain genomic stability, TEs are usually silenced and inactive, owing to DNA and chromatin modifications.

In this study, we describe the isolation and characterization of *GLABROUS GLUME 5* (*GLAG5*), which controls the formation of glabrous glumes and leaves in African rice. This gene was found to encode a homeobox domain transcription factor located on chromosome 5. Furthermore, we show that an *hAT* transposon insertion in the *GLAG5* promoter during African cultivated rice domestication may have repressed gene expression through DNA methylation and led to the development of glabrous glumes.

Results

Comparison of plant phenotypes between W1411 and IRGC104165

The seeds of African wild rice (*O. barthii*) are densely covered by hard (thorn-like) trichomes that are macroscopic and sharp, whereas the hulls of African cultivated rice are glabrous and smooth (Fig. 1A and 1B). The hard trichomes of the wild rice cause the seeds to stick to clothes and penetrate skin, causing inflammation, while soft trichomes cover the adaxial surface of the leaves (Fig. S1), which can cause skin irritation. The hard trichomes are critical for efficient propagation and dissemination of wild rice seeds under natural conditions, serving to protect them from predation and aid dispersal (Fig. S2). However, trichomes decrease the efficiency of grain packaging, while the glabrous glumes of cultivated rice facilitate harvesting and storage (Fig. S3). To further investigate the differences in the hulls and leaves between African wild and cultivated rice, the leaf and grain phenotypes of the African wild rice line W1411 (IRGC100119) and the African cultivated cultivar IRGC104165 were examined using stereomicroscopy and scanning electron microscopy (SEM). Epidermal cells on surface of glumes from wild rice were seen to bulge (Fig. 1C). However, no swollen epidermal cells were found on surface of the

glumes from the cultivated rice (Fig. 1D). The swollen epidermal cells formed basipetally along the glume, and they elongated and sharpened upwardly to form hard (thorn-like) trichomes at the late stage in W1411 (Fig. 1E). However, no barbs formed at any time during glume development in the cultivated rice, resulting in glabrous glumes (Fig. 1F).

Map-based cloning of *GLAG5*

To isolate the gene controlling the transition from the barbed hulls of wild rice to the glabrous glumes of cultivated rice, we carried out a genetic analysis using an F_2 population containing 182 individuals developed from a cross between the African wild rice cultivar W1411 (IRGC100119) and the African cultivated cultivar rice line IRGC104165. Among the 182 individuals, 138 exhibited barbed glumes and 48 revealed smooth glumes. A 3:1 (X^2 = 0.0645, P > 0.05) segregation ratio was determined, indicating the glabrous glume phenotype was controlled by a single recessive gene, designated as *GLAG5*, which was mapped between SSR markers C03 and RM593 on the short arm of chromosome 5.

Using 4500 recessive homozygotes with the glabrous hull phenotype from the F_2 population and 6 newly developed markers, we ultimately delimited *GLAG5* to a 15.9-kb region between the C06 and C02 markers. This region contained only one predicted gene, *OBART05G01300*, in the African wild rice reference genome (https://ensembl.gramene.org/Oryza_barthii/Info/Index). No difference was found in the coding region of *OBART05G01300* between the wild (W1411) and cultivated (IRGC104165) rice. However, a transposon (478 bp) was found to lie in the promoter region of this gene in the IRGC104165 genome (Figs. 2A, 2B, S4). These results suggested that *OBART05G01300* is the *GLAG5* gene controlling trichome development in the hulls of African rice species.

To confirm this hypothesis, we generated a construct (p*GLAG5*) by placing a 4.04-kb genomic fragment from wild rice (W1411), covering the *OBART05G01300* gene only, into the vector pCAMBIA1300. Owing to the recalcitrance of the cultivated line IRGC104165 to regenerating shoots from calli, we introduced the p*GLAG5* construct into another African cultivated rice cultivar (WK21) which also

has glabrous glumes. The 11 independent transgenic lines all grew trichomes on their glumes (Fig. 2C–2F) and demonstrated small bulges on their leaves (Fig. 2G and 2H). These results confirmed that the *OBART05G01300* gene was indeed *GLAG5*.

Subcellular localization and functional analysis of *GLAG5*

The GLAG5 protein was predicted to encode a transcription factor and contains a nuclear localization signal. Further analyses revealed that GLAG5 contains a homeodomain (HD), which is a DNA-binding domain and a *WUSCHEL* domain, indicating that it belongs to the *WUSCHEL*-like homeobox (WOX) family. The GLAG5-green fluorescent protein fusion protein specifically localized to the nuclei in rice protoplast cells (Fig. S5). These results indicated that *GLAG5* encodes a transcription factor.

To further validate the function of *GLAG5*, we knocked out the endogenous *OBART05G01300* gene in W1411 using CRISPR-Cas9 genome editing technology. Using PCR and sequencing analysis, we obtained five heterozygous transgenic plants in the T0 generation by PCR and sequencing analyses. These five transgenic plants were self-pollinated to generate T1 generations and genotyped using the primers flanking the target region. Different knockout lines with homozygous mutations showed the same phenotype and all exhibited glabrous glumes and leaves (Fig. 3A–3F). These results further indicated that the *OBART05G01300* is *GLAG5*.

Interestingly, the homozygous mutations also exhibited other phenotypes. The auricles and ligules of wild rice W1411 carried trichomes, while the auricles of CAS9-GLAG5 was disappeared (Fig. 3G and 3H). Moreover no trichomes appeared on the ligules of the homozygous mutants (Fig. S6H and S6I). In addition, the glumes of CAS9-GLAG5 were malformed and accompanied by sterile, short and smooth awns (Fig. 3I, 3J, S6A–S6G). Compared with W1411, the CAS9-GLAG5 plants were dwarfed (Fig. 3K).

A transposon in the promoter of *GLAG5* led to glabrous glumes in *O. glaberrima*

To identify the mutation responsible for glabrous glumes, we analyzed the 478 bp inserted into the promotor of *GLAG5* that lay 139 bp upstream of the translation

initiation site. On the basis of the TE classification characteristics, there was an 8-bp target-site duplication and 23-bp terminal inverted repeat in the 478-bp insertion. We identified the insertion to be an *hAT* transposon, which is a Class II transposon (Fig. 4A).

Gene expression is usually repressed by the TE insertion into a promotor. To determine the effects of the *hAT* transposon on gene expression, we examined the expression pattern of *GLAG5* using quantitative RT-PCR (qRT-PCR). The TE reduced the expression of *GLAG5* in the leaves, lead sheaths, and young panicles (Fig. 4B). Moreover, GUS activity in transgenic plants expressing the β-glucuronidase reporter gene fused to a *GLAG5* promoter fragment was detected in the young panicles, ligules and leaves (Fig. S7A–S7D).

To further validate the influence of the *hAT* TE, we independently cloned the W1411 and IRGC104165 sequences that differed into reporter constructs upstream of the promoter of the cauliflower mosaic virus, the firefly luciferase (LUC) ORF and the nopaline synthase terminator and compared their effects on gene expression in rice protoplasts. The rice construct containing the *hAT* TE significantly repressed LUC activity relative to the W1411 construct without the *hAT* TE (Fig. 4C).

In general, a TE insertion induces epigenetic modifications of either DNA or histones. To clarify the molecular mechanisms behind the promoter-associated inhibitory activity of *GLAG5*, the sequence around the *hAT* TE insertion region was analyzed using http://www.urogene.org/methprimer/. Two CpG islands in the *hAT* TE insertion region were identified (Fig. S7E). Then, the methylation status of the CpG sites within the two CpG islands was detected using methylation-specific PCR. As shown in Fig. 4D, the *hAT* TE insertion caused DNA methylation of IRGC104165.

In summary, the *hAT* TE insertion caused DNA methylation that reduced the expression of *GLAG5*, resulting in glabrous glumes.

Evidence of positive selection on the *GLAG5* region in *O. glaberrima*

To further determine whether other glabrous African cultivated rice lines are the result of TE insertions, we genotyped this locus in 12 diverse wild and 82 cultivated rice accessions. Interestingly, none of the investigated wild rice accessions carried the

hAT TE, indicating that the TE insertion is most likely a *de novo* mutation that occurred following early domestication (Table S1). Among the 82 cultivated rice varieties, the 58 varieties (approximately 71%) exhibiting glabrous glumes harbored the *hAT* TE; however, the other 24 varieties (approximately 29%) without glabrous glumes lacked the *hAT* TE (Fig. 5A and 5B). These results suggested that the *hAT* TE insertion in the promoter of *GLAG5* was under strong positive selection during the domestication of African cultivated rice.

Selective signatures from domestication include a reduction in nucleotide diversity and altered allele frequencies of the domestication-associated loci (Xu et al., 2011). A favorable mutant allele greatly affects the genetic diversity at the selected gene locus, and it has a strong impact on its surrounding genome regions. To determine the genomic regions resulting from the artificial selection of *glag5* in *O. glaberrima*, we examined and compared the nucleotide diversity (π) within the regions around *GLAG5* between wild and cultivated rice. As shown in Fig. 5C, an approximate 150-kb interval encompassing the *GLAG5* locus was influenced by the *hAT* TE during domestication.

Discussion

Crop domestication is a progress of genetically modifying a wild species to meet human needs (Doebley et al., 2006). The identification of key genes controlling domestication-related traits helps to reveal crop origins. The African cultivated rice *O. glaberrima* was independently domesticated from its wild progenitor *O. barthii* in Africa more than 3000 years ago. Genes reducing seed shattering and leading to erect plant architecture have been identified previously in African cultivated rice (Wu et al., 2017; Hu et al., 2018; Lv et al., 2018), and they provide insights into the domestication process.

Oryza glaberrima is characterized by its glabrous glumes. In this study, we identified the *GLAG5* gene, which controls trichome development in African wild rice, and found a *hAT* TE lying in the promoter region of the *GLAG5* gene. The TE reduced

GLAG5 expression levels in the glumes and leaves, which consequently resulted in their glabrous phenotypes in African cultivated rice. In addition to Africa, glabrous rice is cultivated in the USA and on the Yunnan-Guizhou Plateau in China. In a previous study, Angeles-Shim et al. (2012) found a few SNPs in the CDs between *O. glaberrima* (IRGC104038) and *Oryza sativa* Koshihikari, which resulted in three amino acid substitutions and a serine deletion. However, the sequence analysis showed that both glabrous and pubescent lines of *O. glaberrima* carry the same amino acid substitutions and deletion relative to Koshihikari, which implied that the changes in the CDs did not cause the glabrous glumes. However, the expression of *GLAG5* in glabrous cultivars was lower than in pubescent cultivars and no sequence differences were found in the 2-kb promoter regions of *O. glaberrima* and Koshihikari.

In the present study, the 2.5-kb sequence in the promoter region of *GLAG5* was compared between W14111 and IRGC104165. There was a 478-bp insertion (the *hAT* TE) in the *O. glaberrima* promoter region (Fig. S4). Among the 82 African cultivated rice varieties used in our research, the 59 (approximately 71%) varieties exhibiting glabrous glumes harbored the *hAT* TE; however, the other 23 lines exhibiting pubescent glumes (approximately 29%) lacked the *hAT* TE. The result indicated that the *hAT* TE insertion in the promoter of *GLAG5* led to glabrous glumes in African cultivated rice (Table S1). Two previous studies (Li et al., 2012; Zhang et al., 2012) showed that glabrous glumes and leaves are controlled by a reduction in *NUDA/GL-1* expression in Asian rice; however, no mutations in these CDs or promoters were found between the glabrous and pubescent cultivars of *O. sativa*. Thus, the results indicated that the glabrous glumes in Asian rice (*O. sativa*) and African rice (*O. glaberrima*) are of independent origin, resulting from different mutations.

In many tissues, HD-leucine zippers (HD-Zip) plays important roles during trichome initiation in the epidermis. In *Arabidopsis*, *GL2* is an HD transcription factor that promotes trichome initiation in shoots and mucilage biosynthesis in seeds, but inhibits root-hair formation in roots (Chen and Wang, 2019). In tomato, the HD-Zip transcription factor *Woolly* plays crucial roles in trichome formation and embryo development (Yang et al., 2011). An HD-Zip gene, *Tri*, is involved in multicellular

241 trichome initiation in cucumber (Wang et al., 2016). *GLAG5* contains an HD and a
242 *WUSCHEL* domain, which indicated that it belonged to the *WOX* gene family.
243 *GLAG5* also plays an important role in trichome initiation. In Asian rice, *OsWOX3B* is
244 an essential regulator of trichome initiation, and auxin-related genes are regulated by
245 *OsWOX3B*. Moreover, *HL6* induces the expression of several auxin-related genes that
246 determine trichome elongation, and this induction may be enhanced by *OsWOX3B*
247 interacting with *HL6* (Sun et al., 2017). *GLAG5* and *OsWOX3B* are homologous
248 genes, and, therefore, trichome development in African wild rice may also involve
249 *OsWOX3B*.

250 The *WOX* gene family also plays important roles in regulating the development
251 of plant tissues and organs (Zhang et al., 2010). Interestingly, loss-of-function
252 mutations of *GLAG5* not only cause glabrous glumes and leaves, but also result in
253 auricle disappearance, glabrous ligules, glume malformation, sterility, short and
254 smooth awns and dwarfing (Fig. 3G–3K). Thus, *GLAG5* has pleiotropic effects with
255 regard to glumes and sterility, which are necessary for the normal growth of rice.
256 Either intentionally or fortuitously, ancient humans in West Africa selected for rice
257 that have a TE insertion in the promoter region rather than a change in the gene
258 coding region. The presence of the TE through the DNA methylation of *GLAG5*
259 reduced its expression in the glumes and leaves, leading to glabrous glumes, while
260 maintaining normal spikelets and plant architecture.

261 A CACTA-like TE within the maize *ZmCCT10* promoter and a Harbinger-like
262 TE within the *ZmCCT9* promotor were strongly selected to facilitate the adaptation of
263 maize to higher latitudes after its domestication (Yang et al., 2013; Huang et al.,
264 2018). A TE insertion into the maize domestication gene *tb1* leads to an in increase in
265 apical dominance compared with its wild ancestor, teosinte, by increasing gene
266 expression (Studer et al., 2011). A MITE insertion represses *ZmNAC111* expression
267 through the DNA and histone methylation of its nearby regions, and this is
268 significantly associated with natural variations in maize drought tolerance (Mao et al.,
269 2015). Thus, TE insertions into promotor regions can finely regulate gene expression
270 and selection for the mutated gene promoter is an important aspect of crop

domestication.

Materials and Methods

Plant materials and growth conditions

To clone of *GLAG5*, W1411 and IRGC104165 plants were used. W1411, an African wild rice accession (*O. barthii*), was collected from Sierra Leone. IRGC104165, a cultivar of African rice (*O. glaberrima*), was collected from Guinea. The F_2 segregating population was derived from the cross between W1411 and IRGC104165. For the genetic transformation experimentat, we used an accession of IRGC104038 (*O. glaberrima*), WK21. Details of the 82 *O. glaberrima* accessions and 12 *O. barthii* accessions are listed in Table S1. All plants were grown under field conditions in Beijing or Sanya, Hainan Province, China.

Phenotype observations

Leaves and auricles were dissected, fixed overnight at 4°C in 1:1:18 formalin:glacial acetic:70% ethanol and dehydrated in a graded ethanol series (50%, 70%, 80%, 90%, 95% and 100%). Samples were critical-point-dried, mounted, gold plated, observed and photographed using a Hitachi S-2460 scanning electron microscope.

Vector construction

The complementary vector was generated by placing a 4.044-kb genomic fragment from W1411, which contained the entire *GLAG5* ORF (0.841 kb), the promoter region (2.506 kb) and the 3′ -untranslated region (0.697 kb), into pCAMBIA1300. The complementary vector was transformed into WK21. The genome-editing vector of *GLAG5* was generated using the target ‘tgctggaggagatgtacag’. The genome-editing vectors was transformed into W1411. To construct the *GLAG5* promoter:GUS fusion plasmid, an approximately 2-kb DNA fragment comprising the promoter sequence of *GLAG5* in W1411 was amplified and cloned into the pCAMBIA1301-GUS-nos vector. The pCAMBIA1301-GUS-nos vector was transformed into ZH17.

RNA extraction and qRT-PCR.

Total RNA was extracted using TRIzol reagent (Invitrogen) and purified using an RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's instructions. First-strand cDNA synthesis was carried out using a BcaBEST RNA PCR Kit (TaKaRa). The qRT-PCR was performed using the first-strand cDNA as the template. Endogenous ubiquitin transcripts were used to normalize the expression levels. The qRT-PCR was performed on the ABI Prism 7900 Sequence Detection System (Applied Biosystems) using SYBR Green. Each set of experiments was repeated three times.

Dual-luciferase Reporter Assay System to identify promoter activity

Promoter sequences of 2506 bp and 2984 bp were amplified from W1411 and IRGC104165 genomic DNA, respectively, using the specific pLUCL-F/R primers and transformed into the vector pGreenII0800-LUC using an in-fusion enzyme (Genebank Biosciences) (Hellens et al., 2005) to generate proW1411:LUC and pro IRGC104165:LUC plasmids, respectively. The plasmids were co-infiltrated into rice protoplasts of ZH17. Firefly and Renilla LUC activities were measured using a Dual-luciferase Reporter Assay System (Promega).

Methylation-specific PCR

The DNA was treated using a DNA Bisulfite Conversion Kit. Then, methylation-specific PCR was performed using a Methylation-specific PCR Kit. Both kits were purchased from TIANGEN BIOTECH.

Subcellular localization.

The *GLAG5* CDs of W1411 without the stop codon was fused to the N-terminus of green fluorescent protein driven by the 35S promoter. The constructed vector was co-infiltrated into rice protoplasts of ZH17 and observed using an Olympus FV1000 fluorescence microscope.

Primers

The primer sequences used in this research are listed in Table S2.

The sequencing data

The next-generation sequencing data of 94 *O. barthii* samples, with the GenBank accession numbers SRR1206362–SRR1206455, were downloaded from the SRA of NCBI, and the sequencing data for the 110 *O. glaberrima* samples, with the GenBank accession numbers SRR3231659–SRR3231748 and SRR1206500–SRR1206519, were also downloaded from the SRA of NCBI (Wang et al., 2014; Meyer et al., 2016). A total of 3,619,010 and 12,004,364 SNPs were identified across the 110 *O. glaberrima* accessions and 94 *O. barthii* accessions, respectively, and were used to measure genome-wide levels of π.

DNA sequencing data analysis and SNP calling

The DNA re-sequencing data of all the samples were subjected to quality control using Trimommatic (Bolger et al., 2014), and then, the remaining qualified reads were aligned using BWA to the reference sequence, the *Oryza glaberrima* genome (IRGC104165). The alignments were sorted, and duplicate reads were removed using SamTools (https://www.htslib.org.). Final SNP calling was performed using the Genome Analysis Toolkit (GATK) HaplotypeCaller with the default parameters. For the SNP set from the 204 samples, the filter settings were as follows: DP > 16,200, QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 3.0, MQRankSum < −12.5 and ReadPosRankSum < −8.0.

Selection sweeps analyses

Genome-wide π was calculated using VCFtools V0.1.17 for nonoverlapping 100-kb windows and a 1-kb step size across the genomes of *O. glaberrima* and *O. barthii*. Distributions of π and π_{wild} (π_w) /$\pi_{cultivated}$ (π_c) ratios within the regions associated with the *hAT* TE were determined. The distribution of π_w/π_c along the chromosome was

plotted using an R script.

CRediT authorship contribution statement

Leqin Chang: Investigation, Validation, Writing - Original draft, Review & Editing. **Min Hu**: Data curation, Investigation. **Jing Ning**: Data curation, Investigation. **Wei He**: Data curation, Validation. **Jiayu Gao**: Data curation. **Marie-Noelle Ndjiondjop**: Writing - Review & Editing. **Yongcai Fu**: Resources. **Fengxia Liu**: Writing - Review & Editing. **Hongying Sun**: Data curation. **Ping Gu**: Resources. **Chuanqing Sun**: Resources, Supervision. **Zuofeng Zhu**: Conceptualization, Project administration, Resources, Supervision, Writing - Original draft, Review & Editing.

Conflict of interest

The authors declare no conflict of interests.

Acknowledgments

We thank the International Rice Research Institute, the Japan National Institute of Genetics and the United States Department of Agriculture-Agricultural Research Service for providing the wild rice and cultivated rice samples. This research was supported by the National Natural Science Foundation of China (31925029). The funders had no role in the study design, data collection and analysis, decision to publish or manuscript preparation.

References

Ammiraju, J.S.S., Lu, F., Sanyal, A., Yu, Y., Song, X., Jiang, N., Pontaroli, A.C., Rambo, T., Currie, J., Collura, K., Talag, J., Fan, C.Z., Goicoechea, J.L., Zuccolo, A., Chen, J., Bennetzen, J.L., Chen, M.S., Jackson, S., Wing, R.A., 2008. Dynamic Evolution of Oryza Genomes Is Revealed by Comparative Genomic Analysis of a Genus-Wide Vertical Data Set. Plant Cell. 20, 3191-3209.

Angeles-Shim, R.B., Asano, K., Takashi, T., Shim, J., Kuroha, T., Ayano, M., Ashikari, M., 2012. A WUSCHEL-related homeobox 3B gene, depilous (dep), confers glabrousness of rice leaves and glumes. Rice 5, 1-12.

Atkinson, P.W., Warren, W.D., Obrochta, D.A., 1993. The Hobo Transposable Element of Drosophila Can Be Cross-Mobilized in Houseflies and Excises Like the Ac Element of Maize. Proc. Natl. Acad. Sci. U. S. A. 90, 9693-9697.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Chen, J.F., Huang, Q.F., Gao, D.Y., Wang, J.Y., Lang, Y.S., Liu, T.Y., Li, B., Bai, Z.T., Goicoechea, J.L., Liang, C.Z., Chen, C.B., Zhang, W.L., Sun, S.H., Liao, Y., Zhang, X.M., Yang, L., Song, C.L., Wang, M.J., Shi, J.F., Liu, G., Liu, J.J., Zhou, H.L., Zhou, W.L., Yu, Q.L., An, N., Chen, Y., Cai, Q., Wang, B., Liu, B.H., Min, J.M., Huang, Y., Wu, H.L., Li, Z.Y., Zhang, Y., Yin, Y., Song, W.Q., Jiang, J.M., Jackson, S.A., Wing, R.A., Wang, J., Chen, M.S., 2013. Whole-genome sequencing of Oryza brachyantha reveals mechanisms underlying Oryza genome evolution. Nat. Commun. 4, 1595.

Chen, S.Y., Wang, S.C., 2019. GLABRA2, a Common Regulator for Epidermal Cell Fate Determination and Anthocyanin Biosynthesis in Arabidopsis. Int. J. Mol. Sci. 20, 4997.

Feschotte, C., Jiang, N., Wessler, S.R., 2002. Plant transposable elements: Where genetics meets genomics. Nat. Rev. Genet. 3, 329-341.

Feschotte, C., Pritham, E.J., 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331-368.

Friedli, M., Trono, D., 2015. The Developmental Control of Transposable Elements and the Evolution of Higher Species. Annu. Rev. Cell. Dev. Bi. 31, 429-451.

Fujino, K., Sekiguchi, H., 2011. Transposition behavior of nonautonomous a hAT superfamily transposon nDart in rice (Oryza sativa L.). Mol. Genet. Genomics. 286, 135-142.

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., Laing, W.A., 2005. Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods. 1, 13.

Hu, M., Lv, S.W., Wu, W.G., Fu, Y.C., Liu, F.X., Wang, B.B., Li, W.G., Gu, P., Cai, H.W., Sun, C.Q., Zhu, Z.F., 2018. The domestication of plant architecture in African rice. Plant J. 94, 661-669.

Huang, C., Sun, H.Y., Xu, D.Y., Chen, Q.Y., Liang, Y.M., Wang, X.F., Xu, G.H., Tian, J.G., Wang, C.L., Li, D., Wu, L.S., Yang, X.H., Jin, W.W., Doebley, J.F.,

Tian, F., 2018. ZmCCT9 enhances maize adaptation to higher latitudes. Proc. Natl. Acad. Sci. U. S. A. 115, E334-E341.
Hudson, A.D., Carpenter, R., Coen, E.S., 1990. Phenotypic Effects of Short-Range and Aberrant Transposition in Antirrhinum-Majus. Plant Mol. Biol. 14, 835-844.
Li, J.J., Yuan, Y.D., Lu, Z.F., Yang, L.S., Gao, R.C., Lu, J.G., Li, J.Y., Xiong, G.S., 2012. Glabrous Rice 1, encoding a homeodomain protein, regulates trichome development in rice. Rice. 5, 32.
Li, X.K., Guo, K., Zhu, X.B., Chen, P., Li, Y., Xie, G.S., Wang, L.Q., Wang, Y.T., Persson, S., Peng, L.C., 2017. Domestication of rice has reduced the occurrence of transposable elements within gene coding regions. Bmc. Genomics. 18, 55.
Lim, H.H., Domala, Z., Joginder, S., Lee, S.H., Lim, C.S., Abubakar, C.M., 1984. Rice Millers Syndrome - a Preliminary-Report. Brit. J. Ind. Med. 41, 445-449.
Lv, S.W., Wu, W.G., Wang, M.H., Meyer, R.S., Ndjiondjop, M.N., Tan, L.B., Zhou, H.Y., Zhang, J.W., Fu, Y.C., Cai, H.W., Sun, C.Q., Wing, R.A., Zhu, Z.F., 2018. Genetic control of seed shattering during African rice domestication. Nat. Plants. 4, 331-337.
Ma, X., Fan, J., Wu, Y., Zhao, S., Zheng, X., Sun, C., Tan, L., 2020. Whole-genome de novo assemblies reveal extensive structural variations and dynamic organelle-to-nucleus DNA transfers in African and Asian rice. Plant J. 104, 596-612.
Mao, H.D., Wang, H.W., Liu, S.X., Li, Z., Yang, X.H., Yan, J.B., Li, J.S., Tran, L.S.P., Qin, F., 2015. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nat. Commun. 6, 8326.
Meyer, R.S., Choi, J.Y., Sanches, M., Plessis, A., Flowers, J.M., Amas, J., Dorph, K., Barretto, A., Gross, B., Fuller, D.Q., Bimpong, I.K., Ndjiondjop, M.N., Hazzouri, K.M., Gregorio, G.B., Purugganan, M.D., 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. Nat. Genet. 48, 1083-1088.
Molina, J., Sikora, M., Garud, N., Flowers, J.M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B.A., Bustamante, C.D., Boyko, A.R., Purugganan, M.D., 2011. Molecular evidence for a single evolutionary origin of domesticated rice. Proc. Natl. Acad. Sci. U. S. A. 108, 8351-8356.
Peng, Y., Zhang, Y.Y., Gui, Y.J., An, D., Liu, J.Z., Xu, X., Li, Q., Wang, J.M., Wang, W., Shi, C.H., Fan, L.J., Lu, B.R., Deng, Y.W., Teng, S., He, Z.H., 2019. Elimination of a Retrotransposon for Quenching Genome Instability in Modern Rice. Mol. Plant. 12, 1395-1407.
Purugganan, M.D., 2014. An evolutionary genomic tale of two rice species. Nat Genet 46, 931-932.
Sakai, H., Ikawa, H., Tanaka, T., Numa, H., Minami, H., Fujisawa, M., Shibata, M., Kurita, K., Kikuta, A., Hamada, M., Kanamori, H., Namiki, N., Wu, J.Z., Itoh, T., Matsumoto, T., Sasaki, T., 2011. Distinct evolutionary patterns of Oryza glaberrima deciphered by genome sequencing and comparative analysis. Plant J. 66, 796-805.
Stein, J.C., Yu, Y., Copetti, D., Zwickl, D.J., Zhang, L., Zhang, C.J., Chougule, K.,

Gao, D.Y., Iwata, A., Goicoechea, J.L., Wei, S.R., Wang, J., Liao, Y., Wang, M.H., Jacquemin, J., Becker, C., Kudrna, D., Zhang, J.W., Londono, C.E.M., Song, X., Lee, S., Sanchez, P., Zuccolo, A., Ammiraju, J.S.S., Talag, J., Danowitz, A., Rivera, L.F., Gschwend, A.R., Noutsos, C., Wu, C.C., Kao, S.M., Zeng, J.W., Wei, F.J., Zhao, Q., Feng, Q., El Baidouri, M., Carpentier, M.C., Lasserre, E., Cooke, R., Farias, D.D., da Maia, L.C., dos Santos, R.S., Nyberg, K.G., McNally, K.L., Mauleon, R., Alexandrov, N., Schmutz, J., Flowers, D., Fan, C.Z., Weigel, D., Jena, K.K., Wicker, T., Chen, M.S., Han, B., Henry, R., Hsing, Y.I.C., Kurata, N., de Oliveira, A.C., Panaud, O., Jackson, S.A., Machado, C.A., Sanderson, M.J., Long, M.Y., Ware, D., Wing, R.A., 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza (vol 50, pg 285, 2018). Nat. Genet. 50, 1618-1618.

Studer, A., Zhao, Q., Ross-Ibarra, J., Doebley, J., 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43, 1160-U1164.

Sun, W.Q., Gao, D.W., Xiong, Y., Tang, X.X., Xiao, X.F., Wang, C.R., Yu, S.B., 2017. Hairy Leaf 6, an AP2/ERF Transcription Factor, Interacts with OsWOX3B and Regulates Trichome Formation in Rice. Mol. Plant. 10, 1417-1433.

Van Andel, T., 2010. African Rice (Oryza glaberrima Steud.): Lost Crop of the Enslaved Africans Discovered in Suriname(1). Econ. Bot. 64, 1-10.

Wang, M.H., Yu, Y., Haberer, G., Marri, P.R., Fan, C.Z., Goicoechea, J.L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J.S.S., Cossu, R.M., Maldonado, C., Chen, J., Lee, S., Sisneros, N., de Baynast, K., Golser, W., Wissotski, M., Kim, W., Sanchez, P., Ndjiondjop, M.N., Sanni, K., Long, M.Y., Carney, J., Panaud, O., Wicker, T., Machado, C.A., Chen, M.S., Mayer, K.F.X., Rounsley, S., Wing, R.A., 2014. The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nat. Genet. 46, 982-988.

Wang, Y.L., Nie, J.T., Chen, H.M., Guo, C.L., Pan, J., He, H.L., Pan, J.S., Cai, R., 2016. Identification and mapping of Tril, a homeodomain-leucine zipper gene involved in multicellular trichome initiation in Cucumis sativus. Theor. Appl. Genet. 129, 305-316.

Wu, W.G., Liu, X.Y., Wang, M.H., Meyer, R.S., Luo, X.J., Ndjiondjop, M.N., Tan, L.B., Zhang, J.W., Wu, J.Z., Cai, H.W., Sun, C.Q., Wang, X.K., Wing, R.A., Zhu, Z.F., 2017. A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. Nat. Plants. 3, 17064.

Yang, C.X., Li, H.X., Zhang, J.H., Luo, Z.D., Gong, P.J., Zhang, C.J., Li, J.H., Wang, T.T., Zhang, Y.Y., Lu, Y.E., Ye, Z.B., 2011. A regulatory gene induces trichome formation and embryo lethality in tomato. Proc. Natl. Acad. Sci. U. S. A. 108, 11836-11841.

Yang, Q., Li, Z., Li, W.Q., Ku, L.X., Wang, C., Ye, J.R., Li, K., Yang, N., Li, Y.P., Zhong, T., Li, J.S., Chen, Y.H., Yan, J.B., Yang, X.H., Xu, M.L., 2013. CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. U.

S. A. 110, 16969-16974.
Zhang, H.L., Wu, K., Wang, Y.F., Peng, Y., Hu, F.Y., Wen, L., Han, B., Qian, Q., Teng, S., 2012. A WUSCHEL-like homeobox gene, OsWOX3B responses to NUDA/GL-1 locus in rice. Rice 5, 30.
Zhang, X., Zong, J., Liu, J.H., Yin, J.Y., Zhang, D.B., 2010. Genome-Wide Analysis of WOX Gene Family in Rice, Sorghum, Maize, Arabidopsis and Poplar. J Integr Plant Biol 52, 1016-1026.

Figure Legends

Fig. 1. A phenotypic comparison of wild rice W1411 (*O. barthii)* and cultivated rice IRGC104165 (*O. glaberrima*). **A** and **B**: Stereomicroscope images of the seeds of W1411 and IRGC104165. **C**–**F**: SEM images of the seeds of W1411 and IRGC104165, Scale bars, 24 mm (**A** and **B**); 200 μm (**C** and **D**); 1 mm (**E** and **F**).

Fig. 2. Map-based cloning of *GLAG5*. **A**: The target gene was mapped to the interval between C03 and RM593 on the short arm of chromosome 5. *GLAG5* was finally fine mapped to an interval approximately 15.9-kb between C02 and C06. **B**: The construction of the complementary vector: a 4.04-kb DNA fragment containing the entire *GLAG5* gene region from the W1411 genome was inserted into the binary vector pCAMIA1300 to generate the complementary transformation vector p*GLAG5*. **C**–**H**: The phenotypes of a control plant WK21 and transgenic plant (p*GLAG5*) harboring the complementary transformation vector p*GLAG5*. Stereomicroscope images of the seeds of WK21 and p*GLAG5* (**C** and **D**). SEM images of the seeds (**E** and **F**) and leaves of WK21 and p*GLAG5* (**G** and **H**), respectively. Scale bars, 20 mm (**C** and **D**); 800 μm (**E** and **F**); 7 mm (**G** and **H**).

Fig. 3. The phenotype of gene-editing plant CAS9-GLAG5. **A** and **B**: Stereomicroscope images of the seeds of W1411 and gene-editing plant CAS9-GLAG5, 20 mm (**A** and **B**); **C** and **D**: SEM images of the seeds of W1411 and CAS9-GLAG5, 800 μm (**C** and **D**); **E** and **F**: SEM images of the leaves of W1411 and CAS9-GLAG5, 7 mm (**E** and **F**); **G** and **H**: The auricles (direction of arrow) of W1411 and CAS9-GLAG5, 6 mm (**G** and **H**); **I** and **J**: The glumes of W1411 and CAS9-GLAG5, 30 mm (**I** and **J**); (**K**) Plant architecture of W1411 and CAS9-GLAG5, 18 cm (**K**).

Fig. 4. A transposon in the promoter of *GLAG5* led to glabrous glumes in *O.*

glaberrima. **A**: Sequence analysis of the *hAT* transposon. **B**: Tissue expression of *GLAG5*. The expression of *GLAG5* was detected in root, stem, leaf, leaf sheath and young panicle. Values are means ± SDs (n = 3). **, $P < 0.01$; *, $0.01 < P < 0.05$, (two-tailed Student's t-test). **C**: Dual-LUC reporter assays showing that the promoter activities of *GLAG5* from W1411 and IRGC104165 were affected differently by the *hAT* transposon. Values are means ± SDs (n = 4). **, $P < 0.01$, (two-tailed Student's t-test). **D**: Methylation-specific PCR. SD, standard deviation.

Fig. 5. Evidence of positive selection on the *GLAG5* gene region in *O. glaberrima*. **A**: Proportion of glabrous glumes in wild rice *O. barthii*. **B**: Proportion of glabrous glumes in African cultivated rice *O. glaberrima*. **C**: Distribution of nucleotide diversity (π) and π_{wild} (π_w)/$\pi_{cultivated}$ (π_c) ratios within the regions associated with the *hAT* transposon. The π levels within a 100-kb sliding window with a 1-kb step size in *O. glaberrima* and *O. barthii*. The translucent orange bars at the bottom represent selective sweeps.

Fig 1

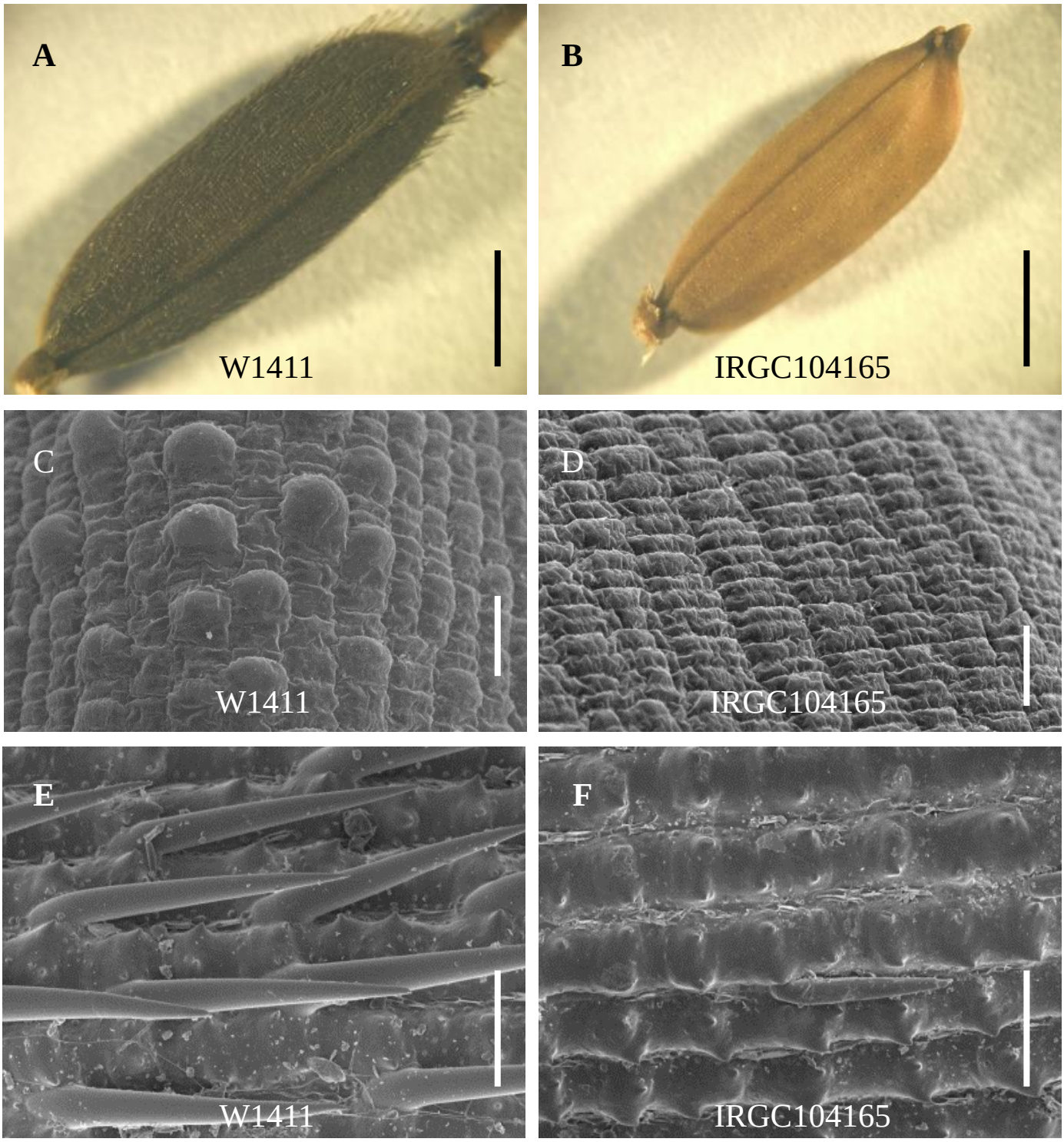
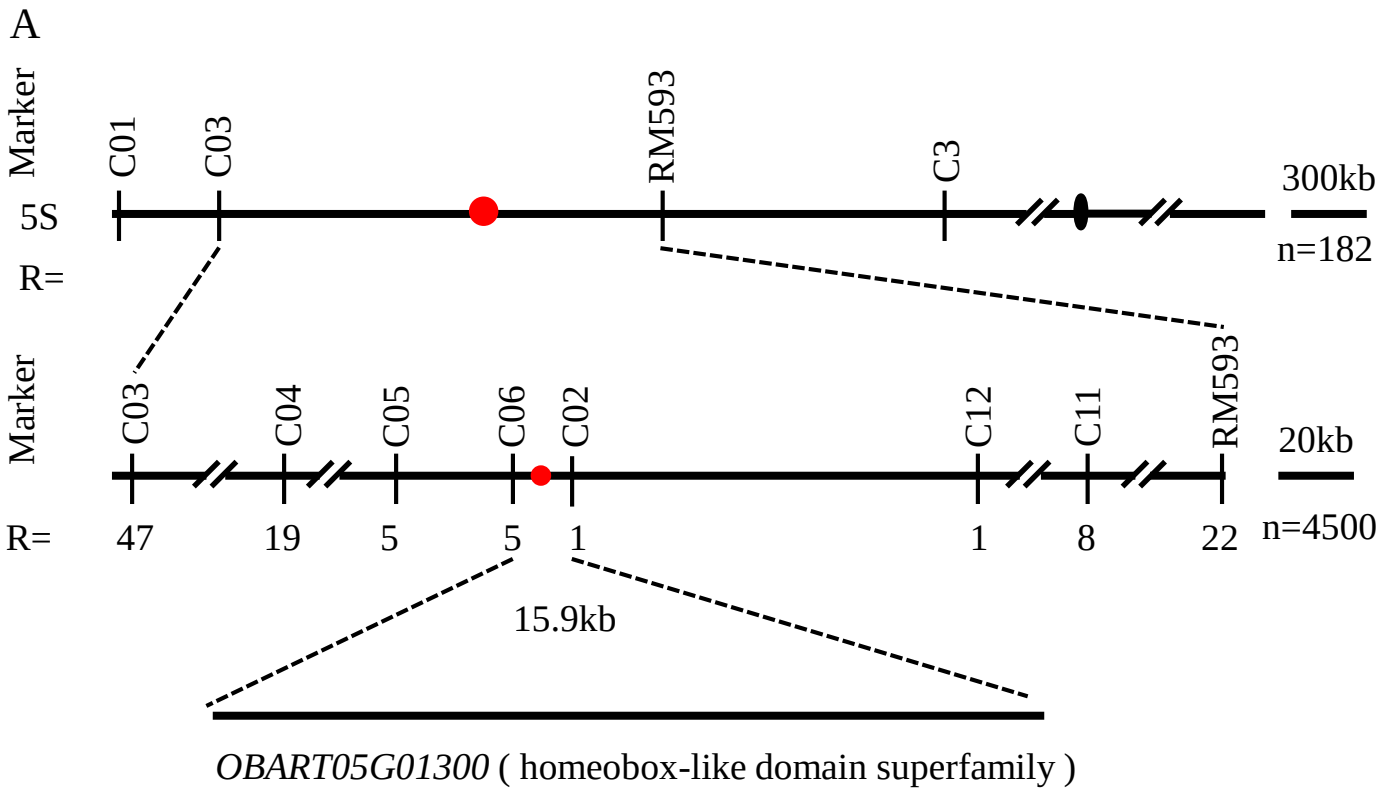
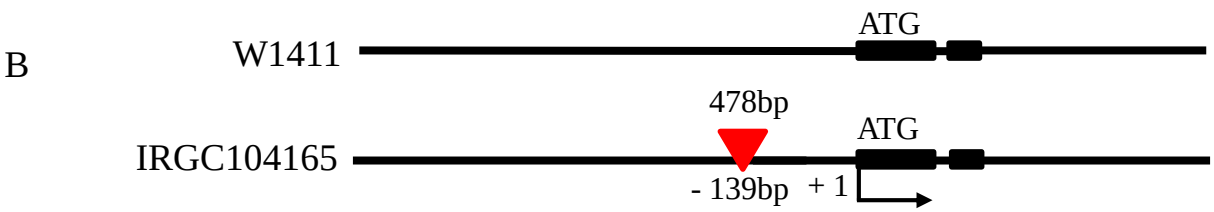

A
W1411
B
IRGC104165
C
W1411
D
IRGC104165
E
W1411
F
IRGC104165

Fig 2


A
Marker
C01
C03
RM593
C3
300kb
5S
n=182
R=
Marker
C03
C04
C05
C06
C02
C12
C11
RM593
20kb
R=
47
19
5
5
1
1
8
22
n=4500
15.9kb
OBART05G01300 (homeobox-like domain superfamily)


B
W1411
ATG
478bp
IRGC104165
ATG
- 139bp
+ 1

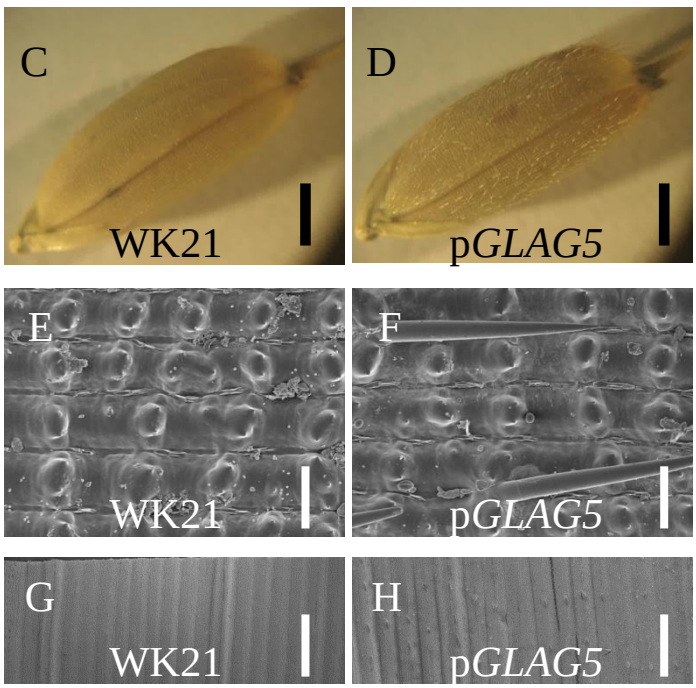

C
WK21
D
pGLAG5
E
WK21
F
pGLAG5
G
WK21
H
pGLAG5

Fig 3

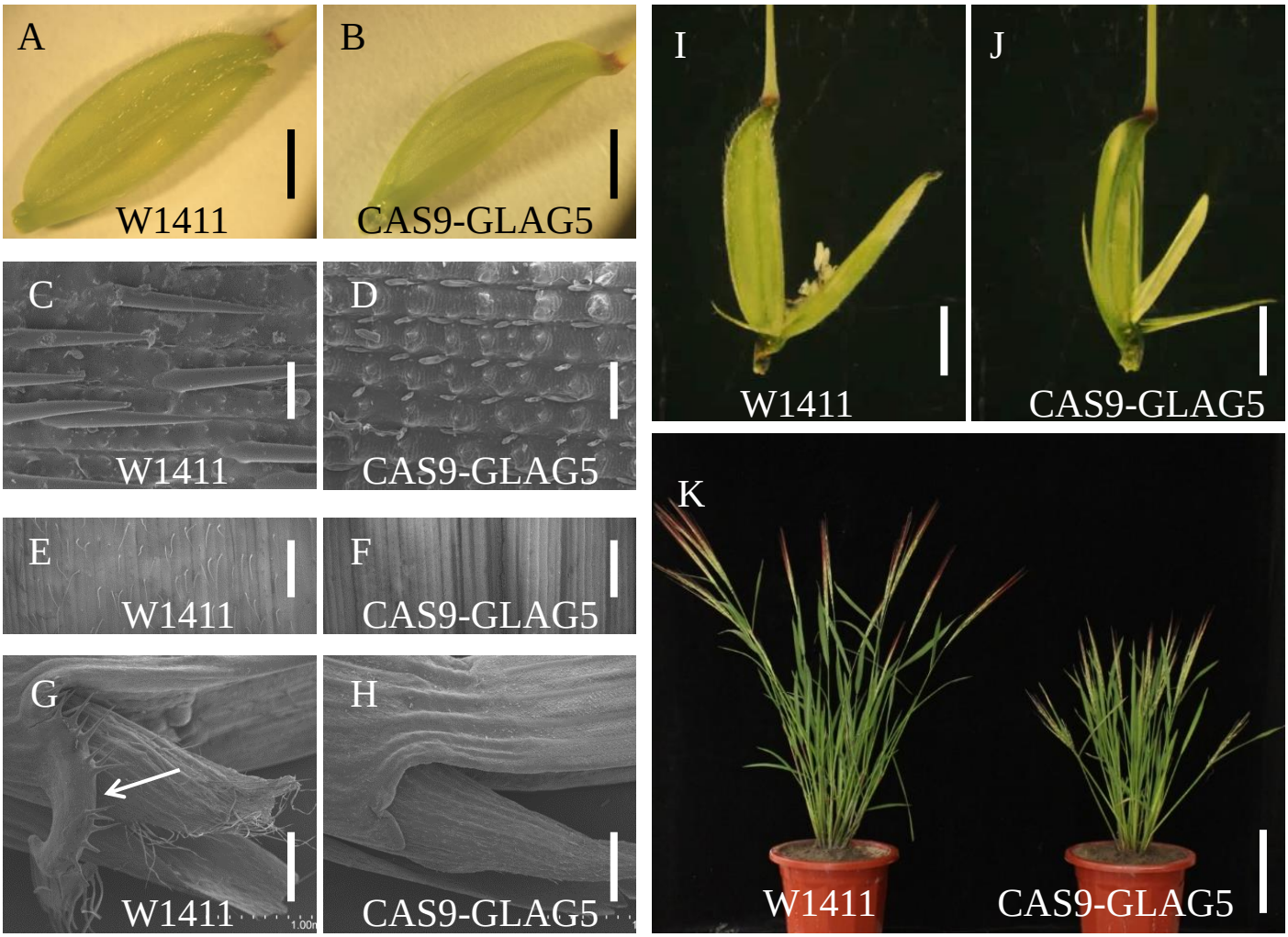

A
W1411
B
CAS9-GLAG5
C
W1411
D
CAS9-GLAG5
E
W1411
F
CAS9-GLAG5
G
W1411
H
CAS9-GLAG5
I
W1411
J
CAS9-GLAG5
K
W1411
CAS9-GLAG5

Fig 4

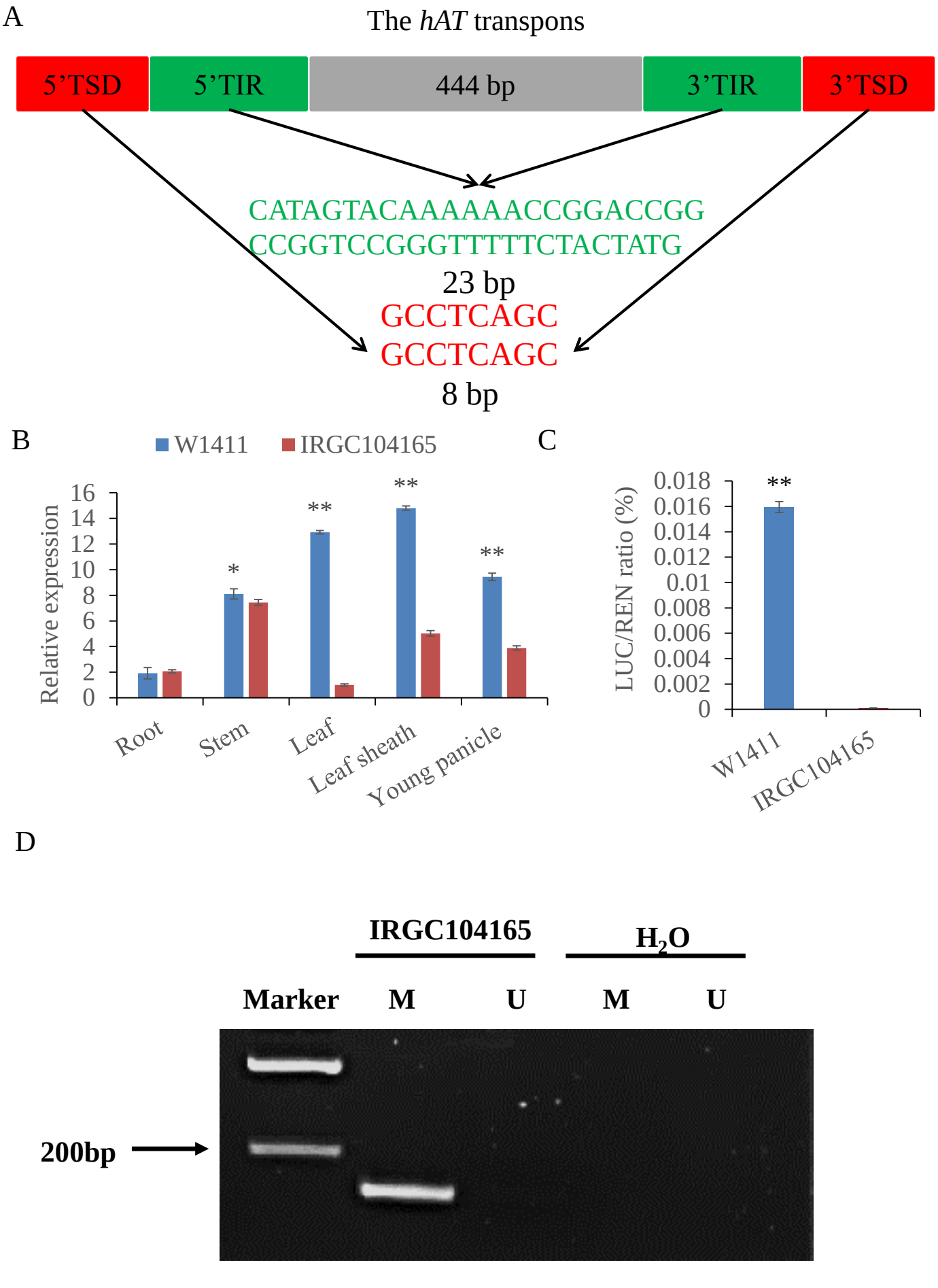
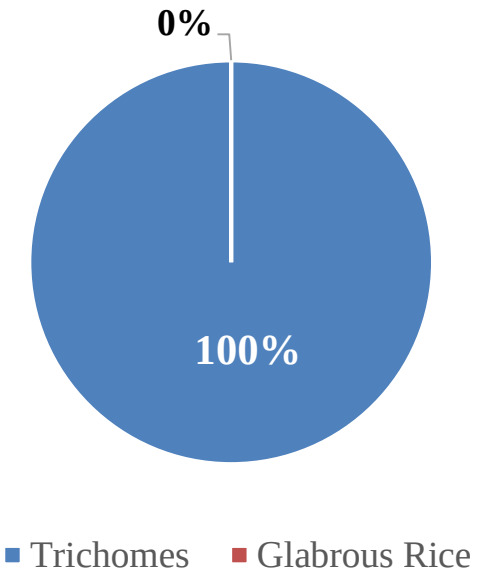

A
The hAT transpons
5'TSD
5'TIR
444 bp
3'TIR
3'TSD
CATAGTACAAAAAACCGGACCGG
CCGGTCCGGGTTTTTCTACTATG
23 bp
GCCTCAGC
GCCTCAGC
8 bp
B
W1411
IRGC104165
Relative expression
16
14
12
10
8
6
4
2
0
*
**
**
**
Root
Stem
Leaf
Leaf sheath
Young panicle
C
LUC/REN ratio (%)
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
**
W1411
IRGC104165
D
IRGC104165
H2O
Marker
M
U
M
U
200bp

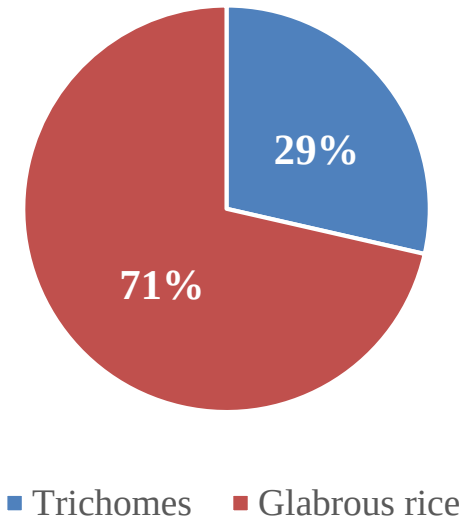
Fig 5

A



O. barthii

B



O. glaberrima

C