Journal of Integrative Agriculture 2026, 25(8): 3118–3125

Available online at ScienceDirect

Journal of Integrative Agriculture



RESEARCH ARTICLE

The genetic control of long sterile lemma during African rice domestication

Liang Luo[1, 2*], Wenkai Luo[1*], Bin Gao[1*], Kaihong Wang[1], Sijing Ma[1], Marie-Noelle Ndjiondjop[3], Chuanqing Sun[4], Zuofeng Zhu[1#], Leqin Chang[1, 2#]

[1] *Frontiers Science Center for Molecular Design Breeding/National Center for Evaluation of Agricultural Wild Plants (Rice)/Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China*
[2] *Sanya Institute, China Agricultural University, Sanya 572025, China*
[3] *Africa Rice Center, Cotonou 01 BP 4029, Benin*
[4] *State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, Beijing 100193, China*

Highlights

- The first long sterile lemma gene *lsl3* of African cultivated rice was reported.
- The geographical distribution analysis of the *lsl3* haplotypes provides a reference for the molecular evolution of African cultivated rice.
- The key mutations comparison of *lsl3* between African cultivated rice and Asian cultivated rice provides new evidence for the parallel domestication of rice and the independent origin of African cultivated rice.

Abstract

The inflorescence and spikelet structural units of rice significantly impact grain development. Among the grasses, the sterile lemma represents a unique spikelet organ exclusive to rice. As most studies on sterile lemma genes primarily focus on Asian cultivated rice (*Oryza sativa* L.), genes influencing the sterile lemma phenotype in African rice remain unreported. This study identified *lsl3*, a gene located on the short arm of chromosome 7 that controls the sterile lemma length in African rice. The key mutation of *lsl3*, leading to long sterile lemmas in African rice, was discovered and differed from that in Asian rice. The investigation of *lsl3* not only provides a reference for the molecular evolution of cultivated rice in Africa but also offers new evidence supporting the parallel domestication of rice and the independent origin of *Oryza glaberrima*.

Keywords: *Oryza glaberrima* Steud., *Oryza sativa* L., long sterile lemma, independent origin, domestication

1. Introduction

Rice serves as the staple food for over half of the global population. Two distinct rice cultivated species (*Oryza sativa* L. and *Oryza glaberrima* Steud.) have been identified, originating independently from Asia and Africa, respectively (Zhang *et al.* 2014; Huang *et al.* 2015; Meyer *et al.* 2016). These two cultivated rice species offer a unique opportunity to investigate the parallel domestication of crops. African wild rice (*Oryza barthii*) was domesticated into African cultivated rice (*O. glaberrima* Steud.) approximately 3,000 years ago (Purugganan 2014). Throughout long-term natural selection and domestication, *O. glaberrima* has accumulated a wealth of genetic resources and possesses numerous crucial genes that have been identified, including those responsible for high temperature and drought tolerance, reduced seed shattering, and erect plant architecture (Linares 2002; Wambugu *et al.* 2013; Li X M *et al.* 2015; Wu *et al.* 2017; Hu *et al.* 2018). Consequently, *O. glaberrima* warrants further exploration and utilization.

Floral organs represent a crucial indicator of the transition from vegetative growth to reproductive growth in flowering plants. Through genetic studies on the flower organs of various dicotyledonous plants, particularly the model species *Arabidopsis thaliana* and *snapdragon*, scientists have proposed the ABCDE model of flower development (Coen and Meyerowitz 1991; Pelaz *et al.*

Received 8 July 2024; Received in revised form 18 September 2024; Accepted 8 October 2024; Available online 13 November 2024
[#]Correspondence Zuofeng Zhu, E-mail: zhuzf@cau.edu.cn; Leqin Chang, E-mail: changleqin@cau.edu.cn
[*] These authors contributed equally to this study.

© 2026 CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer review under responsibility of Editorial Board of *Journal of Integrative Agriculture*.
doi: 10.1016/j.jia.2024.11.031

2000; Theissen 2001; Ditta *et al.* 2004; Theissen and Ditta 2004). This genetic model is applicable not only to eudicots but also to monocots, including certain grass species such as rice (*O. sativa* L.) and maize (*Zea mays*). Rice (*O. sativa* L.) serves as an ideal subject for investigating flower development in monocotyledonous plants. It possesses a unique structural unit consisting of spikelets and florets, which is characteristic of some grasses (Bommert *et al.* 2005; Zanis 2007). Rice spikelet exhibits distinct features compared to other grass crops, comprising two rudimentary glumes, a pair of sterile lemmas, and a fertile floret (Schmidt and Ambrose 1998; Ren *et al.* 2013). The floret of a rice spikelet contains a lemma, a palea, two lodicules, six stamens, and a pistil consisting of an ovary and a stigma (Lombardo and Yoshida 2015). Currently, more than 10 genes have been identified as involved in the development of sterile lemmas. These genes exhibit varying structural characteristics and influence the phenotype of sterile lemmas directly or indirectly through diverse molecular mechanisms. *G1*, *PAP2/OsMADS34*, *NSG1/LRG1*, and *SAD1* inhibit the development of sterile lemmas (Yoshida *et al.* 2009; Kobayashi *et al.* 2010; Li W *et al.* 2015; Xu *et al.* 2020). *OsVIL2*, *OsIG1*, *EG1*, and *ASP1* not only inhibit the development of sterile lemmas but also affect their number (Li *et al.* 2009; Yoshida *et al.* 2012; Yang *et al.* 2013; Yoon *et al.* 2018). *FZP*, *SNB*, *OsIDS1*, and *MFS1* belong to the AP2/EREBP transcription factor family and promote the development of sterile lemmas (Yi *et al.* 2005; Lee *et al.* 2007; Lee and An 2012; Ren *et al.* 2013). However, the genes influencing sterile lemmas are primarily studied using *O. sativa* as the research material, and there are currently no reports on sterile lemma genes in African rice.

This study describes the isolation and characterization of the long sterile lemma gene (*lsl3*), which regulates sterile lemma length in African cultivated rice (*O. glaberrima*). This gene encodes a transcription factor, located on chromosome 7 of the rice genome. Moreover, guanine (G) deletion at nucleotide position 581 induces a frameshift in the protein reading frame, resulting in premature translation termination. This genetic alteration consequently generates the elongated sterile lemma phenotype observed in African rice.

2. Materials and methods

2.1. Plant materials and growth conditions

To clone *lsl3*, W1411, and IRGC104165 plants were utilized. W1411, an African wild rice accession (*O. barthii*), was obtained from Sierra Leone, while IRGC104165, a cultivar of African rice (*O. glaberrima*), was acquired from Guinea. The F_2 segregating population was generated through the cross between W1411 and IRGC104165. The pLSL3 construct was introduced into another African cultivated rice cultivar IRGC102305 (*O. glaberrima*) for the genetic transformation experiment. A comprehensive list of the *O. glaberrima* and *O. barthii* accessions was provided in Appendix A. All plants were cultivated under field conditions in either Beijing or Sanya, Hainan Province, China.

2.2. Phenotype observations

Seeds and sterile lemmas were dissected and fixed overnight at 4°C in a 1:1:18 solution of formalin, glacial acetic acid, and 70% ethanol. Subsequently, the samples were dehydrated using a graded ethanol series of 50, 70, 80, 90, 95, and 100% ethanol concentrations. Following dehydration, the samples underwent critical point drying, a process that replaces ethanol with liquid CO_2 under pressure, which is then evaporated to dryness. The dried samples were mounted and coated with a thin layer of gold using a sputter coater to enhance their conductivity and improve imaging quality. Finally, the prepared samples were observed and photographed using a Hitachi S-2460 scanning electron microscope (SEM), which produces high-resolution images of the sample surface topography by scanning it with a focused beam of electrons.

Young sterile lemmas of W1411 and IRGC104165 were fixed in 4% (w/v) paraformaldehyde, followed by a series of dehydration and infiltration processes. Subsequently, the tissues were embedded in paraffin (Paraplast Plus, Sigma-Aldrich). Utilizing a microtome (Leica RM2145, Leica Microsystems, Germany), the embedded tissues were sectioned into 8 μm-thick slices.

2.3. Molecular cloning of *lsl3*

The *lsl3* gene was mapped using an F_2 population derived from a cross between the W1411 and IRGC104165 rice varieties. The complementary vector was constructed by amplifying the 3,761-bp DNA fragment of the candidate gene *LSL3* from W1411 *via* PCR, which encompassed the 831-bp open reading frame (ORF), a 1,921-bp promoter region, and a 1,009-bp 3′-untranslated region. Subsequently, the PCR product was inserted into pCAMBIA1300. This construct was then transformed into IRGC102305 (*O. glaberrima*). Comprehensive information on the primers utilized in this study is shown in Appendix B.

2.4. RNA extraction and qRT-PCR analyses

Total RNA extraction from various rice tissues was conducted using TRIzol reagent (Invitrogen, Thermo Fisher Scientific, Waltham, USA). Complementary DNA (cDNA) synthesis was performed through reverse transcription utilizing the BcaBEST RNA PCR Kit (TaKaRa, Japan). Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was performed using TOROGreen QPCR Master mix (K23010910, Toroivd, Shanghai, China) and the *Actin1* gene was employed as an internal control. Each qRT-PCR assay was performed in triplicate. The primer sequences used in qRT-PCR are listed in Appendix B.

2.5. Subcellular localization

The coding sequence of *LSL3* from W1411, excluding the stop codon, was fused to the N-terminus of green fluorescent protein under the control of the 35S promoter. This construct was transiently expressed in rice protoplasts of the ZH17 cultivar (*O. sativa*). Fluorescence was visualized using a confocal laser microscope (Fluoview FV1000, Olympus Corp, Tokyo, Japan).

2.6. Single nucleotide polymorphism calling and haplotype analysis

Next-generation sequencing data of 187 *O. barthii* samples and 281 *O. glaberrima* samples were obtained from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) (Wang *et al.* 2014; Meyer *et al.* 2016; Cubry *et al.* 2018). Initially, the raw reads of each accession were filtered using Trimmomatic (Bolger *et al.* 2014). Subsequently, the remaining high-quality reads were aligned to the reference genome (*O. glaberrima* IRGC104165) using the Burrows-Wheeler Aligner (BWA) and then were utilized for single nucleotide polymorphism (SNP) calling with the Genome Analysis Toolkit (GATK) software (Li and Durbin 2010; McKenna *et al.* 2010). Natural variations in the *lsl3* gene were extracted from the final SNP map, and all variants within the coding region were subjected to further analysis. Haplotypes were integrated and classified using custom scripts, and a median-joining haplotype network was constructed using popART (Jessica and David 2015).

3. Results

3.1. Phenotype observations of W1411 and IRGC104165

The sterile lemmas, the primary structural unit of the rice spikelet, exhibited significant differences in their average length between W1411 and IRGC104165. The sterile lemmas on the floret sides of IRGC104165 were abnormally elongated, while those of W1411 were short (Fig. 1-A and B). To further investigate the differences in sterile lemma morphology between African wild and cultivated rice, SEM and paraffin sections were employed to examine the sterile lemma phenotypes of the African wild rice line W1411 and the African cultivated cultivar IRGC104165. The outer and inner cells of the sterile lemma differed between W1411 and IRGC104165. W1411 exhibited long, smooth, and flat outer cells, while IRGC104165 displayed rough outer cells with convex structures arranged in close parallel to axial directions. The inner cells on the veins of W1411 were shorter than those of IRGC104165, but the cells of IRGC104165 between the veins were shorter, wider, and more numerous (Fig. 1-C–H). Paraffin sections of the sterile lemma corroborated the SEM findings. The cell structure of the W1411 glume was simple, consisting of only one type of sclerenchyma cell. In contrast, the outer epidermal cells of IRGC104165 glumes were silicified, with a larger cell volume compared to the wild-type short glumes, and the inner epidermal cells were enlarged. The outer epidermis cells of the lemma of IRGC104165 and W1411 resembled those of the outer epidermis cells of the sterile lemma of IRGC104165 (Fig. 1-I and J; Appendices C and D). These results suggest that the long sterile lemma may be a homologous transformation of the lemma.

3.2. Genetic analysis and mapping of *lsl3*

To identify the gene responsible for the long sterile lemma phenotype, we employed a map-based cloning strategy to map the *lsl3* locus using a mapping population derived from IRGC104165 and W1411. The F_1 generation exhibited a short sterile lemma, and genetic analysis of the F_2 population revealed a segregation ratio of approximately 3:1 (χ^2=0.468, $P_{0.05}$=5.991), indicating that a single recessive locus controlled the long sterile lemma trait. This locus was mapped between the SSR markers RM1210 and RM5752 on the short arm of chromosome 7 (Table 1). By utilizing 1,335 segregating plants with the short sterile lemma phenotype from the F_2 population, we ultimately delimited *lsl3* to a 269.6-kb region between markers BG7-2 and LS-1 (Fig. 2-A).

In the mapping region of the Nipponbare (*O. sativa*) genome, a gene (*G1*) controlling the development of the long sterile lemma was identified. Therefore, we predicted that the allelic gene *OglabAA_FGT23456* (*lsl3*) of *G1* in IRGC104165 was the candidate gene. In accordance with the reference sequence of *G1* in Nipponbare was available on the Gramene website (http://www.gramene.org/), a pair of specific primers,

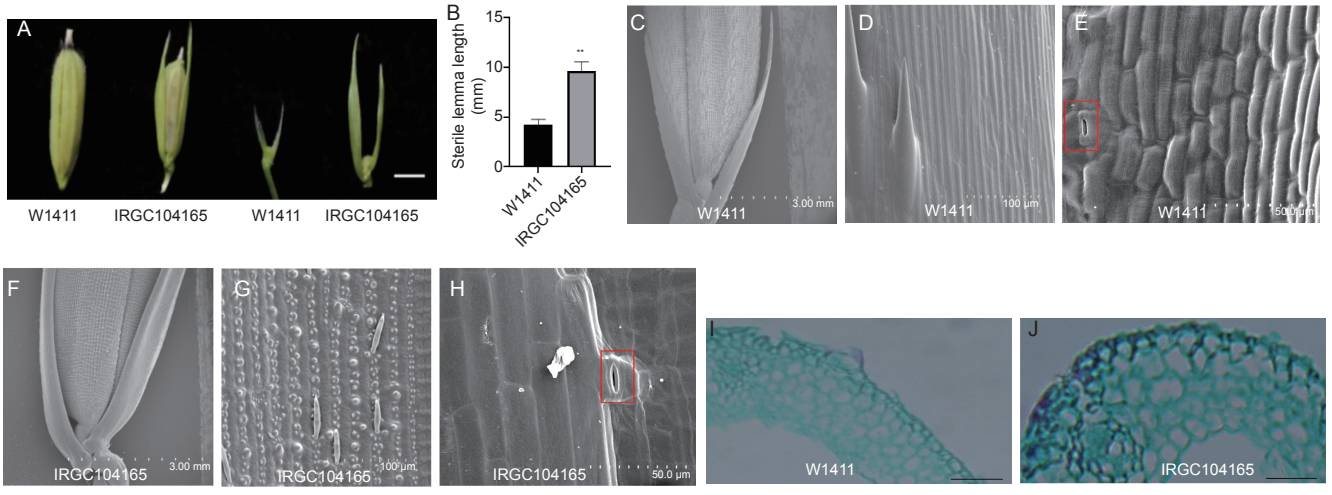


Fig. 1 Phenotypes of W1411 and IRGC104165. A and B, a comparison of sterile lemma length between W1411 and IRGC104165, with a scale bar representing 2 mm. Values represent mean±SD (n=20). **, P<0.01 (two-tailed Student's t-test). C–H, scanning electron microscopy images of the sterile lemma of W1411 and IRGC104165. I and J, Paraffin section morphology of the sterile lemma of W1411 and IRGC104165, with a scale bar of 100 μm.

designated CHY1, was designed within the target region to amplify and sequence the *G1* homologous genes of IRGC104165 and W1411. A comparison of the sequences of IRGC104165 and W1411 revealed the presence of four mutation sites within the coding region, including a single base deletion and three base substitutions. The nucleobase mutations at positions 405 and 525 of the coding region did not result in amino acid alterations. However, the mutation at position 445, which changed from G to T, resulted in an amino acid change from alanine to serine. Furthermore, the deletion of G at position 581 resulted in a frameshift of the protein reading frame and premature translation termination (Fig. 2-B and C). These findings suggest that the *lsl3* gene may be allelic with *G1*.

3.3. Cloning and functional analysis of *lsl3*

To confirm the function of *lsl3*, a construct (pLSL3) was generated by placing a 3.761-kb genomic fragment from W1411, which encompassed the *OglabAA_FGT23456* gene sequence from W1411 only, into the vector pCAMBIA1300. Due to the recalcitrance of the cultivated line IRGC104165 to regenerate shoots from calli, the pLSL3 construct was introduced into another African cultivated rice cultivar (IRGC102305) with long sterile lemmas. This cultivar had a single guanine (G) nucleotide deletion at position 587 in the coding region, resulting in a frameshift of the protein reading frame and premature translation termination. While the mutation positions of IRGC102305 and IRGC104165 in the coding sequence (CDS) differed, the mutation positions in the protein sequence were identical (Appendices E and F). All seven independent transgenic lines displayed the same phenotype, exhibiting short sterile lemmas (Fig. 3-A–C). This result corroborates the hypothesis that the *OglabAA_FGT23456* gene is *lsl3*.

The spatial expression pattern and subcellular localization of *lsl3* were further examined. qRT-PCR was employed to analyze the differential gene expression between two accessions, W1411 and IRGC104165, in the sterile lemma, lemma, and palea. The gene expression level in the sterile lemma of W1411 was lower than that in IRGC104165. Conversely, the expression levels of the lemma and palea in W1411 were higher than those in IRGC104165. The results indicated that the expression of *LSL3* was not the primary factor responsible for the elongation of the sterile lemma. Alternatively, the deletion of a single G nucleotide, resulting in premature translation termination, could have been the underlying cause of the long sterile lemma phenotype (Fig. 3-D).

The LSL3 protein was predicted to encode a transcription factor and to possess a nuclear localization signal. Prior analyses revealed that LSL3 contained a DUF640 domain, which had been named the ALOG (*Arabidopsis LSH1* and *Oryza G1*) domain. To ascertain the subcellular localization of LSL3 experimentally, we constructed a GFP fusion vector and transferred it into protoplasts of ZH17 (*O. sativa*). Our results demonstrated that the LSL3-GFP fusion protein localized to the nucleus of ZH17 protoplasts (Fig. 3-E). This finding suggests that *LSL3* encodes a nuclear protein.

Table 1 Genetic analysis of *lsl3*

Crosses	F_1 phenotype	F_2 population			$X^2_{0.05}$ (3:1)	$P_{0.05}$
		Short sterile lemma plants	Long sterile lemma plants	Total plants		
W1411/ IRGC104165	Short sterile lemma	100	39	139	0.468	5.991

3.4. Molecular evolution of *lsl3*

To gain further insight into the variant types of the *lsl3* gene, we conducted genotyping analysis of the *lsl3* gene based

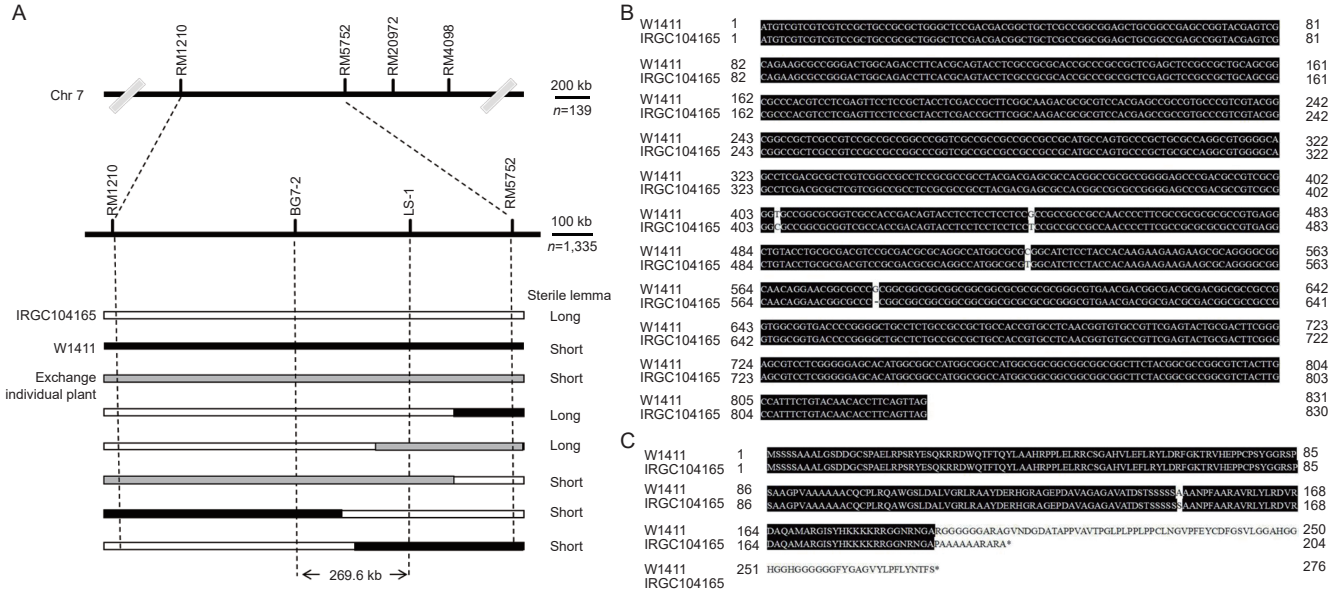


Fig. 2 Molecular mapping of the *lsl3*. A, the fine mapping of the *lsl3*. B, base sequence alignment analysis of *lsl3*. C, amino acid sequence alignment analysis of *lsl3*.

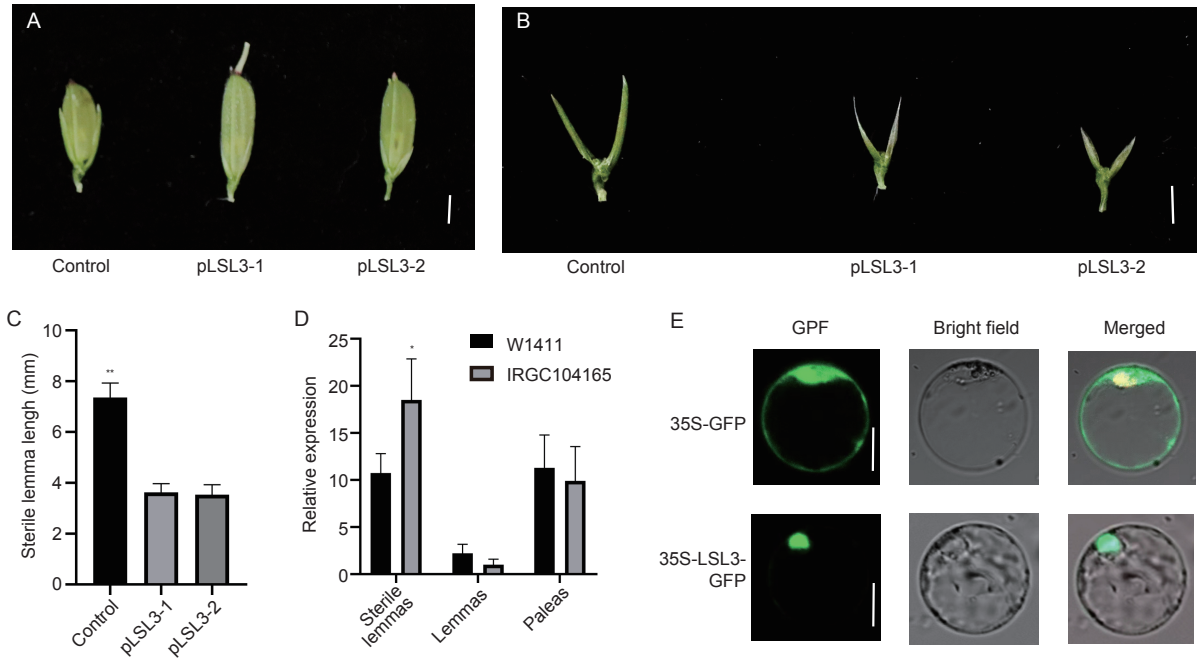


Fig. 3 Functional analysis of *lsl3*. A–C, phenotype of complementary transgenic plants. Bar=3 mm. D, RNA expression analysis of *lsl3*. Values are mean±SD (*n*=3). **, *P*<0.01; *, *P*<0.05; (two-tailed Student's *t*-test). E, the 35S:: LSL3-GFP fusion gene and 35S:: GFP were expressed transiently in rice protoplast cells, with a scale bar of 20 μm.

on the resequencing data from a publicly available database comprising a collection of 187 *O. barthii* and 281 *O. glaberrima* accessions. The results revealed 18 distinct haplotypes, five of which affected sterile lemma length (Fig. 4-A; Appendix F). IRGC104165 was designated as haplotype 3 (H3). W1411 was categorized as H10. Furthermore, IRGC102305 was classified as H4. To explore the relationships between different haplotypes, we constructed a haplotype network for the *lsl3* gene. The results demonstrated that H1 was the most prevalent. The diverse haplotypes that account for different lengths of sterile lemmas could derive from H1 through multiple independent mutations (Fig. 4-B). The comparison of the sequences of H1 revealed three distinct types of causal mutations; all 1-bp deletions at three different positions (579, 581, and 587 bp) within the ORF. Each of these deletions resulted in premature termination of the translated protein, leading to a loss of gene function. H3, H15, and H18 exhibited a shared key mutation base (a deletion at 581 bp in the ORF), leading to their classification as the same type of long sterile lemma genotype variation.

Our dataset revealed that 9.96% (28/281) of African cultivated rice accessions harbored the causal mutations (H3, H15, H18, H4, and H9) of *lsl3*. The distribution of samples with these haplotypes across West Africa was as follows: H3 (H3, H15, and H18) was found in the Guinea Highlands; H4 was present in the Senegambian and Inner Niger Delta regions; H9 was distributed across the Eastern Inland Region. These findings demonstrate that three distinct natural mutations in African rice can result in the long sterile lemma phenotype, suggesting that these causal mutations may have arisen independently in West Africa.

3.5. *lsl3* provided new evidence for parallel evolution

To elucidate the evolutionary characteristics of *lsl3*, a comparison was made between the *lsl3* haplotypes in African rice and those in 10,548 previously sequenced Asian rice accessions (http://www.ricesuperpir.com/web/rspvm) (Wang *et al.* 2023). A total of 67 distinct SNPs or indels were identified in Asian rice accessions, as detailed in Appendix G. Notably, only 11 haplotypes of *lsl3* from Asian rice resulted in the long sterile lemma phenotype, all of which were caused by premature translation termination. Given that the length of the *lsl3* CDS was 831 bp in both Asian and African rice, the positions of the key mutations were compared. In Asian rice, the key mutations were identified at CDS positions 55, 58, 59, 64, 65, 66, 261, 265, 269, 288, and 593. In African rice, they were found at CDS positions 579, 581, and 587 (Fig. 5). The results of the protein sequence comparison demonstrate that the protein variations resulting from the key variations in African cultivated rice and Asian cultivated rice are entirely distinct (Appendix F). These findings indicate that the long sterile lemma trait in Asian rice (*O. sativa*) and African rice (*O. glaberrima*) evolved independently, resulting from distinct mutations. The divergent mutations in the *lsl3* gene between Asian and African rice provide novel evidence of parallel evolution.

4. Discussion

4.1. Cloning of the first sterile lemma development gene in African rice

The sterile lemma is a primary component of the rice spikelet, playing a crucial role in its development (Ren *et al.* 2013). Numerous studies have focused on genes that regulate the trait of sterile lemma (Lee *et al.* 2007; Li *et al.* 2009; Yoshida

A

Haplotypes	*O. barthii*	*O. glaberrima*	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	SNP11	SNP12	SNP13	InDel	Number
H1	126	231	G	G	G	C	C	C	T	T	C	G	G	G	G	TGGCGGCCA	357
H2	31	0	G	G	G	C	C	C	G	C	C	G	G	G	G	TGGCGGCCA	31
H3	6	8	G	G	G	C	C	C	T	T	C	-	G	G	G	TGGCGGCCA	14
H4	0	13	G	G	G	C	C	C	T	T	C	G	-	G	G	TGGCGGCCA	13
H5	6	4	A	G	G	C	C	C	G	C	C	G	G	G	G	TGGCGGCCA	10
H6	2	6	G	G	G	C	C	C	G	T	C	G	G	G	G	TGGCGGCCA	8
H7	3	4	G	G	G	C	G	C	T	T	C	G	G	G	G	TGGCGGCCA	7
H8	0	6	G	G	A	C	C	C	T	T	C	G	G	G	G	TGGCGGCCA	6
H9	0	5	G	G	G	C	C	C	T	T	-	G	G	G	G	TGGCGGCCA	5
H10	4	0	G	G	G	C	C	T	G	C	C	G	G	G	G	TGGCGGCCA	4
H11	0	3	G	A	G	C	C	C	G	T	C	G	G	G	G	TGGCGGCCA	3
H12	2	0	G	G	G	C	C	C	G	C	C	G	G	G	G	-	2
H13	2	0	A	G	G	T	C	C	G	C	C	G	G	G	G	TGGCGGCCA	2
H14	2	0	G	G	G	C	C	C	T	C	C	G	G	G	G	TGGCGGCCA	2
H15	1	0	G	G	G	C	C	C	T	T	C	-	G	A	C	TGGCGGCCA	1
H16	1	0	G	G	G	C	C	T	G	C	C	G	G	G	G	TGGCGGCCA	1
H17	0	1	A	G	A	C	C	C	T	C	C	G	G	G	G	TGGCGGCCA	1
H18	1	0	G	G	G	C	C	C	G	C	C	-	G	G	G	TGGCGGCCA	1
			16	52	89	147	263	405	445	525	579	581	587	612	749	744–752	

B

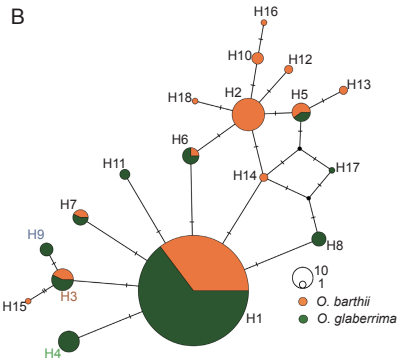


Fig. 4 Haplotypes and geographic distribution of *Isl3*. A, the haplotype analysis of *Oryza barthii* and *O. glaberrima*. B, the haplotype network of *Isl3*.

et al. 2009; Lee and An 2012; Yoshida *et al.* 2012; Li X M *et al.* 2015; Yoon *et al.* 2018; Xu *et al.* 2020). However, most research on sterile lemma has been conducted on Asian rice. African rice (*O. glaberrima*), another cultivated rice species, is a repository of genetic resources, and numerous dominant genes have been identified (Linares 2002; Wambugu *et al.* 2013; Li W *et al.* 2015; Wu *et al.* 2017; Hu *et al.* 2018). Despite this, there is a paucity of research examining the sterile lemma in African rice, and no genes associated with this trait have been cloned, leaving the regulatory mechanisms governing sterile lemma development largely unexplored.

To explore the regulation mechanism of the sterile lemma in African rice, we identified the *Isl3* gene from a mapping population constructed using IRGC104165 and W1411. The results demonstrated that *Isl3* was a homolog of the *G1* gene, exhibiting a novel mutation type compared to that observed in Asian rice. This finding corroborates the existence of a naturally occurring long sterile lemma mutant in African rice. The cloning of *Isl3* represents a significant advancement in our understanding of sterile lemma development in African rice, as it is the first reported gene associated with this trait in the species. The long sterile lemma not only provides physical protection for the internal floral organs but also serves as a valuable marker for breeding purposes. The identification of *Isl3* provides a genetic resource that can be leveraged in future breeding programs.

4.2. Molecular evolution of *Isl3* in African rice

In this study, a long sterile lemma gene, *Isl3*, was cloned from *O. glaberrima*. The haplotypes of *Isl3* were analyzed using publicly available resequencing data from 187 *O. barthii* and 281 *O. glaberrima*. The analysis revealed that the *Isl3* gene in *O. glaberrima* exhibited 18 distinct variation patterns, of which five resulted in premature termination of the ORF, leading to the development of elongated sterile lemmas (Fig. 4-A; Appendix F).

At present, two competing theories exist regarding the origin of *O. glaberrima*. Portères (1962) proposed that the Inner Niger Delta of Mali was the primary center of *O. glaberrima* domestication, with Senegambia and the Guinea Highlands serving as secondary centers. This perspective has been reinforced by the findings of Wang *et al.* (2014) and Cubry *et al.* (2018), who conducted a comprehensive analysis of re-sequencing data from *O. barthii* and *O. glaberrima*. Their findings substantiate the hypothesis that the Inner Niger Delta represents the primary center of domestication. In contrast, Harlan *et al.* (1971) suggested that *O. glaberrima* had no specific center of origin. Choi *et al.* (2019) analyzed resequencing data from 282 rice samples, including both *O. barthii* and *O. glaberrima*, and concluded that *O. glaberrima* was not domesticated in a single origin center but rather resulted from a diffusion process. Despite recent progress in studying the origin of *O. glaberrima*, a unified view has yet to be established. Our results demonstrated that the five haplotypes were divided into three types: H3 (H3, H15, and H18), H4, and H9 (Fig. 4-A; Appendix F). These three types of variations were distributed across the Guinea Highlands, the Senegambia, the Inner Niger Delta, and the Eastern Inland Region, indicating multiple independent original sites for the long sterile lemma and providing a reference for the molecular evolution of cultivated rice in Africa.

African cultivated rice (*O. glaberrima*), having undergone a relatively recent domestication process and lacking the benefits of modern systematic breeding and artificial cultivation management, relies on its elongated sterile lemma to protect the delicate floral organs, thereby ensuring stable production.

4.3. New evidence of parallel evolution between Asian rice and African rice

The parallel domestication of crops represents a pivotal scientific issue within the broader field of plant evolution. Recent studies on the molecular basis of domestication traits in *O. sativa* and *O. glaberrima* have yielded valuable insights into the parallel domestication between these two species (Purugganan 2014). Wu *et al.* (2017) cloned the seed shattering gene *SH4* from African cultivated rice and discovered that its ORF exhibited distinct mutation patterns in *O. sativa* and *O. glaberrima*. Another shattering gene, *SH3*, underwent parallel selection during domestication through distinct variants in sorghum, maize, *O. sativa*, and *O. glaberrima* (Lv *et al.* 2018). Hu *et al.* (2018) found that

PROG7 in *O. glaberrima* and the homeotic gene *PROG1* in *O. sativa* collectively promoted plant architecture evolution, in conjunction with gene clusters comprising closely linked genes. The two gene clusters were selected in parallel during the domestication of *O. sativa* and *O. glaberrima*, respectively. However, research on the parallel domestication of sterile lemma genes in *O. sativa* and *O. glaberrima* is limited.

In this study, we cloned the *lsl3* gene, which regulates sterile lemma length, from *O. glaberrima*. The variation patterns observed in the three types analyzed in *O. glaberrima* differed from those observed in Asian rice (Fig. 5; Appendix F). Based on these findings, we hypothesized that *lsl3* underwent parallel selection during the independent domestication processes of African and Asian cultivated rice. This evidence supports the hypothesis of parallel domestication of rice and the independent origin of *O. glaberrima*.

5. Conclusion

This study identifies the first gene controlling sterile lemma length in African rice. The key mutations of *lsl3*, which lead to long sterile lemma in African rice, differ from those in Asian rice. The investigation of *lsl3* not only provides a reference for the molecular evolution of cultivated rice in Africa but also offers new evidence for the parallel domestication of rice and the independent origin of *O. glaberrima*.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (31925029) and the Ph D Scientific Research and Innovation Foundation of Sanya Yazhou Bay Science and Technology City, China (HSPHDSRF-2023-05-003).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that they did not use AI in the preparation and writing of this manuscript.

Appendices associated with this paper are available at https://doi.org/10.1016/j.jia.2024.11.031

References

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

Bommert P, Satoh-Nagasawa N, Jackson D, Hirano H Y. 2005. Genetics and evolution of inflorescence and flower development in grasses. *Plant Cell Physiology*, **46**, 69–78.

Choi J, Zaidem M, Gutaker R, Dorph K, Singh R, Purugganan M. 2019. The complex geography of domestication of the African rice *Oryza glaberrima*. *PLoS Genetics*, **15**, e1007414.

Coen E S, Meyerowitz E M. 1991. The war of the whorls: Genetic interactions controlling flower development. *Nature*, **353**, 31–37.

Cubry P, Tranchant-Dubreuil C, Thuille A, Monat C, Ndjiondjop M, Labadie K, Cruaud C, Engelen S, Scarcelli N, Rhoné B, Burgarella C, Dupuy C, Larmande P, Wincker P, François O, Sabot F, Vigouroux Y. 2018. The rise and fall of African rice cultivation revealed by analysis of 246 new genomes. *Current Biology*, **28**, 2274–2282.

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky M F. 2004. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Current Biology*, **14**, 1935–1940.

Harlan J R. 1971. Agricultural origins centres and non-centres. *Science*, **174**, 468–474.

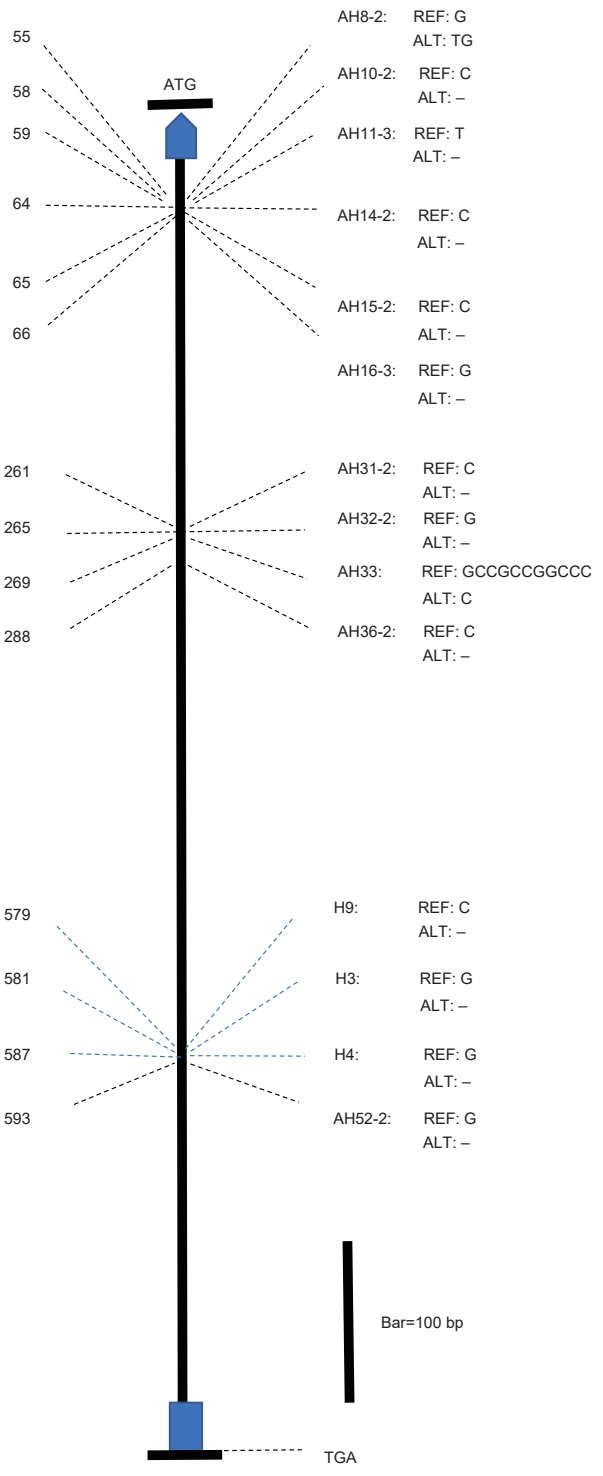


Fig. 5 Key mutation types of *lsl3* between African and Asian rice. The coding DNA sequence position of key mutations in African and Asian rice. The black color represents the key mutations of Asian rice, while the blue color signifies the key mutations of African rice.

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 Journal*, **94**, 661–669.

Huang X, Zhao Q B, Han B. 2015. Comparative population genomics reveals strong divergence and infrequent introgression between Asian and African rice. *Molecular Plant*, **8**, 958–960.

Jessica W, David B. 2015. POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.

Kobayashi K, Maekawa M, Miyao A, Hirochika H, Kyozuka J. 2010. *PANICLE PHYTOMER2* (*PAP2*), encoding a SEPALLATA subfamily MADS-box protein, positively controls spikelet meristem identity in rice. *Plant Cell Physiology*, **51**, 47–57.

Lee D Y, An G. 2012. Two AP2 family genes, supernumerary bract (*SNB*) and Osindeterminate spikelet 1 (*OsIDS1*), synergistically control inflorescence architecture and floral meristem establishment in rice. *Plant Journal*, **69**, 445–461.

Lee D Y, Lee J, Moon S, Park S Y, An G. 2007. The rice heterochronic gene *SUPERNUMERARY BRACT* regulates the transition from spikelet meristem to floral meristem. *Plant Journal*, **49**, 64–78.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.

Li H, Xue D, Gao Z, Yan M, Xu W, Xing Z, Huang D, Qian Q, Xue Y. 2009. A putative lipase gene *EXTRA GLUME1* regulates both empty-glume fate and spikelet development in rice. *Plant Journal*, **57**, 593–605.

Li W, Yoshida A, Takahashi M, Maekawa M, Kojima M, Sakakibara H, Kyozuka J. 2015. *SAD1*, an RNA polymerase I subunit A34.5 of rice, interacts with Mediator and controls various aspects of plant development. *Plant Journal*, **81**, 282–291.

Li X M, Chao D Y, Wu Y, Huang X, Chen K, Cui L G, Su L, Ye W W, Chen H, Chen H C, Dong N Q, Guo T, Shi M, Feng Q, Zhang P, Han B, Shan J X, Gao J P, Lin H X. 2015. Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of african rice. *Nature Genetics*, **47**, 827–833.

Linares O F. 2002. African rice (*Oriza glaberrima*): History and future potential. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16360–16365.

Lombardo F, Yoshida H. 2015. Interpreting lemma and palea homologies: A point of view from rice floral mutants. *Frontiers in Plant Science*, **18**, 61.

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, Zhu Z F. 2018. Genetic control of seed shattering during African rice domestication. *Nature Plants*, **4**, 331–337.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M A. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

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. *Nature Genetics*, **48**, 1083–1088.

Pelaz S, Ditta G S, Baumann E, Wisman E, Yanofsky M F. 2000. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200–203.

Porteres R. 1962. Berceaux agricoles primaries sur le continent Africain. *Journal of African History*, **3**, 195–210.

Purugganan M D. 2014. An evolutionary genomic tale of two rice species. *Nature Genetics*, **46**, 931–932.

Ren D Y, Li Y F, Zhao F M, Sang X C, Shi J Q, Wang N, Guo S, Ling Y H, Zhang C W, Yang Z L, He G H. 2013. *MULTI-FLORET SPIKELET1*, which encodes an AP2/ERF protein, determines spikelet meristem fate and sterile lemma identity in rice. *Plant Physiology*, **162**, 872–884.

Schmidt R J, Ambrose B A. 1998. The blooming of grass flower development. *Current Opinion Plant Biology*, **1**, 60–67.

Theissen G. 2001. Development of floral organ identity: Stories from the MADS house. *Current Opinion Plant Biology*, **4**, 75–85.

Theissen G, Saedler H. 2001. Plant biology-Floral quartets. *Nature*, **409**, 469–471.

Wambugu P W, Furtado A, Waters D L, Nyamongo D O, Henry R J. 2013. Conservation and utilization of African *Oryza* genetic resources. *Rice*, **6**, 29.

Wang M, Yu Y, Haberer G, Marri P R, Fan C, Goicoechea J L, Zuccolo A, Song X, Kudrna D, Ammiraju J S, Cossu R M, Maldonado C, Chen J F, Lee S, Sisneros N, Baynast K D, Golser W, Wissotski M, Kim W, Sanchez P, *et al.* 2014. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nature Genetics*, **16**, 982–988.

Wang T Y, He W C, Li X X, Zhang C, He H Y, Yuan Q L, Zhang B, Zhang H, Leng Y, Wei H, Xu Q, Shi C L, Liu X P, Guo M L, Wang X M, Chen W, Zhang Z P, Yang L B, Lv Y, Qian H G, *et al.* 2023. A rice variation map derived from 10548 rice accessions reveals the importance of rare variants. *Nucleic Acids Research*, **51**, 10924–10933.

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. *Nature Plants*, **8**, 17064.

Xu Q K, Yu H P, Xia S S, Cui Y J, Yu X Q, Liu H, Zeng D L, Hu J, Zhang Q, Gao Z Y, Zhang G H, Zhu L, Shen L, Guo L B, Rao Y C, Qian Q, Ren D Y. 2020. The C2H2 zinc-finger protein lacking rudimentary *GLUME 1* regulates spikelet development in rice. *Science Bulletin*, **65**, 753–764.

Yang J, Lee S, Hang R, Kim SR, Lee YS, Cao X, Amasino R, An G. 2013. *OsVIL2* functions with PRC2 to induce flowering by repressing *OsLFL1* in rice. *Plant Journal*, **73**, 566–578.

Yi G, Choi J, Jeong E, Chon N, Jena K, Ku Y, Kim D, Eun M, Jeon J, Nam M. 2005. Morphological and molecular characterization of a new frizzy panicle mutant, "*fzp-9(t)*", in rice (*Oryza sativa* L.). *Hereditas*, **142**, 92–97.

Yoshida A, Ohmori Y, Kitano H, Taguchi-Shiobara F, Hirano H Y. 2012. *Aberrant spikelet and panicle1*, encoding a TOPLESS-related transcriptional co-repressor, is involved in the regulation of meristem fate in rice. *Plant Journal*, **70**, 327–339.

Yoshida A, Suzaki T, Tanaka W, Hirano H Y. 2009. The homeotic gene long sterile lemma (*G1*) specifies sterile lemma identity in the rice spikelet. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 20103–20108.

Yoon H, Yang J, Liang W, Zhang D, An G. 2018. *OsVIL2* regulates spikelet development by controlling regulatory genes in *Oryza sativa*. *Frontiers in Plant Science*, **6**, 102.

Zanis M J. 2007. Grass spikelet genetics and duplicate gene comparisons. *International Journal of Plant Science*, **168**, 93–110.

Zhang Q J, Zhu T, Xia E H, Shi C, Liu Y L, Zhang Y, Liu Y, Jiang W K, Zhao Y J, Mao S Y, Zhang L P, Huang H, Jiao J Y, Xu P Z, Yao Q Y, Zeng F C, Yang L L, Gao J, Tao D Y, Wang Y J, *et al.* 2014. Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E4954–E4962.

Executive Editor-in-Chief Xueyong Zhang
Managing Editor Ning Wang