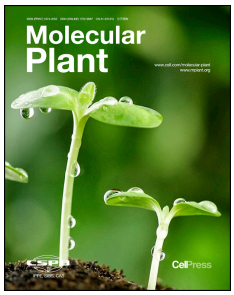
Accepted Manuscript



Elimination of a retrotransposon for quenching genome instability in modern rice

Yu Peng, Yingying Zhang, Yijie Gui, Dong An, Junzhong Liu, Xun Xu, Qun Li, Junmin Wang, Wen Wang, Chunhai Shi, Longjiang Fan, Baorong Lu, Yiwen Deng, Sheng Teng, Zuhua He

PII: S1674-2052(19)30205-9
DOI: https://doi.org/10.1016/j.molp.2019.06.004
Reference: MOLP 800

To appear in: *MOLECULAR PLANT*
Accepted Date: 9 June 2019

Please cite this article as: **Peng Y., Zhang Y., Gui Y., An D., Liu J., Xu X., Li Q., Wang J., Wang W., Shi C., Fan L., Lu B., Deng Y., Teng S., and He Z.** (2019). Elimination of a retrotransposon for quenching genome instability in modern rice. Mol. Plant. doi: https://doi.org/10.1016/j.molp.2019.06.004.

Elimination of a retrotransposon for quenching genome instability in modern rice

Yu Peng[1,7], Yingying Zhang[2,7], Yijie Gui[3], Dong An[1], Junzhong Liu[1], Xun Xu[5], Qun Li[1], Junmin Wang[6], Wen Wang[5], Chunhai Shi[4], Longjiang Fan[4], Baorong Lu[3], Yiwen Deng[1,*], Sheng Teng[1,*], Zuhua He[1,*]

[1]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Chinese Academy of Sciences, Shanghai 200032, China.

[2]The Protected Horticulture Institute, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China.

[3]School of Life Sciences, Fudan University, Shanghai 200433, China.

[4]College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China.

[5]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China.

[6]Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China.

[7]These Authors contributed equally to the study.

*Correspondence: zhhe@sibs.ac.cn; steng@sibs.ac.cn; ywdeng@sibs.ac.cn;

Running Title: **Retrotransposon removal for genetic and epigenetic stability**

Short Summary

A low-copy-number LTR retrotransposon is strictly controlled by the gene silence machinery, and was gradually removed during rice domestication and breeding because its transposition could trigger epigenome turbulence, and decrease agronomic values of yield and disease resistance in modern rice.

ABSTRACT

Transposable elements (TEs) constitute the most abundant portions of plant genomes and can dramatically shape host genomes during plant evolution. They also play important roles in crop domestication. However, whether TEs themselves are also selected during crop domestication remained unknown. Here, we identify an active long terminal repeat (LTR) retrotransposon, *HUO*, as a potential selection target during rice domestication and breeding. *HUO* is a low-copy-number LTR retrotransposon, and displays transposition activity under the natural growth conditions and transmits transpositions through male gametogenesis, preferentially inserting into genomic regions capable of transcription. *HUO* exists in all wild rice accessions, about half of the archaeological rice grains (1200-7000 years ago) and landraces, but is absent in almost all modern varieties, indicating its gradual elimination during rice domestication and breeding. Our data hints that *HUO* is subjected to strict gene silencing through the RNA-directed DNA methylation (RdDM) pathway. Our study also suggests that *HUO* may trigger genomic defense through altering genome-wide DNA methylation and small RNA biogenesis, and changing global gene expression, resulting in decreased disease resistance and yield, which may explain its elimination in rice breeding. Thus, our study reveals that negative selection of an active retrotransposon may be important for genome stability during crop domestication and breeding.

INTRODUCTION

Transposable elements (TEs) are fragments of DNA which can move around their host genomes, and are divided into RNA-based retrotransposons (Class I) and DNA transposons (Class II) (Feschotte et al., 2002). TEs are well-known to play important roles in host genome evolution by changing gene expression or inducing DNA rearrangement (Friedli and Trono, 2015). In vascular plant, TEs constitute a significant portion of the genomes, and have deeply shaped the genomes at the structural and functional levels (Piegu et al., 2006; Marí-Ordóñez et al., 2013). Recent studies have revealed that TEs regulate gene expression genetically and epigenetically, and are involved in crop domestication and environmental adaptation (Studer et al., 2011; Deng et al., 2017). However, whether TEs themselves can be directly selected during domestication and breeding remains elusive.

As the main component of Class I TEs, long terminal repeat (LTR) retrotransposons are predominant in plant genomes (Feschotte et al., 2002; Vitte et al., 2007). Despite their variety in copy numbers, LTR retrotransposons' activities are largely suppressed by the accumulated mutations and host genome defense mechanisms on the time scale of evolution (Ito et al., 2011). LTR retrotransposon's transposition is generally induced by stress such as tissue culture (Hirochika et al., 1996; Picault et al., 2009), DNA methylation-related mutations (Cheng et al., 2015), or interspecies hybridization (Wang et al., 2013b). So far, only a few active LTR retrotransposons, such as *Tos17* (Hirochika et al., 1996) and *Lullaby* (Picault et al., 2009), have been identified in the model crop rice (*Oryza sativa* L.). These TEs are often active in tissue culture but keep silenced under the normal growth conditions.

Plant hosts have developed genomic defense mechanisms to limit the activity of TEs and control their otherwise deleterious propagation, mainly by the transcriptional gene silencing through RNA-directed DNA methylation (RdDM), which involves small interfering RNA (siRNA) biogenesis from TEs (Slotkin and Martienssen, 2007; Marí-Ordóñez et al., 2013; Kim and Zilberman, 2014). Occasionally, TEs are derepressed in DNA methylation-related mutants (Cheng et al., 2015), further supporting that DNA methylation plays a critical role in silencing TEs. Furthermore, TE silencing has proved to be intertwined with other epigenetic phenomena, and TEs are known to directly or indirectly regulate gene expression (Feschotte et al., 2002).

Genome evolution is accelerated during crop domestication, in which cultivated crops with ideal agronomic traits have been selected from their wild relative populations, accompanied by reduced genetic diversity and increased genome stability (Doebley et al., 2006; Wang et al., 2008; Purugganan and Fuller, 2009; Marí-Ordóñez et al., 2013; Meyer and Purugganan, 2013). Many functional genes controlling agronomic traits, such as plant architecture, yield, and resistance, were selected during domestication. However, whether and how TEs were targeted during crop domestication remains unclear. In this study, we report that a new rice active LTR retrotransposon, *HUO*, has low copy number and is rapidly silenced by the host RdDM mechanism when its copy number increases . We provide genetic and molecular evidence that *HUO* is a potential negative target of rice domestication and modern breeding. *HUO* triggers genomic instability and decreases agronomic values, leading to its final elimination from modern rice, likely through genomic defense and/or genetic bottleneck.

RESULTS

HUO is an active LTR retrotransposon under normal growth conditions

During cloning of the rice *ELONGATED UPPERMOST INTERNODE 1* (*EUI1*) gene (Zhu et al., 2006), we identified a mutagenic LTR retrotransposon, hereafter named *HUO* (Chinese meaning “alive” or “flaming”, NCBI accession number: DQ004853). *HUO* inserted in the second exon of *EUI1* gene in the same transcriptional direction and led to the null mutation (**Figure 1A**). *HUO* has a total length of 4889 bp, with identical LTRs of 257 bp at both ends, and produces 5-bp target site duplication (TSD) in the *eui-1*mutant (**Figure 1B** and **Supplemental Figure 1**). The full-length open reading frame (ORF) (4365 bp) of *HUO* encodes a *gag-pol* polyprotein containing functional domains necessary for transposition, such as protease, integrase, reverse transcriptase, and RNase H (**Figure 1B**). Phylogenetic tree analysis reveals that *HUO* is close to the clade containing the active LTR retrotransposons *Tos17* and *Lullaby* (**Supplemental Figure 2**), indicating a *copia*-type LTR retrotransposon. To investigate the origin of *HUO*, we searched the genome database of *Oryza*, including 10 rice genomes (eight AA genome, one FF genome, one BB genome), and found only 6 AA genomes (*O. sativa*, *O. longistaminata*, *O. glumipatula*, *O. meridionalis*, *O. nivara*, *O. rufipogon*) contain highly homology sequence of *HUO* (**Supplemental Figure 3**). Thus, *HUO* could be considered as an active *copia*-type LTR retrotransposon and might have arisen after the evolution of AA genome of *Oryza*.

The spontaneous *eui-1* mutant was obtained from a breeding population (Rutger and Carnahan, 1981), and introduced into the recurrent parent *Indica/Xian* Zhenshan 97 (ZS97) to generate a pair of near-isogenic lines (NIL): 307T (*eui-1* mutation with *HUO*) and 308D (wild-type without *HUO*) (**Supplemental Figure 4A**) (Zhu et al., 2006). Sequence analysis confirmed that *HUO* was originated from the *Japonica*/*Geng* landrace rice Terso (**Supplemental Figure 4B**) through transposition (Rutger and Carnahan, 1981). This recent transposition event in *eui-1* mutant

prompted us to examine whether *HUO* is an active retrotransposon. We first checked progenies derived from different 307T individuals by Southern blot, and found new *HUO* insertions (**Figure 1C**). To determine these new insertion sites, two lines both with a new *HUO* insert were studied in detail by thermal asymmetric interlaced (TAIL) PCR to detect the insertion sites (**Figure 1D**) and were confirmed by multiple endonuclease cleavage assays (**Figure 1E**).

Further, we obtained a total of 16 new *HUO* insertions, all are localized in or nearby expressed genes, hypothetical ORFs, or pseudogenes with the potential to transcribe (**Figure 1F**). Therefore, we conclude that *HUO* is an active retrotransposon under normal paddy field conditions and preferentially inserts in the genomic regions capable of transcription. We found that *HUO* was predominately expressed in young panicles (**Figure 2A**). Similar to other TEs, *HUO* is strongly induced during tissue culture and by various stresses, such as heat, drought, radiation, darkness, salt, and abscisic acid (ABA) treatment (**Figure 2A** and **2B**). Accordingly, we found new insertions in plants regenerated from tissue culture (**Supplemental Figure 5**). Consistently, full-length *HUO* transcripts were detected from rice callus (**Supplemental Figure 6A**).

HUO is transgenerationally transposes through male gametogenesis

We further determined the origin of new *HUO* insertions by tracking the new copies over successive generations. We found that all the new copies were heterozygous, as they were segregated in the subsequent generation (**Figure 2C**). It has been reported that the transposition of LTR retrotransposon occurred before the differentiation of male and female gametophytes (Ito et al., 2011). To investigate the transposition mode

of *HUO*, we performed a reciprocal crossing experiment, and analyzed the F1 individuals to determine the new *HUO* copy. A total of 11 new insertions were detected in 24 F1 plants produced with 307T1 (one *HUO* copy) as the paternal parent, but no new copies were detected in 72 F1 plants produced by either with 307T1 or 307T2 (two *HUO* copies with the original one in the *eui-1* mutant) as the maternal parent (**Figure 2D**). Consistent with these findings, we observed that *HUO* was highly expressed in the anther (**Supplemental Figure 6B** and **6C**). Thus, *HUO* displayed transgenerational transposition through male gametogenesis.

HUO is confined by DNA methylation-mediated suppression

In order to determine the transposition activity of *HUO*, we examined the frequency of new *HUO* insertions. To our surprise, in the progeny of 307T plants with a single *HUO* copy (307T1), 23.3% (21 out of 90) of the individuals contained one or two new insertions; while in the progeny of the 307T plants with two starting copies (307T2), the frequency of new insertion dropped to 14.4% (25 out of 173); in the progeny of the parents with three starting copies (307T3), the frequency further dropped to 6.25% (3 out of 48) (**Figure 3A**). Therefore, *HUO*'s transposition capacity rapidly dropped along with increasing copy number. Notably, the transcription of *HUO* also greatly decreased in 307T3 compared to 307T1 (**Figure 3B**), suggesting the occurrence of gene silencing. Thus, *HUO* is a low copy number TE, and its transcriptional activity greatly decreases when additional copies are generated.

Copy number-dependent transposition is thought to be the result of the host genomic defense machinery preventing uncontrolled genomic propagation, which is otherwise deleterious to host genome stability (Perez-Hormaeche et al., 2008; Friedli and Trono,

2015). To investigate the underlying mechanism of *HUO* silence, we first examined the methylation status of *HUO* by Southern blot. Genomic DNAs were digested with *Hpa*II (sensitive to both CHG and CG methylation) and *Msp*I (sensitive to CHG methylation) respectively. There are 7 "CCGG" restriction sites in the sequence of *HUO* (**Supplemental Figure 7A**). We found that the methylation degree of the *HUO* sequence increased with two (307T2) or three copies (307T3), compared to only one copy (307T1) (**Figure 3C**). We further performed bisulfite-sequencing of the left LTR and the adjacent section of *HUO* in 307T1 and 307T3 plants. The result showed that DNA methylation of these regions was much higher in the 307T3 plants than in the 307T1 plants (**Figure 3D** and **Supplemental Figure 7B**). Relative to the CHH methylation, the increase methylation at CG and CHG sites was more prominent, especially in the promoter region. Therefore, *HUO* expression was repressed by DNA methylation when additional copies of *HUO* were introduced.

RNA-directed DNA methylation confers *HUO* repression

Plant hosts attenuate TE activity mainly through transcriptional gene silencing, which relies on the generation of small interfering RNAs (siRNAs) (Arikit et al., 2013; Marí-Ordóñez et al., 2013; Kim and Zilberman, 2014). To test whether siRNAs were involved in the *HUO* silence, small RNAs were extracted from 307T1 and 307T3 plants and subjected to high throughput RNA-seq. We found higher accumulation of *HUO*-derived 24-nt siRNAs in 307T3 relative to 307T1 around the region of 1300-1900 bp (**Figure 4A**). Most of the siRNA reads mapped to *HUO* are unique to the *HUO* sequence (**Supplemental Figure 8**), suggesting that its repression is controlled by the RdDM pathway. To determine how the RdDM mechanism controls

the *HUO* silence, we introduced RNA interference (RNAi) of *OsDCL3a* and *OsAGO4ab*, which played important roles in the rice RdDM pathway (Deng et al., 2017), into 307T3 plants (**Figure 4B** and **4C).** While **a** control RNAi transgene of the *VOZ1* gene in 307T3 showed no effect on *HUO*'s transposition activity (**Supplemental Figure 9**), *HUO* increased to 5-9 copies in the *OsDCL3a*-RNAi 307T3 and *OsAGO4ab*-RNAi 307T3 plants (**Figure 4D**), with increased RNA levels (**Figure 4E**). Therefore, *HUO* was repressed by RdDM when its copy number increased, resulting in the inhibition of further propagation.

HUO has been removed during rice domestication and modern breeding

HUO originally identified in the *eui-1* mutant was inherited from a landrace, Terso (Rutger and Carnahan, 1981; **Supplemental Figure 4**). However, *HUO* was absent in the sequenced cultivated rice genomes, Nipponbare (*Japonica/Geng*) and 93-11 (*Indica/Xian*). To determine the distribution of *HUO* in rice accessions, we collected and analyzed representative rice germplasm collection from different regions, including 21 wild rice accessions with AA genome (14 *O. rufipogon*, 5 *O. longistaminata*, 1 *O. nivara*, and 1 *O. glumaepatula*), 134 landraces, and 115 modern rice accessions (**Supplemental Table 1**) We first detected the presence of *HUO* by PCR using *HUO*-specific primers. *HUO* is present in all of the wild species (100%) and in 42.5% of the landraces, but absent in almost all of the 115 modern cultivars that were selected by different breeding programs, except, Shuangkezao, which contains one *HUO* copy and was selected in the early 80's of the last century without commercialization due to phenotypic instability, (**Figure 5A** and **Supplemental Table 1**).

Further, we detected 2-5 copies of *HUO* in the wild rice accessions with an average of 3.3 copies, and 1-3 copies in the landraces with an average of 1.2 copies per genome using Southern blot (**Figure 5B, Supplemental Figure 10** and **Supplemental Table 1**). The *HUO* sequences in the wild rice and landraces are the same as that in 307T, with only a few single nucleotide polymorphisms (SNPs) (data not shown). Consistent with the sequence identity, the *HUO* elements were also transcriptionally activated by salt stress in both wild rice and landraces (**Figure 5C**). These results suggested that *HUO* might have been removed from rice cultivars during rice domestication and modern breeding. To support this hypotheses, we were able to detected the *HUO* sequence in 5 out of 10 archaeological rice grains (about 1200-7000-year-old) found in the lower Yangtze River corridor of China (**Figure 5D** and **5E**) (Fan et al., 2011), which was one of the original regions of modern cultivated rice (Fuller et al., 2009; Huang et al., 2012). The sequences of fragments amplified from these archaeological grains showed high sequence identity with the *HUO* sequence (**Supplemental Figure 11**). Moreover, the insertion pattern of *HUO* in the landraces was similar with 1 or 2 copies in comparison with its diverse polymorphism in the wild rice (**Supplemental Figure 10**). Therefore, insertions of *HUO* were gradually reduced from wild rice to landraces. This line of evidence suggests a negative selection against active *HUO* during rice domestication and breeding, leading to its eventual removal in modern rice. However, owing to its low copy number and suppressed situation, we could not exclude the possibility that the genetic bottleneck may also contribute to the *HUO* elimination.

HUO induces host genomic defense

HUO has a low copy number in different rice accessions (**Figure 5B** and

Supplemental Table 1). The genetic instability caused by *HUO* insertion was unlikely the major reason for its elimination. We speculated that the host genomic defense network triggered by *HUO* not only effectively suppressed its transcription and transposition, but also induced epigenetic instability throughout the genome. To test these hypothesis, we first analyzed global DNA methylation in plants with one (307T1), three (307T3) or zero (308D) copies of *HUO* using bisulfite-sequencing analysis. Surprisingly, we found that at the genome-wide resolution, the landscape of DNA methylation, particularly for the CG and CHG sites, was increased in 307T1 and 307T3 in comparison with non-*HUO* 308D (**Figure 6A** and **Supplemental Figure 12**). When we zoomed in at the regions up- and down-stream of repeat sequences (mainly transposons) and transcribed regions, the changes were obvious, 307T1 and 307T3 displayed higher methylation levels than 308D (**Figure 6B**). Similar results were also observed by using methylated-DNA immunoprecipitation combined with high-throughput sequencing of 307T3 and 308D (MeDIP-seq) (**Supplemental Figure 13**). Therefore, *HUO* likely induced global methylation.

Further, we compared genomic methylation in the 307T1 and 307T3 plants with identical background except *HUO* copy number and expression levels. Interestingly, 307T1 displayed slightly higher global methylation than 307T3 (**Figure 6A** and **Supplemental Figure 13**). The vast majority of the differentially methylated regions (DMR) between 307T1 and 307T3 were CG and CHG sites (**Figure 6C**), which were mainly distributed in repeat regions (**Figure 6D**). We proposed that single copy *HUO* was more active and more likely to stimulate genome defense mechanisms, resulting in a global methylation, especially repeat elements. When the number of *HUO* copy is increased in 307T3, the genomic defense machinery might target on specific sites, including *HUO* itself by the RdDM pathway. In support of this hypothesis, whole

genome small RNA-seq revealed the higher ratio of 24-nt sRNAs in 307T3 (29.4%±0.4%) compared to 307T1 (22.2%±3.4%) (**Figure 6E**). Moreover, qRT-PCR assays showed the expression of some genes, such as *OsDRM2* and *AGO4a* functioning in the RdDM pathway, was increased in 307T3 (**Supplemental Figure 14**), consistent with the result that *HUO* was reactivated in the *AGO4*-RNAi lines (**Figure 4D** and **4E**). Therefore, the rice genome likely adopts various DNA methylation pathways to suppress *HUO*, depending on its copy number, similar to the observation in *de novo* silencing of the LTR-element *EVD* in *Arabidopsis* (Lisch, 2013). Further, we compared global gene expression profiling between 307T1 and 307T3 by RNA-seq and revealed significant number of differential expressed genes involved in diverse biological processes includinggrowth, defense, and abiotic stress responses (**Supplemental Figure 15**). However, the relationship between gene expression changes and *HUO*-triggered genomic methylation remain unclear. Nevertheless, the results together suggest that *HUO* likely triggers genomic defense that induces epigenetic regulation of global gene expression.

HUO decreases agronomic values

To investigate the effect of *HUO* elimination during rice domestication and breeding, we first analyzed the agronomic traits of 307T1 and 307T3 lines, which have the same genetic background except for *HUO* copy number. We found that the tilling number and final grain yield of 307T3 were significantly decreased compared with 307T1 (**Figure 7A** and **7B**), probably due to the differential regulation of plant growth and developmental processes (**Supplemental Figure 15**). Moreover, we also found that disease resistance to rice bacterial blight (*Xanthomonas oryzae* pv. *oryzae*), one of major rice diseases, was obviously compromised in 307T3 compared with 307T1

(**Figure 7C** and **7D**). Consistently, qRT-PCR assay showed that many defense genes were down-regulated in 307T3 compared with 307T1 (**Figure 7E** and **7F**). Intriguingly, salt tolerance was significantly enhanced in 307T3 compared with 307T1 (**Figure 7G** and **7H**), associated with the increased induction of the abiotic stress responsive genes (**Figure 7I** and **Supplemental Figure 16)**. Therefore, increase of *HUO* copy number has a negative effect on rice yield and disease resistance (**Figure 7J**), two key agronomic traits of rice cultivation.

DISCUSSION

In this study, a low copy autonomous retrotransposon, *HUO*, was identified in rice genome with potential negative selection during domestication and breeding. *HUO* is a *copia*-type LTR retrotransposon. Similar to other reported retrotransposons (Kumar and Bennetzen, 1999), *HUO* is active during tissue culture or under stress conditions. In contrast to other retrotransposons, *HUO* is also active under the natural growth conditions. However, we could not rule out the possibility that *HUO* is transcriptionally sensitive to environmental stresses in the paddy field.

Genetic analysis of the transposition events of *HUO* indicated that the inheritance of new copies of *HUO* to the next generation was mediated by the male gamete, suggesting that the transpositions occurred during the formation of male gametophyte and then could be effectively inherited. This hypothesis is also supported by the fact that *HUO* highly expressed in male gametophyte, but not in female gametophyte (**Supplemental Figure 6B** and **6C**). It is also consistent with previous discovery that vegetative nucleus (VN) of the pollen grain has decondensed heterochromatin while the sperm cell (SC) nuclei contain tightly condensed chromatin (Calarco et al., 2012;

Ibarra et al., 2012). Small RNAs were reported to be generated in VN and functioned in SC to suppress repeat sequences in *Arabidopsis*, including TEs and centromere (Slotkin et al., 2009; Schoft et al., 2009). In contrast to the dynamic reprogramming in male reproductive tissues, constitutive epigenetic silencing was maintained in female gametes (Stroud et al., 2014; Jullien et al., 2012). Derepression of TEs in the lineages of the vegetative nuclei in pollen (Slotkin et al., 2009) and endosperm (Gehring et al., 2009; Hsieh et al., 2009) has been observed. We also recently reported that miniature transposon (MITE) drives pollen expression of a disease susceptibility gene in rice (Deng et al. 2017). Transposition of *HUO* occurred in male gametogenesis but the detailed mechanism remains unknown. Our results therefore suggested that male reproductive cells might have a specific derepression machinery to ensure TE activation and that *HUO* may provide a perfect model to decipher the mysterious relationship between plant epigenetic regulation and reproductive development.

TEs are usually silenced due to the genomic defense system of the host by the RdDM pathway to prevent uncontrolled genomic propagation that is otherwise deleterious to host genome stability (Slotkin and Martienssen, 2007; Lisch, 2013; Mari Ordonez et al., 2013; Kim and Zilberman, 2014). We found that *HUO* was subjected to strict control by the host genomic defense system and keeps low copy in *Oryza* genus. Therefore, siRNA biogenesis from *HUO* regions might set a threshold to induce strong RdDM along with increasing *HUO* copy number. Similarly, the activity of the tobacco retrotransposon *TNT1* in *Arabidopsis* was also reported to be copy number dependent (Perez-Hormaeche et al., 2008). However, a burst of TE activity might have a generally repressive effect on gene expression and thus reduce the overall fitness of a population or a species (Lisch, 2013). Indeed, *HUO* can trigger strong host genomic defense that results in epigenome instability and decreases plant fitness,

providing a mechanistic explanation for its elimination from modern rice. It is unclear why *HUO*, even with low copy number, induced epigenome turbulence, given that TEs usually only influence nearby gene expression as well as self-silencing (Ito et al., 2011; Lippman et al., 2004). One mechanistic explanation is that *HUO* could induce a strong genome instability that impacts the host epigenetic machinery by altering siRNA production and/or DNA methylation. This same epigenetic guarding machinery may also rapidly silence *HUO* itself. It is also tempting to speculate that *HUO* can form silent nuclear compartments, similar to the human L1 retrotransposons-triggered X chromosome inactivation (Elbarbary et al., 2016).

The distribution of *HUO* in wild rice, landrace and modern rice indicates that *HUO* has been removed from the cultivated genome during the process of artificial selection. It was reported that uniquely mapping siRNAs promoted TE sequence removal in *Arabidopsis* accessions (Wang et al., 2013a). This TE-quenching mechanism may also work in *HUO*. Interestingly, *HUO*-triggered genomic defense decreases agronomic values through negatively impacting on yield and disease resistance. We propose that *HUO*, as an active retrotransposon, might have been beneficial for wild rice to cope with changing environments such as abiotic stress, a possible consequence of siRNA-mediated regulation of stress responsive genes as in *Arabidopsis* (Dowen et al., 2012; McCue et al., 2012) , and *HUO*-mediated disease susceptibility could be attenuated by the long-lasting defense machinery in wild rice. Domestication has rendered the plant completely dependent on humans such that it is no longer capable of propagating itself in nature, with reduced genetic diversity throughout the genome for many crops (Doebley et al., 2006; Meyer and Purugganan, 2013; Olsen and Wendel, 2013). Rice was originally cultivated in environmentally favorable centers such as the lower Yangtze River region of east China and the regions in south Asia

(Fuller et al., 2009; Huang et al., 2012), where humid and flooding environments might have favored disease resistance selection during rice domestication, in addition to domestication for other agronomic traits including grain yield (Wang et al., 2008). With this scenario, tolerance to salt might not be a target during rice domestication and breeding in these areas where salty soil was not a problem for early agricultural activity. Thus, the *HUO*-triggered epigenome instability might be a major target of rice domestication in order to quench genomic defense and thereby benefit the yield and disease resistance, in addition to its secondary effect on genetic instability through transposition. This finding reveals a previously unrecognized mechanism affecting crop domestication and breeding by which selection also attenuated the activity of a retrotransposon and eventually removed it from the cultivated plant genome (**Figure 7J**).

The maize TE *Hopscotch* inserted in a regulatory region of the maize domestication gene *tb1*, driving maize domestication (Studer et al., 2011).In contrast, *HUO* itself most likely was a domestication target. It is unclear whether other active TEs with low copy numbers were also subjected to selection during rice domestication and/or breeding, given that particular TEs could be selectively epigenetically controlled (Mirouze et al., 2009). Our study provides a snapshot of TEs' selection and impact in shaping modern crop genomes, enriching our knowledge of how crop genomes have been modified by selection. The *HUO*-mediated genomic defense might also provide new insight into TE-mediated regulation of the methylome and transcriptome (Law and Jacobsen, 2010; Naito et al., 2009). Further recognition of the genomic defense mechanism that guides *HUO* silence and mediates differential biotic and abiotic stress responses could lead to new insight into TE-triggered epigenetic regulation of plant immunity through extensive experiments that take advantage of available epigenetic

mutants and new molecular technology.

METHODS

Plant materials

A pair of near-isogenic lines (NILs), 307T and 308D, were obtained from the cross between the original spontaneous *eui-1* mutant and an *Indica/xian* rice Zhenshan 97 (ZS97), The ZS97 was used as the recurrent parent in this study. Lineages of the 307T progenies with one (307T1), two (307T2) and three copies (307T3) of *HUO* were selected by Southern blot till all the lines became homozygous and were retained for further experiments. Individual the 307T1, 307T2, and 307T3 lineages were further screened to remove those with copies of *HUO*. In particular, two lines representing 307T2 (307TM and 307TN) were identified and studied in detail to confirm autonomous retrotransposition. These rice lines were planted and harvested in a paddy field under the normal field management conditions. Wild rice species and rice landraces were from Hainan and other provinces, China. Archaeological rice grains were from the lower Yangtze region(Fuller et al., 2009).

Reciprocal crossing experiment

Reciprocal crossing was performed between 307T and 308D to confirm that *HUO* transposes in a germline-specific manner. The cross combination was set as (female/male): 308D/307T1, 307T1/308D, and 307T2/308D, using either 307T1 or 308D as the female parent. Ahead of crossing, the *HUO* copy number of 307T1 or

307T2 individuals, which were used for pollinating, was confirmed by Southern blot. Genomic DNA was prepared from individual F1 plants.

Stress treatment

Seedlings of 307T1 population were grown in 1/2 MS liquid medium at 28/26°C (day/night) and with 12-h day/12-h night photoperiod. One-week-old seedlings were grown in the 1/2 MS medium with 25% polyethylene glycol (PEG) 6000 (Sigma-Aldrich) for 24 h, or 150 mM NaCl for 24 h, or 100 μM abscisic acid (ABA) for 10 h. For heat stress treatment, seedlings were grown at 45°C for 10 h. For UV radiation stress, plants were placed 15 cm away under a 20 W UV light for 4 h. Germinating seeds were grown in the dark for 7 d for darkness stress treatment. The 307T1 grown in 1/2 MS liquid medium was used as the treatment control and the 308D grown in 1/2 MS liquid medium was used as the blank control. Stressed and control plants were stored in liquid nitrogen for RNA preparation. Three biological repeats were performed for each treatment.

Southern and Northern blot

For Southern blot, 5μg DNA from each samples was digested with appropriate restriction endonucleases and separated in 1% agarose gel, and then blotted onto Hybond-N^+ membranes (Amersham). For Northern blot, 2 μg total RNAs from rice tissue culture derived from 307T1 lines was separated in 0.8% agarose gel, and then blotted onto Hybond-N^+ membranes (Amersham). A Digoxin-dUTP-labelled (Roche)237-bp *HUO*-specific fragment probe was made through PCR, using specific

primers (**Supplemental Table 2)**. This labelled probe was used for Southern and Northern blot assays flowing the manufacturer's instructions (Roche).

Identification of flanking sequences of *HUO* transposon

Thermal asymmetric interlaced(TAIL)-PCR was performed to identify the new insertion sites of *HUO* in the 307T plants with more than one copy of *HUO*. The PCR programmes followed have been previously described (Liu and Huang, 1998) and the arbitrary degenerate primers and *HUO* specific primers used listed in **Supplemental Table 2**. The insertion sites were confirmed by sequencing the PCR products of the junction sequences.

DNA sequence analysis

A phylogenetic tree was constructed using the neighbor-joining method by GENEIOUS and saved as a Newick file (Kearse et al., 2012), and then the tree was visualized by uploading the Newick file in iTOL software (Letunic and Bork, 2016). The representative rice LTR retrotransposon sequences were provided by Ning Jiang (Michigan State university, USA), and were retrieved by the LTR_Retriever (Ou and Jiang, 2017).

DNA methylation analysis

For Southern blot-based DNA methylation experiments, 10 µg DNA from a sample was digested with the methylation-sensitive restriction endonucleases *Hpa*II and *Msp*I. Southern hybridization was performed using the Digoxin-dUTP-labelled DNA

probe (**Supplemental Figure 7**). For genome-wide bisulfite sequencing (Bis-seq), genomic DNA was isolated from the same two-week-old 308D, 307T1 and 307T3 seedling samples that were used for RNA-seq and small RNA-seq analysis. The genomic DNA samples (5μg each)were sheared by sonication to fragments of 100-500 bp, and the libraries were constructed following the Illumina protocol. Bisulphite treatment of DNA was performed as described by Meissner et al.(2005), and the high throughput sequencing and data analysis was performed as described (Cokusetal., 2008) by BGI (www.genomics.cn). The Bis-seq data were validated through traditional bisulphite sequencing at some loci. An additional experiment was performed to confirm DNA methylation levels using methylated DNA immunoprecipitation assay combined with the high-throughput sequencing (MeDIP-seq). DNA immunoprecipitation assay with methylated-cytosine antibody and the high throughput sequencing were carried out using the standard Illumina protocol (Gruntman et al., 2008). The reads were mapped to the reference sequences and compared between 308D and 307T3. The complete sets of bisulfite sequencing and MeDIP-Seq data have been deposited in the National Center for Biotechnology Information GEO database (http://www.ncbi.nlm.nih.gov/geo/) under series accession number GSE78903.

RNA analysis

Total RNAs were isolated from the stressed seedlings, tissue cultures, different organs of the spikelet, and wild rice leaves using the TRIzol reagent. Quantitative real-time PCR (qRT-PCR) was performed to analyze gene expression levels using specific primers (**Supplemental Table 2**). Rice RNA-seq and data analysis were performed as

previously described (Lu et al., 2010) by BGI (www.genomics.cn). The RNA samples were isolated from shoots of two-week-old 308D, 307T1 and 307T3 seedlings grown under 12-h-light/12-h-dark at 28°C in a greenhouse. Gene ontology analysis of up and down-regulated genes was performed in 307T1 in comparison with 307T3 using the comprehensive annotation platform CARMO (http://bioinfo.sibs.ac.cn/carmo/). The entire RNA sequencing data can be obtained from the NCBI Gene Expression Omnibus (GEO) database under series accession number GSE78903.

Deep sequencing of small RNAs

Small RNAs were isolated from the same total RNAs f samples for RNA analysis, then reverse-transcribed. Small RNA libraries were constructed and deep-sequenced with an Illumina Hiseq 2000 according to the manufacturer's instructions (Illumina) by BGI. Small RNAs were mapped to the *HUO* and genomic sequences using software Bowtie-0.12.7 (Langmead et al., 2009). Sequences from different samples were normalized based on the number of total reads with perfect genomic matches. Abundances of 21 and 24-nt siRNAs were compared by removing microRNAs. The entire small RNA sequencing data can be obtained from the NCBI Gene Expression Omnibus (GEO) database under series accession number GSE78903.

Development of *DCL3, AGO4,* and *VOZ1*-RNAi transgenic plants

cDNA fragments representing partial sequences of the key genes in the RdDM pathway, *OsDCL3a* (LOC_Os01g68120), *OsAGO4a* and *OsAGO4b* (conserved fragment of LOC_Os01g16870 and LOC_Os04g06770) (Wu et al., 2010) and a

negative control gene *VOZ1* (Kumar et al., 2018), were amplified using gene-specific primers (**Supplemental Table 2**) from the cDNA templates of Nipponbare then individually inserted as inverted repeats into the conventional RNAi vector PTCK303 to conduct the hairpin RNAi constructs. The resulting RNAi vectors were introduced into variety 307T3 via *Agrobacterium*-mediated transformation to generate more than 20 independent RNAi lines. Plants transformed with the empty PTCK303 vector were also produced and used as controls.

Salt tolerance assay

Salt tolerance assay was performed as described previously(Ren et al., 2005). Germinated seeds of 307T1 and 307T3 were incubated in 1/2 MS liquid medium in a growth chamber set at 26°C and with a 13-h light/11-h dark photoperiod. Two-week-old seedlings were transferred into 1/2 MS in a 1/2 MS liquid medium containing 150 mM NaCl for 10 days, followed by a recovery with water for 5 days. Survival rates of the treated seedlings were recorded and statistically analyzed by Student's *t*-test.

Pathogen inoculation

For disease resistance assay, two-month-old rice plants were inoculated with *Xanthomonas oryzae* pv. *oryzae* virulent race P6 (strain PXO99) by the leaf-clipping method. Lesion length was measured after 2 weeks post inoculation as previously reported (Yang et al., 2008) and statistically analyzed by Student's *t*-test.

AUTHOR CONTRIBUTIONS

Y.P., Y.D., Y.Z., B.L. and Z.H conceived and initiated the study. Z.H. designed and supervised the entire study. Y.P., Y.D., Y.Z., Y.G., D.A., J.L., Q.L., L.F. and Z.H. performed experiments and data analysis. Y.D., J.L., X.X. and W.W. performed bioinformatic analysis. Y.P., J.W., S.T. and C.S. grew and prepared rice materials. Y.P., Y.D., Y.Z., L.F., B.L. and Z.H. wrote the manuscript.

ACKNOWLEDGMENTS

We thank S. Wessler for discussion and useful suggestions. X. Cao, J. Wang and X. Song for critical reading and suggestions; N. Jiang for providing rice LTR sequences; L. Zeng for help with rice growth; Y. Jia for providing rice materials; Y. Zhang and X. Li for help with the bioinformatic analysis; Supported by grants from National Natural Science Foundation of China (31330061, 31570269, 31772149 and 31720103913), the Ministry of Science and Technology of China (2016YFD0100600)

REFERENCES

Arikit, S., Zhai, J.X., and Meyers, B.C. (2013). Biogenesis and function of rice small RNAs from non-coding RNA precursors. Curr. Opin. Plant Biol. 16: 170-179.

Cheng, C.Y., Tarutani, Y., Miyao, A., Ito, T., Yamazaki, M., Sakai, H., Fukai, E., and Hirochika, H. (2015). Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. Plant J. 83: 1069-1081.

Calarco, J.P., Borges, F., Donoghue, M.T., Van, Ex. F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., Martienssen, R.A. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell 151: 194-205.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215-219.

Deng, Y., Zhai, K., Xie, Z., Yang, D., Zhu, X., Liu, J., Wang, X., Qin, P., Yang, Y., Zhang, G., et al. (2017). Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. Science 355: 962-965.

Doebley, F., Gaut, S., and Smith, D. (2006). The molecular genetics of crop domestication. Cell 127: 1309-1321.

Dowen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Dowen, J.M., Nery, J.R., Dixon, J.E., and Ecker, J.R. (2012). Widespread dynamic DNA methylation in response to biotic stress. Proc. Natl. Acad. Sci. USA 109: E2183-E2191.

Elbarbary, R.A., Lucas, B.A., and Maquat, L.E. (2016). Retrotransposons as regulators of gene expression. Science 351: aac7247.

Fan, L., Gui, Y., Zheng, Y., Wang, Y., Cai, D., and You, X. (2011). Ancient DNA sequences of rice from the low Yangtze reveal significant genotypic divergence. Chinese Sci. Bull. 56: 3108-3113.

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

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

Fuller, Q., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, A., and Sun, P. (2009). The domestication process and domestication rate in rice: spikelet bases from the lower Yangtze. Science 323: 1607-1610.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447-1451.

Gruntman, E., Qi, Y., Slotkin, R.K., Roeder, T., Martienssen, R.A., and Sachidanandam, R. (2008). Kismeth: analyzer of plant methylation states through bisulfite sequencing. BMC Bioinformatics 9: 371.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996). Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA 93: 7783-7788.

Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451-1454.

Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu,

H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. Nature 490: 497-501.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., Zilberman, D. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337: 1360-1364.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472: 115-119.

Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr. Biol. 22: 1825-1830.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer T., Ashton B., Meintjes P., and Drummond A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Kim, Y., and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. Trends Plant Sci. 19: 320-326.

Kumar, A. and Bennetzen, J.L. (1999). Plant retrotransposons. Annu. Rev. Genet. 33: 479-532.

Kumar, S., Choudhary, P., Gupta, M., and Nath, U. (2018). Vascular plant one-zinc finger1 (VOZ1) and VOZ2 interact with CONSTANS and promote photoperiodic flowering transition. Plant Physiol. 176: 2917-2930.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11: 204-220.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44: W242-W245.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington J.C., Doerge R.W., Colot V., and Martienssen R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471-476.

Lisch, D. (2013). How important are transposons for plant evolution? Nat. Rev. Genet. 14: 49-61.

Liu, Y.G., and Huang, N. (1998). Efficient amplification of insert end sequences from bacterial artificial chromosome clones by thermal asymmetric interlaced PCR. Plant Mol. Biol. Rep. 16: 175-181.

Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y., Li, W., Huang X., and Han B. et al. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. Genome Res. 20: 1238-1249.

Marí-Ordóñez A., Marchais A., Etcheverry M., Martin A., Colot V., and Voinnet O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. Nat. Genet. 45, 1029-1039.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R.

(2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33: 5868-5877.

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. 14: 840-852.

McCue, A.D., Nuthikattu, S., Reeder, S.H., and Slotkin, R.K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS Genet. 8: e1002474.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. Nature 461:427-430.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T., and Wessler, S.R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130-1134.

Olsen, K.M., and Wendel, J.F. (2013). A Bountiful harvest: Genomic insights into crop domestication phenotypes. Annu. Rev. Plant Biol. 64:47-70.

Ou, S., and Jiang, N. (2018). LTR_retriever: A hghly accurate and sensitive program for identification of LTR retrotransposons. Plant Physiol. 176: 1410-1422.

Perez-Hormaeche, J., Potet, F., Beauclair, L., Le Masson, I., Courtial, B., Bouche, N., and Lucas, H. (2008). Invasion of the Arabidopsis genome by the tobacco retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing. Plant Physiol. 147: 1264-1278.

Picault, N., Chaparro, C., Piegu, B., Stenger, W., Formey, D., Llauro, C., Descombin, J., Sabot, F., Lasserre, E., Meynard, D., Guiderdoni E., and Panaud O. (2009). Identification of an active LTR retrotransposon in rice. Plant J. 58: 754-765.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A., and Panaud O. (2006). Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Res. 16: 1262-1269.

Purugganan, D., and Fuller, Q. (2009). The nature of selection during plant domestication. Nature 457: 843-848.

Ren, Z.H., Gao, J.P., Li, L.G., Cai, X.L., Huang, W., Chao, D.Y., Zhu, M.Z., Wang, Z.Y., Luan, S., and Lin, H.X. (2005). A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat. Genet. 37: 1141-1146

Wing, R.A., Purugganan, M.D., Zhang, Q. (2018). The rice genome revolution: from an ancient grain to Green Super Rice. Nat Rev Genet. 19: 505-517.

Rutger, J.N., and Carnahan, H.L. (1981). A fourth genetic element to facilitate hybrid cereal production-recessive tall in rice. Crop Sci. 21: 373-376.

Schoft, V.K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep. 10: 1015-1021.

Slotkin, K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8: 272-285.

Slotkin, K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and rmall RNA silencing of transposable elements in pollen. Cell 136: 461-472.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in

Arabidopsis. Nat. Struct. Mol. Biol. 21: 64-72.

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43: 1160-1163.

Vitte, C., Panaud, O., and Quesneville, H. (2007). LTR retrotransposons in rice (Oryza sativa, L.): recent burst amplifications followed by rapid DNA loss. BMC Genomics 8: 218.

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., Zhang, L., He, W., Lu, B., Lin, H., Ma H., Zhang G., and He Z. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat. Genet. 40: 1370-1374.

Wang, X., Weigel, D., and Smith, M. (2013a). Transposon variants and their effects on gene expression in Arabidopsis. PLoS Genet. 9: e1003255.

Wang, Z.H., Zhang, D., Bai, Y., Zhang, Y.H., Liu, Y., Wu, Y., Lin, X.Y., Wen, J.W., Xu, C.M., Li, L.F., et al. (2013b). Genomewide variation in an introgression iine of rice-zizania revealed by whole-genome re-sequencing. PloS One 8: e0074479.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., and Qi, Y. (2010). DNA methylation mediated by a microRNA pathway. Mol. Cell 38: 465-475.

Yang, D.L., Li, Q., Deng, Y.W., Lou, Y.G., Wang, M.Y., Zhou, G.X., Zhang, Y.Y., and He, Z.H. (2008). Altered disease development in the eui mutants and Eui overexpressors indicates that gibberellins negatively regulate rice basal disease resistance. Mol. Plant 1: 528-537.

Zhu, Y., Nomura, T., Xu, Y., Zhang, Y., Peng, Y., Mao, B., Hanada, A., Zhou, H., Wang, R., Li, P., Zhu X., Mander L.N., Kamiya Y., Yamaguchi S., and He Z. (2006). *ELONGATED UPPERMOST INTERNODE* encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. Plant Cell 18:442-456.

FIGURE LEGENDS

Figure 1. *HUO* is an active LTR retrotransposon under normal growth conditions.

(**A**) *HUO* was originally identified as an insert in the *eui-1* mutant. (**B**) Schematic of *HUO*. TSD, target site duplication; LTR, long terminal repeat; PBS, primer binding site; GAG, group-specific antigen protein; PRO, proteinase; IN, integrase; RT, reverse transcriptase; RH, RNase H; PPT, polypurine tract. (**C**) Retrotransposition occurred under normal growth conditions. *HUO* inserts were detected by Southern blotin six progeny plants from two 307T plants, plant 1 (left panel) and plant 2 (right panel). Genomic DNAs were digested by *Bam*HI. Stars indicate the original copy in the *eui-1* gene, triangles indicate new insertions. (**D**, **E**) New *HUO* insertions were confirmed in 307T progeny by Southern blot. 307TM line contains *HUO* in *Os02g30910* (*MtN3*), and 307TN line contains *HUO* in *Os06g39360* gene (NBS-LRR). 307TN genomic DNA digested by *Eco*RI produced two nearly equal-length fragments, 12,033 bp and 12,187 bp, as shown in (**E**). Stars indicate the original copy in *eui-1*. E, *Eco*RI; B, *Bam*HI; S, *Sac*I. (**F**) *HUO* preferentially inserted into the genome regions capable of transcription. A total of 16 new insertions were detected by TAIL-PCR in the 307T1 population, with most *HUO* insertions near or in the expressed genes or hypothetical ORFs. The distances between the inserts and the genes are shown. Arrows indicate the direction of *HUO* insertion. The diagrams represent relative positions. For Southern blot, *HUO*-specific sequence was used as probe (**D** and **Supplemental Table 2**).

Figure 2. *HUO* is sensitive to abiotic stresses and retrotransposes through male gametogenesis.

(**A**) Developmental and tissue-specific expression of *HUO* detected by RT-PCR. L, leaf; S, sheath; P, young panicle; I, elongating internode. (**B**) Stress induction of *HUO* expression. Seedlings were treated with 150 mM NaCl (24 h), 40% PEG4000 (24 h), 100 mM ABA (10 h), 45°C (10 h), UV radiation (4 h) or dark (7 d). 308D, negative control without *HUO*. (**C**) Southern blot detection of *HUO* copies in the individuals of continuous generations indicating that *HUO* transposition occurred during gametogenesis. Left panel, twelve individuals from self-fertilized 307T progeny that contained the original *HUO* copy inserted in *eui-1*. Note that four individuals contained one (plants 5, 8 and 12) or two (plant 9) new *HUO* insertions. Right panel, twelve individual next generation plants from plants 5, 8 and 9 in the left panel. (**D**) Transposition of *HUO* occurred during male gametogenesis. The parent plants (307T and 308D) were reciprocally cross-pollinated and 12 or 17 F_1 individuals were detected. New *HUO* insertions were found only when 307T was used as the paternal (left) but not maternal parent either with one (307T1) or two (307T2) *HUO* copies. Triangles denote the new *HUO* copies in the F1 plants. For Southern blot, *HUO*-specific sequence was used as probe and genome DNAs were digested by *Bam*HI.

Figure 3. *HUO* is suppressed by DNA methylation with copy number increasing.

(**A**) Transposition frequency of *HUO* was decreased from 23.3% to 14.4% and 6.2%, when the initial copy number of *HUO* increased from one (307T1) to two (307T2) and three (307T3), respectively. (**B**) Transcription levels of *HUO* were significantly decreased in young panicles of 307T3 compared with 307T1. Values are the means ±

SD; **, statistically significant difference by Student's *t*-test ($P < 0.01$). (**C**) Increased DNA methylation of *HUO* in 307T2 and 307T3 compared with that in 307T1. Genomic DNA was extracted from leaf samples and then digested by the methylation-sensitive enzymes *Hpa*II (H) and *Msp*I (M). Numbers on the right side indicate the length (bp) of digested fragments (**Supplemental Figure 7A**). (**D**) Bisulfite-sequencing analysis of the left LTR and the junction region of *HUO*. Genomic DNA samples were extracted from the leaf samples and the details of the single clones were shown in **Supplemental Figure 7B**.

Figure 4. Small RNAs accumulated from *HUO* sequence and were involved in *HUO* repression.

(**A**) 24-nt sRNA mapping revealed the different accumulation of sRNAs corresponding to the regions of the *HUO* element. Majority of *HUO*-derived sRNAs was 24-nt, which were likely differentially accumulated in 307T1 and 307T3. (**B**) Generation of *DCL3*-RNAi and *AGO4*-RNAi lines. The specific *DCL3a* region and conserved *AGO4a* and *AGO4b* region were inserted into the double-strand RNAi vector pTCK303. Plasmids PTCK303-OsDCL3a-RNAi and PTCK303-OsAGO4ab-RNAi were introduced individually into 307T3 to generate individual RNAi lines. (**C**) qRT-PCR result showed that the expression levels of *DCL3a* and *AGO4ab* were greatly decreased in the RNAi lines. (**D**) Increased copies of *HUO* in the 307T3/*DCL3a*-RNAi and 307T3/*AGO4ab*-RNAi lines, with up to 8 copies in 307T3/*DCL3a*-RNAi compared to the usual 3 copies in the 307T3 progeny, indicating the release of *HUO* silencing. Genomic DNAs from individual plants of 307T3/*DCL3a*-RNAi (4 plants) and 307T3/*AGO4ab*-RNAi (7 plants) were digested with *Eco*RI and subjected to Southern hybridization. Note that some strong bands may

represent multiple *HUO* copies, and weak bands likely indicate possible heterozygosity. (**E**) The expression levels of *HUO* were greatly increased in the siRNA defective 307T3/*DCL3a*-RNAi and 307T3/*AGO4ab*-RNAi lines. **, statistically significant difference by Student's *t*-test ($P < 0.01$) **(C, E)**.

Figure 5. Genetic evidence of *HUO* removal during rice domestication.

(**A**) Ratio of *HUO* presence in wild rice (*O. refupogen*, *O. nivara* and *O. longistaminata*), landraces, and modern rice cultivars with *HUO*. Note that only one modern variety Shuangkezao has one copy of *HUO* (see **Supplemental Table 1** for details of the accessions). (**B**) Average copy numbers of *HUO* in the wild rice and landraces, revealed by Southern blot. The copy number of *HUO* was greatly decreased from wild rice to domesticated landraces (see **Supplemental Table 1** for details). (**C**) Expression of *HUO* was induced by NaCl (150 mM, 24 h) in the leaves of wild rice (*O. refupogen*) and seedlings of two landraces. Transcript levels of *HUO* at 0 h (control) were set as one. (**D**) *HUO* was detected in the archaeological rice grains. Representative archaeological rice grains (~2400-year-old) were photographed immediately after archaeological excavation. The insert at the right-up corner is a magnified archaeological grain (carbonized due to exposure after excavation). (**E**) *HUO* detection showed its presence in half of the archaeological rice grain DNA preparations (C, ~7000-year-old grains; E, ~2400-year-old grains; and T, ~1200-year-old grains) found in the lower Yangtze River corridor in China. PCR was performed using *HUO*-specific primers to produce the 237-bp fragment (2752-2988 bp), with 307T (+) and 308D (-) as controls. Additional PCR assays obtained similar results.

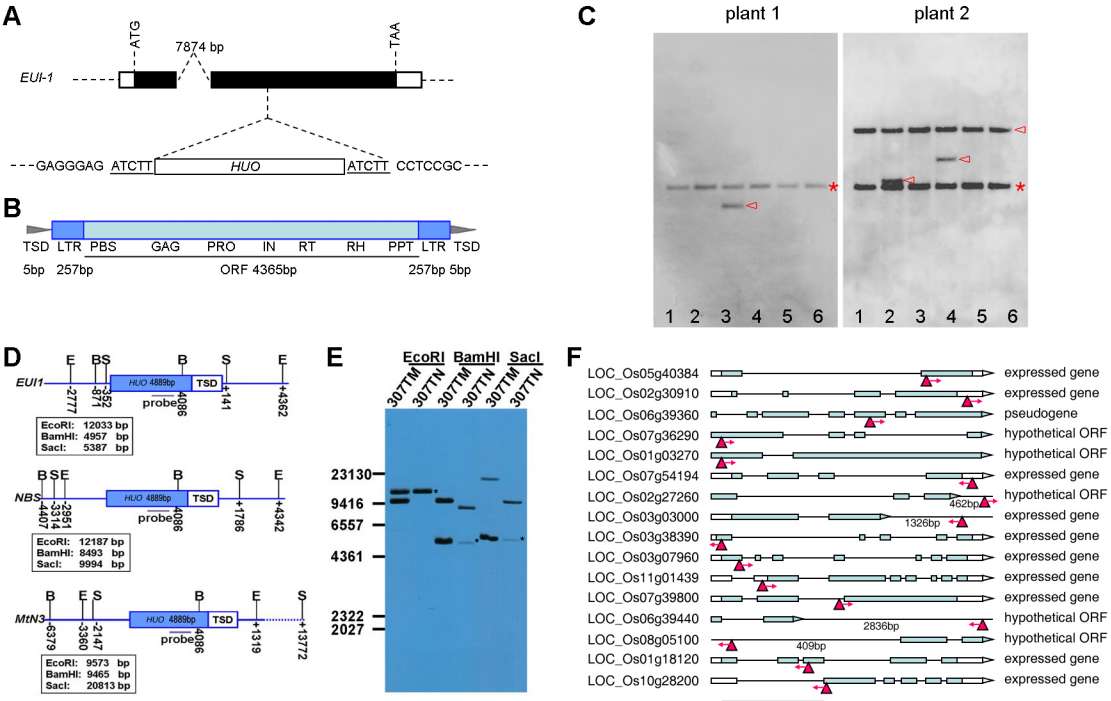
Figure 6. *HUO* induces epigenomic changes.

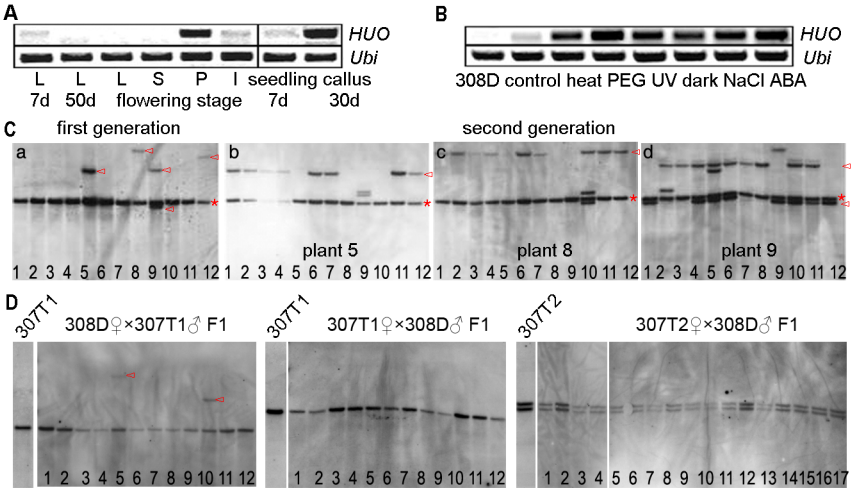
(**A**) Increased genome-wide methylation in 307T1 compared with 307T3 and 308D, as revealed by bisulfite-sequencing. The details are shown in **Supplemental Figure 12**. (**B**) Average methylation levels of the regions 1-kb- upstream to 1-kb downstream of transcribed regions (left panel) or repeat sequences (mainly transposons, right panel). (**C**, **D**) Differentially methylated regions (DMR) between 307T1 and 307T3 mainly occurred in CG and CHG (**C**), and were mostly distributed in repeat regions (**D**). Note that there were only a few CHH DMR in 307T1 compared with 307T3. (**E**) Increased accumulation of 24-nt small RNAs in 307T3 (29.4%±0.4%) compared with 307T1 (22.2%±3.4%), as revealed by deep small RNA sequencing.

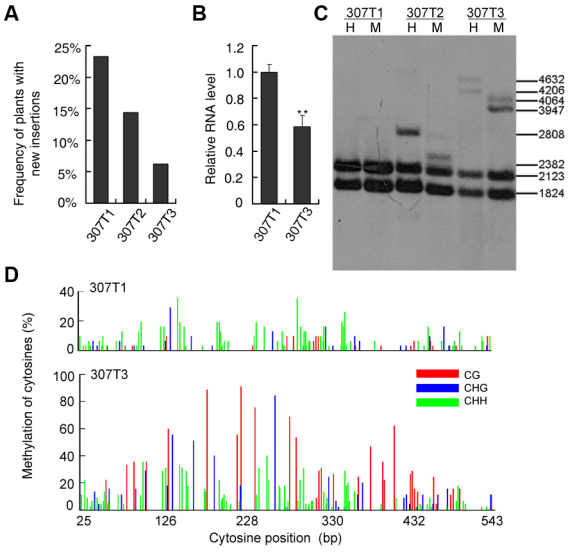
Figure 7. *HUO* decreases agronomic values in cultivated rice.

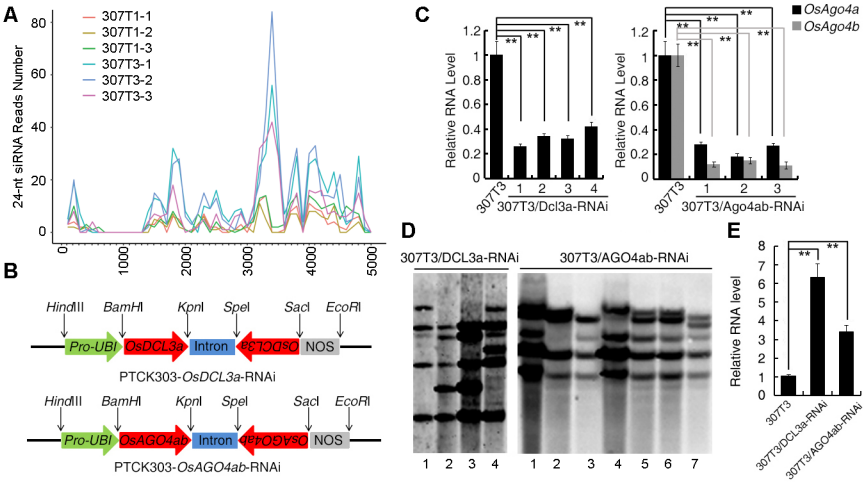
(**A, B**) Yield performance assay showed that the tilling number (**A**) and grain yield per plant (**B**) were significantly decreased in 307T3 in comparison with 307T1. Values are means ± SE (n = 150), statistical significance was determined by Student's *t*-test. Two-seasons' experiments showed similar results. (**C**, **D**) 307T3 plants were more susceptible to bacterial blight (*Xanthomonas oryzae* pv. *oryzae*) than 307T1 plants (**C**), with significantly longer lesions (**D**). Values are means ± SD (n = 30), statistical significance was determined by student's *t*-test. (**E**) qRT-PCR was performed to detect the transcript levels of three predicted disease resistance receptor (*R*) genes, including two NLR (nucleotide-binding leucine-rich repeat) receptor genes (LOC_Os11g38440 and LOC_Os11g38580) and one receptor-like kinase (LOC_Os07g03810), in two-week-old seedlings of 307T3 and 307T1, indicating that the transcriptional levels of the predicted *R* genes decreased in 307T3 in comparison with 307T1. Stars indicate statistically significant difference by Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$). (**F**)

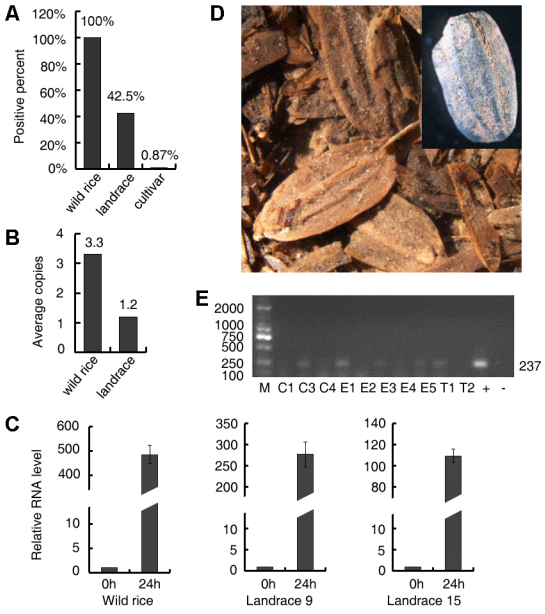
qRT-PCR analysis of the defense genes, *OsPR10*, *OsWRKY77* and *OsWRKY14*. Two-weeks-old seedlings were inoculated with *Xanthomonas oryzae* pv. *oryzae* (strain P6), and leaves were harvested at 0, 12, 24, and 48 h after inoculation. (**G**, **H**) 307T3 plants were more tolerant to salt stress (**G**), and had a significantly higher survival rate than 307T1 during salt treatment (**H**). Survival rates were recorded and statistically analyzed by student's *t*-test. Values are means ± SD (n = 3). *P* value was indicated. (**I**) Induction of the abiotic stress genes *OsHKT4* was enhanced in 307T3 in comparison with 307T1 as revealed by qRT-PCR. Data were shown as means ± SD (n = 3). *OsActin1* was used as an internal control (**E**,**F**,**I**). (**J**) A proposed model for *HUO* removal during rice domestication and breeding. *HUO* triggers genomic defense through DNA methylation that greatly changes expression profiling of genes involved in development and defense, consequently leading to the decrease in yield and disease resistance. *HUO* copy numbers were decreased from wild rice to landraces and finally removed in modern varieties.

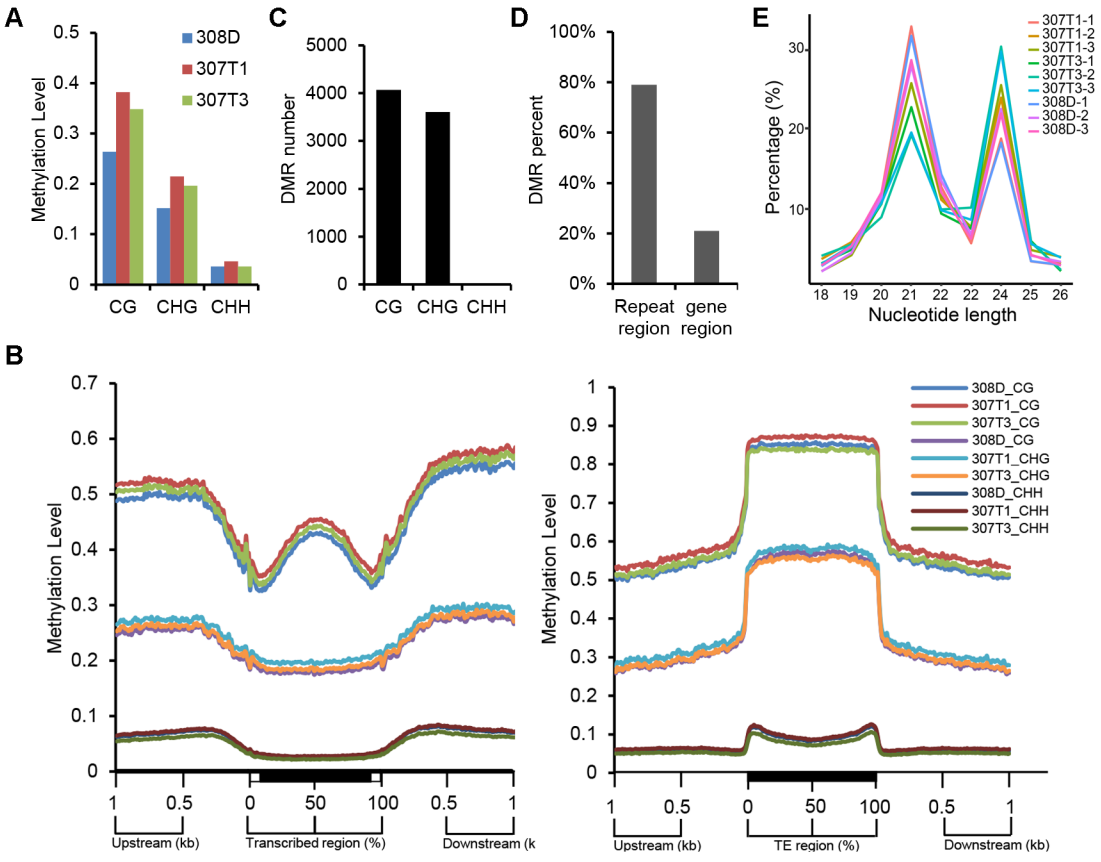


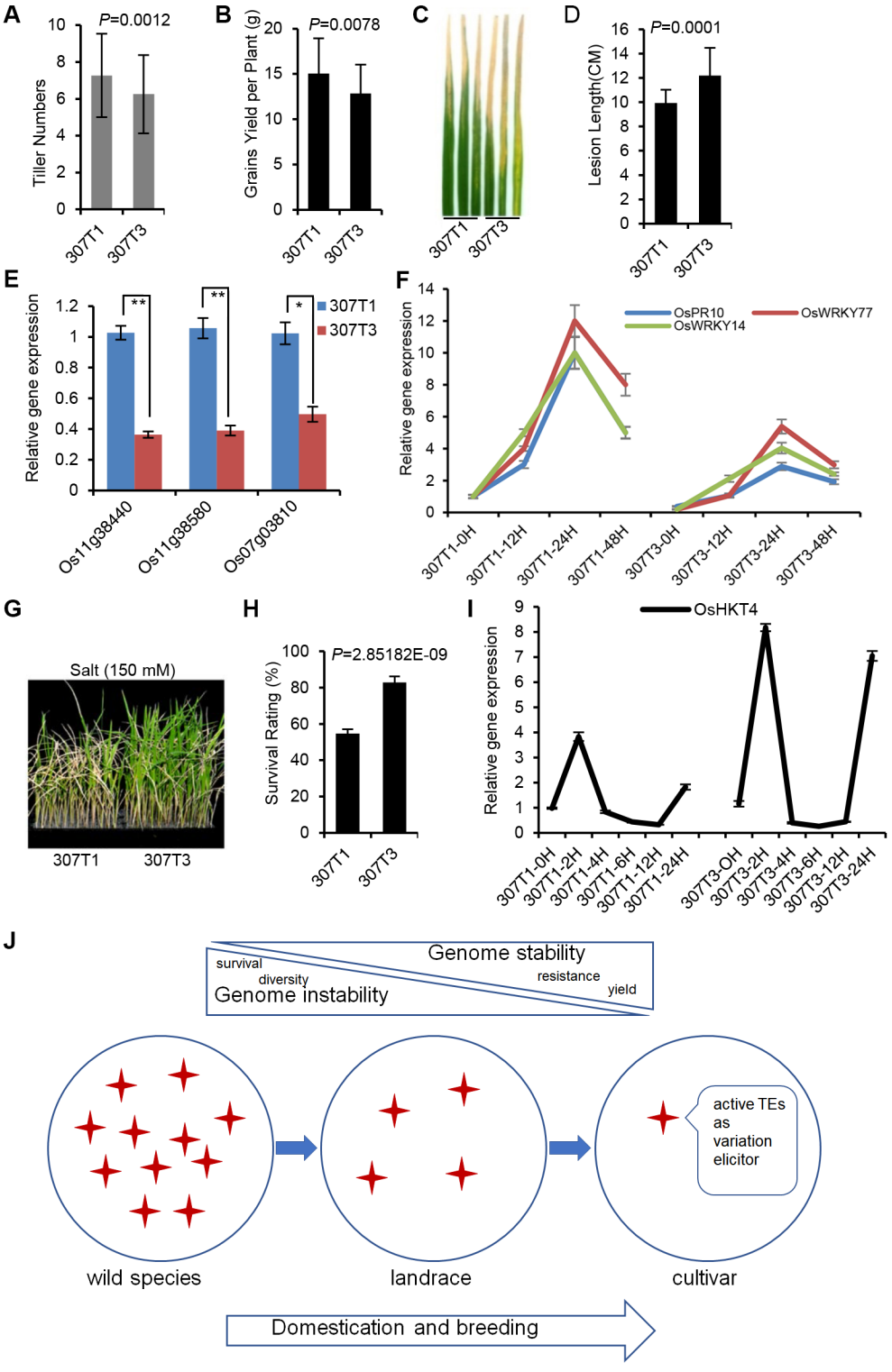











A
P=0.0012
Tiller Numbers
307T1
307T3
B
P=0.0078
Grains Yield per Plant (g)
C
D
P=0.0001
Lesion Length(CM)
E
Relative gene expression
307T1
307T3
Os11g38440
Os11g38580
Os07g03810
F
OsPR10
OsWRKY77
OsWRKY14
307T1-0H
307T1-12H
307T1-24H
307T1-48H
307T3-0H
307T3-12H
307T3-24H
307T3-48H
G
Salt (150 mM)
H
P=2.85182E-09
Survival Rating (%)
I
OsHKT4
307T1-2H
307T1-4H
307T1-6H
307T3-2H
307T3-4H
307T3-6H
J
Genome stability
survival
diversity
resistance
yield
Genome instability
active TEs as variation elicitor
wild species
landrace
cultivar
Domestication and breeding