

# Origination and Establishment of a Trigenic Reproductive Isolation System in Rice

Dear Editor,

Reproductive isolation is both the indicator and a primary force of speciation, and plays a key role in maintaining species identity. Understanding the origin and mechanisms of reproductive isolation is of fundamental importance in evolutionary biology. In recent years, a number of genes that induce reproductive barriers have been identified in several model organisms such as *Drosophila*, rodents, yeast, *Arabidopsis*, rice, and other species, which has greatly advanced our understanding of the mechanisms regulating reproductive isolation (Maheshwari and Barbash, 2011; Ouyang and Zhang, 2013; Chae et al., 2014; Lafon-Placette and Kohler, 2015 for a review). However, little is known about how the genomes evolve to form reproductive barriers and how such systems become established in populations, although such reproductive barriers are widely observed and demand understanding for both biological study of speciation and practical application in crop improvement.

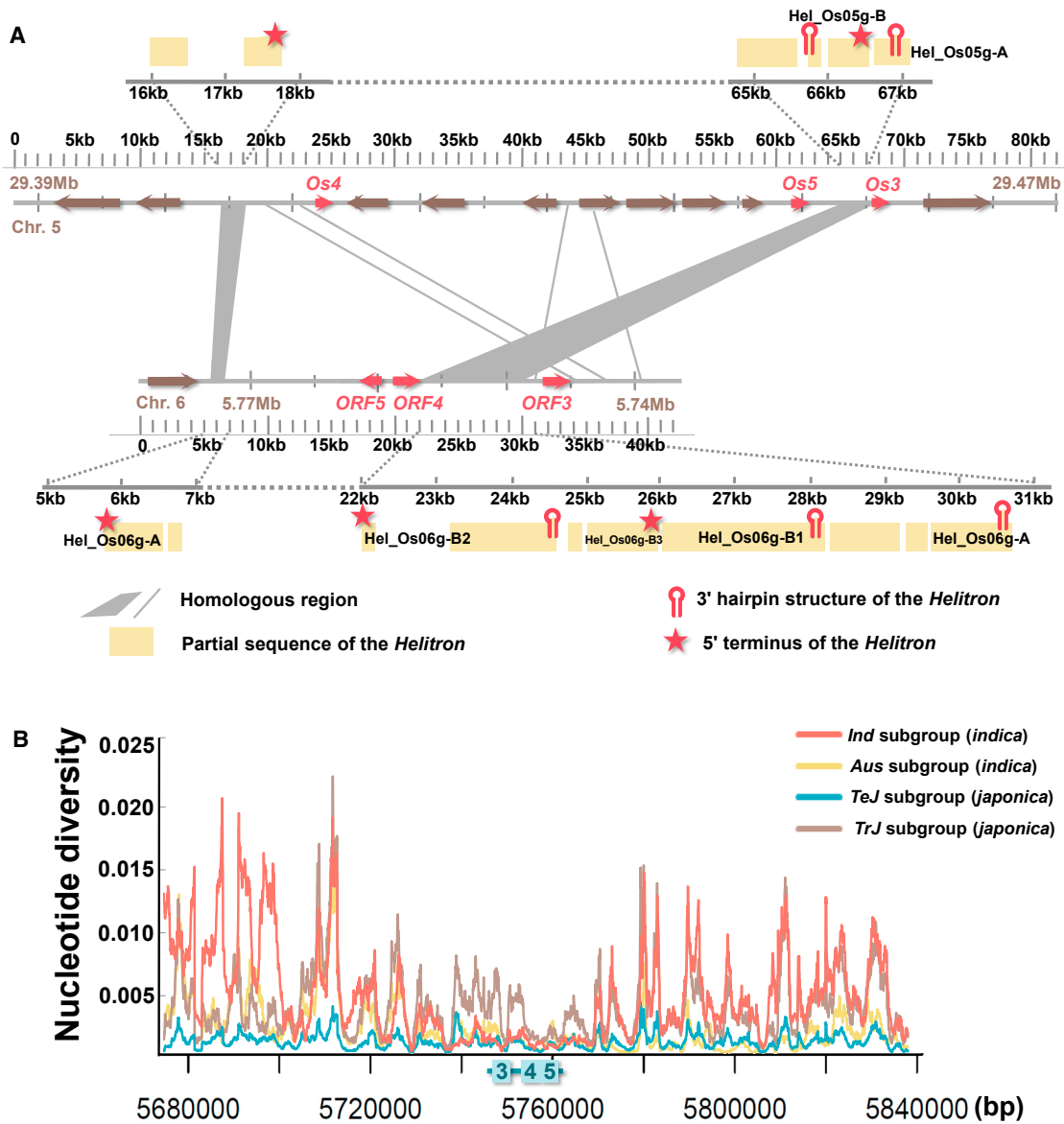
The Asian cultivated rice (*Oryza sativa* L.) comprises two subspecies, *indica* and *japonica*. Hybrid sterility between *indica* and *japonica* represents one of the best-characterized examples of postzygotic reproductive isolation in plants. We previously reported a killer-protector system at the S5 locus composed of three tightly linked genes, *ORF3*, *ORF4*, and *ORF5*, which together regulated the female gamete fertility of *indica-japonica* hybrids (Chen et al., 2008; Yang et al., 2012). Typical *indica* and *japonica* varieties contain the haplotypes of *ORF3+ORF4-ORF5+* and *ORF3-ORF4+ORF5-*, respectively. The killer is made of *ORF4+* and *ORF5+*, which work together to kill the gamete, while *ORF3+* confers protection to the gametes, such that female gametes without *ORF3+* are selectively eliminated, causing hybrid sterility and segregation distortion.

A reciprocal BLASTN search using *ORF3*, *ORF4*, and *ORF5* as queries identified a unique set of three tightly linked genes on chromosome 5 (designated as *Ospara3-5*) showing 73.72%, 67.74%, and 71.37% identities to *ORFs3*, 4, 5 (chromosome 6), respectively (Supplemental Table 1). A TBLASTN search identified another set of three genes located adjacent to each other on chromosome 12 (designated as *Os345-like*), which showed less similarity to either *ORFs3-5* or *Ospara3-5*. Sequences with significant similarity to either *ORFs3-5*, *Ospara3-5*, or *Os345-like* were identified in other species by reciprocal blast (Supplemental Table 1). These similar sequences also located adjacent to each other, although they have frequently changed their relative orientations. One group of similar sequences showed higher similarity with *ORFs3-5* or *Ospara3-5* (labeled as para), and the other group was more similar to *Os345-like* (labeled as like). Therefore, *ORF3-5/Ospara3-5* and *Os345-like* belonged to two different lineages, which led us to focus our attention on the *ORF3-5/Ospara3-5*

pair relationship in subsequent pursuit of the origin of S5. Based on phylogenetic trees, the *Ospara3-5* were more closely related to most hits in the outgroup species than to *ORFs3-5* (Supplemental Figure 1A–1C). Thus, the *Ospara3-5* are likely to be more ancestral than the S5 locus, suggesting that *ORFs3-5* were derived from *Ospara3-5*. Interestingly, we found a TBLASTN hit in *Z. latifolia* (ZIORF4), which belonged to the Oryzae tribe, showed higher similarity to *ORF4* than to other sequences (sequences labeled as para). Therefore, the three-gene fragment of the S5 complex might have originated from *Ospara3-5* after the formation of the Oryzae tribe (Supplemental Figure 2).

A question thus arose: how did the *Ospara3-5* fragment duplicate to form the *ORFs3-5* block? We investigated the sequence similarity between the *ORFs3-5* and *Ospara3-5* fragments, which identified several intergenic regions showing high similarity between the two homologous blocks (Figure 1A, Supplemental Table 2). These regions contained multiple repetitive sequences of the RPO\_OS *Helitron* transposon elements (<http://www.girinst.org/>) (Kapitonov and Jurka, 2007). We detected *Helitron* termini bracketing *Ospara4* and *Ospara5* (Hel\_Os05g-A). And similarly, *ORF4* and *ORF5* were also flanked by *Helitron* termini (Hel\_Os06g-A). This suggested a possibility that the *ORFs3-5* segment was derived by *Helitron* movement, carrying the internal *Ospara4-5* fragment on chromosome 5 and capturing the 3'-terminal *Ospara3*, to insert them on chromosome 6 thus generating a new fragment containing *ORFs3-5* (Supplemental Figure 3). This inference was partly supported by previous observations that *Helitron* transposition mediated by a rolling circle mechanism may fail to recognize the assumed 3' palindrome termination signal and to capture the DNA fragment downstream (Kapitonov and Jurka, 2007). After duplication through *Helitron* transposition, a combination of mutational mechanisms acted together to recreate the three genes of the S5 locus from *Ospara3-5* (Supplemental Figure 4A–4C), thus conferring functional divergence to the newborns giving rise to reproductive barriers.

To learn how the three genes evolved into a reproductive barrier, we analyzed patterns of variation and diversity of *ORFs3-5* in 635 rice accessions from 62 geographically diverse countries, representing two cultivated and 11 wild species (Supplemental Tables 3 and 4). The *ORF3+ORF4+ORF5+* was the most frequent type (56.94%) in the wild species (Supplemental Table 5), and *ORF3+*, *ORF4+*, and *ORF5+* were the most frequent alleles in both *O. sativa* (0.6735, 0.4822, and 0.5797) and wild species (0.9306, 0.7222, and 0.7639) (Supplemental



**Figure 1. Sequence and genetic variation analysis in the S5 region.**

(A) Sequence analysis was performed between S5 and its paralogous regions in rice. The numbers in brown and black indicate genomic positions on the chromosomes (in Mb) and the relative scales (kb) in the analyzed regions, respectively. (B) Nucleotide diversity was calculated based on polymorphism data flanking the S5 locus (4000 SNPs) in 529 rice accessions of *O. sativa*. The position of the S5 region is shown at the bottom. See [Supplemental Methods](#) for details.

Table 6), suggesting that they may be either the ancestral alleles or under selection. We further analyzed the functional Indels/SNPs of *ORFs3–5* in their parent genes. Interestingly, the 11-bp sequence characteristic of *ORF4+* and functional SNP of C819 characteristic of *ORF5+* were identified in *Ospara4* and *Ospara5*, respectively, suggesting that they were the ancestral alleles (Supplemental Figure 4D) (Yang et al., 2012).

The ancestral *ORF3+ORF4+ORF5+* represented a balance between killing and protecting. Therefore, mutation in the protector *ORF3+* alone would be unable to survive because of a functional killer. Thus, the mutation of *ORF3+* to *ORF3–* must have occurred after the emergence of non-functional killer, either

*ORF4–* or *ORF5–* (including *ORF5n*). Based on our data, a mutation that occurred in *ORF4+* in the ancestral populations resulted in *ORF3+ORF4–ORF5+* giving rise to the typical *indica* rice. Similarly, *ORF5–* emerged in the ancient population with *ORF3+ORF4+ORF5+* to generate the intermediate population carrying *ORF3+ORF4+ORF5–*. An important question was whether *ORF3–* arose in the population carrying *ORF3+ORF4–ORF5+* or the one with *ORF3+ORF4+ORF5–*. The absence of *ORF3–ORF4–ORF5+* ruled out the first possibility. In contrast, *ORF3–ORF4+ORF5–* was observed in high frequency indicating that the *ORF3–* allele arose in the population carrying *ORF3+ORF4+ORF5–*, thus generating the typical *japonica* genotype *ORF3–ORF4+ORF5–* and subsequently

spread in the population (Supplemental Figure 5). The inference was well supported by the phylogenetic tree (Supplemental Figure 6) and the co-existence of *ORF3+ORF4+ORF5-* and *ORF3-ORF4+ORF5-* in wild rice accessions from Jiangxi Province of China (Supplemental Table 3). Consequently, a reproductive barrier occurred in hybridization between the populations carrying *ORF3+ORF4-ORF5+* and ones with *ORF3-ORF4+ORF5-*.

An interesting question now is: what are the evolutionary forces that drive the mutually incompatible haplotypes, *ORF3+ORF4-ORF5+* and *ORF3-ORF4+ORF5-*, to high frequencies to establish such at the population level? We investigated variation patterns of 4000 SNPs in regions flanking *S5* (~154 kb) in 529 accessions of *O. sativa* (<http://ricevarmap.ncpgr.cn/django/home/>) (Figure 1B). A dramatic decrease in genetic variation was observed around *S5* region (~44 kb) in the *indica* subgroup, which is in contrast to the patterns observed in other rice groups especially compared with the temperate *japonica* subgroup. We further conducted tests for selection in different rice lineages (Supplemental Table 7). Significant negative values were detected for all three genes at the *S5* locus in *indica* rice using Fay and Wu's *H*, Fu and Li's *D\** and *F\** tests, suggesting that the spread of the *indica* haplotype might be driven by natural selection, possibly due to fitness advantage. We also determined whether demography is responsible for the departure from neutrality of the *S5* locus in *indica* rice. The coalescent simulation results revealed significant values for Fay and Wu's *H*, Fu and Li's *D\** and *F\** tests in *indica* populations for all three genes at the *S5* locus (Supplemental Table 8). Taken together, the results suggested that the *indica* haplotype of the *S5* locus was under strong positive selection during evolution and/or rice breeding. Such positive selection may have provided a driving force leading to the establishment of *ORF3+ORF4-ORF5+* to dominate the *indica* group.

By contrast, the *japonica* haplotype *ORF3-ORF4+ORF5-* also reached a high frequency after origination within a short period of time. However, using the allele frequency spectrum tests and demographic simulations, not all the tests returned significant values for the three genes, suggesting that selection might not be responsible for the spread of the *japonica* haplotype. Therefore, the preservation of the typical *japonica* genotype (*ORF3-ORF4+ORF5-*) might be due to the founder effect resulting from domestication of *japonica* rice. This inference was highly plausible when considering that *japonica* subspecies had experienced a severe bottleneck during domestication of rice (Zhu et al., 2007; Huang et al., 2012).

In summary (Supplemental Figure 7), the *S5* complex originated by duplication from *Ospara3-5* after the formation of the Oryzaeae tribe, most likely through *Helitron* transposition. A combination of mutational steps generated incompatible *indica* and *japonica* alleles in pre-differentiated rice groups, giving rise to the trigenic reproductive isolation system. Natural selection in *indica* rice and the founder effect associated with domestication in *japonica* populations increased the frequencies of incompatible alleles to form a functional reproductive barrier between the *indica* and *japonica* subspecies, eventually resulting in genetic differentiation and restructuring of rice genetic composition. This also suggests the possibility that domestication and

artificial breeding in crops can lead to the early stages of speciation.

## ACCESSION NUMBERS

KX607476 to KX608521.

## SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

## FUNDING

This research was supported by grants from the National Natural Science Foundation of China (31130032 and 31371599), the Natural Science Foundation of Hubei Province (Grant No. 2014CFB226), and the Fundamental Research Funds for the Central Universities (2662016Q002 and 2662015PY037).

## AUTHOR CONTRIBUTIONS

Y.O. and Q.Z. conceived and designed the experiments; Y.O., G.L., J.M., C.X., and H.D. performed the experiments, including the PCR amplification, DNA sequencing, and sequence assembling; Y.O., G.L., J.M., C.X., and C.Z. analyzed the data; W.X., X.L., J.X., and H.S. contributed reagents/materials/analysis tools; Y.O. and Q.Z. wrote the paper.

## ACKNOWLEDGMENTS

We are indebted to Drs. Manyuan Long, Hanhui Kuang, Li Zhang, Bin He, Xiaoming Zheng, and Mr. Muhua Wang for discussion, and Vladimir Kapitonov for critical comments. We thank Wen Yao for extracting the SNP data from accessions of *O. sativa* on the website. We also thank Jiangxi Academy of Agricultural Sciences, Dajian Pan and Zhilan Fan in the Rice Research Institute of Guangdong Academy of Agricultural Sciences, Gongwei Wang and Sibin Yu of Huazhong Agriculture University, and D.S. Brar of IRRI for providing the rice materials. All the authors declare that they have no conflict of interests.

Received: May 30, 2016

Revised: July 13, 2016

Accepted: July 16, 2016

Published: July 28, 2016

Yidan Ouyang<sup>1,3,\*</sup>, Guangwei Li<sup>1,3</sup>,  
Jiaming Mi<sup>1,3</sup>, Conghao Xu<sup>1</sup>, Hongyi Du<sup>1</sup>,  
Chengjun Zhang<sup>2</sup>, Weibo Xie<sup>1</sup>, Xianghua Li<sup>1</sup>,  
Jinghua Xiao<sup>1</sup>, Huazhi Song<sup>1</sup> and Qifa Zhang<sup>1</sup>

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China

<sup>2</sup>Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, No. 132, Lanhei Road, Kunming, Yunnan 650201, China

<sup>3</sup>These authors contributed equally to this article.

\*Correspondence: Yidan Ouyang ([diana1983941@mail.hzau.edu.cn](mailto:diana1983941@mail.hzau.edu.cn))  
<http://dx.doi.org/10.1016/j.molp.2016.07.008>

## REFERENCES

- Chae, E., Bomblies, K., Kim, S.T., Karelina, D., Zaidem, M., Ossowski, S., Martin-Pizarro, C., Laitinen, R.A., Rowan, B.A., Tenenboim, H., et al. (2014). Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* **159**:1341–1351.
- Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J., et al. (2008). A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proc. Natl. Acad. Sci. USA* **105**: 11436–11441.

- Huang, X., Kurata, N., Wei, X., Wang, Z.X., 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.
- Kapitonov, V.V., and Jurka, J.** (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet.* **23**:521–529.
- Lafon-Placette, C., and Kohler, C.** (2015). Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Curr. Opin. Plant Biol.* **23**:39–44.
- Maheshwari, S., and Barbash, D.A.** (2011). The genetics of hybrid incompatibilities. *Annu. Rev. Genet.* **45**:331–355.
- Ouyang, Y., and Zhang, Q.** (2013). Understanding reproductive isolation based on the rice model. *Annu. Rev. Plant Biol.* **64**:111–135.
- Yang, J., Zhao, X., Cheng, K., Du, H., Ouyang, Y., Chen, J., Qiu, S., Huang, J., Jiang, Y., Jiang, L., et al.** (2012). A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science* **337**:1336–1340.
- Zhu, Q., Zheng, X., Luo, J., Gaut, B.S., and Ge, S.** (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol. Evol.* **24**:875–888.