

Research Article

*These authors contributed equally to this study.

Cite this article: Xiang-Yu J, Gu Z, Li H, Lu B-R (2024). Demography as a confounding factor to explain highly diverged loci between cultivated and wild rice. *Plant Genetic Resources: Characterization and Utilization* **22**, 45–49. <https://doi.org/10.1017/S1479262123001132>

Received: 31 July 2023

Revised: 24 December 2023

Accepted: 27 December 2023

First published online: 5 February 2024

Keywords:

artificial selection; F_{ST} ; neutral demography; rice


Corresponding authors:

Haipeng Li and Bao-Rong Lu;

Email: lihaipeng@sinh.ac.cn; brlu@fudan.edu.cn



Demography as a confounding factor to explain highly diverged loci between cultivated and wild rice

Jinggong Xiang-Yu^{1,2,*}, Zhili Gu^{1,*}, Haipeng Li^{1,2}  and Bao-Rong Lu³

¹CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China; ²College of Artificial Intelligence and Big Data for Medical Sciences, Shandong First Medical University & Shandong Academy of Medical Sciences, Qingdao Road 6699, Jinan 250118, China and ³Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, School of Life Sciences, Fudan University, Songhu Road 2005, Shanghai 200438 China

Abstract

The domestication of rice increases the divergence between cultivated rice and its wild progenitor because of artificial selection. However, it remains unknown whether highly diverged loci in rice can be explained by neutral demographic scenarios alone. In this study, we genotyped 45 InDels (insertion/deletion) in two subspecies of Asian cultivated rice (*Oryza sativa* ssp. *japonica* and *Oryza sativa* ssp. *indica*) and their wild progenitor (*O. rufipogon*/*O. nivara*). Among them, 17 loci are highly diverged ($F_{ST} > 0.4$) between rice cultivars and their ancestor. We performed coalescent-based simulations on neutral demographic scenarios and found that neutral demography alone could explain the polymorphic profiles on those highly diverged loci between cultivated and wild rice. Therefore, more signatures of selection should be considered when detecting artificial selection in rice.

Introduction

Artificial selection is a selection process by which human instead of nature select animals or plants with desirable traits to reproduce. After thousands of years, domesticated animals and plants often have distinct characteristics comparing with their wild progenitors. To study the molecular mechanism of artificial selection, it is important to map genes affected by artificial selection during domestication because these genes are commonly related to economic traits or biological processes associated with development and reproduction (Gepts, 2014). To date, there are different methods and software available to detect artificial selection (Li and Stephan, 2006; Sabeti *et al.*, 2006; Li 2011; Lin *et al.*, 2011; Fu and Akey, 2013; Waldmann *et al.*, 2020). When artificial selection occurs in one of populations, it increases the fixation index (F_{ST}) on the selected locus. F_{ST} is a measurement of population differentiation (Hartl and Clark, 1997), and a locus with high F_{ST} indicates that two populations are highly diverged on the locus. Therefore, high F_{ST} has been frequently used as one of the signatures of selection in rice and many other species (Gao, 2004; Yu *et al.*, 2008; Fabian *et al.*, 2012; Huang *et al.*, 2012; Daub *et al.*, 2013; Shen *et al.*, 2014; Kumagai *et al.*, 2016; Islam *et al.*, 2018; Cheng *et al.*, 2019).

It has been well-known that demographic factors, such as population structure and varying population size, may generate selection-like signatures in the genetic variation of domesticated animals or plants (Catriona and Emma, 2006; Li and Stephan, 2006). Thus, it may hinder the efforts to detect artificial selection. To examine the confounding effects of demography, we investigated whether the polymorphism on highly diverged loci (i.e. high F_{ST} loci) could be accounted by neutral demographic scenarios alone, without invoking the hypothesis of artificial selection.

We used Asian rice to examine the confounding effects of demography, including population structure and varying population size. Asian rice (*Oryza sativa* L.) has two subspecies, *indica* and *japonica*, and it domesticated from its wild rice ancestor *O. rufipogon* Griff. (He *et al.*, 2011; House *et al.*, 2014; Zhang *et al.*, 2016). An entangled history of rice domestication has been revealed (Huang *et al.*, 2012). As common variants in rice (Shen *et al.*, 2004), InDels (insertion/deletion) may be good neutral markers and are easy to be genotyped. Polymorphisms of InDels have gained long-term interests in the rice research community (Shen *et al.*, 2004; Liu *et al.*, 2015a; Lu *et al.*, 2015; Moonsap *et al.*, 2019), and these markers were used to conduct QTL mapping (Kim *et al.*, 2016) and study genetic differentiation in rice (Liu *et al.*, 2012; Sahu *et al.*, 2017). In this study, we selected the highly diverged InDels between the rice cultivars (*japonica* or *indica*) and their wild ancestors, *O. rufipogon* and *O. nivara* (also recognized as the annual type of *O. rufipogon*) (Yamanaka *et al.*, 2003; Zheng and Ge 2010; Liu *et al.*, 2015b) and genotyped these loci in 172 accessions of rice collected through the East, South and Southeast Asia. Then we conducted large-scale coalescent-based simulations on neutral demographic scenarios to examine whether the polymorphic profiles on the highly diverged loci can be explained

by demography alone. This study would shed lights on how genes affected by artificial selection could be detected in Asian cultivated rice and other domesticated animals and plants.

Materials and methods

Sampling and InDels genotyping

In this study, total 172 accessions of *japonica* ($n = 63$), *indica* ($n = 66$), *O. rufipogon* ($n = 25$) and *O. nivara* ($n = 18$) were collected through the East, South and Southeast Asia (Table S1). These samples were provided by the International Rice Research Institute, the Shanghai Center for Agricultural Biological Genetics, Zhejiang University and Yunnan Agricultural University, respectively, and most of their genomes have not been sequenced. The 45 InDels were genotyped in the collected accessions (Table S2). Those InDels were initially discovered by comparing the genomic sequences of *indica* (93–11) and *japonica* rice (Nipponbare) (Shen et al., 2004; Lu et al., 2009). They are randomly distributed along the rice genome.

Genomic DNA extraction and InDels genotyping was conducted according to our previous published methods (Lu et al., 2009). Seedlings of rice samples were germinated in an incubator at 37 °C and then transferred into a glasshouse at around 25 °C. About 0.5 g of fresh leaf samples were collected from seedlings at about the 3–4-leaf stage and then placed in a plastic bag containing silica gel for fast drying. Genomic DNA was extracted using the hot CTAB procedure (Murray and Thompson, 1980). PCR analysis and electrophoresis were performed. The primers for the 45 InDels were obtained by the previous research (Shen et al., 2004) and listed in Supplemental Table S3. PCR products were resolved on a 4% denaturing polyacrylamide gel. After electrophoresis, bands were revealed using the silver-staining procedure. The electrophoretic banding patterns were used for genotyping and subsequent calculation of allele frequency.

Heterozygosity and F_{ST} calculation follows the textbook (Hartl and Clark, 1997). Heterozygosity of wild rice, *japonica* and *indica* at each InDel was calculated from allele frequency. To calculate

F_{ST} , we used the following formula: $F_{ST} = \frac{H_T - H_S}{H_T}$, where H_T is expected heterozygosity for total population and H_S is average expected heterozygosity in subpopulations.

Model selection and simulated neutral demographic scenarios

Four demographic models were considered here (Fig. 1). Two of them represent the single evolutionary origin of rice, and the other two models stand for the multiple origins. Here, the single origin means that *O. sativa* was domesticated once and diverged to *japonica* and *indica* (Sang and Ge, 2007; Gross and Zhao, 2014). The multiple origins indicate that *japonica* and *indica* have independent origins from distinct wild rice subpopulations (Oka, 1988).

By using a software *ms* (Hudson, 2002) we simulated the neutral data under the demographic scenarios. The parameters of the four demographic models were obtained from the previous study (Molina et al., 2011). Thus, to simulate the data, the *ms* command lines are as following:

Model I (single origin, *japonica* domesticated first):

```
ms 344 100000000 -s 1 -I 3 86 126 132 -n 2 0.07 -en 1 2 0.01
-ej 1 2 1 -n 3 0.07 -en 0 3 0.01 -ej 0.07 3 2 -m 1 2 2.01 -m
1 3 5.07 -m 2 3 2.35
```

Model II (single origin, *indica* domesticated first):

```
ms 344 100000000 -s 1 -I 3 86 126 132 -n 2 0.07 -en 0 2 0.01
-ej 0.04 2 3 -n 3 0.09 -en 1 3 0.01 -ej 1 3 1 -m 1 2 1.57 -m
1 3 3.84 -m 2 3 2.16
```

Model III (multiple origins, *japonica* domesticated first):

```
ms 344 100000000 -s 1 -I 3 86 126 132 -n 2 0.08 -en 1 2 0.01
-ej 1.01 2 1 -n 3 0.1 -en 0 3 0.01 -ej 0 3 1 -m 1 2 1.85 -m
1 3 4.02 -m 2 3 1.37
```

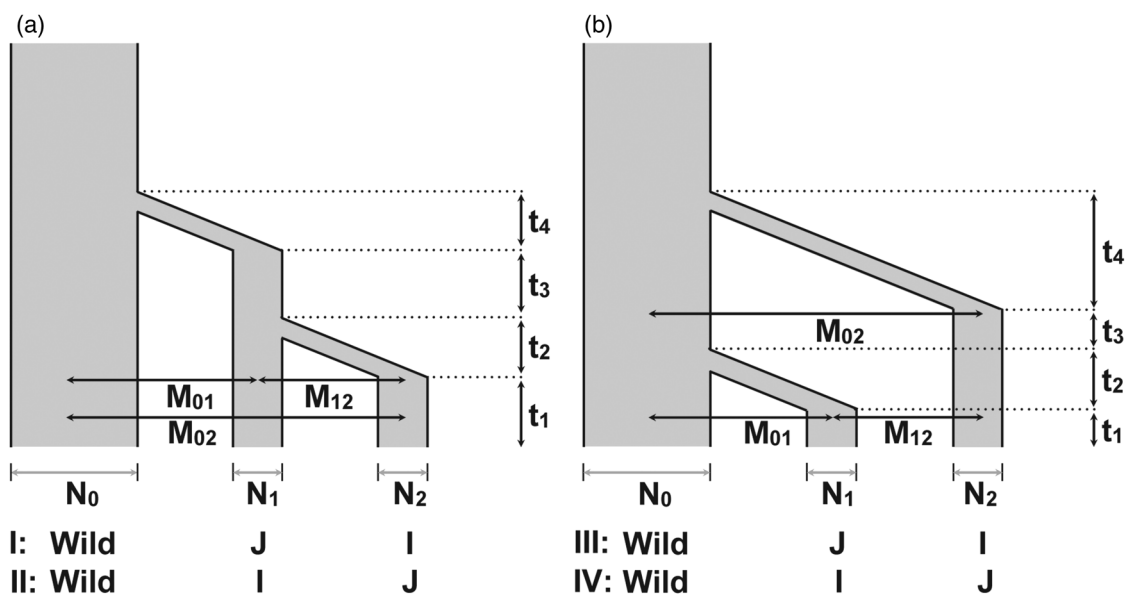


Figure 1. Four demographic scenarios of rice populations considered in this study. (a) The single origin of rice: the model I: *japonica* rice was first domesticated; the model II: *indica* rice was first domesticated. (b) The two origins of rice: the model III: *japonica* rice was first domesticated; the model IV: *indica* rice was first domesticated.

Model IV (multiple origins, *indica* domesticated first):

```
ms 344 100000000 -s 1 -I 3 86 126 132 -n 2 0.1
-en 0 2 0.01 -ej 0.01 2 1 -n 3 0.11 -en 1 3 0.01 -ej
1.01 3 1 -m 1 2 1.41 -m 1 3 3.51 -m 2 3 1.59
```

Neutrality test

The null hypothesis is one of the neutral demographic scenarios described above. The alternative hypothesis is that, a locus evolved neutrally in the progenitor population, but was subject to artificial selection in one of the cultivated rice populations. As expected, artificial selection could cause a high F_{ST} between progenitor-*japonica*, progenitor-*indica* or *japonica-indica*. We pooled simulated results as a sample by rejection-sampling algorithm (Tavare *et al.*, 1997) that has been commonly used to compare simulated and observed data (Li *et al.*, 2016). The neutrality test was conducted as $P(f_j > f_{j,obs} | |f_p - f_{p,obs}| < \epsilon)$ and $P(f_i > f_{i,obs} | |f_p - f_{p,obs}| < \epsilon)$, where f_p , f_j and f_i are the allele frequency of mutant in the progenitor, *japonica* and *indica* population, and ϵ is a fixed tolerance.

When ϵ is very small, the computational load would be very large to estimate the probability, whereas the precision of the probability will be poor when ϵ is large. Our experience suggests that $\epsilon = 0.02$ works well.

Results

In this study, we surveyed the genetic diversity of 129 accessions of Asian cultivated rice (*O. sativa ssp. japonica* and *O. sativa ssp. indica*) and 43 accessions of wild rice (*O. rufipogon* and *O. nivara*). F_{ST} at 45 InDels loci was calculated among wild rice and Asian cultivated rice (Table S2). F_{ST} was widely used to measure the level of population differentiation and detect artificial selection in rice populations. Twelve of the 45 loci between wild rice and *japonica* (Table 1), 5 of the 45 loci between wild rice and *indica* (Table 2), and 35 of the 45 loci between *japonica* and *indica* showed a F_{ST} value greater than 0.4. We defined highly diverged locus between *japonica* (or *indica*) and wild rice with $F_{ST} > 0.4$ as previous study (He *et al.*, 2011). So 17 loci are highly diverged between *japonica* (or *indica*) and wild rice. These loci are R1M7, R1M37, R1M47, R2M10, R2M24, R2M26, R2M50, M3M10, R3M23, R3M30, R4M17, R4M43, R4M50, R6M44, R7M37, R8M23 and R9M42. Because neutral demographic scenarios could affect F_{ST} , we tested whether the polymorphism pattern at these highly diverged loci can be explained by neutral demographic scenarios alone.

The rice domestication process has been studied well (Gross and Zhao, 2014; Choi *et al.*, 2017; Fornasiero *et al.*, 2022; Izawa, 2022; Shang *et al.*, 2022). It was suggested that the first domestication of Asian cultivated rice occurred in the Yangtze River basin at about 9000 years ago (Fornasiero *et al.*, 2022), but the two subspecies (*japonica* and *indica*) originated independently (Choi *et al.*, 2017), indicating that different loci may have different evolutionary histories due to migration and introgression. Therefore, we simulated four different neutral demographic scenarios (the models I, II, III and IV) of the single and the multiple origins (Fig. 1), which represent the possible neutral evolutionary histories of genes in two cultivated rice subspecies.

We then compared the simulated and observed data using the proposed neutrality test (see Materials and Methods). At 17 highly diverged loci in the four models, we only observed

Table 1. Probability value for the test of neutrality in the 12 highly diverged InDel loci between *japonica* and wild rice

Locus	<i>O. sativa ssp. japonica</i>				F_{ST} (<i>japonica</i> – wild rice)
	pI ^a	pII ^b	pIII ^c	pIV ^d	
R1M7	0.129	0.485	0.503	1.000	0.556
R2M10	0.136	0.464	0.490	1.000	0.766
R2M24	0.217	0.481	0.479	0.998	0.763
R2M50	0.161	0.465	0.477	0.954	0.811
R3M10	0.272	0.493	0.476	0.996	0.712
R3M23	0.295	0.496	0.497	0.974	0.479
R3M30	0.101	0.491	0.505	1.000	0.454
R4M43	0.217	0.481	0.479	0.517	0.750
R4M50	0.151	0.463	0.483	0.066	0.793
R6M44	0.042	0.473	0.503	1.000	0.468
R8M23	0.265	0.369	0.292	1.000	0.416
R9M42	0.130	0.464	0.491	1.000	0.755

^aModel I (single origin, *japonica* domesticated first).

^bModel II (single origin, *indica* domesticated first).

^cModel III (multiple origins, *japonica* domesticated first).

^dModel IV (multiple origins, *indica* domesticated first).

Table 2. Probability value for the test of neutrality in the five highly diverged InDel loci between *indica* and wild rice

Locus	<i>O. sativa ssp. indica</i>				F_{ST} (<i>indica</i> – wild rice)
	pI ^a	pII ^b	pIII ^c	pIV ^d	
R1M37	0.860	0.244	0.655	0.312	0.601
R1M47	0.907	0.265	0.664	0.376	0.468
R2M26	0.736	0.244	0.698	0.258	0.572
R4M17	0.786	0.298	0.691	0.288	0.460
R7M37	0.894	0.289	0.646	0.348	0.446

^aModel I (single origin, *japonica* domesticated first).

^bModel II (single origin, *indica* domesticated first).

^cModel III (multiple origins, *japonica* domesticated first).

^dModel IV (multiple origins, *indica* domesticated first).

R6M44 in the model I (single origin, *japonica* being domesticated first) with a marginally significant difference ($p = 0.042$) (Tables 1 and 2). The value becomes insignificantly different when the multiple testing is considered. Other p values are between 0.101 and 1.000. Therefore, the high F_{ST} at 17 loci may have resulted from demography other than selection during *japonica* and *indica* domestication.

Discussion

In this study, we genotyped 45 InDels in the privately collected 172 accessions and identified 17 highly diverged InDels between the rice cultivars and their wild ancestors. Large-scale coalescent-based simulations were conducted on four neutral demographic scenarios to test the neutral evolutionary hypothesis of the 17 InDels. The results of neutrality test suggested that demography alone could explain the polymorphic profiles on the highly diverged InDels, irrespective of which domestication hypothesis

was used and which subspecies was first domesticated. Therefore, F_{ST} is informative for detecting artificial selection in rice, but high F_{ST} should not be used as conclusive evidence that candidate loci are affected by artificial selection.

To examine these 17 highly diverged loci, the new statistical test was developed. It has a number of advantages. First, the test was conducted conditional on a mutation, thus evolutionary rate heterogeneity in the rice genome (Zhao *et al.*, 2018) does not affect our analyses. Second, the test can be used to analyse a single-point mutation because we did not limit the mutation type as InDels (Yu *et al.*, 2022). Third, the test is based on a single locus assuming no recombination occurred within the locus (Yang *et al.*, 2018). Therefore, the test is ready to be applied to other sexual and asexual species.

Though demography alone could explain the polymorphic profiles on the 17 loci with high F_{ST} , our results do not indicate that there is no artificial selection during the domestication of rice. The number of loci examined in this study is small, comparing with those in genome-wide surveys. Our results indicate that it is extremely important to consider the demography as a confounding factor when detecting artificial selection in rice (Gao and Innan, 2008; He *et al.*, 2011; Molina *et al.*, 2011). When genome-wide scans are conducted for searching the signal of artificial selection, the number of windows examined increases and it is very likely to observe a number of loci with high F_{ST} due to random demographic effects, such as demography.

Because the detection of selection may be hindered by the demographic effects, it is further suggested that the combination with different signatures of selection (Lin *et al.*, 2011; Horscroft *et al.*, 2019) could be considered when detecting artificial selection. Moreover, the confounding effects of demography can be addressed by controlling its false-positive rate, which is generally composed by two-step analysis (Li and Stephan, 2006; Koropoulos *et al.*, 2020). Genome-wide DNA polymorphism is used to infer the demography (Speidel *et al.*, 2019; Hu *et al.*, 2023) and the detection of positive selection is followed conditional on the inferred demography. The alternative approaches could be to use neutrality tests that are insensitive to demographic events (Li, 2011; Hunter-Zinck and Clark, 2015). Overall, our results suggest that it should be more cautious about the explanation of highly diverged loci when studying domestication of *indica* and *japonica* rice varieties.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262123001132>.

Author contributions. J. X.-Y., H. L. and B.-R. L. developed the concepts. J. X.-Y. and Z. G. performed the analyses. B.-R. L. provided the resources. J. X.-Y., Z. G. and H. L. wrote the original manuscript. H. L. and B.-R. L. reviewed and edited the manuscript. All authors read and approved the manuscript.

Funding statement. This work was supported by grants from the National Natural Science Foundation of China (nos. 32270674 and 91131010), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDPB17).

Competing interests. None.

Ethics approval. Not applicable.

Consent for publication. Not applicable.

Availability of data and materials. The datasets generated and analysed during the current study are available from the corresponding authors on reasonable request.

References

- Catriona MC and Emma H (2006) Being positive about selection. *Annual Review of Genomics and Human Genetics* **4**, 293–295.
- Cheng L, Kim K-W and Park Y-J (2019) Evidence for selection events during domestication by extensive mitochondrial genome analysis between *japonica* and *indica* in cultivated rice. *Scientific Reports* **9**, 10846.
- Choi JY, Platts AE, Fuller DQ, Hsing YL, Wing RA and Purugganan MD (2017) The rice paradox: multiple origins but single domestication in Asian rice. *Molecular Biology and Evolution* **34**, 969–979.
- Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M and Excoffier L (2013) Evidence for polygenic adaptation to pathogens in the human genome. *Molecular Biology and Evolution* **30**, 1544–1558.
- Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlotterer C and Flatt T (2012) Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology* **21**, 4748–4769.
- Fornasiero A, Wing RA and Ronald P (2022) Rice domestication. *Current Biology* **32**, R20–R24.
- Fu WQ and Akey JM (2013) Selection and adaptation in the human genome. *Annual Review of Genomics and Human Genetics* **14**, 467–489.
- Gao LH (2004) Population structure and conservation genetics of wild rice *Oryza rufipogon* (Poaceae): a region-wide perspective from microsatellite variation. *Molecular Ecology* **13**, 1009–1024.
- Gao LZ and Innan H (2008) Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, demonstrated by multilocus microsatellites. *Genetics* **179**, 965–976.
- Gepts P (2014) The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology* **18**, 51–59.
- Gross BL and Zhao ZJ (2014) Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences of the USA* **111**, 6190–6197.
- Hartl DL and Clark AG (1997) *Principles of Population Genetics*, 3rd Edn. Sunderland: Sinauer Associates.
- He ZW, Zhai WW, Wen HJ, Tang T, Wang Y, Lu XM, Greenberg AJ, Hudson RR, Wu CI and Shi SH (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genetics* **7**, e1002100.
- Horscroft C, Ennis S, Pengelly RJ, Sluckin TJ and Collins A (2019) Sequencing era methods for identifying signatures of selection in the genome. *Briefings in Bioinformatics* **20**, 1997–2008.
- House MA, Griswold CK and Lukens LN (2014) Evidence for selection on gene expression in cultivated rice (*Oryza sativa*). *Molecular Biology and Evolution* **31**, 1514–1525.
- Hu W, Hao Z, Du P, Di Vincenzo F, Manzi G, Cui J, Fu Y-X, Pan Y-H and Li H (2023) Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science* **381**, 979–984.
- Huang XH, Kurata N, Wei XH, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu KY, Lu HY, Li WJ, Guo YL, Lu YQ, Zhou CC, Fan DL, Weng QJ, Zhu CR, Huang T, Zhang L, Wang YC, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan XP, Xu Q, Dong GJ, Zhan QL, Li CY, Fujiyama A, Toyoda A, Lu TT, Feng Q, Qian Q, Li JY and Han B (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501.
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338.
- Hunter-Zinck H and Clark AG (2015) Aberrant time to most recent common ancestor as a signature of natural selection. *Molecular Biology and Evolution* **32**, 2784–2797.
- Islam MZ, Khalequzzaman M, Prince MFRK, Siddique MA, Rashid ESMH, Ahmed MSU, Pittendrigh BR and Ali MP (2018) Diversity and population structure of red rice germplasm in Bangladesh. *PLoS ONE* **13**, e0196096.
- Izawa T (2022) Reloading DNA history in rice domestication. *Plant and Cell Physiology* **63**, 1529–1539.
- Kim SR, Ramos J, Ashikari M, Virk PS, Torres EA, Nissila E, Hechanova SL, Mauleon R and Jena KK (2016) Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-

- genome sequencing strategy to increase yield potential of rice, *Oryza sativa* L. *Rice* **9**, 12.
- Koropoulis A, Alachiotis N and Pavlidis P** (2020) Detecting positive selection in populations using genetic data. *Methods in Molecular Biology* **2090**, 87–123.
- Kumagai M, Kanehara M, Shoda S, Fujita S, Onuki S, Ueda S and Wang L** (2016) Rice varieties in archaic east Asia: reduction of its diversity from past to present times. *Molecular Biology and Evolution* **33**, 2496–2505.
- Li H** (2011) A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Molecular Biology and Evolution* **28**, 365–375.
- Li H and Stephan W** (2006) Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genetics* **2**, 1580–1589.
- Li H, Xiang-Yu J, Dai G, Gu Z, Ming C, Yang Z, Ryder OA, Li W, Fu Y-X and Zhang Y-P** (2016) Large numbers of vertebrates began rapid population decline in the late 19th century. *Proceedings of the National Academy of Sciences of the USA* **113**, 14079–14084.
- Lin K, Li H, Schlotterer C and Futschik A** (2011) Distinguishing positive selection from neutral evolution: boosting the performance of summary statistics. *Genetics* **187**, 229–244.
- Liu P, Cai XX and Lu BR** (2012) Single-seeded InDel fingerprints in rice: an effective tool for *indica-japonica* rice classification and evolutionary studies. *Journal of Systematics and Evolution* **50**, 1–11.
- Liu J, Li JW, Qu JT and Yan SY** (2015a) Development of genome-wide insertion and deletion polymorphism markers from next-generation sequencing data in rice. *Rice* **8**, 27.
- Liu R, Zheng XM, Zhou L, Zhou HF and Ge S** (2015b) Population genetic structure of *Oryza rufipogon* and *Oryza nivara*: implications for the origin of *O. nivara*. *Molecular Ecology* **24**, 5211–5228.
- Lu BR, Cai XX and Jin X** (2009) Efficient *indica* and *japonica* rice identification based on the InDel molecular method: its implication in rice breeding and evolutionary research. *Progress in Natural Science* **19**, 1241–1252.
- Lu Y, Cui X, Li R, Huang PP, Zong J, Yao DQ, Li G, Zhang DB and Yuan Z** (2015) Development of genome-wide insertion/deletion markers in rice based on graphic pipeline platform. *Journal of Integrative Plant Biology* **57**, 980–991.
- Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR and Purugganan MD** (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of the USA* **108**, 8351–8356.
- Moonsap P, Laksanavilat N, Sinumporn S, Tasanasuwan P, Kate-Ngam S and Jantasuriyarat C** (2019) Genetic diversity of Indo-China rice varieties using ISSR, SRAP and InDel markers. *Journal of Genetics* **98**, 80.
- Murray MG and Thompson WF** (1980) Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Research* **8**, 4321–4325.
- Oka HI** (1988) *Origin of Cultivated Rice*, 1st Edn. Amsterdam: Elsevier.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varrilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D and Lander ES** (2006) Positive natural selection in the human lineage. *Science* **312**, 1614–1620.
- Sahu PK, Mondal S, Sharma D, Vishwakarma G, Kumar V and Das BK** (2017) InDel marker based genetic differentiation and genetic diversity in traditional rice (*Oryza sativa* L.) landraces of Chhattisgarh, India. *PLoS ONE* **12**, e0188864.
- Sang T and Ge S** (2007) Genetics and phylogenetics of rice domestication. *Current Opinion in Genetics & Development* **17**, 533–538.
- Shang L, Li X, He H, Yuan Q, Song Y, Wei Z, Lin H, Hu M, Zhao F, Zhang C, Li Y, Gao H, Wang T, Liu X, Zhang H, Zhang Y, Cao S, Yu X, Zhang B, Zhang Y, Tan Y, Qin M, Ai C, Yang Y, Zhang B, Hu Z, Wang H, Lv Y, Wang Y, Ma J, Wang Q, Lu H, Wu Z, Liu S, Sun Z, Zhang H, Guo L, Li Z, Zhou Y, Li J, Zhu Z, Xiong G, Ruan J and Qian Q** (2022) A super pan-genomic landscape of rice. *Cell Research* **32**, 878–896.
- Shen YJ, Jiang H, Jin JP, Zhang ZB, Xi B, He YY, Wang G, Wang C, Qian LL, Li X, Yu QB, Liu HJ, Chen DH, Gao JH, Huang H, Shi TL and Yang ZN** (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiology* **135**, 1198–1205.
- Shen DF, Bo WH, Xu F and Wu RL** (2014) Genetic diversity and population structure of the Tibetan poplar (*Populus szzechuanica* var. *tibetica*) along an altitude gradient. *BMC Genetics* **15**, S11.
- Speidel L, Forest M, Shi S and Myers SR** (2019) A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics* **51**, 1321–1329.
- Tavare S, Balding DJ, Griffiths RC and Donnelly P** (1997) Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505–518.
- Waldmann P, Pfeiffer C and Meszaros G** (2020) Sparse convolutional neural networks for genome-wide prediction. *Frontiers in Genetics* **11**, 25.
- Yamanaka S, Nakamura I, Nakai H and Sato YI** (2003) Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. *Genetic Resources and Crop Evolution* **50**, 529–538.
- Yang Z, Li J, Wiehe T and Li H** (2018) Detecting recent positive selection with a single locus test bipartitioning the coalescent tree. *Genetics* **208**, 791–805.
- Yu YC, Tang T, Qian Q, Wang YH, Yan MX, Zeng DL, Han B, Wu CI, Shi SH and Li JY** (2008) Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell* **20**, 2946–2959.
- Yu D, Yang X, Tang B, Pan Y-H, Yang J, Duan G, Zhu J, Hao Z-Q, Mu H, Dai L, Hu W, Zhang M, Cui Y, Jin T, Li C-P, Ma L, Su X, Zhang G, Zhao W and Li H and Language Translation T** (2022) Coronavirus GenBrowser for monitoring the transmission and evolution of SARS-CoV-2. *Briefings in Bioinformatics* **23**, 1–10.
- Zhang FT, Xu T, Mao LY, Yan SY, Chen XW, Wu ZF, Chen R, Luo XD, Xie JK and Gao S** (2016) Genome-wide analysis of Dongxiang wild rice (*Oryza rufipogon* Griff.) to investigate lost/acquired genes during rice domestication. *BMC Plant Biology* **16**, 103.
- Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Huang T, Wang Y, Fan D, Zhao Y, Wang Z, Zhou C, Chen J, Zhu C, Li W, Weng Q, Xu Q, Wang Z-X, Wei X, Han B and Huang X** (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics* **50**, 278–284.
- Zheng X-M and Ge S** (2010) Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Molecular Ecology* **19**, 2439–2454.