Identification of Selection Signatures in Commercial Asian Rice Subspecies Deciphered Selection Footprints for Four Overrepresented Biological Processes

Siavash Salek Ardestani

Dalhousie University Department Animal Science and Aquaculture

Mahmoud Amiri Roudbar (✉ mahmood.amiri225@gmail.com)

Safiabad-Dezful Agricultural and Natural Resources Research and Education Center

Mohammad Hossein Banabazi

ASRI: Animal Science Research Institute

Seyedeh Fatemeh Mousavi

University of Kurdistan

Madhav Bhatta

University of Wisconsin-Madison

Mehdi Momen

University of Wisconsin-Madison

Research article

Keywords: Asian commercial rice, genome, *Oryza sativa*, signature of selection

Posted Date: December 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-121958/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Selective breeding pressures have led to gradual genomic changes in Asian commercial rice, which have shaped selection footprints on its genome level. Tracing genomic selection footprints might be illuminative for better understanding of recent selection breeding objectives, and how breeding strategies have formed the Asian commercial rice genome.

Results

In this study, the genotypic information (HDRA 700K) of four Asian commercial rice subspecies including *Indica* (n=498), *Aus* (n=187), *Temperate japonica* (n=241), and *Tropical japonica* (n=361) were downloaded from Rice Diversity Project database (<http://www.ricediversity.org>) to detect selection signatures by employing the Z-transformed of fixation index and Tajima's D test, based on a sliding window approach. Although we could not identify overrepresented genomic regions underlying selection pressure among all aforementioned Asian commercial rice subspecies, interestingly, our findings revealed four overrepresented biological processes underlying selection pressure including proteolysis (GO:0006508), phosphorylation (GO:0016310), protein catabolic process (GO:0030163), and transmembrane transport (GO:0055085) that might be associated with immunity, senescing leaves, transporting, and absorption of ions.

Conclusions

These results can provide knowledge on how breeding efforts shaped the Asian commercial rice subspecies genome, and which genomic regions of these subspecies have been targeted in recent decades.

Introduction

Asian rice (*Oryza sativa*.) was domesticated in China sometime between 7,500-12,500 BP [1] and have been adapted to different environmental conditions [2] and human needs [3]. Natural selection and intensive selective breeding programs have played a key role in genetically, morphologically, and physiologically evolving of various commercial Asian rice subspecies, such as *Indica*, *Aus*, *Temperate japonica*, and *Tropical japonica* [3, 4]. These selection pressures can be associated with selection footprints on the genome level which is known as selection signatures [5, 6]. Tracing selection footprints in the genome can be helpful for better understanding the evolutionary trend [7, 8], genetic background of recent selective breeding [3, 9], and breeding programs, and for optimizing genomic evaluation [10]. Recent swift technical advances of cost-effective genotyping have provided an opportunity to reveal the genomic selection signatures in plant [3, 11] and animal species [12-16]. Genomic signatures of selection studies on rice have been illuminative in identifying genomic regions associated with productivity [3], adaptation [2, 17], and domestication [8]. Xie et al. (2015) detected six

candidate genes involving in nitrogen assimilation pathway by employing a cross population signatures of selection approach and a genome-wide association study. He et al. (2017) revealed 107 genes as selective signals associated with biotic stress pathway using whole-genome sequences of Korean weedy rice.

Although a number of literatures are available on reported important selection signatures in rice [2, 3, 8], our knowledge of the genomic changes associated with domestication remains limited. For instance, there is a strong controversy about division of the domestication process into single or independent origins [8]. Evaluating the genetic structure of *O. sativa* is very important to provide more evidence for the important selected regions and use this knowledge into the breeding of rice for sustainable agriculture.

In this study, our main goals were: 1) investigation into the genetic structure of *Indica*, *Aus*, *Temperatejaponica*, and *Tropical japonica* populations, 2) identifying and investigating selection signals in these populations. The results of this study can be useful in revealing novel candidate genes underlying recent selection pressures of the populations. Additionally, it may be illuminative for better understanding of the evolutionary trend in Asian rice.

Materials And Methods

Plant materials and SNP array data

The used data in this study were retrieved from three large density panels including RDP1 (representing 343 accessions), RDP2 (representing 880 accessions) and the National Institute of Agrobiological Sciences (NIAS) (representing 61 accessions) from many landraces which originated from 108 different countries, more details about the data were presented previously [18]. The germplasm collections used in this study are summarized in Supplementary Data 1. Briefly, the subpopulation distribution of lines in our studied population consisted of 14.6% (*Aus*), 38.7% (*Indica*), 18.7% (*Temperate-japonica*), and 28.0% (*Tropical-japonica*). All data are publicly available in the Rice Diversity Project (<http://www.ricediversity.org>). Each germplasm was genotyped by using the high-density rice array (HDRA) comprising of 700,000 single-nucleotide polymorphisms (SNPs). The HDRA represents a very high-density genotyping array with one SNP every 0.54 kb across the rice genome [18].

Quality control

Quality control was performed using the PLINK 1.9 software [19]. First, SNPs with call rate less than 80% were removed from each subpopulation, separately, and 598739, 619563, 661469 and 658713 number of SNPs remained for *Aus*, *Indica*, *Temperate japonica*, and *Tropical japonica* populations were 598739, 619563, 661469 and 658713, respectively. A total number of 513,703 SNPs remained in the merged rice subpopulations. After applying minor allelic frequency < 0.05, 286,183 SNPs remained for the subsequent analyses. Individuals with the sample call rate more than 80% were used in the final data. Finally, the missing genotypes were imputed using the FImpute 2.2 [20].

Population genetic structure

To investigate the genetic diversity pattern in the studied populations, nucleotide diversities (π) [21] of each 10 Kb per chromosome were calculated and visualized in VCFtools 0.1.15 package (<http://vcftools.sourceforge.net/index.html>) [22] and Circos 0.69.6 software [23], respectively. In this study, three methods including Neighbor-joining, Bayesian model-based approach exemplified by ADMIXTURE at $K=2$, $K=3$ and $K=4$, and Principal component analysis (PCA), were used to assess the population genetic structure of the studied populations. The neighbor-joining analysis was performed by applying VCF-kit 0.1.6 (<https://vcf-kit.readthedocs.io/en/latest>) [24]. In order to visualize the results of neighbor-joining analysis, FigTree 1.4.3 software was employed (<http://tree.bio.ed.ac.uk/software/figtree>). The admixture analysis and PCA were performed using ADMIXTURE 1.3 (<http://software.genetics.ucla.edu/admixture>) and SNP1101 1.0 [25], respectively. Additionally, a linkage disequilibrium estimation analysis across all chromosomes was performed and visualized for each population using PopLDdecay 1.01 software [26] and a perl script (<https://github.com/BGI-shenzhen/PopLDdecay>), respectively.

Selection signals and gene ontology

In this study, two methods including Z-transformed fixation index ($Z(F_{st})$) [27] and Tajima's D test based on a sliding window approach (100 kb with a step size of 25 kb) were implemented using VCFtools 0.1.15 [22] and VCF-kit 0.1.6 [24], respectively. The Z-transformation of F_{st} values was calculated using "scale" function in R software (<https://cran.r-project.org>). The windows in the top 1% of $Z(F_{st})$ with a negative Tajima's D value in populations were considered as selection signals in each scenario. The Gene ontology analyses were conducted using Gene Ontology Consortium (<http://geneontology.org>). The overlapped windows underlying selection pressures and overlapped biological processes among selection signatures scenarios visualized via Venn Painter 1.2.0 [28].

Results

Population genetic structure and nucleotide diversity

The PCA, admixture, and neighbor-joining were performed to visualize a comprehensive investigation into the genetic separation of four commercial Asian rice subspecies. More than 37% of the genetic variation was explained by the top three PCs (Fig. 1). PC1 separates the *Indica/Aus* and two *japonica* varietal groups. The *Indica* and *Aus* clustered in two separate clusters at PC2. Finally, the PC3 classified the *Temperatejaponica* and *Tropical japonica* subpopulations into two separate groups. Similar results were achieved by admixture analysis (Fig. 2). The admixture analysis at $K=2$ and $K=3$ showed both *Temperatejaponica* and *Tropical japonica* in one cluster; and also, at $K=2$, it showed *Indica* and *Aus* together in one group each of which was grouped into a cluster by admixture at $K=4$. Similar to PC3, the admixture at $K=4$ separated *Temperatejaponica* and *Tropical japonica* subpopulations. Same interpretation concluded from the neighbor-joining analysis (Fig. 3).

Study of linkage disequilibrium patterns (mean of r^2) in populations can be informative for having an overview of population demography background, due to the effective role of linkage disequilibrium patterns in demographic forces and evolutionary patterns [29]. Our linkage disequilibrium results illustrated a rapid drop at approximately 5 Kb in all groups (Fig. 4). The means of r^2 at 300 Kb for *Indica*, *Aus*, *Temperate japonica*, and *Tropical japonica* groups were 0.04, 0.05, 0.10 and 0.07, respectively.

In this study, values (in 10Kb windows) were calculated to evaluate genomic diversities in studied subpopulations (Fig. 5). The values were ranged from 5.35E-7 to 0.0015 across all subpopulations. The mean of values demonstrated higher nucleotide diversities in *Aus* (0.16E-3) and *Indica* (0.17E-3) in comparison with *Temperate japonica* (7.97E-5) and *Tropical japonica* (1.09E-4) groups.

Selective signals and gene ontology

In this study, based on windows located in the top 1% of Z (Fst) values with Tajima's D values < 0, six scenarios were considered to detect genomic regions underlying selection pressures including: *Tropical japonica* vs. *Temperate japonica*, *Temperate japonica* vs. *Indica*, *Tropical japonica* vs. *Indica*, *Indica* vs. *Aus*, *Tropical japonica* vs. *Aus* and *Aus* vs. *Temperate japonica*.

Tropical japonica vs. *Temperate japonica*

A total number of 66 windows were identified in the top 1% of Z (Fst) values (Fig. 6) with Tajima's D values < 0, in the *Tropical japonica* vs. *Temperate japonica* scenario (Supplementary data 2). These windows contain 145 genes related to 40 biological processes (Supplementary data 2 and Supplementary information, Table S1); however, these biological processes were not significantly enriched. Additionally, we detected three important candidate genes including: *Os05g0405000*, *Os05g0402700* and *Os03g0318500* related to different economical traits, such as temperature response [30], response to stresses [31], starch metabolism, structure [32], root growth, and development. The *Glucose-6-phosphate1-dehydrogenase* gene is the ortholog of *Os03g0318500*, the important role of this gene was shown in response to stresses, such as metal toxicity [33], osmotic stress [34] and pathogenesis [35]. The ortholog of *Os05g0405000* is *PPDK1*, which in rice endosperm can play a key role in starch metabolism and structure by regulating starch synthesis-related enzymes expression [32]. The temperature tolerance has been one of the important traits in rice breeding strategies [36, 37]. The *Os05g0405000* gene is also related to temperature response. A gene expression study revealed that high temperature can decrease the *PPDK* enzyme activity [30]. The ortholog of *Os05g0402700* gene is *Fructose-bisphosphate aldolase 1*; Konishi et al (2005) demonstrated the pivotal role of *Fructose-bisphosphate aldolase* enzyme activity in root growth of rice [38].

Temperate japonica vs. *Indica*

In *Temperate japonica* vs. *Indica* scenario, a total number of 104 windows including 274 genes related to 87 biological processes were screened in the top 1% of Z(Fst) values (Fig. 6) with Tajima's D values < 0 (Supplementary data 3 and Supplementary information, Table S2). The detected biological processes

were not significant similar to *Tropical japonica* vs. *Temperate japonica* scenario. Additionally, there were not overlapped genomic regions underlying selection pressures among *Tropical japonica* vs. *Temperate japonica* and *Temperate japonica* vs. *Indica* scenarios (Fig. 7). In this scenario, we screened some candidate genes such as, *Os01g0894000* (*Calcineurin-related phosphoesterase-like*), and *Os10g0430200* (*CAD1*) related to drought tolerance [39], and lignification development [40] in rice, respectively. A former study revealed that lignification development can be associated with intensification of drought tolerance in rice [41]. The drought tolerance, as a complex trait, plays a key role in saving water and sustainable agriculture development, which have been the objectives of *Indica* breeding strategies [42], as they are broadly cultivated in tropical regions with warm climate [43].

Tropical japonica* vs. *Indica

A genome scan for signatures of selection between *Tropical japonica* and *Indica* subpopulations showed a total of 87 windows including 213 genes related to 64 biological processes in the top 1% of Z(Fst) values (Fig. 6) with a negative Tajima's D value (Supplementary data 4 and Supplementary information, Table S3). This scenario had 61 and one overlapped windows with *Temperate japonica* vs. *Indica* and *Tropical japonica* vs. *Temperate japonica* scenarios, respectively (Fig. 7). The *Os08g0559400* (*OsCyp3*), *Os02g0462900*, *Os02g0332200* (*T-complex protein 1 subunit delta*), *Os02g0782500* (*18.6 kDa class III heat shock protein*), and *Os05g0533900* genes involve in protein folding (GO:0006457) as a significant biological process ($q\text{-value} = 0.027$) in *Tropical japonica* vs. *Indica* scenario. This biological process is related to heat shock proteins [44]. The heat shock proteins have a fundamental role in response to abiotic stressful conditions through helping to refold proteins which were damaged by heat stresses [44, 45]. The higher thermo- tolerability of *Indica* compared to the japonica subspecies was revealed in a comparative former study [46].

Indica* vs. *Aus

The number of signatures of selection windows in *Indica* vs. *Aus* scenario was 17, which included 55 genes related to 27 biological processes in the top 1% of Z(Fst) values (Fig. 6) with a negative Tajima's D value (Supplementary data 5 and Supplementary information, Table S4). Among 17 detected windows underlying selection pressures, two overlapped windows (chromosome 4: 34.47- 34.57 Mb and chromosome 4: 34.50- 34.60 Mb) were observed in both *Indica* vs. *Aus* and *Temperate japonica* vs. *Indica* scenarios (Fig. 7). These genomic regions include seven genes (*Os04g0676800*, *Os04g0675700*, *Os04g0676000*, *Os04g0676700*, *Os04g0676800*, *Os04g0677100*, and *Os04g0677200*). There is a lack of information about biological processes related to these genes except for *Os04g0676700*, and *Os04g0677100*. The *Os04g0676700* gene participates in response to salicylic acid (GO:0009751). Salicylic acid is a well-known molecule which involves in protection under biotic and abiotic stress responses [47, 48]. The *Os04g0677100* involves in proteolysis (GO:0006508) and protein catabolic process (GO:0030163). These GOs can play important roles, such as senescing leaves [49] and releasing trans-membrane proteins [50].

Tropical japonica* vs. *Aus

A total number of 91 windows containing 254 genes were detected in the top 1% of Z(Fst) values (Fig. 6) with Tajima's D values < 0 (Supplementary data 6). Our GO analyses showed that these genes are related to 77 biological processes (Supplementary information, Table S5). Among genes underlying selection pressures in this scenario, we detected two candidate genes including *Os01g0266600* (*Peroxisredoxin-2F* or *PRXIIF*), *Os03g0710500* (*Heat shock 70 kDa protein BiP2* or *BiP2*) which are related to regulating root growth [51], folding and secretion of pathogenesis-related (PR) proteins [52] and stress response [53]. In the presence of salicylhydroxamic acid and cadmium, the *PRXIIF* gene can contribute to regulating root growth of rice [51]. In *Arabidopsis*, Wang et al (2005) showed the participation of *BiP2* gene in folding and secretion of PR proteins during systemic acquired resistance; the PR proteins are classified as a conserved protein family which plays multiple roles in immune responses [54], plant development, and biotic and abiotic stresses [55]. Here, we also detected 6 genes related to phosphorylation (GO:0016310) including *Os03g0755000*, *Os03g0756200*, *Os03g0686800*, *Os02g0241600*, *Os02g0769700*, and *Os05g0521300*. Although this GO term was not significant in our gene ontology analysis, former studies revealed the effective role of phosphorylation in immunity of plants [56, 57]. For example, Matsui et al (2015) indicated that the phosphorylation of pto-interacting protein 1a can be important for activating rice immunity.

Aus* vs. *Temperate japonica

In *Aus* vs. *Temperate japonica* scenario, we screened a total number of 80 windows including 234 genes in the top 1% of Z(Fst) values (Fig. 6) with Tajima's D values < 0 (Supplementary data 7 and Supplementary information, Table S6). These genes were related to 61 biological functions. However, the biological functions in this scenario were not significant in our gene ontology analysis. Additionally, our findings showed three genes including *Os10g0430200*, *Os08g0463900*, *Os03g0810800* that are related to lignin biosynthetic process (GO:0009809), thylakoid membrane organization (GO:0010027) and response to water deprivation (GO:0009414), respectively. Lignin as a natural phenolic polymer is an important component of the cell wall structure in plants. The biosynthesis of lignin can widely participate in different traits such as responses to biotic and abiotic stresses through lignin accumulation which results in the accumulation of reactive oxygen species in plants, under such stresses [58]. The thylakoid membrane has an effective role in photosynthesis and electron transportation [59]. The chloroplast thylakoid membranes are dispersed in the stroma, where photochemical reactions take place in green algae and higher plants. Additionally, rice thylakoid membrane complexes can play a pivotal role in growth and crop production [60].

Overrepresented genes and GOs among the selection signature scenarios

In this study, we could not identify any overrepresented genomic region among all adopted selection signature scenarios. Conversely, gene ontology analyses screened four biological processes including proteolysis (GO:0006508), phosphorylation (GO:0016310), protein catabolic process (GO:0030163), and transmembrane transport (GO:0055085) which are underlying selection pressures (Fig. 7). Although, several involved candidate genes were intriguingly different from aforementioned biological processes

among scenarios (Supplementary Table S2), which can be due to the different breeding programs or environmental conditions.

Discussion

Investigation of the population genetic structure can provide an opportunity to determine the effect of evolutionary processes, such as selection [3, 61, 62], on populations. Former studies suggested that employing different population genetic structure analysis methods and comparing their results can be informative for having a better overview of genetic distances [9, 17, 61]. These results were in concordance with those of former studies [3, 4, 18]. A visual appraisal of the three approaches indicated that the genetic distances of the samples within each *Indica* and *Aus* clusters were more than *Temperate* and *Tropicaljaponica* subpopulations.

The pattern of nucleotide diversities may explain the results of a high genetic distance within *Aus* and *Indica* groups, as well as a low genetic distance within *Temperatejaponica* and *Tropicaljaponica*, which were obtained from our PCA, admixture and neighbor-joining assessments. The most diverse window was located on chromosome 9: 6.33-6.33 Mb of *Indica*, *Aus*, and across all population groups. This region includes a protein coding gene (Os09g0285900). In *Temperate* and *Tropical japonica* groups the most diverse windows were located on chromosome 4:1.36-1.37 Mb and chromosome 11:28.00-28.01 Mb, respectively. The chromosome 11:28.00-28.01 Mb contains OsRH40, Os11g0689500, and Os11g0689650 genes. The chromosome 4:1.36-1.37 Mb region includes Os04g0122200, Os04g0122300, and two non-protein coding genes (Os04g0122750 and Os04g0122750). The serine/threonine kinase activity, and ATP binding proteins encoded by the Os04g0122200 gene; the serine/threonine kinase plays a pivotal role in resistance to rice stripe disease [63]. It can support that the recombination rate and nucleotide diversity in genomic regions related to disease resistance can potentially be higher than other genomic regions, such as MHC and CNL gene groups in mammals [64] and barley [65], respectively.

The previous studies on genomic signatures of selection using different tools revealed a large number of genomic regions underlying selection pressures which are related to productivity [3], domestication [2, 17, 66] and adaptation [8] in various rice subspecies during recent decades. Therefore, identifying the genomic selection signatures in Asian commercial subspecies can reveal novel candidate genomic regions related to economic traits. Furthermore, it can be helpful to understand the recent selection patterns in these subspecies. We found the four overrepresented biological processes, including GO:0006508, GO:0016310, GO:0030163 and GO:0055085 which can contribute in some important traits such as immunity (phosphorylation) [56], senescing leaves (proteolysis and protein catabolic) [49], and root membrane transporting (transmembrane transport) which can be highly important for absorption, and transporting of ions in rice [67]. Additionally, there were overlapped biological processes underlying selection pressures among several scenarios which can have effective roles in economical traits of Asian rice subspecies; For example, developmental process was identified as an overrepresented biological process among three scenarios (*Tropical japonica* vs. *Temperate japonica*, *Tropical japonica* vs. *Indica*, and *Tropical japonica* vs. *Indica*) which can have a fundamental role in growth and development of rice

seeds [68]. Moreover, selection for these biological processes can be reasonable considering the effective role of phosphorylation, proteolysis and protein catabolic process in immunity [56], senescing leaves [49] and releasing of transmembrane proteins [50], respectively.

Conclusion

In this study, we detected genomic regions underlying selection pressures in four Asian commercial rice subspecies genome. By employing two selection signatures methods, we could not identify any overrepresented genomic region among all adopted selection signatures scenarios; however, the gene ontology analyses indicated the four overrepresented biological processes underlying selection pressures among these scenarios which can participate in important traits such as immunity (phosphorylation), senescing leaves (proteolysis and protein catabolic), and root membrane transporting (transmembrane transport). These results can be helpful for better understanding of the genetic background of recent selective breeding efforts. Additionally, identified genes as selection signatures can be considered for optimizing future SNP panels in rice and rice breeding.

Abbreviations

NIAS: National Institute of Agrobiological Sciences

HDRA: High-density rice array

SNP: Single-nucleotide polymorphism

PCA: Principal component analysis

Z(Fst): Z-transformed fixation index

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Author Contribution Statement:

MAR and SSA conceived and planned the experiments. SSA carried out the experiments with support from MAR, MHB and SFM. MAR and SFM wrote the manuscript. MB and MM contributed to the experimental design of the study and participated in the manuscript writing. All authors read and approved the final draft of the manuscript.

References

1. Zheng Y, Crawford GW, Jiang L, Chen X: Rice domestication revealed by reduced shattering of archaeological rice from the Lower Yangtze valley. *Scientific Reports* 2016, 6:28136.
2. Gilani SA, Kawase M, Ohsawa R, Watanabe KN: Tracking selection signatures based on variation in OsLEA27 within Myanmar landraces of upland and dryland rice. *American Journal of Plant Sciences* 2015, 6(12):1937.
3. Xie W, Wang G, Yuan M, Yao W, Lyu K, Zhao H, Yang M, Li P, Zhang X, Yuan J: Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proceedings of the National Academy of Sciences* 2015, 112(39):E5411-E5419.
4. Wang C, Zheng X, Xu Q, Yuan X, Huang L, Zhou H, Wei X, Ge S: Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity* 2014, 112(5):489-496.
5. Nielsen R: Molecular signatures of natural selection. *Annu Rev Genet* 2005, 39:197-218.
6. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD: Interrogating a high-density SNP map for signatures of natural selection. *Genome research* 2002, 12(12):1805-1814.
7. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg Ö, Jern P, Jørgensen CB: Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences* 2012, 109(48):19529-19536.
8. Cheng L, Nam J, Chu S-H, Rungrapa P, Min M-h, Cao Y, Yoo J-m, Kang J-S, Kim K-W, Park Y-J: Signatures of differential selection in chloroplast genome between japonica and indica. *Rice* 2019, 12(1):65.
9. Salek Ardestani S, Aminafshar M, Zandi Baghche Maryam MB, Banabazi MH, Sargolzaei M, Miar Y: Whole-Genome Signatures of Selection in Sport Horses Revealed Selection Footprints Related to Musculoskeletal System Development Processes. *Animals* 2020, 10(1):53.
10. Chang L-Y, Toghiani S, Hay EH, Aggrey SE, Rekaya R: A weighted genomic relationship matrix based on Fixation Index (Fst) prioritized SNPs for genomic selection. *Genes* 2019, 10(11):922.
11. Marrano A, Micheletti D, Lorenzi S, Neale D, Grando MS: Genomic signatures of different adaptations to environmental stimuli between wild and cultivated *Vitis vinifera* L. *Horticulture research* 2018, 5(1):1-12.
12. Salek Ardestani S, Aminafshar M, Zandi Baghche Maryam MB, Banabazi MH, Sargolzaei M, Miar Y: Signatures of selection analysis using whole-genome sequence data revealed novel candidate genes for pony and light horse types. *Genome* 2020.
13. Ghoreishifar SM, Eriksson S, Johansson AM, Khansefid M, Moghaddaszadeh-Ahrabi S, Parna N, Davoudi P, Javanmard A: Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. *Genetics Selection Evolution* 2020, 52(1):1-15.
14. Ablondi M, Viklund Å, Lindgren G, Eriksson S, Mikko S: Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC genomics* 2019, 20(1):717.
15. Salek Ardestani S, Aminafshar M, Zandi Baghche Maryam MB, Banabazi MH, Sargolzaei M, Miar Y: A genome-wide signatures of selection study of Welsh ponies and draft horses revealed five genes associated with horse type variation. *Gene Reports* 2020:100833.

16. Ghoreishifar SM, Moradi-Shahrbabak H, Fallahi MH, Jalil Sarghale A, Moradi-Shahrbabak M, Abdollahi-Arpanahi R, Khansefid M: Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, *Bubalus bubalis*. *BMC Genetics* 2020, 21(1):16.
17. He Q, Kim KW, Park YJ: Population genomics identifies the origin and signatures of selection of Korean weedy rice. *Plant biotechnology journal* 2017, 15(3):357-366.
18. McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P *et al*: Open access resources for genome-wide association mapping in rice. *Nature Communications* 2016, 7:10532.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ: PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 2007, 81(3):559-575.
20. Sargolzaei M, Chesnais JP, Schenkel FS: A new approach for efficient genotype imputation using information from relatives. *BMC genomics* 2014, 15(1):478.
21. Nei M, Li W-H: Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 1979, 76(10):5269-5273.
22. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST: The variant call format and VCFtools. *Bioinformatics* 2011, 27(15):2156-2158.
23. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: Circos: an information aesthetic for comparative genomics. *Genome research* 2009.
24. Cook DE, Andersen EC: VCF-kit: assorted utilities for the variant call format. *Bioinformatics* 2017, 33(10):1581-1582.
25. Sargolzaei M: SNP1101 User's guide. In.: Version; 2014.
26. Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L: PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 2018, 35(10):1786-1788.
27. Weir BS, Cockerham CC: Estimating F-statistics for the analysis of population structure. *evolution* 1984, 38(6):1358-1370.
28. Lin G, Chai J, Yuan S, Mai C, Cai L, Murphy RW, Zhou W, Luo J: VennPainter: a tool for the comparison and identification of candidate genes based on venn diagrams. *PloS one* 2016, 11(4).
29. Slatkin M: Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* 2008, 9(6):477.
30. Wang Z-m, Li H-x, Liu X-f, He Y, Zeng H-l: Reduction of pyruvate orthophosphate dikinase activity is associated with high temperature-induced chalkiness in rice grains. *Plant Physiology and Biochemistry* 2015, 89:76-84.
31. Fu-Yun H, HUANG J, Ju-Fei L, Zhou-Fei W, ZHANG H-S: Isolation and expression analysis of plastidic glucose-6-phosphate dehydrogenase gene from rice (*Oryza sativa* L.). *Acta Genetica Sinica* 2006,

33(5):441-448.

32. Zhang L, Zhao L, Lin L, Zhao L, Liu Q, Wei C: A Novel Mutation of OsPPDKB, Encoding Pyruvate Orthophosphate Dikinase, Affects Metabolism and Structure of Starch in the Rice Endosperm. *International journal of molecular sciences* 2018, 19(8):2268.
33. Van Assche F, Cardinaels C, Clijsters H: Induction of enzyme capacity in plants as a result of heavy metal toxicity: Dose-response relations in *Phaseolus vulgaris* L., treated with zinc and cadmium. *Environmental pollution* 1988, 52(2):103-115.
34. Nemoto Y, Sasakuma T: Specific expression of glucose-6-phosphate dehydrogenase (G6PDH) gene by salt stress in wheat (*Triticum aestivum* L.). *Plant Science* 2000, 158(1-2):53-60.
35. Šindelář L, Šindelářová M, Burketová L: Changes in activity of glucose-6-phosphate and 6-phosphogluconate dehydrogenase isozymes upon potato virus Y infection in tobacco leaf tissues and protoplasts. *Plant Physiology and Biochemistry* 1999, 37(3):195-201.
36. Andaya V, Tai T: Fine mapping of the qCTS12 locus, a major QTL for seedling cold tolerance in rice. *Theoretical and Applied Genetics* 2006, 113(3):467-475.
37. Sales E, Viruel J, Domingo C, Marqués L: Genome wide association analysis of cold tolerance at germination in temperate japonica rice (*Oryza sativa* L.) varieties. *PloS one* 2017, 12(8).
38. Konishi H, Kitano H, Komatsu S: Identification of rice root proteins regulated by gibberellin using proteome analysis. *Plant, Cell & Environment* 2005, 28(3):328-339.
39. Swamy BM, Vikram P, Dixit S, Ahmed H, Kumar A: Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. *BMC genomics* 2011, 12(1):319.
40. Park HL, Kim TL, Bhoo SH, Lee TH, Lee S-W, Cho M-H: Biochemical characterization of the rice cinnamyl alcohol dehydrogenase gene family. *Molecules* 2018, 23(10):2659.
41. Bang SW, Lee DK, Jung H, Chung PJ, Kim YS, Choi YD, Suh JW, Kim JK: Overexpression of OsTF1L, a rice HD-Zip transcription factor, promotes lignin biosynthesis and stomatal closure that improves drought tolerance. *Plant biotechnology journal* 2019, 17(1):118-131.
42. Lenka SK, Katiyar A, Chinnusamy V, Bansal KC: Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant biotechnology journal* 2011, 9(3):315-327.
43. Lee JS, Torollo G, Ndayiragije A, Berchmans Bizimana J, Choi IR, Gulles A, Yeo US, Jeong OY, Venkatanagappa S, Kim BK: Genetic relationship of tropical region-bred temperate japonica rice (*Oryza sativa*) plants and their grain yield variations in three different tropical environments. *Plant breeding* 2018, 137(6):857-864.
44. Shan J, Cai Z, Zhang Y, Xu H, Rao J, Fan Y, Yang J: The underlying pathway involved in inter-subspecific hybrid male sterility in rice. *Genomics* 2019, 111(6):1447-1455.
45. Jung YJ, Nou IS, Kang KK: Overexpression of Oshsp16. 9 gene encoding small heat shock protein enhances tolerance to abiotic stresses in rice. *Plant Breeding and Biotechnology* 2014, 2(4):370-379.

46. Lin M-y, Chai K-h, Ko S-s, Kuang L-y, Lur H-S, Charng Y-y: A positive feedback loop between HEAT SHOCK PROTEIN101 and HEAT STRESS-ASSOCIATED 32-KD PROTEIN modulates long-term acquired thermotolerance illustrating diverse heat stress responses in rice varieties. *Plant physiology* 2014, 164(4):2045-2053.
47. Raskin I: Role of salicylic acid in plants. *Annual review of plant biology* 1992, 43(1):439-463.
48. Ganesan V, Thomas G: Salicylic acid response in rice: influence of salicylic acid on H₂O₂ accumulation and oxidative stress. *Plant Science* 2001, 160(6):1095-1106.
49. Kang S-M, Titus JS: Increased proteolysis of senescing rice leaves in the presence of NaCl and KCl. *Plant physiology* 1989, 91(3):1232-1237.
50. Ehlers MR, Riordan JF: Membrane proteins with soluble counterparts: role of proteolysis in the release of transmembrane proteins. *Biochemistry* 1991, 30(42):10065-10074.
51. Gho Y-S, Park S-A, Kim S-R, Chandran AKN, An G, Jung K-H: Comparative Expression Analysis of rice and Arabidopsis peroxiredoxin genes suggests conserved or diversified roles between the two species and leads to the identification of tandemly duplicated rice peroxiredoxin genes differentially expressed in seeds. *Rice* 2017, 10(1):30.
52. Wang D, Weaver ND, Kesarwani M, Dong X: Induction of protein secretory pathway is required for systemic acquired resistance. *Science* 2005, 308(5724):1036-1040.
53. Gorantla M, Babu P, Reddy Lachagari V, Reddy A, Wusirika R, Bennetzen JL, Reddy AR: Identification of stress-responsive genes in an indica rice (*Oryza sativa* L.) using ESTs generated from drought-stressed seedlings. *Journal of experimental botany* 2007, 58(2):253-265.
54. Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B: Plant 'pathogenesis-related' proteins and their role in defense against pathogens. *Biochimie* 1993, 75(8):687-706.
55. Wu J, Kim SG, Kang KY, Kim J-G, Park S-R, Gupta R, Kim YH, Wang Y, Kim ST: Overexpression of a pathogenesis-related protein 10 enhances biotic and abiotic stress tolerance in rice. *The plant pathology journal* 2016, 32(6):552.
56. Matsui H, Takahashi A, Hirochika H: Rice immune regulator, OsPti1a, is specifically phosphorylated at the plasma membrane. *Plant signaling & behavior* 2015, 10(3):e991569.
57. Park C-J, Caddell DF, Ronald PC: Protein phosphorylation in plant immunity: insights into the regulation of pattern recognition receptor-mediated signaling. *Frontiers in plant science* 2012, 3:177.
58. Liu Q, Luo L, Zheng L: Lignins: biosynthesis and biological functions in plants. *International journal of molecular sciences* 2018, 19(2):335.
59. Kobayashi K: Role of membrane glycerolipids in photosynthesis, thylakoid biogenesis and chloroplast development. *Journal of plant research* 2016, 129(4):565-580.
60. Shao J, Zhang Y, Yu J, Guo L, Ding Y: Isolation of thylakoid membrane complexes from rice by a new double-strips BN/SDS-PAGE and bioinformatics prediction of stromal ridge subunits interaction. *PLoS One* 2011, 6(5).

61. Yang J, Li W-R, Lv F-H, He S-G, Tian S-L, Peng W-F, Sun Y-W, Zhao Y-X, Tu X-L, Zhang M: Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular biology and evolution* 2016, 33(10):2576-2592.
62. Zhang H, Yin L, Wang M, Yuan X, Liu X: Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Frontiers in Genetics* 2019, 10.
63. Lee K-J, Kim K: The rice serine/threonine protein kinase OsPBL1 (ORYZA SATIVA ARABIDOPSIS PBS1-LIKE 1) is potentially involved in resistance to rice stripe disease. *Plant growth regulation* 2015, 77(1):67-75.
64. Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T: SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research* 2000, 10(10):1579-1586.
65. Andersen EJ, Ali S, Reese RN, Yen Y, Neupane S, Nepal MP: Diversity and evolution of disease resistance genes in barley (*Hordeum vulgare* L.). *Evolutionary Bioinformatics* 2016, 12:EBO. S38085.
66. Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y: A map of rice genome variation reveals the origin of cultivated rice. *Nature* 2012, 490(7421):497-501.
67. Chen X, Ouyang Y, Fan Y, Qiu B, Zhang G, Zeng F: The pathway of transmembrane cadmium influx via calcium-permeable channels and its spatial characteristics along rice root. *Journal of experimental botany* 2018, 69(21):5279-5291.
68. Saichuk J: Louisiana rice production handbook: Louisiana State University Agricultural Center; 2014.

Figures

Figure 1

Principle component (PC) analysis showing global genetic variation in Aus, Indica, Temperate japonica and Tropical japonica subpopulations. Left: first and second PCs; right: second and third PCs.

Figure 1

Principle component (PC) analysis showing global genetic variation in Aus, Indica, Temperate japonica and Tropical japonica subpopulations. Left: first and second PCs; right: second and third PCs.

Figure 2

The admixture analysis at K=2, K=3 and K=4 for Aus, Indica, Temperate japonica and Tropical japonica subpopulations

Figure 2

The admixture analysis at K=2, K=3 and K=4 for Aus, Indica, Temperate japonica and Tropical japonica subpopulations

Figure 3

The neighbor-joining phylogenetic tree of four commercial Asian rice subspecies including Aus (orange lines), Indica (green lines), Temperate japonica (red lines) and Tropical japonica (blue lines)

Figure 3

The neighbor-joining phylogenetic tree of four commercial Asian rice subspecies including *Aus* (orange lines), *Indica* (green lines), *Temperate japonica* (red lines) and *Tropical japonica* (blue lines)

Figure 4

Linkage disequilibrium decay for Aus (black line), Indica (purple line), Temperate japonica (red line) and Tropical japonica (blue line) groups

Figure 4

Linkage disequilibrium decay for Aus (black line), Indica (purple line), Temperate japonica (red line) and Tropical japonica (blue line) groups

Figure 5

Circos plot of $\theta\pi$ values in 10Kb windows (points) for Temperate japonica (purple points, $4.17\text{E-}7 \leq \theta\pi \leq 0.0013$), Aus (black points, $5.35\text{E-}7 \leq \theta\pi \leq 0.0013$), Tropical japonica (green points, $2.78\text{E-}7 \leq \theta\pi \leq 0.0011$), Indica (red points, $2.01\text{E-}7 \leq \theta\pi \leq 0.0014$) and across all populations (sky blue points, $9.55\text{E-}6 \leq \theta\pi \leq 0.0015$). The rules show the position and length of each chromosome. The outermost circle shows the ideograms of each chromosome (Chr 1-12).

Figure 5

Circos plot of $\theta\pi$ values in 10Kb windows (points) for Temperate japonica (purple points, $4.17\text{E-}7 \leq \theta\pi \leq 0.0013$), Aus (black points, $5.35\text{E-}7 \leq \theta\pi \leq 0.0013$), Tropical japonica (green points, $2.78\text{E-}7 \leq \theta\pi \leq 0.0011$), Indica (red points, $2.01\text{E-}7 \leq \theta\pi \leq 0.0014$) and across all populations (sky blue points, $9.55\text{E-}6 \leq \theta\pi \leq 0.0015$). The rules show the position and length of each chromosome. The outermost circle shows the ideograms of each chromosome (Chr 1-12).

Figure 6

The distribution of $Z(F_{st})$ values of 12 rice chromosomes in six different scenarios including: (A) Tropical japonica vs. Temperate japonica scenario (blue line, $Z(F_{st}) \geq 3.21$), (B) Temperate japonica vs. Indica scenario (blue line, $Z(F_{st}) \geq 2.18$), (C) Tropical japonica vs. Indica scenario (blue line, $Z(F_{st}) \geq 2.10$), (D) Aus vs. Indica scenario (blue line, $Z(F_{st}) \geq 2.98$), (E) Tropical japonica vs. Aus scenario (blue line, $Z(F_{st}) \geq 2.17$) and (F) Aus vs. Temperate japonica scenario (blue line, $Z(F_{st}) \geq 2.16$). The data points above the horizontal line are the top 1% of $Z(F_{st})$ values.

Figure 6

The distribution of $Z(F_{st})$ values of 12 rice chromosomes in six different scenarios including: (A) Tropical japonica vs. Temperate japonica scenario (blue line, $Z(F_{st}) \geq 3.21$), (B) Temperate japonica vs. Indica scenario (blue line, $Z(F_{st}) \geq 2.18$), (C) Tropical japonica vs. Indica scenario (blue line, $Z(F_{st}) \geq 2.10$), (D) Aus vs. Indica scenario (blue line, $Z(F_{st}) \geq 2.98$), (E) Tropical japonica vs. Aus scenario, (blue line, $Z(F_{st}) \geq 2.17$) and (F) Aus vs. Temperate japonica scenario (blue line, $Z(F_{st}) \geq 2.16$). The data points above the horizontal line are the top 1% of $Z(F_{st})$ values.

Figure 7

Venn diagram of the number of (A) overlapped genomic regions, and (B) overlapped biological processes underlying selection pressures among the six studied scenarios

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData1.csv](#)
- [SupplementaryData1.csv](#)
- [SupplementaryData2.csv](#)
- [SupplementaryData2.csv](#)
- [SupplementaryData3.csv](#)
- [SupplementaryData3.csv](#)
- [SupplementaryData4.csv](#)
- [SupplementaryData4.csv](#)
- [SupplementaryData5.csv](#)
- [SupplementaryData5.csv](#)
- [SupplementaryData6.csv](#)
- [SupplementaryData6.csv](#)
- [SupplementaryData7.csv](#)
- [SupplementaryData7.csv](#)
- [Supplementaryinformation.docx](#)
- [Supplementaryinformation.docx](#)