Assessment of Genetic Diversity in Traditional Landraces and Improved Cultivars of Rice

Nageen Zahra
National Institute of Genomics and Advanced Bio-Technology (NIGAB), National Agriculture Research Center, Islamabad

Muhammad Kashif Naeem
National Institute of Genomics and Advanced Bio-Technology (NIGAB), National Agriculture Research Center, Islamabad

Bilal Saleem
National Institute of Genomics and Advanced Bio-Technology (NIGAB), National Agriculture Research Center, Islamabad

Muhammad Aqeel
National Institute of Genomics and Advanced Bio-Technology (NIGAB), National Agriculture Research Center, Islamabad

Wajya Ajmal
National Institute of Genomics and Advanced Bio-Technology (NIGAB), National Agriculture Research Center, Islamabad

Syed Adeel Zafar
National Institute of Genomics and Advanced Bio-Technology (NIGAB), National Agriculture Research Center, Islamabad

Shahzad Amir Naveed
Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Muhammad Ramzan Khan (drmrkhan_nigab@yahoo.com)
PARC: Pakistan Agricultural Research Council https://orcid.org/0000-0001-9167-6556

Research article

Keywords: Rice, landraces, improved cultivars, genetic diversity, population genetics

Posted Date: December 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-125856/v1

Abstract

Background: Rice is the staple food for more than half of the world's population. Rice cultivation needs expansion to meet the increasing food demands across the globe. Genetic diversity is desired for crop breeding because it serves as the backbone for improving cultivars. The process of domestication and modern plant breeding technologies applied to rice has contributed to the erosion of genetic diversity. Current breeding programs have extensively shaped the genetic diversity of elite rice cultivars to no small extent.

Results: We explored the genetic diversity of traditional landraces and improved cultivars by inspecting the whole genome SNP markers of 20 rice accessions. We found a higher number of genetic variations (76.70%) and observed heterozygosity (0.024) in landraces than improved cultivars. The principal component analysis also revealed the higher genetic diversity among the landraces. While population structure based on the phylogenetic tree suggested the population's structure according to rice subspecies. The genetic diversity parameter, $F_{ST,}$ was applied to estimate the genetic differentiation of rice, which revealed week genetic differentiation (0.121) and nucleotide diversity (0.314) in modern rice cultivars. Genome-wide genetic differentiation (F_{ST}) analysis identified the two domesticated genes: *Kala4 (*pericarp color*)* and *Ghd7 (*heading date*),* and eight improvement genes: *Sd1, Ghd8*, *GW2*, *NRT1.1b*, *GW6a*, and *Hd3a,* that coincide with the candidate selective sweeps. Inbreeding depression (0.68617) among the modern cultivars suggests no genetic gain in future breeding efforts and compels exotic material utilization in the breeding programs.

Conclusion: These findings demonstrate that modern cultivars have a narrow genetic base compared to landraces. Therefore, exploring the genome of landraces at a large scale to identify the genes responsible for stability and adaptation to abiotic stresses can help design varieties that can survive vulnerable climates.

Background

Rice is one of the important staple food crops worldwide. Rice is cultivated in the broad sphere of ecological and climatic conditions across the globe. The Asian cultivated rice (*Oryza sativa*) is the primary food crop that satisfies the food demands of more than half of the world's population [1]. To create food surpluses for the rapidly growing world population is one of the biggest concerns nowadays. According to the prediction, this current production rate is insufficient for the projected global community in 2035. To feed the expanding population: a further 116 million tons of rice will be required [2]. Rice farming needs expansion to fulfill the demand for rice desired across the globe. Since the onset of agriculture, crop plants have undergone a series of genetic manipulation to meet the expanding world population [3]. Currently, rice breeders focused on both the quality and yield of the major food crops that fulfill human beings' dietary needs. Achieving future production goals is not straightforward due to narrow genetic diversity within the breeding stocks.

Genetic diversity plays an essential role in the evolution of species [4]. It favors the improvement of crops and allows them to adapt to the changing environmental conditions. Diversity in crop plants presents the plant breeders with the opportunity to cultivate improved varieties that have favorable traits like high yield, large grain size, and resilient biotic and abiotic factors. Artificial selection and rice's adaptability to various habitats has ended in a diverse range of improved varieties. Almost 780,000 rice accessions are presented in the gene banks worldwide [5]. To explore the genetic diversity in rice (*Oryza sativa*): an international effort was made to re-sequence the 3000 rice accessions from different parts of the world and made the data publicly accessible to the scientific community [6]. This single nucleotide polymorphism (SNPs) dataset of genome sequences opens a new way to outline the genetic diversity within the spectrum of plant germplasm, including traditional landraces and improved cultivars and wild ancestors. Information from genomic sequences provides an opportunity for breeders to select desired diversity for improved farming varieties. This dataset of genetic variations grants the ground for the characterization of population diversity, population structure, and species [7]. The 3000 rice genome data has been employed recently for the study of structural variants [8], genetic variations, population structure and diversity [9], and detection of transposable elements insertion in rice [10].

During domestication and plant breeding, technologies have contributed to the erosion of genetic diversity, resulting in making the crop plants defenseless against the dynamic climate conditions with lower genetic potentials in the future. Such as the Irish potato famine [11] and southern corn leaf blight [12] are examples of this. Modern breeding practices also resulted in a narrow genetic base of advanced lines due to artificial selection pressure for improvement related genes. Hence, the present study examines the genetic diversity erosion phenomena and artificial selection footprints. Pooling many accessions together and using shallow genetic variation data provided limited information [13]. Therefore, a small sample size with deep sequencing is a more reliable strategy [14]. To get more in-depth information, we collected the whole-genome re-sequencing genetic variant data from 20 diverse accessions belong to four main rice groups, i.e., *indica*, *japonica*, aus, and admixed. We used the whole-genome re-sequencing data of 20 rice accessions to (a) assess the population structure using distance-based methods and principal component analysis, (b) estimate the genetic diversity among traditional landraces and improved cultivars through population genetics analysis, and (c) examine the footprints of genetic erosion and artificial selection.

These findings from population genetic analysis provide insight into genetic diversity within the traditional landraces and improved cultivars and identify the variants highly variable between the populations and are associated with important traits. These variants can be useful in marker-assisted selection for modern breeding programs.

Results

Genotyping and variant calling

Genetic diversity is the prime objective of improving the genetic gain of the crop. This study estimates the genetic diversity between traditional landraces and improved cultivars. We first downloaded the whole genome sequencing data of 20 rice accessions (Fig. 1). Using the GATK variant calling pipeline, we identified approximately 4.91 million single nucleotide polymorphisms (SNPs) and 0.80 million indels (insertions/deletions). These single nucleotide variations were further subjected to hard filtering. Hard filtering reduced the numbers to 3.80 million SNPs and 0.67 million indels. A total of 2.47 million SNPs and 0.42 million indels were identified in improved cultivars, while 2.88 million SNPs and 0.49 million indels were identified in traditional landraces. A large quantity (2.47 million SNPs, 85.7%) of these SNP shared among improved cultivars and landraces, suggesting that most genetic variations in improved cultivars are derived from the variation in landraces. While the remaining 14.3% SNPs are specific to landraces. This is consistent with the previous studies that landraces have a diverse genetic pool than the improved varieties [9, 14, 15, 16] and may contain valuable genetic resources for rice improvement. We selected 3.80 million SNPs for our downstream analysis.

Population structure

Large-scale SNPs that are uniformly distributed across the genome provide the chance to improve the population structure and genetic diversity analysis. To investigate the population structure: whole-genome SNPs from the dataset were used to perform phylogenetic tree construction and principal component analysis. Unrooted neighbor-joining tree divided all rice accessions into three clusters. 11 samples belonging to the *indica* group formed one cluster. The other two clusters represented *japonica* and aus varieties. One sample belonging to the intermediate (admixed) type was clustered close to the *japonica* group (Fig. 2). Interestingly, this distribution of rice accessions displays the same grouping pattern as previously identified by Wang et al. (2018) [9] and Wang et al. (2016) [17].

We analyzed the whole genome biallelic SNPs for two subpopulations; traditional landraces and improved cultivars, to perform principal component analysis. Principal component analysis has clustered the 20 rice accessions into three groups (Fig. 3). Interestingly, the clustering of samples is the same as observed in the phylogenetic tree. Traditional landraces are scattered along the axis of the PCA plot, indicating the higher genetic diversity among the landraces compared to improved cultivars. The first PCA explains 21%, and the second PCA explains 12% of the total variance. Hence, first, two PCA add up to explained 33% of total SNP variation, which is higher than the previous studies [14, 17, 18]. PCA scores of SNPs were analyzed in correlation to the axis. Two principle coordinates are enough to epitomize the total variance between the two populations. SNPs in the 1st principal coordinate explained more variance than the 2nd coordinates. 1st principal coordinate is more differentiating between the populations. Based on 1st principal coordinate plotted top 1000 SNPs are plotted with the highest variance values in the overall population (Fig. 4). PCA allows identifying the contribution of SNPs in structuring the population by using F-statistics. Spatial dependencies of SNPs find regions within the genome that are responsible for structuring the populations and can be identified as selection signals [19].

Estimation of population genetics

We further investigated genetic divergence between traditional landraces and improved cultivars. To measure the genetic differentiation in selected samples, we calculated pairwise nucleotide diversity and F-statistics, expected heterozygosity, and inbreeding coefficient for each SNP locus. To investigate the sub-populations' genetic diversity: we calculated nucleotide diversity (π) in each group. The ratio of nucleotide polymorphism in landraces and varieties is 0.314 and 0.321, with an unremarkable difference. The inbreeding coefficient F_{IS} values in landraces and varieties are 0.68173 and 0.68617; shows an insignificant discrepancy. But these F_{IS} values suggest high inbreeding in selected rice genotypes.

We further investigated the genetic differentiation between the landraces and improved cultivars. For this purpose, we calculated the fixation index value (F_{ST}) at the whole-genome level between the two subpopulations, which showed weaker genetic differentiation of 0.121.

Traditional landraces show higher genetic diversity concerning polymorphic loci of 76.70% and private alleles 847986. The proportion of polymorphic loci in improved varieties is 75.19%, and the number of private alleles is 786955. The observed heterozygosity between the landraces and improved cultivars are 0.024 and 0.017, respectively (Table 1). In comparison, observed homozygosity between the landraces and improved cultivars is 0.975 and 0.983, respectively. The pattern of homozygosity and heterozygosity between the landraces and improved cultivars also suggested a slight decline in variability and increased homozygosity in varieties (Table 1). These findings are consistent with the general expectation of traditional landraces possesses more genetic diversity than modern cultivars.

Table 1
Homozygosity and heterozygosity ratios in landraces and varieties

Pedigree Type	Observed homozygosity	Expected homozygosity	Observed heterozygosity	Expected heterozygosity
Traditional landraces	0.97556	0.72126	0.02444	0.27874
Improved cultivars	0.98329	0.71253	0.01671	0.28747

Selective signatures affected by selection

During selection for crop improvement, the breeder selected for favorable traits and unselected the regions linked to yield-limiting factors. Selective signature causes a reduction in nucleotide diversity, and the choice of favorable characteristics increases the allele frequency in the modern cultivars [14, 20]. Whole-genome sequencing data from landraces and improved cultivars provide an opportunity to identify the selective regions. We calculated the ratio of genetic diversity in modern cultivars to the diversity in landraces ($\pi = \pi_{\text{improved cultivars}}/\pi_{\text{landraces}}$) in the non-overlapping window of 10 kb along the entire genome (Fig. 5). To determine the candidate selective sweeps: the top 10% of the ratio of the genetic differentiation between improved cultivars and landraces was selected. To support these results, we

further estimated the genetic differentiation between improved cultivars and landraces using the same non-overlapping window of 10 kb along the entire genome. Use similar top10% threshold criteria to select the candidate selective sweeps from genetic differentiation (F_{ST}) results. We noticed many regions with strong selection signals where F_{ST} between modern cultivars and landraces were extremely low. To identify the vital selection signals, we selected SNPs that were in domestication and improvement related genes. We selected the most 13 well-characterized domesticated genes, including *Prog1* (tiller angle) [21], *Rc* (pericarp color) [22], *qSH1* (seed shattering) [23], *sh4* (reduce seed shattering) [24], *Ghd7* (heading date) [25], *LABA1* (barbless awns) [26], *Kala4* (pericarp color) [27], *LG1* (grain width) [28], *OsLG1* (Alteration in the laminar joint and ligule development forming closed panicles) [29], *GW5* (grain width) [30], *Bh4* (hull color) [31], *An-1* (awn length) [32] *and GAD1* (awn length) [33].

While, 27 improvement genes, including *OsSPL14* (plant architecture) [34], *DEP1* (dense and erect panicle) [35], *TAC1* (tiller angle) [36], *GW2* (grain width) [37], *GS5* (grain length and width) [38], *GW6a* (grain length) [39], *TGW6* (grain width) [40], *GW7* (grain length and width) [41], *GLW7* (grain length) [42], *GW8* (grain width) [43], *Hd3a* (heading date) [44], *Ghd7* (heading date) [25], *Ghd8* (heading date) [45], *Hd1* (heading date) [46], *Sd1* (plant height) [47], *Tms5* (thermosensitive male sterility) [48], *Pigm* (blast resistance) [49], *Bph6* (brown planthopper resistance) [50], *Sub1A* (submergence tolerance) [51], *SNORKEL1* (deep water adaptation) [52], *SNORKEL2* (deep water adaptation) [52], *TT1* (thermotolerance) [53], *NRT1.1B* (nitrate uptake) [54], *Dro1* (deeper rooting) [55], *Chalk5* (grain quality) [56], *Waxy* (grain quality) [57], *ALK* (starch gelatinization) [58] were noted. Among these domesticated and improvement genes, we identified *Kala4, Ghd7, Sd1, Ghd8*, *GW2*, *NRT1.1b*, *GW6a*, and *Hd3a* genes coincide with the candidate selective sweeps (Fig. 6).

Discussion

Rice is one of the most ancient and extensively consumed staple food crops. Its cultivation and domestication have a significant role in the rise of agricultural civilization in Asia. Rice is considered to have been domesticated from Asian wild rice, O. rufipogon, 10,000 years ago [10, 59, 60, 61]. The split between two progenitors, *indica*, and *japonica* from which both cultivated types originated, occurred 800,000 years ago [10]. This separation shows long before the origin of agriculture. While aus/boro lineage split from *indica* appears to be more recent as ~ 540,000 years ago [10].

During the process of domestication, rice has experienced significant phenotypic changes like grain size, color, shattering, seed dormancy, and tillering, as recently identified and verified through quantitative trait loci mapping [21, 22, 62, 63]. Since the domestication of rice, a series of artificial selection procedures have been applied in rice breeding programs that have led to a decline in genetic variability [64]. After the domestication, rice breeders mainly focused on selecting lines with long grain, more tillering, and high yield potential, except for other biotic and abiotic stress tolerant and quality traits. Such unidirectional selection of varieties resulted in a narrow genetic base among the modern cultivars. The present study assessed 20 rice accession genetic diversity, including landraces and modern cultivars, using SNP markers.

Our results revealed that modern cultivars have a narrow genetic base compared to landraces. Like the previous study, the genetic bottleneck caused a limited diversity in cultivated varieties [15]. There is an urgent need to harnessing genetic variation for further improvement and enhance the crop yield's genetic gains. Higher heterozygosity was observed in landraces (H_T = 0.02444) than improved varieties (H_T =0.01671), as also observed by Alvarez et al. (2007) [65]. Low F_{ST} values between these sub-populations and increased observed homozygosity in varieties suggest high inbreeding depression. Thus, the genetic diversity within landraces will be significant for designing new commercial varieties to broaden the new genotypes' variability.

Genetic diversity is desired for crop breeding because it serves as the backbone for improving cultivars. It assists in designing varieties capable of coping with changing climatic conditions by manipulating genetic makeup [66]. Developing elite rice cultivars with increased genetic variability has become a leading challenge for crop breeders, which can implicate recent advances in breeding technologies. The collection of diverse and valuable germplasm in the gene bank is one of the keys to enhancing genetic diversity [14, 67].

Modern crop breeding techniques and advances in crop management practices significantly improve the annual gain of 0.8–1.2% in crop productivity [68]. Genomic breeding is one of the modern breeding technologies, integrating diverse accessions, genomic resources, and molecular technology and breeding tools. Large scale dense genotyping of various germplasm resources has become an essential part of crop germplasm characterization and its further utilization. Based on the genetic and morphological characterization of germplasm, additional help dissect the genetic basis of quantitative traits and identify the novel genes [16, 41, 69]. Utilization of critical genetic loci and pyramiding of these loci through breeding, leading to the development of new germplasm. This advanced breeding approach is named "genome-based breeding by design." Genome-based breeding by design strategy successfully develops green super rice (GSR) cultivars [70, 71, 72].

Genomic selection is one of the most crucial breeding strategies to increase genetic gains and have advantages over the traditional approaches [73]. It can be improved by incorporating the high-throughput SNP chips and next-generation sequencing (NGS)-based platforms and high-throughput phenotyping technologies. This advanced technology helps identify suitable parents for breeding programs, ultimately resulting in a genetic gain of future crops. A genome-wide association study is another genomic strategy using in rice crops [14, 15, 17]. This technique is used to decipher the genetic basis of important quantitative traits, identifying the novel genes underlying study traits, and providing knowledge about the valuable haplotypes. With the implementing such beneficial haplotypes in breeding program result in a genetic gain of advanced cultivars. A haplotype is consisting of two or more SNPs with strong linkage disequilibrium. Sometimes one SNP linked with an undesirable trait causes linkage drag. In this regard, gene-editing technology plays a vital role in reducing linkage drag and regulating critical genes' gene expression. CRISPR-Cas9 tool enabled multiplex-gene editing and was considered a non-GMO approach [74]. This advanced genome editing technique helps breed the new cultivars by activating the homeo-alleles of a gene or deactivating the alleles causing the linkage drag.

In the future, diverse germplasm collection, especially early domesticated cultivars (landraces), enriches the gene pools with multiple genetic backgrounds. Traditional landraces are a rich source of genetic variability and adaptable to stressful environmental conditions. Therefore, exploring the genome of landraces at a large scale to identify the genes responsible for stability and adaptation to abiotic stresses can help design varieties that can survive vulnerable climates. Further, effective implementation of these advanced breeding techniques at one platform will bring the next-generation crops with higher genetic gains. These next-generation crops will help to meet the food security demands for the projected global population in 2050.

Conclusion

The present study based on whole-genome SNPs assessed the genetic diversity in traditional landraces and improved varieties of rice. Crop breeding requires genetic diversity for developing new cultivars and improving varieties. Genetic diversity estimation suggests that there is an immediate appeal to include more diverse donor parents in the breeding programs for improvement in varieties to broaden the genetic basis.

Abbreviations

SNP: single nucleotide polymorphism

F_{ST}: the population differentiation statistics

PCA: principal component analysis

F_{IS}: the population inbreeding coefficient

H_T: heterozygosity

GSR: green super rice

NGS: next-generation sequencing

CRISPR: clustered regularly interspaced short palindromic repeats repetitive

GMO: genetically modified organisms

Methods

Genetic Data Collection and Imputation

Diverse 20 rice accession, including ten traditional landraces and ten improved cultivars belonging to 12 countries (Pakistan, Japan, South Korea, Bangladesh, Philippines, India, Thailand, Srilanka, Taiwan,

Malaysia, Indonesia, and China) were selected from 3K Rice Genome Project. Whole-genome re-sequencing data of these accessions were retrieved from NCBI SRA.

Raw reads were aligned to reference assembly Nipponbare IRGSP-1.0 using BWA-MEM (version 0.7.17-r1188) [75]. The SNP/Indel calling was performed using the GATK pipeline (version v4.1.6.0) [76]. Variant calling was carried out using the GATK HaplotypeCaller in two steps: (a) on each sample individually, (b) joint variant calling on all samples in GVCF mode. Variants identified were subjected to hard filtering following the GATK VariantFiltration procedure with the criteria of 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0' for SNPs and 'QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0' for indels. Furthermore, variants that passed the criterion of DP > 2 and variant quality more than 50 were selected for downstream analysis.

Structure analysis

Principal component analysis PCA was performed to investigate the genetic structure of populations and relationships in association with SNPs, using the "adegenet" package in R [77].

Phylogenetic analysis

Phylogenetic analysis was performed for the integrity of the variant calling pipeline. SNPs from the whole genome were selected for NJ tree construction. The distance matrix based on the p-distance model was calculated for all SNPs, and an unrooted neighbor-joining tree was constructed using TASSEL (standalone v.5.0) [78] and visualized using Interactive Tree of Life (iTOL) [79].

Estimation of Genetic Diversity

Genetic analysis was performed to estimate the genetic diversity between and within the traditional landraces and improved cultivars population. Populations program from Stacks [80] was used for F-statistics to measure pairwise genetic differentiation F_{ST} [81], nucleotide diversity π [82], and heterozygosity for genetic variability, and percentage of polymorphic loci across the genome.

Genomic Fingerprints for Selective Sweeps

For selective sweeps identification, we employed two approaches: (1) identify the genomic regions which are lost during selection procedure from landraces to modern varieties resulted in narrow genetic-base and (2) fingerprint the selection pressure-related areas, which are highly selected for crop improvement. For this purpose, two population statistics methods, i.e., genetic diversity (π) and genetic differentiation (F_{ST}), were employed. The genetic diversity in the landraces and modern varieties ($\pi_{landrace}$ / $\pi_{variety}$) was measured. Window-based π was calculated using vcftools [83] with the window size of 100 kb, and candidate selective sweeps were selected based on the top 10% of values. Genetic differentiation (F_{ST}) was also calculated with the window size of 100 kb by using vcftools.

Declarations

Acknowledgments

The authors are highly grateful to Director NIGAB and NARC for providing all the necessary facilities.

Funding

This work was part of the “Prime Minister Agriculture Emergency Program”; project entitled “Green Super Rice in Pakistan” (PSDP) in NIGAB-National Agriculture Research Center, Islamabad, Pakistan. The project’s fund was utilized for the completion of the project work. The funds were utilized for study design, data analyses, manuscript writing, and interpretation of the results. No externally aided fund was received for this study.

Availability of data and materials

The datasets analyzed during the current study are available for download from NCBI SRA at PRJEB6180.

Authors' contributions

MRK and MKN design the experiment. NZ and BS retrieved the 20 lines' data from the 3K repository. NZ and MA performed the data analysis. MKN and NZ interpreted the results and wrote down the result and discussion section. NZ, SAN, WA, and BS completed the first draft of the paper. MRK and SAZ finalized the final version of the manuscript. All authors read and approved the final manuscript.

Ethics declarations

Ethics approval and consent to participate

The authors declare that this study complies with the current laws of the countries in which the experiments were performed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Lu BR. Taxonomy of the genus *Oryza*(Poaceae): historical perspective and current status. Int Rice Res Notes. 1999;24:4–8.
2. Seck PA, Diagne A, Mohanty S, Wopereis MCS. Crops that feed the world 7: Rice. Food Sec. 2012;4:7–24.

3. Bhandari H, Bhanu AN, Srivastava K, Singh M, Shreya, Hemantaranjan A. Assessment of genetic diversity in crop plants - an overview. Advances in Plants & Agriculture Research. 2017;Volume 7 Issue 3. doi:15406/apar.2017.07.00255.
4. Reed DH, Frankham R. Correlation between Fitness and Genetic Diversity. Conservation Biology. 2003;17:230–7.
5. Allender C. The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome: Food and Agriculture Organization of the United Nations (2010), pp. 370, US$95.00, ISBN 978-92-5-106534-1. Experimental Agriculture. 2011;47:574.
6. The 3 000 rice genomes project. The 3,000 rice genomes project. GigaScience. 2014;3:7.
7. Onda Y, Mochida K. Exploring Genetic Diversity in Plants Using High-Throughput Sequencing Techniques. Curr Genomics. 2016;17:358–67.
8. Fuentes RR, Chebotarov D, Duitama J, Smith S, De la Hoz JF, Mohiyuddin M, et al. Structural variants in 3000 rice genomes. Genome Res. 2019;29:870–80.
9. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature. 2018;557:43–9.
10. Carpentier M-C, Manfroi E, Wei F-J, Wu H-P, Lasserre E, Llauro C, et al. Retrotranspositional landscape of Asian rice revealed by 3000 genomes. Nature Communications. 2019;10:24.
11. Mizubuti ESG, Fry WE. Potato late blight. In: COOKE BM, JONES DG, KAYE B, editors. The Epidemiology of Plant Diseases. Dordrecht: Springer Netherlands; 2006. p. 445–71. doi:1007/1-4020-4581-6_17.
12. Ullstrup AJ. The Impacts of the Southern Corn Leaf Blight Epidemics of 1970-1971. Annual Review of Phytopathology. 1972;10:37–50.
13. Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. Genetic Structure and Diversity in Oryza sativa L. Genetics. 2005;169:1631–8.
14. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature Biotechnology. 2012;30:105–11.
15. Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, et al. A map of rice genome variation reveals the origin of cultivated rice. Nature. 2012;490:497–501.
16. Xie W, Wang G, Yuan M, Yao W, Lyu K, Zhao H, et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. Proc Natl Acad Sci USA. 2015;112:E5411-5419.
17. Wang H, Xu X, Vieira FG, Xiao Y, Li Z, Wang J, et al. The Power of Inbreeding: NGS-Based GWAS of Rice Reveals Convergent Evolution during Rice Domestication. Mol Plant. 2016;9:975–85.
18. Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. Genome Res. 2017;27:1029–38.

19. Laloë D, Gautier M. On the genetic interpretation of Between-Group PCA on SNP data. Research Report. auto-saisine; 2012. https://hal.archives-ouvertes.fr/hal-01193689. Accessed 3 Sep 2020.
20. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, et al. The effects of artificial selection on the maize genome. Science. 2005;308:1310–4.
21. Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, et al. Control of a key transition from prostrate to erect growth in rice domestication. Nat Genet. 2008;40:1360–4.
22. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S. Caught Red-Handed: Rc Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. The Plant Cell. 2006;18:283–94.
23. Magwa RA, Zhao H, Yao W, Xie W, Yang L, Xing Y, et al. Genome wide association analysis for awn length linked to the seed shattering gene qSH1 in rice. J Genet. 2016;95:639–46.
24. Lin Z, Griffith ME, Li X, Zhu Z, Tan L, Fu Y, et al. Origin of seed shattering in rice (Oryza sativa L.). Planta. 2007;226:11–20.
25. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nature Genetics. 2008;40:761–7.
26. Hua L, Wang DR, Tan L, Fu Y, Liu F, Xiao L, et al. LABA1, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice. The Plant Cell. 2015;27:1875–88.
27. Oikawa T, Maeda H, Oguchi T, Yamaguchi T, Tanabe N, Ebana K, et al. The Birth of a Black Rice Gene and Its Local Spread by Introgression. Plant Cell. 2015;27:2401–14.
28. Shi C, Ren Y, Liu L, Wang F, Zhang H, Tian P, et al. Ubiquitin Specific Protease 15 Has an Important Role in Regulating Grain Width and Size in Rice. Plant Physiology. 2019;180:381–91.
29. Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, Htun TM, et al. OsLG1 regulates a closed panicle trait in domesticated rice. Nat Genet. 2013;45:462–5, 465e1-2.
30. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, et al. Isolation and initial characterization of GW5 , a major QTL associated with rice grain width and weight. Cell Res. 2008;18:1199–209.
31. Zhu B-F, Si L, Wang Z, Zhou Y, Zhu J, Shangguan Y, et al. Genetic control of a transition from black to straw-white seed hull in rice domestication. Plant Physiol. 2011;155:1301–11.
32. Luo J, Liu H, Zhou T, Gu B, Huang X, Shangguan Y, et al. An-1 encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. Plant Cell. 2013;25:3360–76.
33. Jin J, Hua L, Zhu Z, Tan L, Zhao X, Zhang W, et al. GAD1 Encodes a Secreted Peptide That Regulates Grain Number, Grain Length, and Awn Development in Rice Domestication. Plant Cell. 2016;28:2453–63.
34. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, et al. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nature Genetics. 2010;42:541–4.
35. Xu H, Zhao M, Zhang Q, Xu Z, Xu Q. The DENSE AND ERECT PANICLE 1 (DEP1) gene offering the potential in the breeding of high-yielding rice. Breed Sci. 2016;66:659–67.
36. Yu B, Lin Z, Li H, Li X, Li J, Wang Y, et al. TAC1, a major quantitative trait locus controlling tiller angle in rice. The Plant Journal. 2007;52:891–8.

37. Lee KH, Park SW, Kim YJ, Koo YJ, Song JT, Seo HS. Grain width 2 (GW2) and its interacting proteins regulate seed development in rice (Oryza sativa L.). Bot Stud. 2018;59. doi:1186/s40529-018-0240-z.
38. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet. 2011;43:1266–9.
39. Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, et al. Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. PNAS. 2015;112:76–81.
40. Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, et al. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. Nature Genetics. 2013;45:707–11.
41. Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, et al. The OsSPL16 - GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat Genet. 2015;47:949–54.
42. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, et al. OsSPL13 controls grain size in cultivated rice. Nat Genet. 2016;48:447–56.
43. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, et al. Control of grain size, shape and quality by OsSPL16 in rice. Nat Genet. 2012;44:950–4.
44. Ishikawa R, Aoki M, Kurotani K-I, Yokoi S, Shinomura T, Takano M, et al. Phytochrome B regulates Heading date 1 (Hd1)-mediated expression of rice florigen Hd3a and critical day length in rice. Mol Genet Genomics. 2011;285:461–70.
45. Yan W-H, Wang P, Chen H-X, Zhou H-J, Li Q-P, Wang C-R, et al. A Major QTL, Ghd8, Plays Pleiotropic Roles in Regulating Grain Productivity, Plant Height, and Heading Date in Rice. Molecular Plant. 2011;4:319–30.
46. Zhang B, Liu H, Qi F, Zhang Z, Li Q, Han Z, et al. Genetic Interactions Among Ghd7, Ghd8, OsPRR37 and Hd1 Contribute to Large Variation in Heading Date in Rice. Rice. 2019;12:48.
47. Wu B, Hu W, Ayaad M, Liu H, Xing Y. Intragenic recombination between two non-functional semi-dwarf 1 alleles produced a functional SD1 allele in a tall recombinant inbred line in rice. PLOS ONE. 2017;12:e0190116.
48. Zhou H, Zhou M, Yang Y, Li J, Zhu L, Jiang D, et al. RNase Z S1 processes Ub L40 mRNAs and controls thermosensitive genic male sterility in rice. Nature Communications. 2014;5:4884.
49. Ning X, Yunyu W, Aihong L. Strategy for Use of Rice Blast Resistance Genes in Rice Molecular Breeding. Rice Science. 2020;27:263–77.
50. Guo J, Xu C, Wu D, Zhao Y, Qiu Y, Wang X, et al. Bph6 encodes an exocyst - localized protein and confers broad resistance to planthoppers in rice. Nature Genetics. 2018;50:297–306.
51. Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, et al. Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond. Ann Bot. 2009;103:151–60.
52. Hattori Y, Nagai K, Furukawa S, Song X-J, Kawano R, Sakakibara H, et al. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature. 2009;460:1026–30.

53. Li X-M, Chao D-Y, Wu Y, Huang X, Chen K, Cui L-G, et al. Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. Nature Genetics. 2015;47:827–33.
54. Hu B, Wang W, Ou S, Tang J, Li H, Che R, et al. Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. Nature Genetics. 2015;47:834–8.
55. Uga Y, Okuno K, Yano M. Dro1, a major QTL involved in deep rooting of rice under upland field conditions. J Exp Bot. 2011;62:2485–94.
56. Li Y, Fan C, Xing Y, Yun P, Luo L, Yan B, et al. Chalk5 encodes a vacuolar H + -translocating pyrophosphatase influencing grain chalkiness in rice. Nature Genetics. 2014;46:398–404.
57. Zhang C, Zhu J, Chen S, Fan X, Li Q, Lu Y, et al. Wxlv, the Ancestral Allele of Rice Waxy Gene. Molecular Plant. 2019;12:1157–66.
58. Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y, et al. ALK, the Key Gene for Gelatinization Temperature, is a Modifier Gene for Gel Consistency in RiceF. Journal of Integrative Plant Biology. 2011;53:756–65.
59. Kovach MJ, Sweeney MT, McCouch SR. New insights into the history of rice domestication. Trends Genet. 2007;23:578–87.
60. Sang T, Ge S. Genetics and phylogenetics of rice domestication. Current Opinion in Genetics & Development. 2007;17:533–8.
61. Fuller DQ, Sato Y-I, Castillo C, Qin L, Weisskopf AR, Kingwell-Banham EJ, et al. Consilience of genetics and archaeobotany in the entangled history of rice. Archaeol Anthropol Sci. 2010;2:115–31.
62. Li C, Zhou A, Sang T. Rice Domestication by Reducing Shattering. Science. 2006;311:1936–9.
63. Jin J, Huang W, Gao J-P, Yang J, Shi M, Zhu M-Z, et al. Genetic control of rice plant architecture under domestication. Nature Genetics. 2008;40:1365–9.
64. Viana VE, Pegoraro C, Busanello C, Costa de Oliveira A. Mutagenesis in Rice: The Basis for Breeding a New Super Plant. Front Plant Sci. 2019;10. doi:3389/fpls.2019.01326.
65. Alvarez A, Fuentes JL, Puldón V, Gómez PJ, Mora L, Duque MC, et al. Genetic diversity analysis of Cuban traditional rice (Oryza sativa L.) varieties based on microsatellite markers. Genetics and Molecular Biology. 2007;30:1109–17.
66. Alam MdA, Juraimi AS, Rafii MY, Hamid AA, Arolu IW, Latif MA. Application of EST-SSR marker in detection of genetic variation among purslane (Portulaca oleracea L.) accessions. Braz J Bot. 2015;38:119–29.
67. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun. 2011;2:467.
68. Li H, Rasheed A, Hickey LT, He Z. Fast-Forwarding Genetic Gain. Trends in Plant Science. 2018;23:184–6.
69. Wang J, Zhou L, Shi H, Chern M, Yu H, Yi H, et al. A single transcription factor promotes both yield and immunity in rice. Science. 2018;361:1026–8.
70. Zhang Q. Strategies for developing Green Super Rice. Proc Natl Acad Sci U S A. 2007;104:16402–9.

71. Ali J, Jewel ZA, Mahender A, Anandan A, Hernandez J, Li Z. Molecular Genetics and Breeding for Nutrient Use Efficiency in Rice. Int J Mol Sci. 2018;19. doi:3390/ijms19061762.
72. Wing RA, Purugganan MD, Zhang Q. The rice genome revolution: from an ancient grain to Green Super Rice. Nature Reviews Genetics. 2018;19:505–17.
73. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci. 2017;22:961–75.
74. Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nature Communications. 2017;8:14261.
75. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. https://arxiv.org/abs/1303.3997v2. Accessed 26 Aug 2020.
76. Ga V der A, Mo C, C H, R P, G DA, A L-M, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1-11.10.33.
77. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–5.
78. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633–5.
79. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44 Web Server issue:W242–5.
80. CATCHEN J, HOHENLOHE PA, BASSHAM S, AMORES A, CRESKO WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22:3124–40.
81. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38:1358–70.
82. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. PNAS. 1979;76:5269–73.
83. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156.

Figures

Fig. 1

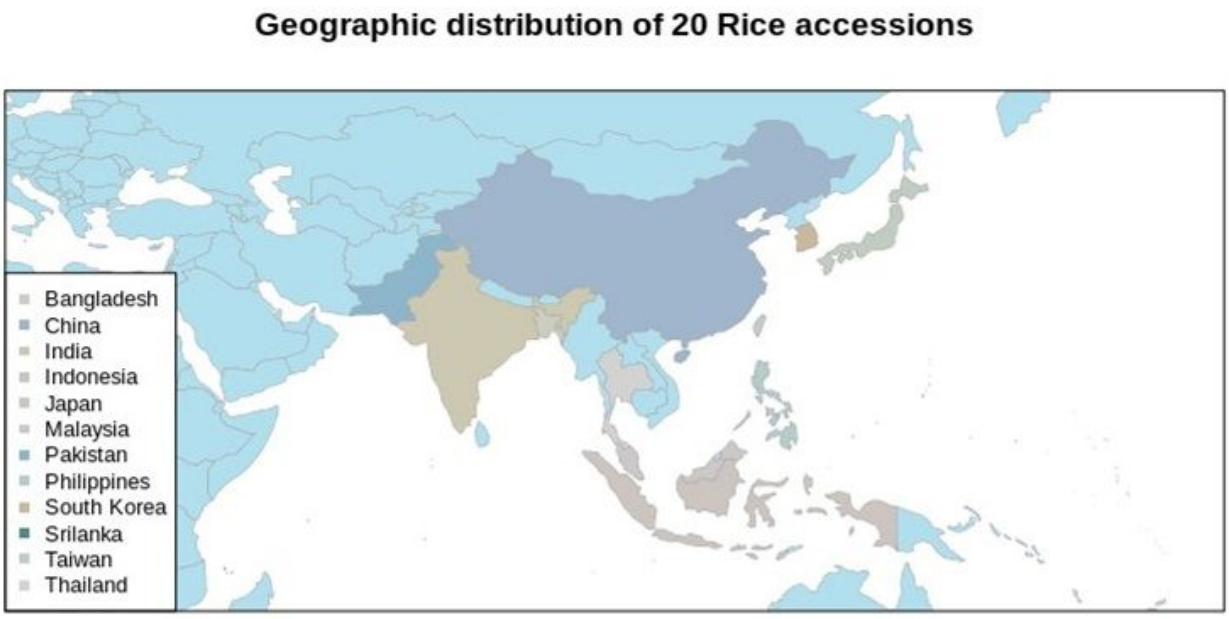

Figure 1

Geographical distribution of 20 rice accession across the world. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Fig. 2

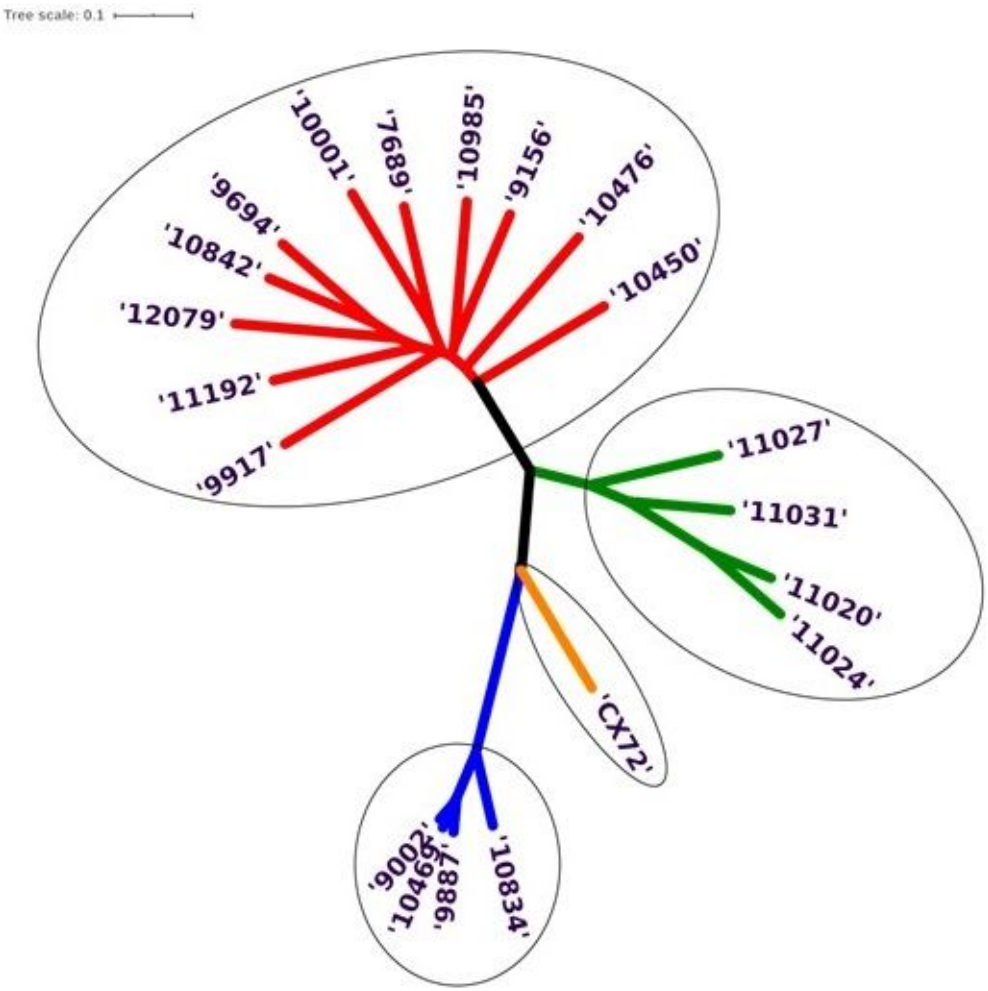

Figure 2

Population structure analysis. The NJ tree built using 3.8 million SNPs. Branch color depicts major groups of Oryza sativa. Red color shows indica cultivars, the blue color represents japonica varieties, green color depicts aus group, and a single orange branch demonstrates an admixed type.

Fig. 3

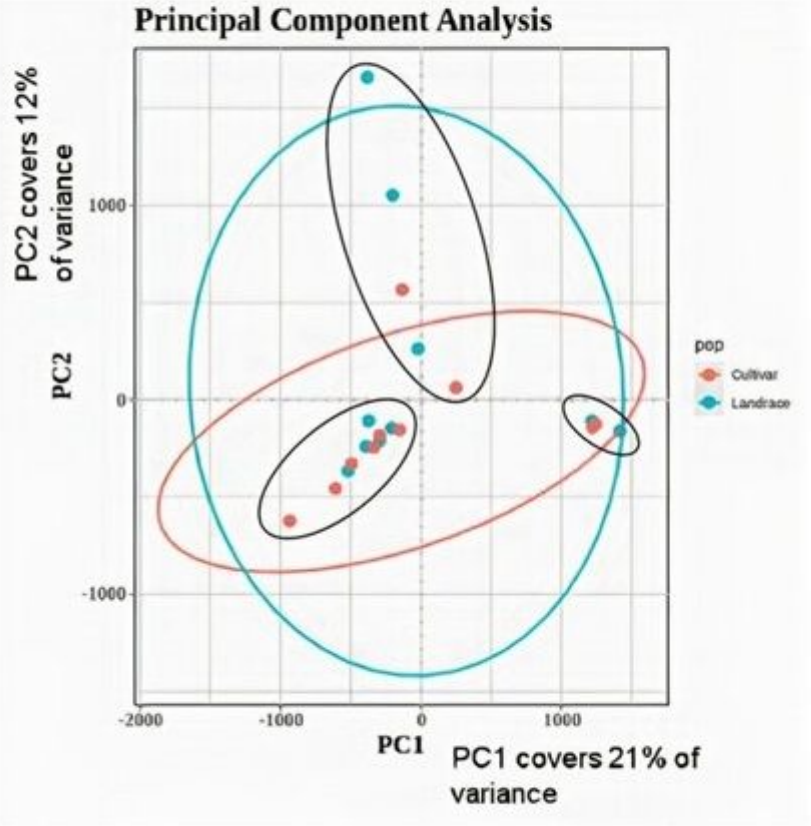

Figure 3

Principal component analysis plot of the first two components of 20 rice accessions. The black circle demonstrates the three main clusters. The dot's color indicates the population group like red color exhibits cultivars, and green color represents landraces. The size of red and green circles shows diversity within the cultivars and landraces population, respectively. PC1 represents a maximum variance of 21%, while PC2 explains a 12% variance.

Fig. 4

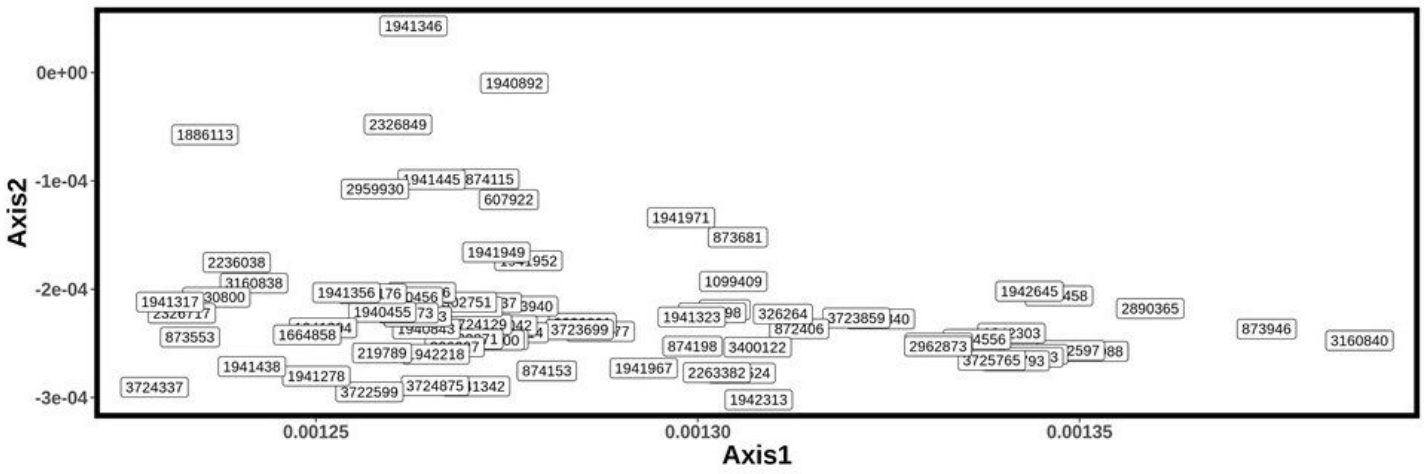

Figure 4

The first 1000 SNPs' PCA plot. Plot shows the highest values across the first coordinate and accounts for the highest variance in the population.

Fig. 5

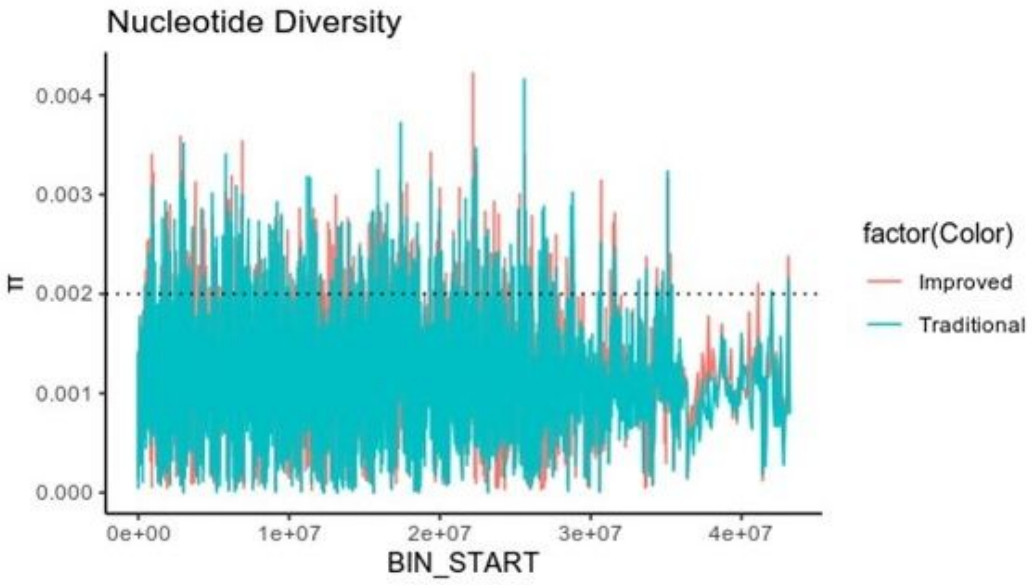

Figure 5

Nucleotide diversity (π landraces/π improved cultivars) by sliding window analysis with a bin size of 100 kb. The X-axis contains the nucleotide position, and Y-axis shows π landraces/π improved cultivars values. The green color represents the landraces, while the red color indicates the improved cultivars.

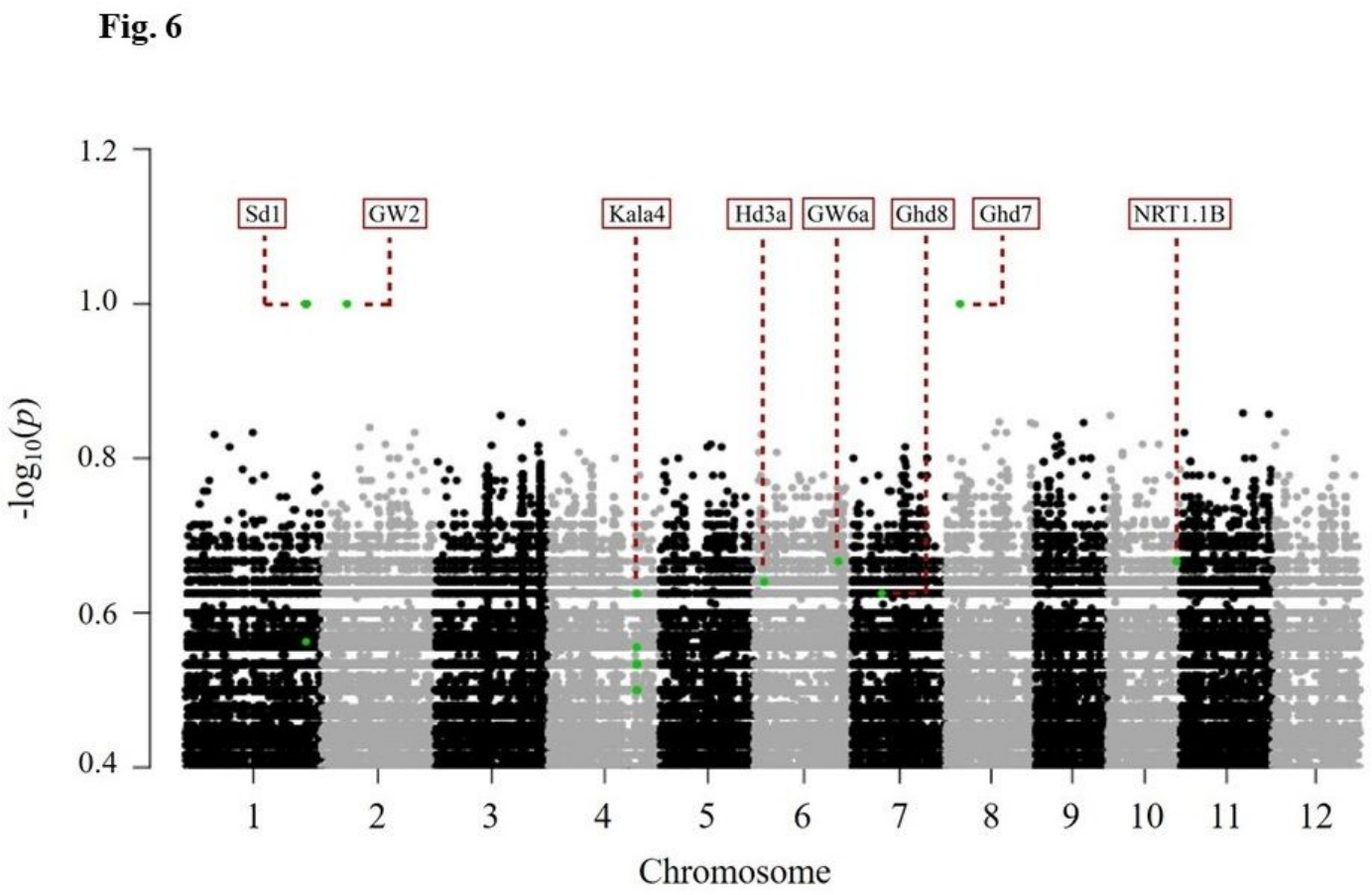

Figure 6

Genome-wide pairwise FST values plotted against 12 chromosomes. Pairwise FST values show a higher differentiation between the populations. The X-axis represents the chromosomes, while Y-axis demonstrates the differentiation value between the modern cultivars and landraces. Green dots indicate the known regions related to domestication and improvement regions.