IMAM KHOMEINI INTERNATIONAL UNIVERSITY
IJGPB
Iranian Journal of Genetics and Plant Breeding
Print ISSN : 2251-9610
Online ISSN : 2676-346X

Fundamentals, applications, and challenges of the 3K rice genomes project

Mojtaba Kordrostami[1]*, Ali Akbar Ghasemi-Soloklui[1], Mehdi Rahimi[2]

[1]Nuclear Science and Technology Research Institute (NSTRI), Nuclear Agriculture Research School, Karaj, Iran.
[2]Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran.
**Corresponding author, 0000-0003-0314-1354. Email: mkordrostami@aeoi.org.ir.*

ABSTRACT INFO

Review Paper

Received: 08 Sep 2022

Accepted: 29 Apr 2023

ABSTRACT

For more than half of the world's population, rice (*Oryza sativa* L.) is the primary source of nutrition. Rice production must expand by at least 25 percent by 2030 to feed the world's ever-increasing population. As climate change and arable land loss take a tremendous toll on the world's food supply, genetic advancements in rice would be crucial for alleviating the yield gap. We must first obtain extensive information on the genetic diversity of the *Oryza* spp. gene pool, the association between diverse alleles and critical rice characteristics, and the systematic exploitation of the rich genetic diversity using approaches that employ expertise in rice breeding procedures. The vast genetic diversity of rice cannot be represented by a single genome. To produce variants that are more tolerant to adverse weather conditions, a multi-national rice genome sequencing project was launched on May 28, 2014. Genes associated with drought tolerance, disease resistance, and pest resistance in rice could be identified using the above mentioned genetic information. Using diverse germplasm resources and high-throughput genome sequencing projects, rice genomics has made great progress toward applying basic research advances to understanding agronomic traits. It is important to remember that many important genes are missing from the previous sequencing projects, and many useful genes are present in native and traditional populations that cannot be retrieved without gene sequencing.

Key words: 3K rice genomes project, Genetic structure, Genetic variation, Structural variants.

How to cite this article:
Kordrostami M., Ghasemi-Soloklui A. A., and Rahimi M. (2021). Fundamentals, applications, and challenges of the 3K rice genomes project. *Iranian Journal of Genetics and Plant Breeding,* 10(2): 57-69.
DOI: 10.30479/IJGPB.2023.17775.1328

INTRODUCTION

It is estimated that rice accounts for 30% of the calories in the diets of the poorest people in the world (Thangadurai *et al*., 2020). However, a growing human population means a steady demand for foods such as rice (Seck *et al*., 2012; Kordrostami *et al*., 2017). It is predicted that rice productivity will increase by at least 25% during the next decade (Islam and Karim, 2019). The green revolution led to the usage of plant genetic diversity by rice breeders to improve grain yield in addition to selective breeding (Anwar *et al.,* 2022). Research is also being conducted to find methods for reducing the size of farms (Kordrostami and Mafakheri, 2021). Using less land and water can increase their efficiency and resilience to extreme environmental pressures induced by climate change (Mafakheri and Kordrostami, 2020; Shahzad *et al.,* 2021; Rivero *et al.,* 2022). Cereals must continue to grow by undergoing genetic changes that will allow them to increase yield, quality, and capacity flexibly (3000 Rice Genomes Project, 2014). To begin this process, accurate information about the gene pool of the genus *Oryza* needs to be obtained, correlations between alleles and rice traits must be established, and the rich genetic diversity obtainable must be used (Nie and Peng, 2017).

Next-generation sequencing (NGS) technologies have revolutionized the field of rice genomics, enabling the rapid and cost-effective sequencing of large genomes and the identification of genetic variation within and between different rice varieties (Bommisetty *et al.,* 2020). NGS technologies first emerged in the mid-2000s, around the same time that the first rice genome was sequenced (Wang *et al.,* 2019). The International Rice Genome Sequencing Project, launched in 1998, was a collaboration between researchers in several countries to sequence and analyze the genome of the *Japonica* rice variety Nipponbare (Purugganan and Jackson, 2021). The project relied on Sanger sequencing, a laborious and expensive method that involves the sequencing of individual DNA fragments one at a time (Wei and Huang, 2019). The completion of the first rice genome sequence in 2005 marked a major milestone in rice genomics, providing a reference genome for subsequent studies of rice biology and genetics (Olsen, 2022). However, the high cost and slow pace of Sanger sequencing limited the ability of researchers to sequence large numbers of rice genomes and identify genetic variation across different varieties (Pervez *et al.,* 2022).

NGS technologies, including Illumina sequencing and other platforms, have since enabled the rapid and cost-effective sequencing of large numbers of genomes, including those of rice (Kumar *et al.,* 2021). This has led to the development of projects such as the 3000 Rice Genomes Project, which aims to sequence and analyze the genomes of 3,000 rice varieties to better understand rice diversity and improve breeding efforts (Lu *et al.,* 2021). The application of NGS technologies has also enabled the identification of genetic variation underlying important traits in rice, such as disease resistance, yield, and quality (Table 1) (Salgotra and Chauhan, 2023). This information can be used to develop new rice varieties with improved characteristics, contributing to global food security (Lenaerts *et al.,* 2019). Overall, the development and application of NGS technologies have greatly accelerated progress in rice genomics, enabling researchers to explore the genetic basis of rice diversity, biology, and agronomic traits in unprecedented detail (Gu *et al.,* 2022).

The 3000 Rice Genomes Project is a remarkable achievement in the field of crop genomics, and its impact on rice breeding and food security cannot be overstated. By sequencing the genomes of 3,000 rice varieties, the project has provided a comprehensive view of the genetic diversity of this essential crop, and has identified numerous genetic variations that can be used to improve rice breeding programs (Gupta *et al.,* 2021). One of the project's major achievements is the development of the Rice SNP-Seek Database, which provides a user-friendly interface for researchers to explore the genetic variations and their distribution across the rice genomes (Katara *et al.,* 2021). This database has become an invaluable resource for rice researchers and breeders, allowing them to develop new rice varieties with improved traits, such as disease resistance, drought tolerance, and higher yields (Paul, 2020). The project has also provided insights into the genetic basis of rice adaptation and evolution, which can inform conservation efforts and enhance our understanding of the history and diversity of this essential crop. By identifying the genetic variations that have allowed rice to adapt to different environments and stress conditions, the project has paved the way for the development of new rice varieties that can thrive under changing climatic conditions (Sakhale *et al.,* 2023). Moreover, the project's collaborative and open approach has fostered international cooperation in rice research and has set a precedent for large-scale sequencing projects for other crops. The project's success demonstrates the potential of genomics and data sharing to drive progress in agricultural research

Table 1. Some key events and milestones in the history of rice genomics (Junliang and Dilin, 2022).

Year	Event
1986	The first rice genome map was created using restriction fragment length polymorphism (RFLP) markers.
1991	The first rice microsatellite (simple sequence repeat, SSR) markers were developed.
1994	The International Rice Genome Sequencing Project (IRGSP) was initiated by Japan.
1997	The first physical map of the rice genome was created.
1999	The first draft sequence of the rice genome was published.
2002	The IRGSP completed the sequencing of the rice genome.
2005	The Rice Annotation Project Database (RAP-DB) was established to provide annotation for the rice genome sequence.
2007	The first high-density rice SNP map was published.
2013	The 3,000 Rice Genomes Project was initiated to provide a comprehensive view of genetic variation in rice.
2015	The Rice Diversity Project was launched to study the genetic diversity of rice and its relationship to phenotype.
2018	The pan-genome of rice was published, revealing the genomic variation among different rice varieties.
2021	The Rice Genome Atlas was published, providing a comprehensive analysis of the functional elements of the rice genome.

and development (Fuentes *et al.,* 2019). Overall, the 3000 Rice Genomes Project represents a major milestone in rice research, providing a valuable resource for researchers, breeders, and policymakers to improve rice production and enhance global food security. The project's impact will continue to be felt for years to come, as its findings are applied to develop new rice varieties that can better meet the challenges of feeding a growing global population in a sustainable and equitable manner (Gupta *et al.,* 2021).

Rice genomics and the 3K Rice Genomes Project are closely related, as the latter is a large-scale research initiative focused on exploring the genetic diversity of rice at the whole-genome level (Song *et al.,* 2018). The 3K Rice Genomes Project aims to sequence the genomes of 3,000 diverse varieties of rice from around the world and use this information to gain insights into the genetic basis of rice traits, as well as its evolutionary history and diversity. The 3K Rice Genomes Project builds on the foundation laid by the International Rice Genome Sequencing Project (IRGSP), which completed the sequencing of the rice genome in 2002 (Gupta *et al.,* 2021). While the IRGSP focused on the reference genome of a single *Japonica* variety of rice, the 3K Rice Genomes Project aims to sequence the genomes of 3,000 diverse varieties of rice, including both *Japonica* and *Indica* varieties, as well as wild and cultivated varieties from different regions of the world. By sequencing the genomes of these diverse rice varieties, researchers can identify genetic variations that are associated with desirable traits, such as higher yields, disease resistance, and tolerance to environmental stresses. This information can then be used to develop new rice varieties with improved traits through breeding and genome editing. Moreover, the 3K Rice Genomes Project provides a comprehensive dataset that can be used to study the evolutionary history and diversity of rice (Gu *et al.,* 2022). This information is essential for understanding the genetic basis of adaptation to different environments and for the conservation of genetic resources. Overall, rice genomics and the 3K Rice Genomes Project are critical for improving our understanding of the genetics of rice and for developing new varieties that can help address global food security challenges.

What will be implemented in this project, and what the current status is?

The sequencing of 3,000 rice genomes was completed in this project, with an average depth of 14×. Several mega-varieties of rice (Table 2) were included in this study, including those that grow in large areas of different ecosystems throughout Asia (Yu *et al.,* 2003). Genetic mapping populations and popular rice cultivar parental lines were also examined (Li and Zhang, 2013). Rice researchers have unprecedented access to the project's sequencing data, housed in GigaScience Journal GigaDB (3000 Rice Genomes Project, 2014). We will be able to find genetic diversity related to crucial characteristics for breeding thanks to the massive quantity of data offered by this repository, which

Table 2. Some key information about the 3,000 rice genomes (Gupta *et al.,* 2021).

Project name	3,000 rice genomes project
Objective	Analyze the genetic diversity of rice
Number of rice genomes sequenced	3,000
Species of rice	*Oryza sativa* (cultivated rice)
Geographic origin of rice samples	Worldwide
Project duration	2011-2018
Funding source(s)	Bill and Melinda Gates Foundation, Chinese Academy of Agricultural Sciences, and others
Key findings	Identification of genetic markers associated with important traits, such as disease resistance and yield potential; insights into the genetic history and evolution of rice; development of genomic resources for rice breeding and improvement

will help us better understand rice selection history (natural or artificial) (Li and Zhang, 2013). Scientists are excited about this project's potential impact, but integrating phenotypic, genetic, and genomic data from various rice sources poses many challenges. Here are some potential challenges (Kumawat *et al.,* 2022; Rana *et al.,* 2020):

1. Data heterogeneity: Rice data can be collected from various sources, such as field experiments, greenhouse experiments, and genotyping platforms. These data can differ in terms of quality, format, and measurement standards, which can make it difficult to integrate them into a single analysis.
2. Missing data: Phenotypic, genetic, and genomic data can sometimes have missing values due to experimental errors, technical limitations, or other reasons. Integrating data with missing values can result in biased or incomplete analysis results.
3. Data volume: With the advent of high-throughput sequencing and phenotyping technologies, large amounts of rice data can be generated in a short period of time. Integrating such large amounts of data can pose computational challenges, as well as challenges in terms of data storage and management.
4. Data standardization: Phenotypic and genomic data can be collected using different methods and standards across different rice sources, which can make it challenging to integrate and compare data across studies.
5. Data interpretation: Integrating phenotypic, genetic, and genomic data can result in complex data structures and relationships, which can make data interpretation and inference challenging.

Addressing these challenges requires careful data curation, standardization, and integration using appropriate statistical and computational methods. Collaboration between researchers and data providers can also help to ensure the quality and accuracy of the integrated data. Information portals for rice diversity, as well as large-scale gene/trait identification and allele extraction, are also required (McNally *et al.,* 2014) and hopefully provide strategies to preserve rice genetic resources (McCouch *et al.,* 2012). There are several steps needed for the 3000 Rice Genome Project to be practical: 1) Identification of local and global population types; 2) Construction of novel high-quality and advanced reference genomes that represent the main groups; 3) Analysis of linkage disequilibrium and recombination breakpoints for the development of haplotype maps; 4) the detection within and between populations of indels, structural variants, and single nucleotide polymorphisms (Nie and Peng, 2017; Ramadas, 2018; Palanisami *et al.,* 2019). The IRRI and CAAS work together in phenotyping to improve plant growth and resistance to abiotic and biotic stresses, grain quality-related traits, and grain yield-related traits. The ability to link phenotypes with sequence information is greatly enhanced by high-performance phenomics that utilizes image capture and sensor data from controlled environments and field-based substrates (Kumar *et al.,* 2015). The results of this combined effort provide deeper insights and broader applications than previously reported (Roitsch *et al.,* 2019). This project is expected to provide useful

information even before phenotypic data are available, and more extensive sampling with more sequence depth will be guided by population structure analysis (Kordrostami and Rabiei, 2019; Kordrostami and Rahimi, 2015). More SNPs need to be discovered in the pan-rice genome as more high-quality reference genomes are developed. High-density maps may be employed to find genes and extract alleles (Yu *et al.,* 2014). As a result, new population genotyping arrays are created that are useful for genetic and breeding purposes. Secondly, it is essential to reveal the demographic structures formed by evolution, domestication, and selection. By identifying and analyzing unique genomic structural groups in the rice genome, we can better understand their contribution to the grouping of previously identified varieties in rice (Kordrostami *et al.,* 2017). We can also use the lines used in mapping and breeding programs to validate genes directly to improve breeding populations' traits by including them in gene validation. Rice breeding programs can benefit from implementing, testing, and developing novel breeding methods such as recurrent and genomic selection based on these replicated lines (Yu *et al.,* 2003).

Rice breeding is fundamental to its success due to its exceptional intra-species and intra-genus genetic diversity (Rabiei *et al.,* 2015). A wealth of genetic diversity can be found in over 230,000 *Oryza* germplasms stored in gene banks worldwide (mainly from Asia) (Kebriyaee *et al.,* 2012). Research into rice's genetic variation at the DNA sequence level has been a pipe dream for rice scientists up until very recently. However, this dream has become a reality with the launch of the 3kR genome project (Wing *et al.,* 2018). The 3K Rice (3KR) genome research is a significant step in showing the wide range of rice germplasm worldwide. Substantial resources must be committed to assessing plant performance under various situations for this ambitious effort to be helpful outside the scientific community. Understanding plant biology may be improved by identifying phenotype-genotype connections and using data management techniques.

The 3K Rice Genomes Project is a large-scale research initiative aimed at sequencing and analyzing the genomes of 3,000 diverse varieties of rice from around the world. The project seeks to identify genetic variations that are associated with desirable traits, such as higher yields, disease resistance, and tolerance to environmental stresses, and use this information to develop new rice varieties with improved traits (Lu *et al.,* 2021).

To achieve this goal, the 3K Rice Genomes Project involves several key steps, including (Sun *et al.,* 2017; Fuentes *et al.,* 2019; Gupta *et al.,* 2021):

1. Sample collection: The project involves the collection of diverse rice varieties from different regions of the world, including both *Japonica* and *Indica* varieties, as well as wild and cultivated varieties.
2. DNA sequencing: The collected rice samples are then subjected to high-throughput DNA sequencing, which involves reading the genetic code of each sample and generating massive amounts of genomic data.
3. Genome assembly: The raw genomic data generated through DNA sequencing are then assembled into complete genomes for each of the 3,000 rice varieties.
4. Genome annotation: Once the genomes are assembled, the project involves the annotation of the genomes, which involves identifying the function and location of genes and other important genomic features.
5. Data analysis: The annotated genomes are then analyzed to identify genetic variations that are associated with desirable traits, such as higher yields, disease resistance, and tolerance to environmental stresses.
6. Trait identification: The identified genetic variations are then used to develop new rice varieties with improved traits through breeding and genome editing.

Overall, the 3K Rice Genomes Project is a comprehensive effort to understand the genetic diversity of rice and develop new rice varieties that can help address global food security challenges (Table 3). According to the 3K Rice Genomes Project website, the project had sequenced over 4,000 rice genomes by September 2021, exceeding its original goal of sequencing 3,000 genomes. The project has produced a large dataset of genomic information, including data on genetic variation, gene expression, and epigenetic modifications, which is publicly available to researchers around the world. The project has also resulted in several publications in high-impact scientific journals, reporting new insights into the genetic basis of rice traits, such as yield, quality, and adaptation to different environments. For example, one study published in Nature Genetics in 2020 identified genetic variations associated with drought tolerance

Table 3. The achievements of the 3,000 Rice Genomes Project (Gupta *et al.,* 2021; Jia *et al.,* 2021).

Achievement	Description
Genetic diversity analysis	The project analyzed the genetic diversity of rice by sequencing the genomes of 3,000 rice varieties from around the world, allowing researchers to understand the genetic basis of important traits and to identify useful genetic variations for breeding
Identification of genetic markers	The project identified genetic markers associated with important traits, such as disease resistance, yield potential, and grain quality, providing valuable information for rice breeding and improvement
Development of genomic resources	The project generated a wealth of genomic resources for rice research, including a pan-genome for rice, a high-quality genome assembly for the variety Minghui 63, and a dataset of genetic variations in rice
Insights into rice evolution	The project shed light on the evolutionary history of rice by revealing the relationships between different rice varieties and identifying genomic regions associated with domestication and adaptation
Impact on rice breeding	The project's genomic resources and genetic markers have already been used in rice breeding programs around the world to develop new rice varieties with improved traits, such as disease resistance and drought tolerance

in rice, providing new targets for breeding drought-tolerant varieties (Gupta *et al.,* 2021).

Sequencing platform and germplasm used for the 3000 Rice (3KR) Genome Project

The 3,000 Rice (3KR) Genome Project utilized a variety of sequencing platforms and germplasm to generate a comprehensive dataset of genomic information for rice. The project used multiple sequencing platforms, including Illumina, PacBio, and Oxford Nanopore technologies, to sequence the genomes of 3,010 diverse varieties of rice. The use of different sequencing platforms allowed for the generation of high-quality genome assemblies with different levels of completeness and accuracy (Gu *et al.,* 2022; Kumawat *et al.,* 2022).

The germplasm used in the project included both cultivated and wild varieties of rice, representing the diversity of rice from different regions of the world (Figure 1). The germplasm included both *Indica* and *Japonica* sub-species, as well as a number of wild and hybrid varieties (Gupta *et al.,* 2021).

The project also used bioinformatics tools and pipelines to analyze the genomic data generated from the different sequencing platforms and germplasm. This allowed researchers to identify genetic variations, including single nucleotide polymorphisms (SNPs), insertions/deletions (indels), and structural variations (SVs), which were then used to study the genetic basis of rice traits and diversity (Fuentes *et al.,* 2019). The use of multiple sequencing platforms and diverse germplasm in the 3KR Genome Project allowed for the generation of a comprehensive dataset of genomic information for rice, providing new insights into the genetics of this important crop and its potential for improving global food security.

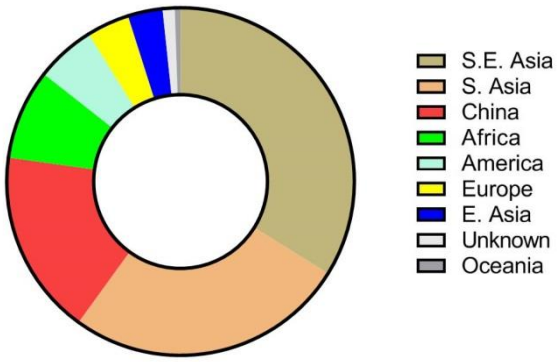

Figure 1. The geographic distribution of the 3000 rice genome project's genotypes (Source: 3000 rice genome 2014).

The size of genomes, genetic diversity, and population structure

Through biotechnology, rice breeding aims to increase yield and improve grain quality characteristics (Kordrostami *et al.,* 2021). Rice breeding programs have made significant progress and accomplishments (Thangadurai *et al.,* 2020). Biotechnology has the potential to significantly enhance rice breeding efforts, by allowing scientists to manipulate the genetic material of rice plants in ways that would be impossible using traditional breeding methods alone (Ahmar *et al.,* 2020). One biotechnology tool that is

commonly used in rice breeding programs is genetic engineering. This involves introducing specific genes into rice plants in order to confer desirable traits such as resistance to pests or tolerance to herbicides. Genetic engineering can also be used to enhance the nutritional content of rice, for example by increasing the amount of vitamin A in so-called "golden rice." (Hasan *et al.,* 2021). Another biotechnology tool that is increasingly being used in rice breeding is genomics. Genomics involves analyzing the entire DNA sequence of an organism in order to identify genes that are associated with particular traits. By comparing the genomes of different rice varieties, scientists can identify the genetic factors that underlie desirable traits such as yield, disease resistance, and nutritional quality (Adlak *et al.,* 2019). This knowledge can then be used to develop new rice varieties with these traits. In addition to genetic engineering and genomics, biotechnology tools such as marker-assisted selection (MAS) and genomic selection (GS) are also being used in rice breeding programs. MAS involves identifying genetic markers that are associated with a particular trait, and then using these markers to selectively breed rice plants that have the desired trait. GS takes this approach a step further by using complex statistical models to predict the performance of rice plants based on their genetic makeup (Anilkumar *et al.,* 2022). Overall, the use of biotechnology in rice breeding programs has the potential to significantly enhance our ability to develop new rice varieties with improved yield, disease resistance, and nutritional quality. However, it is important to ensure that these technologies are used responsibly and with appropriate safeguards in place to protect human health and the environment. This has resulted in the development and distribution a diversified and novel rice variety that produces greater yields and higher quality like size and shape, whiteness, long and thin uncooked grains (Tu Anh *et al.,* 2018; Gouda *et al.,* 2020).

Genetic diversity is essential to rice breeding to meet current nutritional needs. In previous studies, the more genetic diversity in the population, the more valuable it was to breeding programs (Tu Anh *et al.,* 2018). It is possible to detect genetic diversity in plants using DNA markers and genetic engineering (Ram *et al.,* 2007). As well as distinguishing between individuals and accessions at the molecular level, they can determine the properties of new germplasms, which can be used to breed plants (Edzesi *et al.,* 2016; Gouda *et al.,* 2020).

The 3000 Rice Genome Project collected rice samples from a wide range of geographic regions to ensure that the genetic diversity of rice was well represented in the dataset. The samples were collected from 89 countries and regions around the world, including Asia, Africa, North and South America, Europe, and Oceania. Within Asia, which is the center of origin and diversity of rice, the samples were collected from countries such as China, India, Vietnam, Indonesia, and the Philippines, among others. The project also included samples from other regions where rice is cultivated, such as Africa (including Madagascar, Nigeria, and Tanzania), Latin America (including Brazil, Colombia, and Costa Rica), and Oceania (including Australia and Papua New Guinea). Overall, the geographic distribution of the 3000 Rice Genome Project's genotypes was designed to capture the full diversity of rice, including both cultivated and wild varieties, from different regions of the world. This allowed researchers to study the genetic basis of important traits such as yield, disease resistance, and adaptation to different environments, and to identify new targets for rice breeding and improvement (Song *et al.,* 2018; Wang *et al.,* 2018; Gutaker *et al.,* 2020).

It has been found that the 3,000 rice genome study (Li *et al.,* 2014) has identified 3,000 individuals as belonging to the *Indica* and *Japonica* varieties, as well as two smaller kinds, aus/boro and basmati/Sadri, and an additional population of 134 genotypes of mixed forms. *Indica* group with 1760 members (58.2%) was the largest group in five different subgroups with different genetic backgrounds. Among 843 populations (27.9%), 388 were in the temperate *Japonica* subgroup, and 455 were in the tropical *Japonica* subgroup. A total of 215 populations belong to the aus/boro group, mainly related to the *Indica* group, whereas 68 populations belong to the aromatic basmati/Sadri group, predominantly from South Asia (Li *et al.,* 2014). According to Wang *et al.* (2018) further research on population structure and diversity of genotypes of the Rice Genome Project led to nine main subpopulations, most of which have similar geographical origins. The XI clusters were divided into four (XI-1A from East Asia, XI-1B from modern improved varieties with diverse roots, XI-2 from South Asia, and XI-3 from Southeast Asia). GJ consists of three groups (mainly temperate East Asian cultivars (called GJ-tmp), subtropical Southeast Asian cultivars (called GJ-sbtrp), and southeastern tropical Asia (called GJ-trp)), and individual groups for accessions mainly derived from South Asia cA and cB. It was determined that groups in clusters XI and GJ that had fewer than 0.65 mixed individuals were referred to as "XI-adm" and "GJ-adm," respectively (Wang *et al.,* 2018).

According to Wang et al. (2018), the role of genome design in the 3K Rice (3KR) genome project was, on average, 92% (74.6-98.7%) through the alignment of the 3K Rice (3KR) genome with the Nipponbare reference genome. Over 29 million SNPs were also found in the study, almost all of which were diallelic. Using filtering to reduce the data set to 17 million SNPs, most SNPs (more than 99.9%) had a minor allelic frequency (MAF) of more than 25%. Many large effect NPs were present in nearly all transposable elements (91%) and half of the non-transposable elements (56%). SNPs with a minimum allelic frequency (MAF) greater than 10% show the types of demographic and adaptation events. Among the subpopulations of cA and cB, private alleles (alleles occurring in only one population out of a range of populations) were more prevalent. As a result of continuing gene transmission via natural hybridization and breeding, the IX populations had fewer "private alleles" than other populations. This investigation identified a similar doubleton sharing pattern across subpopulations and within them (Wang *et al.,* 2018).

Furthermore, the link disequilibrium decay rates appear to be higher for hybrid subpopulations in XI compared to GJ, with slight variations between the two GJ subpopulations. All nine subpopulations, however, had different link disequilibrium decay rates while the link disequilibrium decay rates were much higher in the XI-3 and XI-2 subpopulations than in I-1B and IX-1A. Studying the number of genes presents in parts of the genome where gene diversity is limited and where these genes are subject to little restriction was the goal of this investigation. Across all subpopulations, the *Sh419* gene, which controls non-shattering, showed similar diversity trends, indicating a much longer selection process for this gene than the *qSH120* gene (Konishi *et al.,* 2006). The *sd121* gene locus showed a reduction in genetic diversity in all major branches of the phylogenetic tree, similar to the *qSH1* gene (Wang *et al.,* 2018). In groups XI, cA, and cB, there was more diversity in the 100 kb zones, while GJ groups had reduced diversity, which indicates a past associated with the "green revolution". Other significant gene loci also showed a reduction in diversity. The *Wx23* gene locus (Wang *et al.,* 1995), which affects amylose content and viscosity in cooking, and the *Badh*2.1 gene locus, which affects aroma, in classes; XI, cA, and cB, were highly diverse, indicating complex genetic backgrounds (Wang *et al.,* 2018). In all clusters studied, the diversity of Rc25 loci (Sweeney *et al.,* 2006) was very low.

Overall, the 3K RGP has provided a more comprehensive view of rice diversity and population structure than was previously available, allowing for new insights into the genetic basis of important traits and the potential for rice improvement and breeding. The project's large dataset of genomic information is publicly available and continues to be a valuable resource for researchers worldwide. The 3000 Rice Genome Project (3K RGP) is expected to add significantly to our understanding of rice genetics, evolution, and diversity. Here are some potential scientific benefits of the project (Sun *et al.,* 2017; Fuentes *et al.,* 2019; Gupta *et al.,* 2021):

1. Better crop improvement: The 3K RGP will provide a comprehensive database of rice genetic variations, including rare and novel alleles. This information can be used to develop new rice varieties with improved yield, quality, and resilience to biotic and abiotic stresses.
2. Comparative genomics: The 3K RGP will enable comparative analysis of rice genomes across different ecotypes, geographic regions, and breeding programs. This will help identify conserved genomic regions, evolutionary trends, and functional genetic elements that are critical for rice adaptation and domestication.
3. Functional genomics: The 3K RGP will facilitate functional genomics studies, such as gene expression profiling, epigenetic analysis, and gene editing. This will help uncover the regulatory mechanisms of rice development, metabolism, and stress response, and provide potential targets for genetic engineering.
4. Biogeography and conservation: The 3K RGP will shed light on the biogeography and conservation of rice germplasm, especially in regions with high rice biodiversity and endemism. This will help develop strategies for preserving and utilizing rice genetic resources, and promote sustainable agriculture and food security.

Overall, the 3000 Rice Genome Project is expected to have a significant impact on rice research and crop improvement, as well as on our understanding of plant genetics and evolution.

Variations in structural features

Genetic differences can be identified through phenotypic diversity by detecting and analyzing variations between and within individuals. Various methods and technologies have been used to identify these genetic variations, and strategies based on next-generation sequencing technologies, due to their enhanced resolution and accuracy in identifying genetic diversity (Futschik and Schlötterer, 2010),

have been widely used in the past few years. The genomes of different organisms contain different types of genetic variations. Two classes of these variations, called Indels and SNPs to the genomes, are more predominant than the others (Mullaney *et al.,* 2010). Because SNPs are so widely distributed throughout the genome, this type of genotype variation is commonly used in various genomic studies, including genomic scans to identify loci affecting multiple genetic traits. Indels are secondary in density in the genome (Mills *et al.,* 2011). SVs, which have been discovered and characterized, have revolutionized how we think about genetic variation between species. Generally, a structural variant is defined as a distinct genomic variation (based on a reference genome) with a variable number of copies and a specified chromosomal location (Escaramís *et al.,* 2015; Fuentes *et al.,* 2019). The term structural variation is used to explain variations in base pairs in the genome rather than differences in the types of SNPs. In plants, structural variants have been studied less extensively. Even though structural variants are less common than single nucleotide polymorphisms (SNPs), their bigger size and potential to modify gene composition, dosage, or placement on chromosomes mean that they have a more significant impact on gene activity than SNPs (Fuentes *et al.,* 2019). When structural genetic variability in the human genome was discovered, various investigations were conducted to detect structural variants in some relevant species, including extinct creatures (Escaramís *et al.,* 2015; Fuentes *et al.,* 2019). Traditional structural variant identification through SNPs is slowing due to a shortage of high-quality reference genomes (Escaramís *et al.,* 2015; Fuentes *et al.,* 2019) and accurate methodologies, both of which are required to identify and genotype structural variations (Francia *et al.,* 2015). Corn is the first crop to have been thoroughly studied and found to have hundreds of structural variants. Plant research has previously shown that structural variations and phenotypic are linked (Żmieńko *et al.,* 2014). According to another study (Würschum *et al.,* 2015), late and early flowering are due to increased copying of *Vrn-A1* and *Ppd-B1* genes, respectively. Structure variant analysis has recently focused on only 453 sequences with mapping depths >15× and sequencing depths >20× because genome coverage was established only when sequencing depths greater than 20× were applied to the 3K Rice (3KR) genome project. There are 93683 structural variants, with 582 larger than 500 kb. In each genome, 12178 structural variants were identified. The structural variants in the XI - GJ groups were very well differentiated in the 3,000 rice genome project. Regarding structural variants, on average, the XI, cA, and cB groups had 14754, 12997, and 7892 differences from the Nipponbare reference genome (Wang *et al.,* 2018).

Challenges facing this project

Creating a comprehensive database and advanced tools for accelerating rice breeding is only the first step toward completing the sequencing and genome analysis of these 3,000 rice genotypes. There must be a long-term worldwide effort to integrate genetic diversity with phenotypic variation and environmental compatibility in functional rice genomics. Identifying rare genes/alleles that are highly influential is not the only approach to better understanding the genetic diversity of *O. sativa* but also to identifying new allelic compounds that underlie complex traits. Based on various reports, the 3000 Rice Genome Project (3K RGP) has faced several challenges throughout its implementation, including (Wing *et al.,* 2018; Kaur *et al.,* 2021):

1. Data analysis: One of the biggest challenges of the 3K RGP has been the analysis of the vast amounts of genomic data generated by the project. The analysis of such large datasets requires the development of new bioinformatics tools and pipelines, as well as significant computational resources. Additionally, the integration and comparison of data generated by different sequencing platforms and germplasm sources can be challenging.
2. Germplasm collection and maintenance: Collecting and maintaining the diverse germplasm used in the 3K RGP is a significant logistical challenge. Many of the germplasm sources are rare or endangered, and there can be legal and regulatory barriers to collecting and transporting them across international borders. Additionally, the long-term storage and maintenance of germplasm requires significant resources and infrastructure.
3. Quality control: Ensuring the quality and accuracy of the genomic data generated by the project is essential for its usefulness in downstream analyses. Quality control measures, such as the identification and removal of sequencing errors and the detection of mislabeled or contaminated samples, can be time-consuming and resource-intensive.
4. Ethical considerations: The use of genetic information, including that generated by the 3K RGP, raises ethical considerations related to privacy, informed consent, and the

equitable distribution of benefits. Addressing these considerations requires collaboration between researchers, policymakers, and local communities.

Despite these challenges, the 3K RGP has made significant progress in advancing our understanding of rice genomics and diversity. Addressing these challenges will be crucial for realizing the full potential of the project's data and for ensuring that its benefits are shared equitably among researchers and communities worldwide (Wing *et al.,* 2018; Saad *et al.,* 2022).

CONCLUSION

The 3000 Rice Genome Project (3K RGP) has been a landmark effort in the field of rice genomics, generating a vast dataset of genomic information on 3,010 diverse varieties of rice. The project has provided new insights into the genetic basis of important traits, as well as the diversity and population structure of rice. The 3K RGP has also provided a valuable resource for rice breeding and improvement efforts, enabling the identification of novel alleles and the development of new molecular markers for use in marker-assisted selection. Furthermore, the project's data has facilitated comparative genomic analyses between different varieties and related species, providing new insights into the evolution and domestication of rice. However, the 3K RGP also faces challenges related to data analysis, germplasm collection and maintenance, quality control, and ethical considerations. Addressing these challenges will require ongoing collaboration and investment in infrastructure, resources, and policy frameworks. To overcome these challenges and successfully complete the 3000 Rice Genome Project, researchers will need to collaborate closely with one another, employ the latest tools and technologies for data analysis, and secure funding from a variety of sources. Additionally, they will need to be patient, persistent, and willing to adapt their approach as needed in order to achieve their goals. Despite these challenges, the 3K RGP has already had a significant impact on rice genomics and breeding, and its data and resources are likely to continue to be a valuable asset for researchers worldwide. The project's focus on global collaboration and open data sharing has helped to build a community of researchers and stakeholders dedicated to advancing our understanding of rice biology and improving global food security.

Author contributions

All authors contributed to the study conception, and Mojtaba Kordrostami had the idea for the article. Mojtaba Kordrostami, Mehdi Rahimi and Ali Akbar Ghasemi Soloklui performed the literature search, wrote and/or critically revised the work.

Data Availability

No data were used to support this study.

Conflict of interest

The authors declare no conflict of interest.

Consent for publication

All the authors have read the manuscript and have approved this submission.

REFERENCES

3000 Rice Genomes Project. (2014). The 3,000 rice genomes project. *GigaScience*, 3(1): 2047-217X-3-7. DOI: https://doi.org/10.1186/2047-217X-3-7.

Adlak T., Tiwari S., Tripathi M., Gupta N., Sahu V. K., Bhawar P., and Kandalkar V. (2019). Biotechnology: an advanced tool for crop improvement. *Current Journal of Applied Science and Technology,* 33(1): 1-11. DOI: 10.9734/cjast/2019/v33i130081.

Ahmar S., Gill R. A., Jung K.-H., Faheem A., Qasim M. U., Mubeen M., and Zhou W. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. *International Journal of Molecular Sciences,* 21(7): 2590. DOI: 10.3390/ijms21072590.

Anilkumar C., Sah R. P., Muhammed A. T., Sunitha N., Behera S., Marndi B., Sharma T. R., and Singh A. K. (2022). Genomic selection in rice: current status and future prospects. In book: Genomic Selection in Plants A Guide for Breeders, 68-82. DOI: 10.1201/9781003214991-4.

Anwar K., Joshi R., Morales A., Das G., Yin X., Anten N. P., Raghuvanshi S., Bahuguna R. N., Singh M. P., Singh R. K., Zanten M., Sasidharan R., Singla-Pareek S. L., and Pareek A. (2022). Genetic diversity reveals synergistic interaction between yield components could improve the sink size and yield in rice. *Food and Energy Security,* 11(2): e334. DOI: 10.1002/fes3.334.

Bommisetty R., Chakravartty N., Bodanapu R., Naik J. B., Panda S. K., Lekkala S. P., Lalam K., Thomas G., Mallikarjuna S., Reddy G. E., Murty G., Bollineni S. N., Issa K., Srividhya A., Srilakshmi C., Hariprasadreddy K., Rameshbabu P., Sudhakar P., Gupta S., Lachagari R. V. B., and Vemireddy L. (2020). Discovery of genomic regions and candidate genes for grain weight employing next generation sequencing based QTL-seq approach in rice (*Oryza sativa* L.). *Molecular Biology Reports,* 47(11): 8615-8627. DOI: 10.1007/s11033-020-05904-7.

Edzesi W. M., Dang X., Liang L., Liu E., Zaid I. U., and Hong D. (2016). Genetic diversity and elite allele mining for grain traits in rice (*Oryza sativa* L.) by association mapping. *Frontiers in Plant Science,* 7: 787. DOI: 10.3389/fpls.2016.00787.

Escaramís G., Docampo E., and Rabionet R. (2015). A

decade of structural variants: description, history and methods to detect structural variation. *Briefings in Functional Genomics,* 14(5): 305-314. DOI: 10.1093/bfgp/elv014.

Francia E., Pecchioni N., Policriti A., and Scalabrin S. (2015). CNV and structural variation in plants: prospects of NGS approaches. In: Sablok G., Kumar S., Ueno S., Kuo J., and Varotto C. (Eds), Advances in the understanding of biological sciences using next generation sequencing (NGS) approaches. Springer, Cham, 211-232. DOI: https://doi.org/10.1007/978-3-319-17157-9_13.

Fuentes R. R., Chebotarov D., Duitama J., Smith S., De la Hoz J. F., Mohiyuddin M., Wing R. A., McNally K. L., Tatarinova T., and Grigoriev A. (2019). Structural variants in 3000 rice genomes. *Genome Research,* 29(5): 870-880. DOI: 10.1101/gr.241240.118.

Futschik A., and Schlötterer C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics,* 186(1): 207-218. DOI: 10.1534/genetics.110.114397.

Gouda G., Gupta M. K., Donde R., Kumar J., Parida M., Mohapatra T., Dash S. K., Pradhan S. K., and Behera L. (2020). Characterization of haplotypes and single nucleotide polymorphisms associated with Gn1a for high grain number formation in rice plant. *Genomics,* 112(3): 2647-2657. DOI: 10.1016/j.ygeno.2020.02.016.

Gu H., Liang S., and Zhao J. (2022). Novel sequencing and genomic technologies revolutionized rice genomic study and breeding. *Agronomy,* 12(1): 218. DOI: 10.3390/agronomy12010218.

Gupta M. K., Gouda G., Donde R., Selvaraj S., Dash G. K., Nambi R., Ponnana M., Pati P., Rathore S. K., Ramakrishna V., and Behera L. (2021). 3000 genome project: a brief insight. In book: Bioinformatics in Rice Research: Theories and Techniques, 89-100. DOI: 10.1007/978-981-16-3993-7_5.

Gutaker R. M., Groen S. C., Bellis E. S., Choi J. Y., Pires I. S., Bocinsky R. K., Slayton E. R., Wilkins O., Castillo C. C., and Negrão S. (2020). Genomic history and ecology of the geographic spread of rice. *Nature Plants*, 6(5): 492-502. DOI: 10.1038/s41477-020-0659-6.

Hasan N., Choudhary S., Naaz N., Sharma N., and Laskar R. A. (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *Journal of Genetic Engineering and Biotechnology*, 19(1): 1-26. DOI: 10.1186/s43141-021-00231-1.

Islam S. M. F., and Karim Z. (2019). World's demand for food and water: The consequences of climate change. In book: Desalination-challenges and opportunities, 1-27. DOI: 10.5772/intechopen.85919.

Jia L., Xie L., Lao S., Zhu Q.-H., and Fan L. (2021). Rice bioinformatics in the genomic era: Status and perspectives. *The Crop Journal,* 9(3): 609-621. DOI: 10.1016/j.cj.2021.03.003.

Junliang Z. V., and Dilin L. (2022). Rice genetics, genomics and breeding revolutionized by next-generation sequencing. *Next-Generation Sequencing and Agriculture*, 12: 39-58.

Katara J. L., Parameswaran C., Devanna B., Verma R. L., Anilkumar C., Patra B., and Samantaray S. (2021). Genomics assisted breeding: The need and current perspective for rice improvement in India. *ORYZA-An International Journal of Rice,* 58(1): 61-68. DOI: 10.35709/ory.2021.58.spl.1.

Kaur B., Sandhu K. S., Kamal R., Kaur K., Singh J., Röder M. S., and Muqaddasi Q. H. (2021). Omics for the improvement of abiotic, biotic, and agronomic traits in major cereal crops: Applications, challenges, and prospects. *Plants*: 10(10): 1989. DOI: https://doi.org/10.3390/plants10101989.

Kebriyaee D., Kordrostami M., Rezadoost M. H., and Lahiji H. S. (2012). QTL analysis of agronomic traits in rice using SSR and AFLP markers. *Notulae Scientia Biologicae,* 4(2): 116-123. DOI: 10.15835/nsb427501.

Konishi S., Izawa T., Lin S. Y., Ebana K., Fukuta Y., Sasaki T., and Yano M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science,* 312(5778): 1392-1396. DOI: 10.1126/science.1126410.

Kordrostami M., and Mafakheri M. (2021). Consequences of water stress and salinity on plants/crops: physiobiochemical and molecular mitigation approaches. In: Handbook of Plant and Crop Physiology, CRC Press, 789-814.

Kordrostami M., Mafakheri M., and Chaleshtori M. H. (2021). Characteristics of grain quality in rice: Physiological and molecular aspects. In: Handbook of Plant and Crop Physiology, CRC Press, 147-157.

Kordrostami M., and Rabiei B. (2019). Major databases to study genes expression in plants (A case study: Rice). *Cereal Research*, 9(1): 83-101. DOI: 10.22124/c.2019.13568.1499.

Kordrostami M., Rabiei B., and Kumleh H. H. (2017). Different physiobiochemical and transcriptomic reactions of rice (*Oryza sativa* L.) cultivars differing in terms of salt sensitivity under salinity stress. *Environmental Science and Pollution Research,* 24(8): 7184-7196. DOI: 10.1007/s11356-017-8411-0.

Kordrostami M., and Rahimi M. (2015). Molecular markers in plants: concepts and applications. *Genetics in the 3rd Millennium,* 13: 4024-4031.

Kumar J., Pratap A., and Kumar S. (2015). Plant phenomics: an overview. In book: Phenomics in Crop Plants: Trends, Options And Limitations, 1-10. DOI: 10.1007/978-81-322-2226-2_1.

Kumar P., Choudhary M., Jat B., Kumar B., Singh V., Kumar V., Singla D., and Rakshit S. (2021). Skim sequencing: an advanced NGS technology for crop improvement. *Journal of Genetics,* 100: 1-10. DOI: 10.1007/s12041-021-01285-3.

Kumawat S., Raturi G., Dhiman P., Sudhakarn S., Rajora N., Thakral V., Yadav H., Padalkar G., Sharma Y., and Rachappanavar V. (2022). Opportunity and challenges for whole-genome resequencing-based genotyping in plants. Genotyping by Sequencing for Crop Improvement, 38-51. DOI: https://doi.org/10.1002/9781119745686.ch3.

Lenaerts B., Collard B. C., and Demont M. (2019). Improving global food security through accelerated plant breeding. *Plant Science,* 287: 110207. DOI: 10.1016/j.plantsci.2019.110207.

Li J.-Y., Wang J., and Zeigler R. S. (2014). The 3,000 rice genomes project: new opportunities and challenges for future rice research. *GigaScience,* 3(1): 2047-2217X-2043-2048.

Li Z.-K., and Zhang F. (2013). Rice breeding in the post-genomics era: from concept to practice. *Current Opinion in Plant Biology*, 16(2): 261-269. DOI: 10.1016/j.pbi.2013.03.008.

Lu J., Wang C., Zeng D., Li J., Shi X., Shi Y., and Zhou Y. (2021). Genome-wide association study dissects resistance loci against bacterial blight in a diverse rice panel from the 3000 rice genomes project. *Rice,* 14(1): 1-13. DOI: 10.1186/s12284-021-00462-3.

Mafakheri M., and Kordrostami M. (2020). Role of Molecular Tools and Biotechnology in Climate-Resilient Agriculture. In Hasanuzzaman M. (Ed.), Plant Ecophysiology and Adaptation Under Climate Change: Mechanisms and Perspectives II: Mechanisms Of Adaptation and Stress Amelioration, Springer Singapore, 491-529. DOI: https://doi.org/10.1007/978-981-15-2172-0_17.

McCouch S. R., McNally K. L., Wang W., and Sackville Hamilton R. (2012). Genomics of gene banks: a case study in rice. *American Journal of Botany,* 99(2): 407-423. DOI: 10.3732/ajb.1100385.

McNally K. L., Jackson M., Ford-Lloy B., and Parry M. (2014). Exploring 'omics' of genetic resources to mitigate the effects of climate change. *Plant Genetic Resources and Climate Change*. DOI: 10.1079/9781780641973.0166.

Mills R. E., Walter K., Stewart C., Handsaker R. E., Chen K., Alkan C., Abyzov A., Yoon S. C., Ye K., and Cheetham R. K. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature,* 470(7332): 59-65. DOI: 10.1038/nature09708.

Mullaney J. M., Mills R. E., Pittard W. S., and Devine S. E. (2010). Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics,* 19(R2): R131-R136. DOI: 10.1093/hmg/ddq400.

Nie L., and Peng S. (2017). Rice production in China. In: Rice production worldwide, Springer, 33-52.

Olsen K. M. (2022). The rice pangenome branches out. *Cell Research,* 32(10): 867-868. DOI: 10.1038/s41422-022-00699-7.

Palanisami K., Kakumanu K. R., Nagothu U. S., and Ranganathan C. (2019). Climate Change and future rice production in India. India Studies in Business and Economics. ISBN: 978-981-13-8362-5.

Paul A. (2020). Sequencing the rice genome: gateway to agricultural development. Rice Research for Quality Improvement: Genomics and Genetic Engineering, Volume 1, Breeding Techniques and Abiotic Stress Tolerance, 109-157.

Pervez M. T., Abbas S. H., Moustafa M. F., Aslam N., and Shah S. S. M. (2022). A comprehensive review of performance of next-generation sequencing platforms. *BioMed Research International,* 2022(107): 1-12. DOI: 10.1155/2022/3457806.

Purugganan M. D., and Jackson S. A. (2021). Advancing crop genomics from lab to field. *Nature Genetics,* 53(5): 595-601. DOI: 10.1038/s41588-021-00866-3.

Rabiei B., Kordrostami M., Sabouri A., and Sabouri H. (2015). Identification of QTLs for yield related traits in Indica type rice using SSR and AFLP markers. *Agriculturae Conspectus Scientificus,* 80(2): 91-99.

Ram S. G., Thiruvengadam V., and Vinod K. K. (2007). Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. *Journal of Applied Genetics,* 48(4): 337-345. DOI: 10.1007/BF03195230.

Ramadas V. S. (2018). Consumption pattern of traditional and non-traditional rice consumers and their health status. *Asian Journal of Multidimensional Research (AJMR),* 7(1): 67-72.

Rana N., Rahim M. S., Kaur G., Bansal R., Kumawat S., Roy J., Deshmukh R., Sonah H., and Sharma T. R. (2020). Applications and challenges for efficient exploration of omics interventions for the enhancement of nutritional quality in rice (*Oryza sativa* L.). *Critical Reviews in Food Science and Nutrition,* 60(19): 3304-3320. DOI: 10.1080/10408398.2019.1685454.

Rivero R. M., Mittler R., Blumwald E., and Zandalinas S. I. (2022). Developing climate-resilient crops: improving plant tolerance to stress combination. *The Plant Journal,* 109(2): 373-389. DOI: 10.1111/tpj.15483.

Roitsch T., Cabrera-Bosquet L., Fournier A., Ghamkhar K., Jiménez-Berni J., Pinto F., and Ober E. S. (2019). New sensors and data-driven approaches—A path to next generation phenomics. *Plant Science,* 282: 2-10. DOI: 10.1016/j.plantsci.2019.01.011.

Saad N. S. M., Neik T. X., Thomas W. J., Amas J. C., Cantila A. Y., Craig R. J., Edwards D., and Batley J. (2022). Advancing designer crops for climate resilience through an integrated genomics approach. *Current Opinion in Plant Biology,* 67: 102220.

Sakhale S., Yadav S., Clark L., Lipka A., Kumar A., and Sacks E. (2023). Genome-wide association analysis for emergence of deeply sown rice (*Oryza sativa*) reveals novel aus-specific phytohormone candidate genes that conferred adaption to dry-direct seeding in the field. *Authorea Preprints*. DOI: 10.22541/essoar.167407907.78496270/v1.

Salgotra R. K., and Chauhan B. S. (2023). Genetic diversity, conservation, and utilization of plant genetic resources. *Genes,* 14(1): 174. DOI: 10.3390/genes14010174.

Seck P. A., Diagne A., Mohanty S., and Wopereis M. (2012). Crops that feed the world 7: Rice. *Food Security,* 4(1): 7-24. DOI: https://doi.org/10.1007/s12571-012-0168-1.

Shahzad A., Ullah S., Dar A. A., Sardar M. F., Mehmood T., Tufail M. A., Shakoor A., and Haris M. (2021). Nexus on climate change: Agriculture and possible solution to cope future climate change stresses. *Environmental Science and Pollution Research,* 28: 14211-14232.

Song S., Tian D., Zhang Z., Hu S., and Yu J. (2018). Rice genomics: over the past two decades and into the future. *Genomics, Proteomics & Bioinformatics,* 16(6): 397-404. DOI: 10.1016/j.gpb.2019.01.001.

Sun C., Hu Z., Zheng T., Lu K., Zhao Y., Wang W., Shi J., Wang C., Lu J., Zhang D., Li Z., and Wei Ch. (2017). RPAN: rice pan-genome browser for~ 3000 rice genomes. *Nucleic Acids Research,* 45(2): 597-605. DOI: 10.1093/nar/gkw958.

Sweeney M. T., Thomson M., Pfeil B. E., and McCouch S. (2006). Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *The Plant Cell,* 18(2): 283-294. DOI: 10.1105/tpc.105.038430.

Thangadurai D., Kordrostami M., Islam S., Sangeetha J., Al-Tawaha A. R. M. S., and Jabeen S. (2020). Genetic Enhancement of Nutritional Traits in Rice Grains Through Marker-Assisted Selection and Quantitative Trait Loci. In: Rice Research for Quality Improvement: Genomics and Genetic Engineering, Springer, 493-507.

Tu Anh T. T., Khanh T. D., Dat T. D., and Xuan T. D. (2018). Identification of phenotypic variation and genetic diversity in rice (*Oryza sativa* L.) mutants. *Agriculture,* 8(2): 30.

Wang B., Kumar V., Olson A., and Ware D. (2019). Reviving the transcriptome studies: an insight into the emergence of single-molecule transcriptome sequencing. *Frontiers in Genetics,* 10: 384. DOI: 10.3389/fgene.2019.00384.

Wang W., Mauleon R., Hu Z., Chebotarov D., Tai S., Wu Z., Li M., Zheng T., Fuentes R. R., and Zhang F. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature,* 557(7703): 43-49. DOI: 10.1038/s41586-018-0063-9.

Wang Z. Y., Zheng F. Q., Shen G. Z., Gao J. P., Snustad D. P., Li M. G., Zhang J. L., and Hong M. M. (1995). The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. *The Plant Journal,* 7(4): 613-622. DOI: 10.1046/j.1365-313X.1995.7040613.x.

Wei X., and Huang X. (2019). Origin, taxonomy, and phylogenetics of rice. In: Rice, Elsevier, 1-29. DOI: https://doi.org/10.1016/B978-0-12-811508-4.00001-0.

Wing R. A., Purugganan M. D., and Zhang Q. (2018). The rice genome revolution: from an ancient grain to Green Super Rice. *Nature Reviews Genetics,* 19(8): 505-517. DOI: https://doi.org/10.1038/s41576-018-0024-z.

Würschum T., Boeven P. H., Langer S. M., Longin C. F. H., Leiser W. L. (2015). Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. *BMC Genetics,* 16(1): 1-8. DOI: 10.1186/s12863-015-0258-0.

Yu H., Xie W., Li J., Zhou F., and Zhang Q. (2014). A whole-genome SNP array (RICE 6 K) for genomic breeding in rice. *Plant Biotechnology Journal,* 12(1): 28-37. DOI: 10.1111/pbi.12113.

Yu S., Xu W., Vijayakumar C., Ali J., Fu B., Xu J., Jiang Y., Marghirang R., Domingo J., and Aquino C. (2003). Molecular diversity and multilocus organization of the parental lines used in the International Rice Molecular Breeding Program. *Theoretical and Applied Genetics,* 108(1): 131-140. DOI: 10.1007/s00122-003-1400-3.

Żmieńko A., Samelak A., Kozłowski P., and Figlerowicz M. (2014). Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics,* 127(1): 1-18. DOI: 10.1007/s00122-013-2177-7.