Data in Brief 57 (2024) 111051

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on whole genome resequencing of selected Malaysian rice accessions with opposing response to salinity stress

Shahril Ab Razak[a,b,*], Rabiatul-Adawiah Zainal-Abidin[b], Asmuni Mohd Ikmal[a], Norfarhan Mohd-Assaad[c], Noraziyah Abd Aziz Shamsudin[a]

[a] *Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
[b] *Biotechnology & Nanotechnology Research Centre, MARDI Headquarter, 43400 Serdang, Selangor, Malaysia*
[c] *Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*

ARTICLE INFO

Article history:
Received 23 April 2024
Revised 4 October 2024
Accepted 11 October 2024
Available online 18 October 2024

Dataset link: Whole Genome Re-sequencing of selected rice accession with contrast responses to salinity stress (Original data)

Keywords:
Rice genomics
Salinity
Single nucleotide polymorphisms (SNPs)
Whole genome sequencing
Oryza sativa L

ABSTRACT

The genomics and genetic information of Malaysian rice (*Oryza sativa* L.) is currently limited. It was necessary to conduct genome resequencing of these rice accessions exhibiting different responses to salinity stress. The sequencing was carried out using the Illumina NovaSeq X platform with 30× sequencing coverage to pinpoint variants between salinity tolerant and sensitive rice accessions. The discovery of single nucleotide polymorphisms (SNPs) is crucial for the development of DNA markers associated with salinity tolerance traits. The genome sequence data (FASTQ format) for these accessions have been deposited to the European Nucleotide Archive (ENA) database under the accession number PRJEB71716.

* Corresponding author at: Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.
E-mail address: shahrilf@mardi.gov.my (S. Ab Razak).

https://doi.org/10.1016/j.dib.2024.111051

Specifications Table

Subject	*Agricultural sciences*
Specific subject area	*Plant Science*
Type of data	*Clean raw reads of genomic sequence data*
Data collection	The total genomic DNA was extracted from the young leaves of three-week-old seedlings of each rice accession using the DNeasy Plant Pro DNA extraction kit. The DNA library was prepared using the NEBNext® Ultra™ II DNA Library Prep Kit, and sequencing was performed on the Illumina NovaSeq X sequencing platform. The sequencing data was filtered using Fastp v0.20.0.
Data source location	*City/Town/Region: Serdang, Selangor* *Country: Malaysia* *Latitude and longitude (and GPS coordinates) for collected samples/data:]* *2°59′21"N, 101°42′00"E*
Data accessibility	The clean raw reads of these five accessions have been deposited at European Nucleotide Archive (ENA). The information of data accessibility are: *if general: Repository name: Whole Genome Re-sequencing of selected rice accession with contrast responses to salinity stress Data identification number: PRJEB71716 Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/PRJEB71716 1) Repository name: Whole genome sequencing of Bidor accession Data identification number: ERR12421769 Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/ERR12421769 2) Repository name: Whole genome sequencing of Pulut Siding accession Data identification number: ERR12421776 Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/ERR12421776 3) Repository name: Whole genome sequencing of Chatek Kuning accession Data identification number: ERR12421775 Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/ERR12421775 4) Repository name: Whole genome sequencing of Sangam accession Data identification number: ERR12421778 Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/ERR12421778 5) Repository name: Whole genome sequencing of Rengan Wangi accession Data identification number: ERR12421782 Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/ERR12421782

1. Value of the Data

- The genome data provide information for the genomic database of Malaysian rice which can be utilized in understanding their genetic basis.
- The genetic variations identified from the genome will be of significant benefit in crop improvement, particularly in the development of a salt-tolerant rice variety using a marker-assisted breeding approach.
- The comparative genome analysis will provide a basic understanding of genes related to salinity tolerant in rice.
- These genome data are valuable for diversity and evolutionary study.

2. Background

Salinity represents a significant environmental constraint that poses a threat to global food security by affecting crop growth, development, and productivity through reduced water uptake and increased salt concentration in plants [1]. Given the ongoing challenges of climate change and poor irrigation practices, salinity is expected to continue as a major hindrance to crop production. Therefore, focusing on enhancing the adaptation of major crop plants to saline conditions and improving irrigation management practices are essential and practical strategies for increasing global food production [2]. In order to improve rice productivity in saline environments, it is essential to develop rice varieties that can withstand salt stress. The first step

Table 1
Salinity scoring of five rice accessions at 10 dS m^{-1} during the early vegetative stage using a hydroponic approach. Pokkali and IR29 were used as tolerant and susceptible controls, respectively.

Accessions	Observation	Score	Tolerant
Bidor	Nearly normal growth, but the leaf tips or a few leaves whitish and rolled	3.0	Tolerant
Pulut Siding	Nearly normal growth, but the leaf tips or a few leaves whitish and rolled	3.5	Tolerant
Chatek Kuning	Nearly normal growth, but the leaf tips or a few leaves whitish and rolled	3.0	Tolerant
Sangam	Almost all plants dead or dying	9.0	Highly susceptible
Rengam Wangi	Almost all plants dead or dying	9.0	Highly susceptible
Pokkali	Nearly normal growth, but the leaf tips or a few leaves whitish and rolled	3.0	Tolerant
IR29	Almost all plants dead or dying	9.0	Highly susceptible

The salinity performance was scored according to the IRRI Standard Evaluation System (SES) [3].

Table 2
Data summary of selected Malaysian rice accessions.

Accessions	Raw bases(bp)	Clean bases(bp)	Q20(%)	Q30(%)	ENA data identification no
Bidor	14,533,526,100	14,128,032,900	97.12	94.59	ERR12421769
Pulut Siding	14,068,041,000	13,898,026,500	98.09	96.36	ERR12421776
Chatek Kuning	14,628,870,000	14,414,520,000	97.75	95.62	ERR12421775
Sangam	14,871,189,600	14,651,475,900	97.81	95.75	ERR12421778
Rengam Wangi	14,676,031,200	14,457,752,400	97.78	95.66	ERR12421782

Q20 (%): The percentage of bases with higher Phred score than 20.
Q30 (%): The percentage of bases with higher Phred score than 30.

towards achieving this is to screen and evaluate genetic resources in order to identify potential breeding materials. We have successfully identified five rice accessions that exhibit different responses to salinity stress. Among these accessions, three (Bidor, Pulut Siding, and Chatek Kuning) have shown tolerance to salinity stress, while two (Sangam and Rengan Wangi) have been found to be sensitive. By identifying the specific genes and their variants, we can gain a comprehensive understanding of the mechanisms behind salt tolerance. This knowledge will also be valuable for rice breeding activities through marker-assisted selection.

3. Data Description

The genomes of the five rice accessions were sequenced using the Illumina NovaSeq X sequencing platform, with a read length of 150 bp on each end. These five accessions were selected from a set of 182 landrace rice accessions based on their performance for salinity tolerance at 10 dS m^{-1} during the early vegetative stage using a hydroponic approach. Pokkali and IR29 were used as tolerant and susceptible controls, respectively. The details of the accessions are summarized in Table 1. A sequencing coverage depth of 30x was employed, resulting in approximately 72.8 Gb of raw sequenced data from the genomes of these five rice genotypes. After filtering out low-quality data, 71.6 Gb of clean reads were obtained and used for subsequent analyses, including reads mapping and SNPs calling. The summary of raw and clean reads generated from the genome resequencing of these rice accessions is presented in Table 2. The clean reads of all rice genotypes have been deposited in the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home) under the accession number PRJEB71716. This whole-genome resequencing of rice germplasm provides valuable genomic resources that will expedite both genetic analysis and molecular breeding of agronomically significant traits.

4. Experimental Design, Materials and Methods

4.1. Plant materials

The seeds of these five accessions were acquired from the Malaysian Agricultural Research and Development Institute (MARDI) Rice Gene Bank. Prior to DNA extraction, the seeds were allowed to germinate for a period of three weeks.

4.2. DNA extraction and quality control

Extraction of the genomic DNA from the leaves of three-week-old seedlings was carried out for each accession using the DNeasy Plant Pro DNA extraction kit (Qiagen, Hilden, Germany). The integrity of the DNA was evaluated using a 0.8 % agarose gel, and the quantity and quality of the DNA were assessed using a Nano Drop spectrophotometer (Thermo Scientific, Waltham, MA, USA).

4.3. DNA library preparation and sequencing

The DNA samples were sent to the sequencing service provider at the Apical Scientific Sdn Bhd. (Selangor, Malaysia). The construction of paired-end sequencing libraries using NEBNext® Ultra™ IIDNA Library Prep Kit (Cat No. E7645), with insert sizes of 350 bp, adhered to the standard protocol provided by New England Biolabs (Massachusetts, United States) and Illumina (San Diego, CA, USA). The sequencing process was carried out using the Illumina NovaSeq X sequencing platform, with a read length of 150 bp at each end.

4.4. Raw data filtering and processing

The original fluorescence image files from the sequencer were transformed to raw reads by base calling and were recorded in FASTQ format [4], which contains sequence information and corresponding sequencing quality information. The raw reads were subsequently filtered using Fastp version 0.20.0 [5] to remove the adapter sequences, contamination and low-quality reads. Filtering raw reads was conducted as follow: 1) discard a paired reads if either one read contains adapter contamination ($>$ 10 nucleotides aligned to the adapter, allowing $\leq$ 10 % mismatches); 2) discard a paired reads if more than 10 % of bases are uncertain in either one read; 3) discard a paired reads if the proportion of low quality (Phred quality $<$5) bases is over 50 % in either one read.

Limitations

Not Applicable.

Ethics Statement

The authors declare that this research did not involve either human or animal subjects. This manuscript is an original work and has not been published or submitted for publication elsewhere.

CRediT Author Statement

Shahril Ab Razak: Conceptualization, Investigation, Writing – original draft, Writing – review & editing; Rabiatul Adawiah Zainal Abidin: Formal analysis, Writing – original draft, Writing – review & editing; Asmuni Mohd-Ikmal: Supervision, Validation, Writing – review & editing.; Norfarhan Mohd-Assaad: Supervision, Validation, Writing – review & editing; Noraziyah Abd Aziz Shamsudin: Supervision, Validation, Writing – review & editing.

Data Availability

Whole Genome Re-sequencing of selected rice accession with contrast responses to salinity stress (Original data) (ENA).

Acknowledgments

The authors wish to extend their appreciation to the Malaysian Government for their financial support through the Ministry of Agriculture and Food Security (MAFS) under the RMK12 Development Fund PRB502.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P.K. Subudhi, R. Shankar, M. Jain, Whole genome sequence analysis of rice genotypes with contrasting response to salinity stress, Sci. Rep. 10 (1) (2020) 21259.
[2] M.A. Rahman, M.J. Thomson, M. Shah-E-Alam, M. de Ocampo, J. Egdane, A.M. Ismail, Exploring novel genetic sources of salinity tolerance in rice through molecular and physiological characterization, Ann. Bot. 117 (6) (2016) 1083–1097.
[3] G.B. Gregoria, D. Senadhira, R.D. Mendoza, Screening Rice for Salinity Tolerance, International Rice Research Institute, Manila, Philippines, 1997 IRRI discussion paper series no. 22.
[4] P.J. Cock, C.J. Fields, N. Goto, M.L. Heuer, P.M. Rice, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res. 38 (6) (2010) 1767–1771.
[5] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.