Molecular Biology Reports
https://doi.org/10.1007/s11033-020-05794-9

ORIGINAL ARTICLE

Genome- wide structural and functional variant discovery of rice landraces using genotyping by sequencing

Smitha Kunhiraman Vasumathy[1] · **Maya Peringottillam**[1] · **Krishna T. Sundaram**[2] · **S. Hari Krishna Kumar**[1] · **Manickavelu Alagu**[1]

Received: 7 June 2020 / Accepted: 28 August 2020

Abstract

Rice landraces are vital genetic resources for agronomic and quality traits but the undeniable collection of Kerala landraces remains poorly delineated. To effectively conserve, manage, and use these resources, understanding the genomic structure of germplasm is essential. Genotyping by sequencing (GBS) enables identification of an immense number of single nucleotide polymorphism (SNP) and insertion deletion (InDel) from 96 rice germplasm. In the present study, a total of 16.9×10^7 reads were generated, and among that 16.3×10^7 reads were mapped to the indica reference genome. Exploring GBS data unfolded a wide genomic variations including 82,59,639 SNPs and 1,07,140 Indels. Both neighbor-joining tree and principal coordinate analysis with InDel markers revealed the selected germplasm in this study as highly diverse in structure. We assembled unmapped reads which were further employed for gene ontology analysis. These unmapped sequences that are generally expelled from subsequent studies of GBS data analysis may exist as an unexplored resort for several novel significant biological findings. The discovery of SNPs from the haplotyping results of *GS3* and *GIF1* genes provided insight into marker- assisted selection based on grain size and yield and can be utilized for rice yield improvement. To our knowledge, this is the first report on structural variation analysis using the GBS platform in rice landraces collected from Kerala. Genomic information from this study endows with valuable resources for perceptive rice landrace structure and can also facilitate sequencing-based molecular breeding.

Keywords Rice landraces · Genotyping by sequencing · SNP · InDel · Unmapped sequences · Gene ontology

Introduction

Rice (*Oryza sativa* L.) is grown in more than a 100 countries, with a production of 700 million tons per annum from a total harvested area of about 158 million hectares globally. As it is one of the major diets consumed by more than half of the people in the world, its production level has to be increased considerably to feed the mounting population. Under the sway of farmer's desire and environmental circumstances, the rice originated in South East Asia expanded into 88,681 different varieties within which 1,171 are wild races, 55,615 are landraces, and 32,895 are other varieties [1]. Rice grown under diverse ecologies endures severe yield loss due to several biotic and abiotic constraints. Rice landraces are lineages evolved via selective breeding by farmers in a time of long term domestication [2]. These are genetically dynamic and can get accustomed to new environments irrespective of their origin [3]. The improved varieties which were the pinnacle outcome of the green revolution have obliquely enthused erosion of wild varieties and landraces of rice [4]. In India, the northern part of Kerala is bestowed with a large number of rice landraces but the prevalent information about landraces is flawed, and therefore collecting and conserving rice landraces are far-reaching. Each landrace plays a prominent role in accomplishing local food security with meticulous characteristics like early maturity, adaptation to particular soil types, tolerance to biotic and abiotic stresses,

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-020-05794-9) contains supplementary material, which is available to authorized users.

✉ Manickavelu Alagu
amanicks@cukerala.ac.in

[1] Department of Genomic Science, Central University of Kerala, Kasaragod, Kerala 671316, India

[2] South Asia hub, ICRISAT Campus, International Rice Research Institute, Secundarabad, Telangana, India