

SEQUENCING OF DIVERSE RICE GENOMES DRIVES BREEDING INNOVATION

MCNALLY K.*, ZHOU Y.**, CHEBOTAROV D.*, TALAG J.***, COPETTI D.***, ZUCCOLO A.**,
LORIEUX M.****, ZHANG J.*****, SCHMUTZ J.*****, WING R.**

*) International Rice Research Institute (IRRI), Rice Breeding Innovations, Los Baños, Laguna, Philippines

**) Center for Desert Agriculture, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

***) Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona, USA

****) DIADE, University of Montpellier, Cirad, IRD, Montpellier, France

*****) National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

*****) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA

Oryza sativa, genetic diversity, sequencing, reference genome

Asian domesticated rice (*Oryza sativa*) is one of the most diverse crop species grown globally with over half of humanity depending on it for sustenance. At IRRI, the International Rice Genebank Collection conserves over 137,000 accessions of *O. sativa* and affiliated species, but the overwhelming majority have not contributed to released varieties through conventional breeding. Population growth and climate change call for accelerated rice improvement for yield and resilience to abiotic and biotic stresses. Detailed exploration of rice diversity and discovery of novel alleles, donors, and haplotypes for critical traits will be crucial to these efforts.

In 2005, the International Rice Genome Sequencing Project released the reference genome of the temperate japonica variety Nipponbare, the first sequenced crop genome. This was a game-changer for rice genetics and genomics serving as the foundation for numerous studies. In 2009, IRRI and partners reported SNPs discovered from 20 varieties from Nipponbare tiling arrays in the OryzaSNP project. This technology was soon overtaken by next generation sequencing (NGS). Then, IRRI worked with Cornell, USDA, and others to develop the 700K high-density rice array (HDRA) for SNPs from NGS data analyzed relative to Nipponbare. HDRA results for genotyping 1571 accessions were published by McCouch et al (2016). Concurrently, IRRI, BGI-Shenzhen, and CAAS undertook the 3000 Rice Genomes Project (3K RGP) with public release of data in 2014, and the first version of the SNP-Seek database (<https://snp-seek.irri.org>) for curating the genotyping data in 2015. Detailed analyses of the 3K RG, involving more collaborators, was published in 2018. Together, the OryzaSNP, HDRA, 3K RGP, and SNP-Seek efforts have led to hundreds of publications, substantial use of the genotyped/sequenced germplasm, and release of new varieties derived from novel alleles and donors.

Yet, these studies were limited to SNP discovery relative just to Nipponbare while rice has 5 major variety groups comprised of 15 or more subpopulations. To address this, we have produced new reference genomes for 15 subpopulations of *O. sativa* using PacBio long-read sequencing (Zhou et al, 2020). Currently,

mapping of the 3K RG reads is underway to these new references. While almost 5000 accessions have been interrogated among prior studies, there are gaps of coverage in terms of origin and sampling of types. For improved coverage of diversity and increased depth of NGS, we are currently resequencing 10K additional genomes that will be analyzed relative to multiple references. These ongoing efforts will substantially broaden the knowledge of rice diversity with spillover effects where further novel haplotypes, donors, and low frequency alleles can be identified and delivered to rice breeders. Lastly, they bring us closer to the realization of a digital rice genebank with all accessions sequenced.