

[Print](#)

Big data on 3,000 rice genomes available on the AWS Cloud



Sample seeds from among the 127,000 rice varieties and accessions stored in the International Rice Genebank at the International Rice Research Institute.

The 3,000 Rice Genomes Project (3K RGP) is a collaborative, international research program that has sequenced 3,024 rice varieties from 89 countries. This massive dataset is a powerful resource for understanding natural genetic variation in rice as well as for large-scale discovery of new genes associated with economically important traits. It will help accelerate the pace of developing improved rice varieties around the globe to feed a growing population, estimated to reach more than 9.6 billion by 2050, with half of humanity relying on rice for sustenance and livelihood.

The [International Rice Genebank](#) of the [T.T. Chang Genetic Resources Center](#) at the [International Rice Research Institute](#) (IRRI) in the Philippines contains more than 127,000 rice varieties and accessions from all over the world. These accessions hold a virtually untapped reservoir of genes/traits that can be used to make rice cultivation more sustainable, with a smaller environmental footprint. Traits targeted for improvement include higher nutritional quality; tolerance of pests, diseases, and environmental stresses, such as flood and drought; and reduced greenhouse gas emissions.

Three research institutions—the [Chinese Academy of Agricultural Sciences](#) (CAAS), the [Beijing Genomics Institute \(BGJ\) Shenzhen](#), and IRRI—collaborated to sequence the genomes of 3,024 rice varieties and lines housed in the IRRI (82%) and the CAAS (18%) genebanks. The sequencing and initial analysis was funded by grants from the [Bill & Melinda Gates Foundation](#) and the Chinese Ministry of Science and Technology. This dataset contains millions of genomic sequences from a diverse set of rice varieties that, when combined with phenotyping observations, gene expression, and

other information, provides an important step in establishing gene-trait associations, building predictive models, and applying these models to breeding.

Through funding from the [Global Rice Science Partnership](#), the 3,024 genomes were re-analyzed against five popular varieties that represent the three main subgroups of cultivated rice—indica, japonica, and aus. This new 3K RGP data analysis set is massive at 120 terabytes, which is well beyond the computing capacities of most research institutions. However, these new results are now publicly available online, as an [Amazon Web Services](#) (AWS) Public Data Set. Accessing the data is free, and use is governed by the stipulations for data analysts and users from the Toronto Statement.

[Dr. Rod Wing](#), director of the [Arizona Genomics Institute](#) at the University of Arizona and a pioneer in rice genome sequencing, remarked: “The dataset provides access to millions of genetic markers that can be used to design sustainable crops for the future, that is, ones that are high-yielding and more nutritious while at the same time requiring less water, fertilizer, and pesticides.”

The great thing about the release of this dataset is that it is immediately useable,” said Dr. Kenneth McNally, senior scientist in IRRI’s T.T. Chang Genetic Resources Center and a project team member. “It comes with tools to help researchers visualize and analyze genetic information.”

Data access and analysis tools are being made available for the 3K RGP dataset through the [International Rice Informatics Consortium](#) (IRIC), which promotes collaboration in bioinformatics analysis of rice data and provides computational tools to facilitate rice improvement via discovery of new gene-trait associations and accelerated breeding. One of the tools, the [SNP-Seek database](#), is designed to provide user-friendly access to a type of genetic marker called single nucleotide polymorphisms (SNPs) identified from this data. Another tool in SNP-Seek, the JBrowse genome browser, displays chromosome-specific SNP data derived from the set.

“The 3K RGP dataset is a powerful tool that will unite researchers from around the world to help drive the next green revolution,” Wing concludes.

Video

IRRI's Internati...



IRRI references

- [Genetic resources](#)
- [International Rice Genebank](#)
- [International Rice Informatics Consortium \(IRIC\)](#)

Rice Today features

- [Dawn of a new era in rice improvement](#)

Media releases and news items

- [3,000 rice genome sequences made publicly available on World Hunger Day](#)

For inquiries, please contact:

scicomm@irri.org

+63 2 580 5600 ext 2840

[Media inquiries page](#)