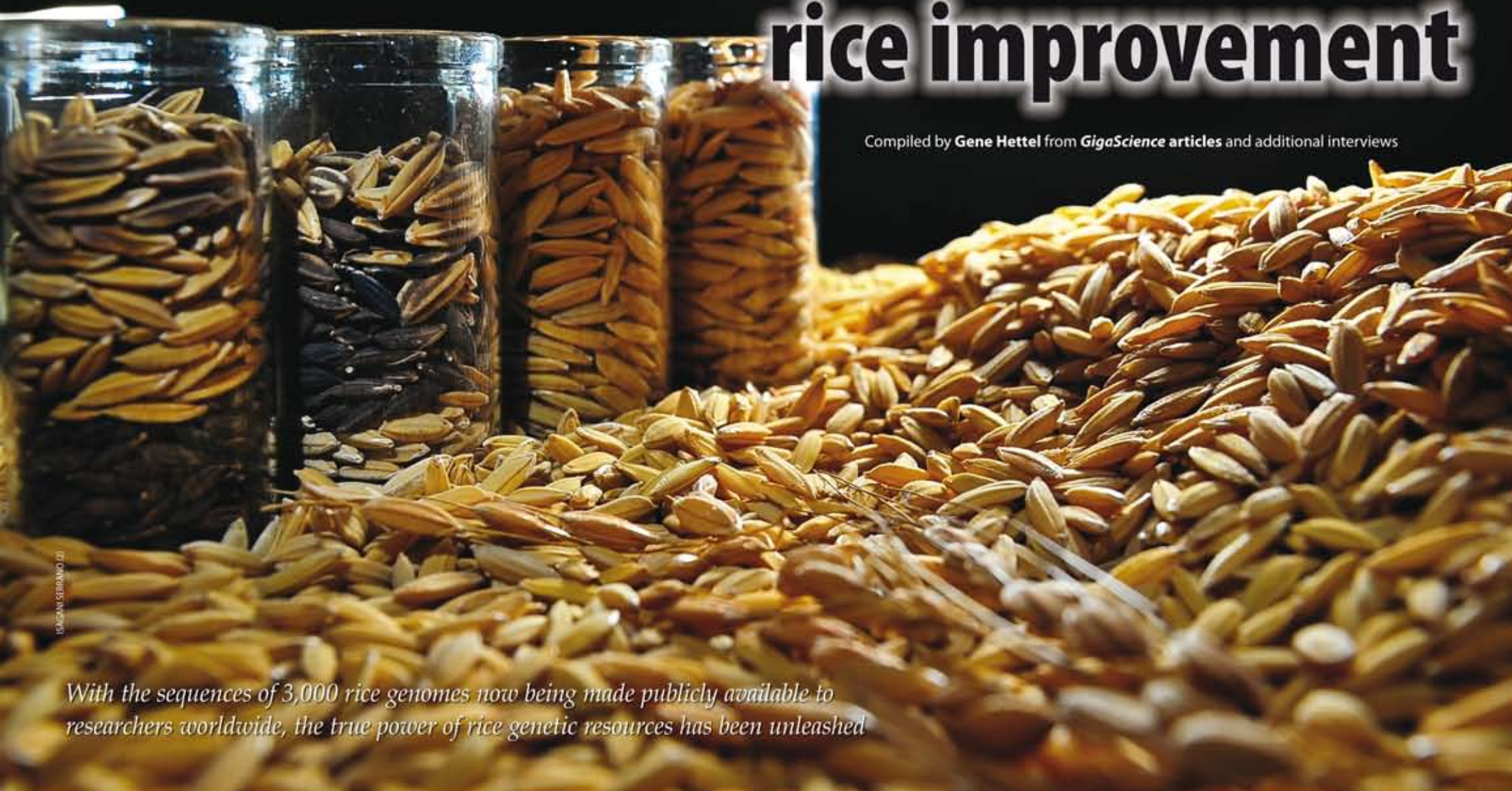


Dawn of a new era in rice improvement

Compiled by Gene Hettel from *GigaScience* articles and additional interviews



With the sequences of 3,000 rice genomes now being made publicly available to researchers worldwide, the true power of rice genetic resources has been unleashed

Traditional rice varieties encompass a huge range of potentially valuable genes. These can be used to develop superior varieties for farmers to take part in the uphill battle of feeding an ever-increasing world population (estimated to reach 9.6 billion by 2050). The genes linked to valuable traits can help breeders create new rice varieties that have improved yield potential, higher nutritional quality, better ability to

grow in problem soils, and improved tolerance of pests, diseases, and the stresses, such as flood and drought, that will be inevitable with future climate change.

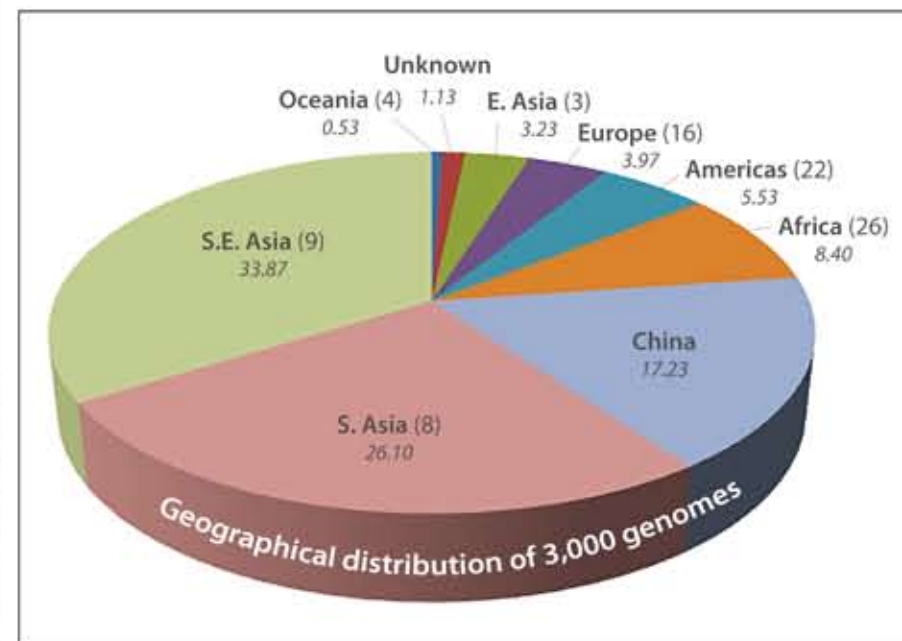
Much of this diversity is conserved within the International Rice Genebank Collection (IRGC) at the International Rice Research Institute (IRRI). According to Ruairaidh Sackville Hamilton, head of IRRI's T.T. Chang Genetic Resources Center (TTC GRC), the IRGC now

holds a global collection of more than 121,000 types or accessions of rice. Yet, breeders have harnessed only less than 5% of the germplasm collection there in active breeding efforts. "But all that is changing with a very exciting development to make the IRGC useful beyond our wildest dreams," he exclaims.

The 3,000 Rice Genomes Project

A single genome¹ cannot reveal the large stockpile of genetic diversity

in rice and hence many potentially important genes are not present in the handful of lines that have been sequenced over the last decade. So, to drastically change this dynamic, IRRI—in collaboration with BGI in Shenzhen, China, and the Chinese Academy of Agricultural Sciences (CAAS)—has completed the sequencing of 3,000 rice genomes of varieties and lines representing 89 countries (see figure) now housed in the IRGC (82%) and CAAS's genebank



GEOGRAPHIC DISTRIBUTION of 3,000 representative rice accessions whose genomes have been sequenced. Numbers in parentheses indicate the number of countries in each region.

(18%). "This is an unparalleled development in plant science for a major food crop," says Ken McNally, senior scientist in the TTC GRC and a project team member.

On top of that, the open-access, open-data journal *GigaScience* (produced by BGI and BioMed Central) published, on World Hunger Day 2014 (28 May), a data note² and commentary³ on the 3,000 Rice Genomes Project (3K RGP). Most significantly, released at the same time was the project's entire 13.4-terabyte dataset in a citable format in the journal's affiliated open-access database, GigaDB (see box), which instantly quadrupled the previous amount of publicly available rice sequence data. The 3K RGP is funded by the Bill & Melinda Gates Foundation (BMGF) and the Chinese Ministry of Science and Technology.

According to Zhikang Li, project director at CAAS, the 3K RGP is part of an ongoing effort to provide resources specifically for poverty-stricken farmers in Asia and Africa. "We are aiming to reach at least 20 million rice farmers in 16 target countries (eight on each continent),"

he says. "With decreasing water and land resources, food security is—and will be—the most challenging issue in these countries." Echoing Dr. Sackville Hamilton, Dr. Li says, "As a scientist in rice genetics, breeding, and genomics, it is a dream come true for me to help solve this problem."

"The population boom and the worsening climate crisis have presented big challenges on global food shortage and safety," adds Jun Wang, BGI director. "BGI is dedicated to applying genomics technologies to make a fast, controllable, and highly efficient molecular breeding model possible. The 3K RGP opens a new way to carry out agricultural breeding. Joining forces with CAAS, IRRI, and BMGF, we have made a step forward in big-data-based crop research and digitalized breeding. We believe this will get us closer to the ultimate goal of improving the well-being of the human race."

According to Robert Zeigler, IRRI director general, access to the sequence data of these 3,000 rice genomes will tremendously accelerate the progress of breeding programs.

¹ A genome is the genetic material of an organism (in the case here for each of the 3,000 selected rice accessions), which is encoded in its DNA and includes the genes and the noncoding sequences of the DNA.

² www.gigasciencejournal.com/content/3/1/7.

³ www.gigasciencejournal.com/content/3/1/8.

Publication in *GigaScience* includes storage of relevant associated data in the journal's affiliated database, GigaDB (<http://gigadb.org/dataset/200001>), where every dataset is provided with a digital object identifier (DOI), making it possible to cite, find, and track data in standard scientific literature, which serves as a strong incentive for researchers to more rapidly release expensive and work-intensive datasets for community use.

On top of hosting this wealth of supporting data in GigaDB to provide the most extensive availability to the community, the sequence reads for this project have also been submitted to the repository of the European Nucleotide Archive: www.ebi.ac.uk/ena/data/view/PRJEB6180.

Enquires can be directed to the following members of the 3K RGP consortium:

Gengyun Zhang, BGI, Shenzhen, China

Zhikang Li, CAAS, Beijing, China

Tian-Qing Zheng, CAAS, Beijing, China

Kenneth L. McNally, IRRI, Los Baños, Philippines

"The collaborative 3K RGP," he says, "will add an immense amount of knowledge to rice genetics and enable detailed analysis by the global research community to ultimately benefit the poorest farmers who grow rice under the most difficult conditions."

To reach their goals, the three institutes have not only released the large volume of data, they are also making available through the IRGC seeds for each rice accession that has been sequenced. "Having available banked seeds is essential to make full use of this germplasm and the data about it," says Dr. Sackville Hamilton.

Speeding the breeding

A major part of the project is to directly link the genetic information (genotype) to the physical traits (phenotype) of these different accessions. This will require careful assessment and curation of each accession for the agriculturally important traits mentioned earlier, which breeders can then link to genetic markers in the now available genome sequences.

Current breeding practices, which have essentially remained the same since the development of

agriculture, typically employ the observation of apparent physical traits to guide parent selection for making crosses with the hope that the offspring will show a new combination and an improvement of the desired traits. However, the underlying genetic makeup of the offspring can often confound breeders' expectations. So, they often have to resort to time-consuming trial and error involving multiple successive generations.

The 3K RGP data will provide a major step forward by helping breeders to take advantage of the natural trait variation that is found across the selected genebank accessions in the project. Knowing fully the genetic makeup of a particular rice accession will allow researchers to identify genetic

markers related to specific traits, and better understand how different genetic interactions affect plant phenotypes. With this information, breeders will be able to make more intelligent choices in accession selection for making crosses, resulting in more rapid development of rice varieties for deploying in poor and environmentally stressed locations around the world.

According to Hei Leung, IRRI principal scientist and team member, the IRRI 3K RGP team is committed to moving quickly to take the sequence data and make detailed analyses. "Through the Global Rice Science Partnership (GRISP), IRRI is leading the development of an informal global effort—the International Rice Phenotyping Network—to systematically evaluate

the 3K RGP sequenced lines and to connect plant performance to specific genes," he says. "By closely integrating these resources into breeding programs based on modern molecular breeding and selection strategies, varietal development in hundreds of rice breeding programs will be accelerated over the next five years, delivering improved varieties to farmers and consumers at a faster pace than before."

AXA-endowed scientist to join the team

"In addition to working closely with GRISP, BGI, and CAAS," Dr. Leung adds, "we have a head start with the timely arrival of Dr. Rod Wing, who was recently appointed to IRRI as an AXA-endowed scientist for genome biology and evolutionary genomics."

The AXA Research Fund, which encourages scientific research that would contribute to understanding and preventing environmental, life, and socioeconomic risks, currently supports more than 400 research projects like the 3K RGP in 30 countries.

"Dr. Wing, an internationally recognized leader in the fields of plant genomics, bioinformatics, and evolutionary biology and particularly interested in genome data, is now part of the IRRI team and he will greatly enhance our capability to analyze the 3K RGP sequence data for gene and trait discovery," says Dr. Leung.

During a seminar that concluded an exploratory visit to IRRI in April-May 2014, Dr. Wing said, "Enormous challenges exist in bridging the gap from the discovery of genotypic and functional diversity to the release of superior varieties to farmers. So, it is a privilege to be at IRRI over the next five years because I want to help translate the rice genome biology of the 3K RGP into practical solutions for rice farmers. This will involve developing high-quality reference genomes for the five distinct varietal groups of rice, namely, tropical japonicas, indicas, aus/boro, basmati/sadri, and intermediates."

Beyond the 3K RGP, Dr. Wing is also interested in working with IRRI scientists to tap into the potential gold mine of wild rice diversity by analyzing genome sequence data of some of the more than 6,000 unexploited accessions of these species in the IRGC.

More genomes can be sequenced if necessary

"After reviewing the results coming out of the 3,000 genomes currently sequenced, we will determine whether we can identify a significant number of critical genes for use in rice improvement," concludes Dr. Leung. "At that point, we will decide how many more of the remaining nearly 180,000 accessions in the IRGC and CAAS genebank we may need to sequence and analyze."

Mr. Hettel is editor-in-chief of Rice Today.

See a video clip about AXA scientist Rod Wing at <http://youtu.be/J4oUQV0gFPA>.



THE CORE 3K RGP team at IRRI involves (left vertical row from top) Nikolai N. Alexandrov, Ma. Elizabeth B. Naredo, Myla Christy Rellosa, Flora C. de Guzman, and Ruairaidh Sackville Hamilton; (right vertical row from top) Renato A. Reaño, Hei Leung, Grace Lee S. Capili, Ramil P. Mauleon, Kenneth McNally, and Sheila Mae Q. Mercado. Jauhar Ali is not pictured.



DR. WING, a professor at the University of Arizona and director of the Arizona Genomics Institute, has joined IRRI as an AXA-endowed scientist working with the 3K RGP team.