

# DECIPHERING THE CODE

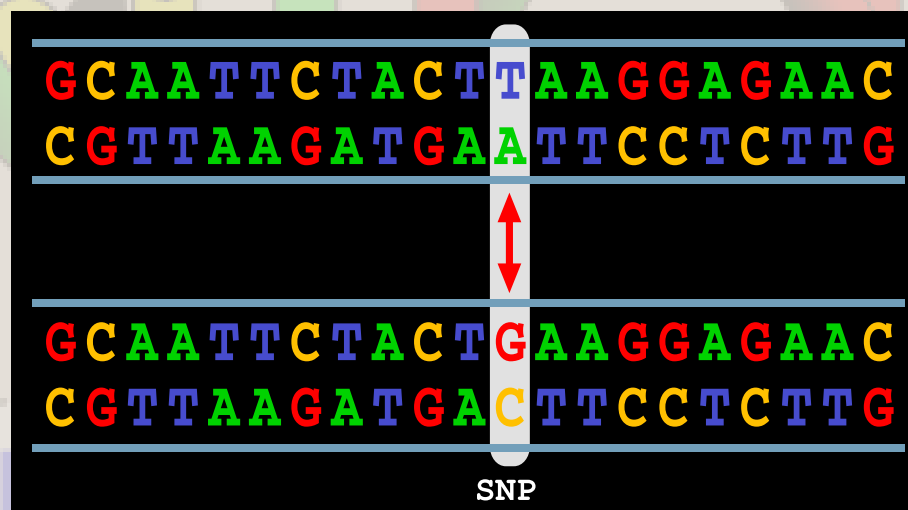
by Jill E. Cairns and Ken McNally

An international drive to generate data on tiny genetic differences will help scientists develop high-yielding, high-quality rice varieties that can better withstand pests, diseases, and environmental stresses

Variation is the spice of life. In humans, plants, and animals, genetic variation within species is immense. This variation, or genetic diversity, is a result of different mixes of genes and is fundamental to the adaptation and survival of species. When the environment of a species changes—because of the occurrence of a new disease or fluctuations in rainfall patterns, for example—the small gene variations occurring in a population are essential for its adaptation. The more diverse the population, the more likely that individuals will exist that can cope with a changed environment or withstand a new pest or disease.

Low genetic diversity was a contributing factor of the infamous 1840s potato famine in Ireland. At this time, Irish farmers were reliant on, essentially, a single potato variety. The lack of diversity had a devastating consequence when an outbreak of potato blight fungus hit Ireland, destroying the vast majority of the potato crop and causing an estimated 1 million deaths from starvation and disease. With approximately the same number emigrating, Ireland's population fell by around 20%.

The fundamental basis of genetic diversity lies in one particular class of difference in the sequence of the nucleotides—named adenine (A), cytosine (C), guanine (G), and thymine (T), these are also referred to as bases—that form the DNA within the cells of organisms. These tiny differences in DNA sequence, known as single nucleotide polymorphisms (SNPs, pronounced “snips”), are the smallest level of genetic variation.



A SINGLE nucleotide polymorphism (SNP) is a difference between one pair of DNA nucleotides.

A difference in only one DNA base (see figure, above) can cause a large difference in physical appearance or behavior (this physical manifestation of an organism's genetics is known as its phenotype). For example, in humans, a single SNP in one gene is responsible for red hair associated with people from Celtic descent. SNPs in the genetic code of other genes have been associated with the stickiness of ear wax and patterns of baldness.

Genetic diversity provides a potential gold mine for scientists to explore. With the steadily increasing human population, decreasing land availability, climate change, and outbreaks of pests and diseases, there is enormous pressure to create high-yielding, high-quality rice varieties with increased tolerance of abiotic stresses (such as drought, flooding, and salinity) and biotic stresses (pests and diseases). The challenge is tough, yet within the rice gene pool lies incredible DNA

diversity that can be exploited to develop new varieties that can help not only feed the billions of people dependent on rice as their staple but also lift millions out of poverty.

The challenge is to decipher the code and identify key alleles (different forms of a gene) that can increase yield, improve quality, and provide resistance to pests and tolerance of harsh environments. Rice has an estimated 50,000 genes encoded in its genome, which comprises 400 million base pairs. If translated into text, the rice genome would take up the equivalent space of 130 copies of Tolstoy's epic *War and peace* (one copy has around 500,000 words). Detecting these minute, yet important, differences in DNA sequence is an immense undertaking.

In 2003, an ambitious research project was undertaken by International Rice Research Institute (IRRI) molecular geneticist Ken McNally to look at genetic variation

DAVID HALL (GRINGER)



DR. CAIRNS (left) and Dr. McNally test rice plants for desirable traits such as drought tolerance.

within 10 genes related to plant performance under drought stress. Lack of adequate water during rice production is a major factor reducing yield, particularly in poorer areas, where farmers rely on rainfall. Worldwide, research to understand the genetic basis of drought tolerance in rice and other species has identified many genes (known as candidate genes) thought to influence plant response to water shortages. Dr. McNally's vision was to relate variation (in the form of SNPs) within key genes to plant performance in the field and to link these genes to yield, the most important trait for farmers.

Over 3 years, a team of 15 researchers worked together at IRRI to document DNA variation within 10 genes in more than 1,500 diverse rice varieties and measure simultaneously performance under drought stress in the field in the same varieties. The results have linked specific sequences of genes to performance under field stress, confirming the potential of such an approach to identify alleles suitable for improving drought tolerance. For example, unique variations within the gene that codes for the protein trehalose-6-phosphatase (a regulator of sugar metabolism) were discovered within a few plants and, in some cases, were found to be associated with improved performance under water-limited conditions.

For Dr. McNally, these results confirmed the potential of such an approach but realized an important

limitation. “Over 3 years, we were only able to look at 1/6250th of the estimated total rice genes,” he explains. “At this rate, we would need 18,750 years to complete the task!”

Not being a man to avoid a challenge, Dr. McNally, along with collaborators Hei Leung, David Mackill, Richard Bruskiewich, Jan Leach, and C. Robin Buell, organized an international consortium, named OryzaSNP. The consortium set about raising the funds and applying the proven “array-based” technology of Perlegen Sciences, Inc. to sequence a significant proportion of the unique (non-repetitive) regions of the rice genome within key rice varieties. By mid-2007, the sequencing of 20 rice varieties had been completed.

With the results delivered by Perlegen, and by applying additional tools for SNP prediction from the teams of Detlef Weigel and Gunnar Raescht at the Max Planck Institute in Germany, the rice community now has access to a database containing more than 400,000 SNPs. For the 100 million base pairs of the rice genome, the project identified an average of 4.2 SNPs per thousand DNA base pairs. Further, OryzaSNP is collaborating with Carlos Bustamante's team at Cornell University in the USA to identify regions of shared SNPs between different types of rice and other data that can help uncover population structure and genetics of rice.

While the results of this project,

representing about 30 gigabytes of data, were remarkable, the ability to identify unique SNPs associated with high yield and performance under abiotic and biotic stresses is limited because it includes results from only 20 of rice's tens of thousands of varieties. Nevertheless, these data provide a foundation upon which a second phase of SNP genotyping will be laid. An international team of partners under the International Rice Functional Genomics Consortium is being formed to use the OryzaSNP and other sources of SNP data to create a high-density SNP genotyping platform similar to technology used in human genetics and capable of identifying millions of SNPs. This technology will be used to genotype a collection of more than 2,500 rice varieties representative of global rice diversity. The resulting dataset will create a rice “hapmap”—a map of the regions in the rice genome that pinpoint key genetic differences. Additionally, these materials will be phenotyped in great detail for as many traits as possible. Together, the two data sources will open a treasure trove of information, enabling the discovery of specific alleles and genetic regions associated with particular traits. The knowledge will boost breeding programs worldwide.

What started as an ambitious idea to link tiny differences in a handful of important genes to performance in the field is surpassing all expectations. The stakes have been raised to look at millions of genotypes in more than 2,500 rice varieties involving a large international team of researchers. The potential is enormous, both in terms of improving rice production and simply learning more about the grain that feeds half the world. It brings to Dr. McNally's mind a quote by Albert Einstein: “The important thing is not to stop questioning. Curiosity has its own reason for existing.”

Dr. McNally works in IRRI's T.T. Chang Genetic Resources Center and Dr. Cairns, a plant physiologist, works in IRRI's Crop and Environmental Sciences Division.