Title

THE RICE GENOME: The Most Precious Things Are Not Jade and Pearls...

Journal

Science, 296(5565)

ISSN

00368075 10959203

Authors

Ronald, P.
Leung, Hei

Publication Date

2002-04-05

DOI

10.1126/science.1071337

Peer reviewed

their talent and their genius. All of that, however, requires liberating the mind from the tyranny of intolerance, bigotry, and fear, and opening the doors to free inquiry, tolerance, and imagination.

With centers of excellence in the developing world, there can be real partnerships between North and South. The promise of science can be fulfilled to make the new century one free of hunger and of absolute poverty, accurately described as a condition beneath any definition of human decency. All of that, however, requires our joint commitment as scientists to work for the benefit of the entire human family, not just the privileged minority who are lucky enough to live in the most advanced industrial societies. These tasks are enormous. But the longest journey starts with a single step. So let us start. If not us, who? If not now, when?

References

1. J. Bronowski, *Science and Human Values* (Harper and Row, New York, 1956).
2. R. D. Putnam, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton Univ. Press, Princeton, NJ, 1993).

PERSPECTIVES: THE RICE GENOME

The Most Precious Things Are Not Jade and Pearls...

Pamela Ronald and Hei Leung

"The most precious things are not jade and pearls but the five grains." The five grains referred to in this Chinese saying are most likely to be rice, wheat, millet, sorghum, and maize (*1*). These cereal grains account for up to 60% of the calories consumed by people in the developing world (*2*). We could also apply this saying to the valuable genetic information that cereals contain—especially rice. With a genome significantly smaller than those of other cereals, rice is an excellent model for genetic and molecular studies (*3*). The publication of draft genome sequences of two major subspecies of rice (*indica* and *japonica*) on pages 79 and 92 of this issue (*4*, *5*), provides a rich resource for understanding the biological processes of plants and promises to positively impact cereal crop production.

If the world's population continues to grow as predicted for the next 20 years, global cereal yield must increase 80% over the 1990 average to feed these additional people (*6*). Compounding the problem is that areas of productive farmland continue to be lost through urbanization and degradation of existing agricultural soils (*7*). Although achieving food security will require a multitude of social and economic solutions, the new knowledge derived from genomics research will make an important contribution. The challenge ahead for the plant research community is to design efficient ways to tap into the wealth of rice genome sequence information to address production constraints in an environmentally sustainable manner.

Taxonomically, all cereals belong to one of the two major groups of flowering plants: the monocotyledonous plants (monocots). Completed in 2000, the genome sequence of the weed *Arabidopsis thaliana* provided our first complete view of the genome of a dicotyledonous plant (*8*). With the availability of the rice genome sequence, we can now directly compare the genome of a monocot to that of a dicot and to genomes of other sequenced organisms. A significant observation is that over 80% of *Arabidopsis* genes have close counterparts (homologs) in rice whereas only 50% of rice genes have homologs in *Arabidopsis*, suggesting that all rice genes are essentially a superset of *Arabidopsis* genes (*4*). Furthermore, at a significant level of similarity, 85% of proteins examined in cereals have a related protein in rice (*3*). This observation poses some interesting questions regarding what the additional rice genes do. Assuming functional conservation, the extensive DNA sequence similarity between rice and other cereals will provide a short cut to the isolation of genes of agronomic importance in cereals as well as in other crop species. Thus, genomewide analyses affirm that rice is indeed a model species for cereal research with practical applications in both monocots and dicots (see the Perspective by Bennetzen on page 60).

Comparative genomic analysis enables biologists to assign a tentative function to a gene according to what that gene does in another species. For instance, the rice genome sequence reveals a network of genes encoding phosphate transporters, first identified in yeast, that are likely to be important for uptake of this macronutrient from soils (*5*). Genes controlling disease resistance, tolerance to abiotic stresses, or synthesis of essential vitamins can also be predicted by comparative genome analysis (*8*–*10*). This information facilitates the formulation of clearly defined hypotheses regarding which genes govern specific biochemical and metabolic pathways. Experiments can then be designed to determine whether the gene of interest has the predicted contribution to that pathway. For example, the presence of candidate sequences for phosphate transporters can be tested for correlation with phosphate-uptake efficiency in rice populations exhibiting variability for this trait (*11*).

The task of the protein encoded by the candidate gene can be further validated by whole plant approaches, typically by overexpressing the gene of interest (by hooking it up to a strong regulatory domain) or by knocking out its function, a field of study called reverse genetics (see the figure). For example, if a gene is hypothesized to govern disease resistance, overexpression of the gene or disruption of its activity should lead to a detectable alteration in resistance to disease, thus confirming the original prediction. Large collections of

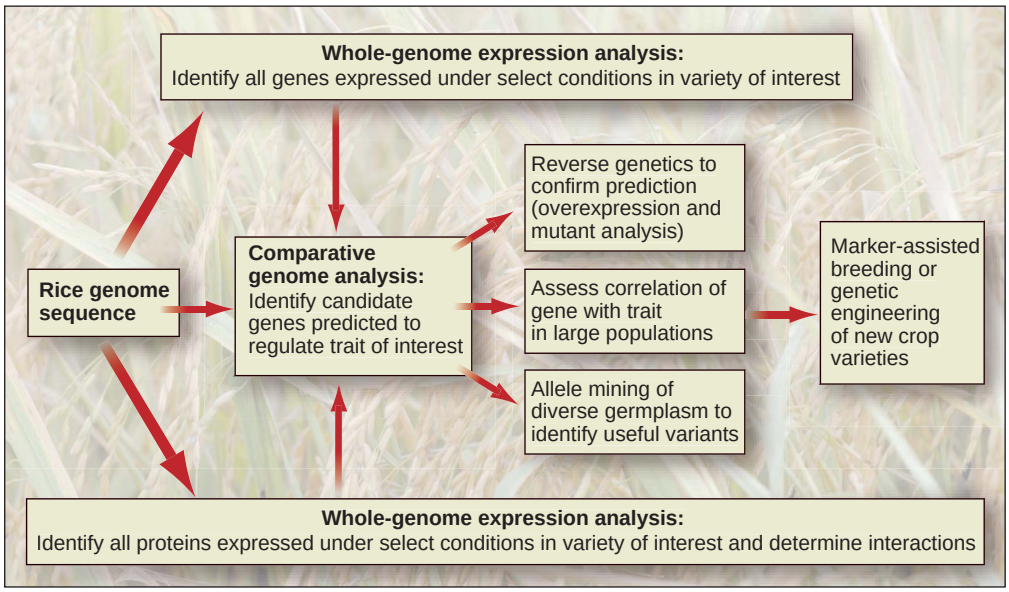
P. Ronald is in the Department of Plant Pathology, University of California Davis, Davis, CA 95616, USA. E-mail: pcronald@ucdavis.edu H. Leung is at the International Rice Research Institute, DAPO Box 7777, Manila, Philippines.

mutant lines of rice and *Arabidopsis* are available for such studies (*12*–*16*). Sequence-based databases are rapidly becoming available for several of these collections such that seed stocks carrying an overexpressed or mutated gene can easily be identified by an electronic search with sequences corresponding to the gene. Such sequence-guided hypothesis testing may well be the engine for future biological discoveries.

In practice, plant improvement will continue to rely largely upon the accumulation of genes with moderate effects. Once the function of a gene is verified, new plant varieties can be developed by introduction of the gene through traditional breeding in combination with marker-assisted selection or direct engineering of the gene into rice or other grains (see the figure). There are also opportunities to achieve major changes in gene expression through manipulation of plant transcription factors. The cold/salinity/drought tolerance and accelerated generation time achieved by controlling gene expression of the *Arabidopsis* CBF1/DREB1 and LEAFY transcription factor proteins, respectively, are good examples (*17*–*20*). There are about 1300 proteins in rice and 1700 in *Arabidopsis* with similarity to known classes of transcription factors that could be targeted for this approach (*5*, *8*). Understanding how transcription factors regulate gene expression in plants offers exciting possibilities for engineering control of seemingly complex biological processes.

Finally, knowing the sequence of specific genes will allow us to tap into the natural genetic variation of crop species. In rice, there are over 100,000 accessions of traditional rice varieties and wild species (together referred to as germplasm) collected from a broad range of geo-climates and held in trust in the International Rice Research Institute Genebank (see the Perspective by Cantrell and Reeves on page 53). These rice seeds serve as a pool of "natural variants" with the advantage that some of these variants (alleles) have already been "tested" through years of natural or artificial selection under different environmental conditions (*21*). To date, this wealth of information has remained largely untapped owing to the difficulty of identifying agronomically important genes. Now, if a gene has been proven to contribute to a trait of agronomic importance, alleles of this gene can be examined from multiple varieties for their relative usefulness. For this purpose, the highly accurate genome sequence (<1/10,000 error) being generated by the International Rice Genome Sequencing Project (IRGSP) will be invaluable (*22*). The convergence of the different versions of the genome sequence by the end of 2002 will yield great insights into the relation between sequence diversity and functional diversity in a wide variety of germplasm, the foundation on which agricultural productivity depends.

The availability of the rice genome sequence will now permit identification of the function of every rice gene. Once such a catalog is complete, a greater challenge will be to analyze the behavior of the encoded proteins in particular contexts and to determine their interactions with relevant cellular machinery to "generate function at a higher level" (*23*). For example, why do genes encoding C4 enzymes in rice (a C3 plant) behave differently from their counterparts in photosynthetically more efficient plants such as maize and sorghum? Is it due to differences in the control of gene expression, compartmentalization of the proteins, or interactions with other cellular components? Answers to these and other questions are now within our reach. Applying the information to food production will require creative integrated approaches using diverse germplasm, traditional breeding, modern technologies, and future knowledge from comparative genomics.



Improving crops using genomic information. The rice genome sequence provides a starting point for dissecting the contribution of individual genes to a particular cellular function. This process requires integration of genomewide analytical tools, genetic resources, and biological knowledge of the traits of interest. High-density arrays of oligonucleotides corresponding to all of the predicted genes can be constructed. Such DNA chips can be used to identify genes expressed at a particular point in time, the transcriptome (*24*). In combination with mass spectrometry, sequence information can be used to systematically identify components of multiprotein complexes (the proteome) and localize them to cellular compartments (*25*, *26*). The function of proteins encoded by candidate genes identified by these approaches or by comparative genome analyses can be verified by reverse genetics. Comparison of multiple alleles can provide information on agronomically useful variants. Finally, this information can be used to introduce the gene of interest (or engineered variants with improved characteristics) into cultivated plant lines through breeding or genetic engineering.

References

1. *The Economist* (9 March 1991), pp. 83–84
2. FAO, *Food Balance Sheet* (United Nations Food and Agriculture Organization, Rome, Italy, 1996).
3. D. M. Devos, M. D. Gale, *Plant Cell* **12**, 637 (2000).
4. J. Yu *et al.*, *Science* **296**, 79 (2002).
5. S. A. Goff *et al.*, *Science* **296**, 92 (2002).
6. I. Serageldin, *Science* **285**, 387 (1999).
7. P. A. Sanchez *et al.*, *Science* **295**, 2019 (2002).
8. The *Arabidopsis* Genome Initiative, *Nature* **408**, 796 (2000).
9. D. Della Penna, *Science* **285**, 375 (1999).
10. J. Yale, H. J. Bohnert, *J. Biol. Chem.* **276**, 15996 (2001).
11. M. Wissuwa, N. Ae, *Plant Breed.* **120**, 43 (2001).
12. J. S. Jeon, G. An, *Plant Sci.* **161**, 211 (2001).
13. P. Greco *et al.*, *Plant Physiol.* **125**, 1175 (2001).
14. H. Hirochika, *Curr. Opin. Plant Biol.* **4**, 118 (2001).
15. X. Li *et al.*, *Plant J.* **27**, 235 (2001).
16. www.arabidopsis.org/TAIR
17. K. R. Jaglo-Ottosen *et al.*, *Science* **280**, 104 (1998).
18. M. Kasuga *et al.*, *Nature Biotechnol.* **17**, 287 (1999).
19. D. Weigel, O. Nilsson, *Nature* **377**, 495 (1995).
20. Z. He *et al.*, *Transgenic Res.* **9**, 223 (2000).
21. S. D. Tanksley, S. R. McCouch, *Science* **277**, 1063 (1997).
22. Japan Rice Genome Program. http://rgp.dna.affrc.go.jp/index.html.
23. D. Noble, *Science* **295**, 1678 (2002).
24. D. J. Lockhart, E. Winzeler *Nature* **405**, 827 (2000).
25. Y. Ho *et al.*, *Nature*, **415**, 180 (2002).
26. A.-C. Gavin *et al.*, *Nature* **415**, 141 (2002).