

Comment

Grain of truth

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.
E-mail: petsko@brandeis.edu

Published: 25 April 2002

Genome Biology 2002, **3(5)**:comment1007.1–1007.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/5/comment/1007>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Three billion people depend on rice for the greater part of their daily caloric intake. With the possible exception of wheat, which has been the cause of several major wars including, in all probability, the Trojan War (it was wheat, not Helen's face, that really launched those thousand ships), no other foodstuff has played a comparable role in human survival. In some countries the cultivation of rice has taken on almost a religious - or at least a patriotic - significance, and rice farmers enjoy political influence vastly out of proportion to their numbers. Little wonder, then, that the announcement in early April that two groups - one publicly funded and one a private company - had completed draft genome sequences of two closely related subspecies of rice made the front pages of newspapers worldwide. But I think the significance of this achievement lies not only in the scientific and agricultural consequence of knowing the first genome sequence of a cereal. There are also profound - some would say disturbing - consequences for the sociology of science in the post-genomic world.

Over a billion years ago, the eukaryotic kingdom diverged into plants, fungi and animals. Some time after that - the best guess is about 200 million years ago - the flowering plants diverged into dicotyledonous plants like *Arabidopsis thaliana* and monocotyledonous plants; the monocot cereals sorghum, rice, wheat, corn and barley diverged from their common ancestor about 60 million years ago. The monocots became the great staples of the human diet, but because of their differences in appearances most people don't appreciate how closely related they are. Gene-mapping experiments have shown that not only are most genes from any one cereal very similar in sequence to the corresponding genes from any of the others, but in most cases gene order is conserved as well. This observation was very exciting to agricultural scientists, because it suggested that beneficial properties in any one cereal - whether found naturally in subspecies or engineered - might be easy to transfer to one or more of the others.

Rice was the first of the great grains to have its genome sequenced, not because of its importance - sequencing efforts for corn, in particular, are well-established - but simply because it has the smallest genome. Its 430 million base pairs code for about the same number of genes (around 50,000) as are estimated to be in the corn and wheat genomes, which are 3 billion and 16 billion base pairs in size, respectively. For reasons that are unclear, rice is more compact by far.

It has escaped no one's attention that the number of genes in these higher plants is comparable to - and very probably exceeds - the number in the human genome. Perhaps our language needs revision: to refer to someone as being as dumb as a plant should no longer be considered a disparaging remark. Still, if the early history of genomics has taught us anything, it should be that genome size and number of genes is a poor indicator of the real complexity of an organism. In the absence of precise data concerning alternative splicing in higher plants we cannot be sure that the rice genome will give rise to as large a set of gene products as the human genome clearly does. Nevertheless, there are a lot of genes in rice. About 50% of them have homologs in *Arabidopsis*, whereas 80% of *Arabidopsis* genes have rice homologs. If this discrepancy is real, and not an artifact of annotation, it suggests that dicot genes are essentially a subset of the genes in rice (and indeed in all monocot cereals, since 98% of proteins examined in other grains have a related protein in rice). More than 50% of rice genes code for proteins whose function is unknown, so trying to find reasons for the large number of genes may be premature, but the most common explanation proffered by commentators is that plants are immobile, cannot evade predators, and so need to synthesize a host of toxic substances as defensive measures.

This seems sensible, and is probably true for many plants, but I'm not sure it's true for rice and the other cereals.

(These are, after all, edible.) I think a more likely explanation is that since plants are immobile they cannot forage for food or move to a better environment if the conditions around them deteriorate, so they need a large complement of genes that allow them to scavenge nutrients, shift their metabolism, respond to various stresses, and go into quiescence until, for example, water becomes available again. I would not be surprised to find, when more functions of rice genes become known, that these activities are much expanded in higher plants.

More than most genome sequences, that of rice has immediate relevance to the quality of life in much of the world. Possible applications include enhancing nutritional content, improving crop yield, and adding resistance to diseases and pests. These have been discussed at length in the various commentaries, in the popular press as well as in the scientific literature, that have accompanied the announcements of the completed draft sequences. What I want to consider here are the political and sociological implications of what happened.

Two groups released draft rice genome sequences at the same time (*Science* 2002, **296**:79-92 and 92-100). In an editorial, Donald Kennedy, the Editor-in-Chief of *Science*, remarks that this reflects a spirit of cooperation “too often absent in an enterprise in which competition sometimes dominates collegiality”. The sequence of the *japonica* subspecies produced by the private company Syngenta (Torrey Mesa Research Institute, San Diego, USA) is proprietary and was not deposited in GenBank at the time of publication, as is normally required for all published genome sequences. An exception was made in view of the importance of the sequence and the authors’ promise to make the data available to the scientific community over the worldwide web. The other sequence, of the *indica* subspecies, was done by a public effort centered in China and has been fully deposited in GenBank.

The decision by Kennedy and *Science* to publish the Syngenta results has been criticized by many scientists who argue that it constitutes a slippery slope for accepted standards (although the slide, if there is one, began earlier, when *Science* made the same exception for Celera’s ‘private’ human genome sequence). Kennedy defends his decision on the grounds that the good of having the information outweighs this potential hazard, but I think there’s an even stronger argument that he was right. I think the accepted standards need to be reconsidered.

Private sequencing efforts are often faster and more cost-efficient than public ones. If a suitable business model can be found for sequence-oriented companies - and this is far from certain - there are likely to be as many sequences of important genomes coming from the private sector as there are from the public sector. Lest genomics become a house of secrets, some mechanism must be found to get that information out into the community at large.

Allowing the results to be published, but with the proviso that they be made available by some convenient mechanism, even if that mechanism is not deposition in a public database, is one way to encourage such distribution. For-profit users of the data could be required to pay a license fee to access the data, while academic users would be granted free access. It would be easy to add a requirement for GenBank deposition as well, after some waiting period (six months, perhaps, or a year) that would allow the companies in question to retain some small measure of control over the results of their efforts. This is not much different philosophically from allowing companies to patent discoveries and inventions: it is easy to forget that patenting has a dual purpose, to allow the world to use the fruits of creativity and research as well as to provide exclusivity of profit for the originators. Without patent protection, companies would keep discoveries such as PCR a secret, to allow them to retain advantages over their competitors, and we would all be losers. Like all compromises, Kennedy’s decision displeased many people, but even though his rice policy goes against the grain, it contains a kernel of the wisdom we need to deal with the complex and changing world that genomics has given us.