

The Use of the Monsanto Draft Rice Genome Sequence in Research

Gerard F. Barry*

Genomics, Monsanto, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167

Monsanto recently produced a draft of a rice (*Oryza sativa*) genome in support of its ongoing internal research programs in genomics and crop improvement. The sequence data were produced primarily in the laboratories of Dr. Leroy Hood (University of Washington, Seattle) under contract for Monsanto. In addition to being one of the most important world food crops, rice serves as a useful research model. The syntenic relation between the rice genome and those of other cereal grasses has been widely recognized and the availability of detailed information about the rice genome will likely advance global efforts to improve other major food crops in the grass family, including corn (maize), wheat, barley, sorghum, millet, and others. The increased research in rice may be expected to have many direct and indirect benefits for this critical crop and its relatives, and ultimately may contribute to an improvement in food production in many world areas. The International Rice Research Institute (Los Banos, The Philippines) estimates that 1 billion new rice consumers will be added in Asia by 2020. By that date, 4 billion people—more than one-half the world's population—will depend on rice as a staple of their diet.

The Monsanto draft rice genome sequence was produced from rice cv Nipponbare, a japonica variety that has been used widely in genetic studies in rice—the same cultivar used by the International Rice Genome Sequencing Project (IRGSP; <http://rgp.dna.affrc.go.jp/Seqcollab.html>). Formally launched in February 1998 to complete and publish the entire genome sequence of rice, IRGSP members have agreed to use a single germplasm, to share materials and information, to immediately make public completed portions of the genome, to achieve certain accuracy standards, and to coordinate their work.

In early 2000, Monsanto decided to share the draft sequence data with the IRGSP to benefit this 11-member consortium in reaching its primary objective sooner—the production of a complete and finished sequence of the genome of rice cv Nipponbare. Monsanto's sequence data and supporting materials have been made available to the IRGSP through the Japanese Ministry of Agriculture, Forestry, and Fisheries, the lead agency in the IRGSP, which distributes the company's data to members according to their assigned chromosome. The work of the IRGSP is ex-

pected to continue, incorporating data from Monsanto's draft sequence, until the entire rice genome sequence has been completed. As each segment (in units of P1 artificial chromosomes or bacterial artificial chromosomes [BACs]) of the sequence is completed, it will be placed in public databases, in accordance with existing IRGSP policy. The availability of the draft sequence and the related materials is expected to reduce both time and cost for the IRGSP.

Monsanto has established a website, <http://www.rice-research.org>, that provides registered researchers around the world access to the company's rice genome sequence data. The Monsanto Rice Genome Sequencing project produced approximately 399 Mb of sequence from 3,391 BACs. Small amounts of contaminating sequence were removed from this "initial" dataset and redundant sequences were assembled into longer contigs. This process produced the "draft" dataset totaling 259 Mb of assembled sequence data. Some redundancy may remain, for example, where sequence quality at the ends of contigs was not high enough to determine if overlaps were present. Both the "initial" dataset and the "draft" dataset are available for performing BLAST searches. The "initial" dataset consists of the original contigs of the BAC sequencing project that have not been assembled; therefore, the contig lengths are shorter than in the "draft" dataset. In addition, the "initial" dataset still contains redundancy and contaminants.

The data are provided at no charge. To access the <http://www.rice-research.org> database, principal investigators and their institutions complete a Database Registration Agreement (available on <http://www.rice-research.org>). User names and passwords are assigned for use by principal investigators and their lab members. Because Monsanto supports global agricultural research to improve rice and other important food crops and such advances are aided by the dissemination and exchange of information, Monsanto encourages the publication of results derived from the use of this rice genome sequence database and research results obtained from the use of <http://www.rice-research.org>, including information obtained from a search of the database. These results may be published in scientific publications or presented in open public forums. There are more than 300 registered users and the site has around 1,000 visits per day.

* E-mail gerard.f.barry@monsanto.com; fax 314-694-1671.

Monsanto recently completed an analysis of its rice genome sequence and produced a file containing all of the di-, tri-, and tetranucleotide simple sequence repeats of 24 bp (or greater) in the draft sequence. These simple sequence repeats are the parts of the genome predicted to be most immediately useful in rice breeding. All of these potential molecular breeding tools (around 6,650 in number and flanked by 100 bp of unique sequences on each side) have been included in this file and may be freely downloaded from <http://www.rice-research.org>. These sequences have also been deposited in GenBank. These new data are expected to help expand knowledge of rice genetics and accel-

erate breeding research in rice in research labs around the world. It is hoped that these data will benefit efforts in marker discovery and mapping in rice.

ACKNOWLEDGMENTS

I wish to acknowledge the contributions of the large number of Monsanto and University of Washington (Seattle) scientists who have contributed to the Monsanto Rice Genome Project and of these and many other Monsanto staff who have worked to make this scientific achievement a valuable resource for the public research community.