The publication of draft genome sequences for the two major subspecies of rice is a milestone for agricultural research; it could also be critical for an international project hoping to produce a finished sequence

Rice: Boiled Down to Bare Essentials

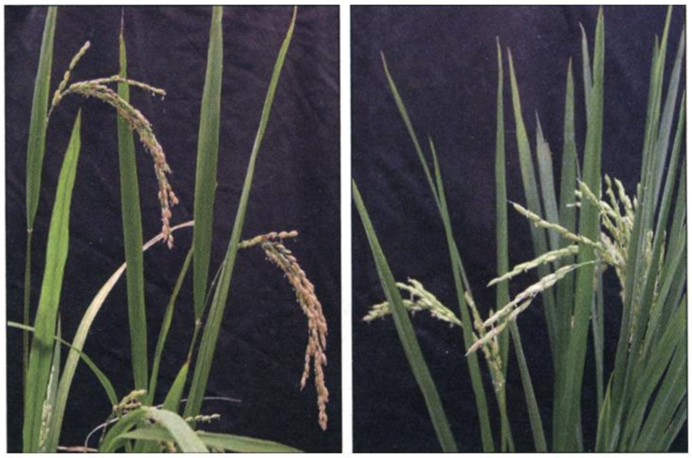
THE RICE GENOME

For more than a decade, groups around the world have been working to sequence the rice genome. The following articles trace those efforts and relate how a team in China produced a draft sequence in record time.

BEIJING, TOKYO, AND WASHINGTON, D.C.—When two groups simultaneously published rough drafts of the human genome sequence just over a year ago, the achievement was hailed as the "beginning of a new era of biology." This issue of *Science* contains two research articles that herald a similar transformation for the agricultural sciences.

On page 79, Yang Huanming of the Beijing Genomics Institute (BGI) and colleagues describe a draft sequence of the *indica* subspecies of rice, the most widely cultivated subspecies in China and most of the rest of Asia. And on page 92, Stephen Goff of the Switzerland-based agrobiotechnology giant Syngenta and his team report a similar achievement for the *japonica* rice subspecies, which is favored in Japan and other countries with temperate climates. Both sequences are works in progress: They contain many gaps and errors. But they provide the first detailed look at the genetic blueprint of a crop that is a staple for more than half the world's population. These sequences, moreover, will produce key insights into the genetics of other major cereal crops, including maize, barley, and wheat, and they will help researchers interpret the sequence of the only other plant whose genetic code has been spelled out: *Arabidopsis*, a favorite organism of plant scientists. Unraveling the rice genome, says Michael Freeling, a plant geneticist at the University of California, Berkeley, "is a big deal."

The publication of these two drafts is only part of the story. With so much at stake, many other groups around the world have been working to sequence rice. And, as with the sequencing of the human genome, these efforts have been marked by rivalries, fears of commercial control over basic genetic data, and a controversy over the conditions under which *Science* is publishing one of the draft sequences (see sidebar, p. 34). Unlike the bitter disputes that characterized the human genome sequencing, however, there has been a high level of cooperation between the public and private groups.



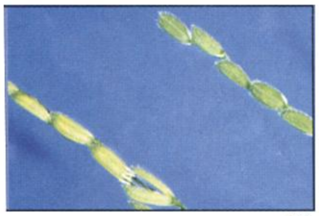
Sister genomes. Draft sequences of the *japonica* (*left*) and *indica* (*right*) subspecies of rice will boost plant comparative genomics.

CREDITS: S. MCCOUCH AND LOIS SWALES/CORNELL UNIVERSITY

Paving the way for these developments has been the International Rice Genome Sequencing Project (IRGSP), an international consortium of publicly funded labs, somewhat similar to the Human Genome Project. Led by Japanese researchers, it has been plugging away at the sequence of *japonica* for nearly 5 years. The consortium had the field largely to itself until 2 years ago, when the U.S.-based agrobiotech company Monsanto announced that it had completed a rough draft of the *japonica* sequence (*Science*, 14 April 2000, p. 239). Now, the Syngenta and Beijing groups have stolen a march by taking a bold approach to sequencing, called whole-genome shotgun, that enabled them to complete rough drafts at lightning speed.

Their achievement has, however, sown some nervousness among members of the public consortium, who worry that their effort could be jeopardized if funders believe that the sequence is now essentially in hand

Rice Genomes Timeline



IRRI

1991
Japan reorganizes a rice genome mapping project into the Rice Genome Research Program, a projected 7-year, $25 million effort largely funded by horse-racing proceeds funneled through the Japan Racing Association.

1991
Colinearity among cereal genomes is established; Michael Gale of the John Innes Centre in Norwich, U.K., concludes that "wheat is rice."

May 1993
China sets up a 5-year, $3.8 million rice genome mapping project directed by Hong Guofan at the National Center for Gene Research in Shanghai.

September 1997
Participants in a plant molecular biology meeting in Singapore agree to form an international consortium to sequence the rice genome. The Nipponbare cultivar of the *japonica* subspecies is chosen for sequencing.

February 1998
Representatives from Japan, the United States, the United Kingdom, China, and South Korea finalize policies for the consortium, now called the International Rice Genome Sequencing Project (IRGSP). Target completion date: 2008.

(see Letters, p. 45). "It needs to be understood that this is not the end," warns Rod Wing, a molecular biologist at Clemson University in South Carolina. He emphasizes that even though commercial benefits and research insights are already being wrung from the data, the drafts are just that: drafts, not finished sequences. The consortium itself hopes to complete a much higher quality draft by the end of the year and a finished sequence, with few gaps and errors, by 2005 if funding for the work continues. "For the benefit of the world, it's important that we get the [complete] genome [sequence] out to everyone as soon as possible," says Wing.

Eyeing the finish. Japan's Takuji Sasaki (in black) and his colleagues want to fill in all the gaps in their rice genome sequence.

Sequencers' beginnings

It has been a long and difficult road to get to this point. Independently, researchers in China and Japan began probing the rice genome almost 20 years ago, and Cornell University geneticists Steven Tanksley and Susan McCouch began mapping the genome in the early 1980s. But at that time, aside from a few academic labs, there was little interest in the United States and Europe, where rice is not a major crop. That began to change in the early 1990s, when researchers realized that rice is "the Rosetta stone" of cereals, says Stanford University molecular biologist Chris Somerville.

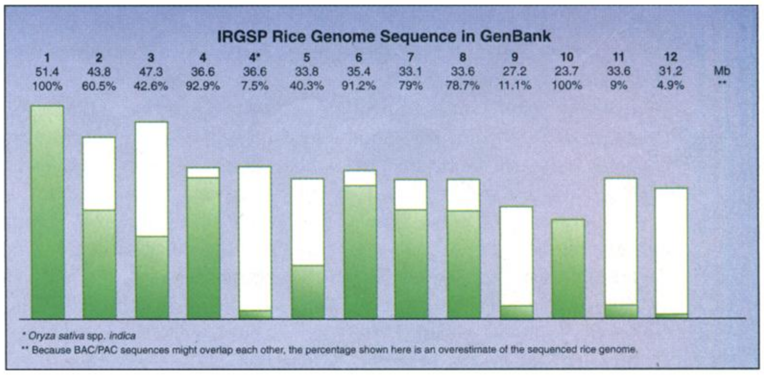
Although there is a huge disparity in the size of their genomes—rice has 430 million bases, corn 3 billion, and wheat a whopping 16 billion—cereals tend to have the same genes in the same order. That synteny, as this matchup is called, helped spark interest by U.S. and U.K. researchers, who began to view the more tractable rice genome as a tool for unlocking genetic secrets in other cereals. For the same reason, companies began to study rice more closely. "Nobody makes money on rice seeds," says Somerville. Companies are interested in rice because of the potential payoff in the sizable markets for maize, barley, sorghum, and wheat seeds, he adds.

Japan, fearing it might be scooped on something so basic to the national diet and culture, in 1991 reorganized an existing mapping project into the Rice Genome Research Program (RGP). It quickly emerged as the world leader in rice genome work, thanks to a steady flow of funds from the Japan Racing Association, which by law must donate some profits from horse racing to agricultural research. Takuji Sasaki, who took over as RGP director in 1994, proved to be a staunch supporter of an international sequencing effort.

China, meanwhile, was moving on a parallel course. In 1993, the National Center for Gene Research in Shanghai, part of the Chinese Academy of Sciences, received $3.8 million for a 5-year effort involving five labs to develop physical maps for *indica* and to prepare a library of clones for sequencing. Center director Hong Guofan said he hoped to move on eventually to actual sequencing.

The idea of an international consortium to sequence rice emerged from informal discussions in early 1997. That September, with support from the Rockefeller Foundation, researchers got together at an international plant molecular biology meeting to map out a strategy. Even though more people eat *indica*, the group agreed to focus on the Nipponbare cultivar of the *japonica* subspecies, because RGP had already done much of the preparatory work.

The first official IRGSP meeting was held just 5 months later in Tsukuba, Japan. "Everybody just seemed to be ready for this," says Ben Burr, a plant geneticist at Brookhaven National Laboratory in Upton, New York. Representatives from five participating countries—Japan, the United States, the United Kingdom, South Korea, and China—hammered out guidelines and divided up rice's 12 chromosomes. Participants agreed to release all data to public databases and to sequence at least 1 megabase a year.



* *Oryza sativa* spp. *indica*

** Because BAC/PAC sequences might overlap each other, the percentage shown here is an overestimate of the sequenced rice genome.

In the United States, the Rockefeller Foundation funded preliminary work by several academic researchers, including Wing, who received a grant to develop a library of bacterial artificial chromosomes (BACs), each one carrying a small bit of rice DNA. Subsequently, Novartis, which later became part of Syngenta, awarded Wing $3 million to ready those BACs for sequencing. The BACs and their mapping information were all made public.

But getting money for sequencing was another matter. With a supportive govern-

CREDITS: (TOP TO BOTTOM) IRGSP; SOURCE: TIGR

April 2000
Monsanto announces that, in conjunction with the University of Washington, Seattle, it has produced a draft of the rice genome. The company promises to share the data with individual researchers and IRGSP.

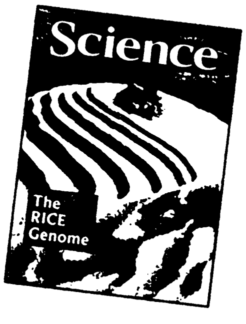
May 2000
The Beijing Genomics Institute (BGI) announces that it plans to sequence the *indica* subspecies using the whole-genome shotgun technique. It promises a draft within 2 years.

January 2001
Syngenta, working with Myriad Genetics, announces that it has used the whole-genome shotgun approach to sequence Nipponbare rice and has produced a draft with six times coverage.

October 2001
BGI completes its sequencing and makes raw data freely available.

December 2001
IRGSP decides to aim for a 10 times coverage draft by the end of 2002 as an interim step.

April 2002
***Science* publishes Syngenta and BGI drafts.**



ment, RGP was in the strongest position, receiving $10 million in 1998 for the first year of a projected 10-year effort. But no government sequencing funds were forthcoming in the United States until 1999, when the U.S. Department of Agriculture (USDA), the National Science Foundation (NSF), and the Department of Energy finally came up with $12.5 million to fund two groups to work on chromosomes 10 and 3. "The thing that slowed us down the most was the late entry of the U.S. funding agencies," says Burr. Other countries fared even worse. The U.K. and Canadian groups never won funding. And Thai researchers joined the consortium, contributed a small amount of sequence data, but then withdrew to concentrate their scarce funding on gene discovery.

Japan, however, picked up some of the slack, and France, Taiwan, South Korea, India, and Brazil stepped in to share the burden. Finally, by the beginning of 2000, IRGSP seemed on its way toward its goal of completing the *japonica* rice genome sequence by 2008, possibly earlier. But the international consortium soon had company.

The tortoise and the hare

In April 2000, Monsanto announced that it had sequenced the *japonica* genome. Working with researchers at the University of Washington and the Institute for Systems Biology in Seattle, Monsanto had produced an incomplete, but very informative, version that the company promised to share with individual academic researchers and with IRGSP. The news shocked, then worried, but ultimately delighted the community, because the data promised to speed up the IRGSP effort.

GENOME SIZES

Species	No. of bases
Arabidopsis	125 million
Drosophila	180 million
Rice	430 million
Maize	3000 million
Human	3000 million
Wheat	16,000 million

"The Monsanto data has been very helpful and very valuable," says Machi Dilworth, who oversees plant genomic programs at NSF. The French group sequencing chromosome 12 turned to Monsanto for 75 of the 109 clones it is now sequencing. And the Shanghai group set aside its *indica* project, picked up the Monsanto materials, and finished a draft of chromosome 4. The Monsanto material "is now proving to be of value in speeding up the IRGSP sequencing and the cost effectiveness of the overall project," says Ed Kaleikau, a plant biologist at USDA's Cooperative State Research, Education, and Extension Service.

The Monsanto boost was not decisive, it turns out. In January 2001, Syngenta reported that it, too, had sequenced *japonica*. It had contracted Myriad Genetics in Salt Lake City, Utah, to work on rice

A Deal for the Rice Genome

For the second time in just over a year, *Science* is at the center of a debate over public access to the data behind a major genome paper it is publishing. The issue: Should journals refuse to publish any DNA sequence paper unless the authors make the data freely available through a public database such as GenBank?

On page 92, a team from the Switzerland-based agricultural biotechnology giant Syngenta describes a draft sequence of the *japonica* subspecies of rice. Under an agreement reached with *Science*, the company is making the data publicly available through its own Web site (tmri.org) or on a CD-ROM, rather than through GenBank. Scientists can use the partially assembled raw genome sequence without strings for research, and Syngenta will permit researchers to publish papers and have Syngenta deposit a gene's worth of DNA data in GenBank without negotiation. (The raw data include minimal notes, an official says, such as labels on DNA likely to be "nonrice in origin.") Larger amounts will require a specific agreement. The company seeks no "reach-through" intellectual property rights, but scientists doing commercial work must negotiate their own data-access agreements.

Last year, *Science* touched off a furor when it struck a similar deal with Celera Genomics of Rockville, Maryland, as a condition of publishing Celera's draft of the human genome (*Science*, 16 February 2001, p. 1304). Celera gives noncommercial researchers free access to raw DNA sequence but charges a fee for access to its annotated gene database. Criticism in a more muted form surfaced again several weeks ago when word of a possible Syngenta agreement with *Science* began to spread in the genomics community. A score of leading researchers—including Michael Ashburner of Cambridge University, U.K., David Botstein of Stanford University, and Maynard Olson of the University of Washington, Seattle—circulated a letter arguing that failure to insist that the sequence be deposited in GenBank constituted a "very serious threat" to genomics research.

"We understand that concern," says *Science* Editor-in-Chief Donald Kennedy, noting that it would be ideal to have "one-stop shopping" for all genomic data at GenBank. But, Kennedy said at a press briefing last week, the company would have been unwilling to publish its raw data if it had been required to deposit the sequence in GenBank. "We think that the public benefit of bringing this important science out of trade secret status greatly outweighs" the cost of granting an exception, Kennedy said.

The arrangement has not so far prompted the intense reaction that greeted the Celera agreement. One reason is that Syngenta has promised to work closely with publicly funded groups to produce more complete drafts of the rice genome (see Letters, p. 45). Monsanto of St. Louis, Missouri, which produced its own draft of the *japonica* sequence 2 years ago but hasn't published it, is also cooperating in this endeavor. Members of the public consortium working with Monsanto say that 30% of the data they have released to GenBank originated from the company.

The Syngenta sequence will be useful in refining draft sequences. "Thanks to Syngenta, I don't think it will be so hard" to close gaps between the more than 100,000 fragments in the draft sequence of the *indica* subspecies—also being published this week (p. 79)—says Wong Gane Ka-Shu, a leader of the research team that sequenced *indica*. (The team's draft sequence has been deposited in GenBank.)

As a result, much of the Syngenta sequence is likely to end up in GenBank over the next "12 to 18 months," mingled with data the public groups will be depositing, says Steven Briggs, head of Syngenta's Torrey Mesa Research Institute in San Diego, California, which oversaw the company's sequencing project. Asked why Syngenta is not prepared to deposit its sequence in GenBank now, Briggs said last week that Syngenta believes it has "a significant commercial advantage" and isn't ready to permit unrestricted use of its data by its competitors.

Susan McCouch, a rice genome researcher at Cornell University, is disappointed that Syngenta's data are not going directly to GenBank. This would have made whole-genome comparisons "easy," she says, enabling more rapid discovery of gene function. Despite the decision not to deposit data in GenBank, Rod Wing of Clemson University in South Carolina has concluded that the new data-sharing terms look "very good," particularly because there are "no reach-through terms" seeking to patent scientists' discoveries.

—ELIOT MARSHALL

and other cereals for $30 million. Steven Briggs, head of Syngenta's Torrey Mesa Research Institute in San Diego, California, which managed the effort, says that IRGSP's target completion date of 2008 was too far off. "We needed the data immediately," he says.

The Beijing group, meanwhile, was eager to move ahead quickly as well, although for different reasons. In May 2000, BGI director Yang announced plans to sequence the entire genome of *indica*. Yang says that with most of the other efforts focused on *japonica*, "there was a feeling that China should sequence its own rice." The team, which was separate from the group in Shanghai tackling chromosome 4, promised to deliver a draft of *indica* within 2 years (see p. 36).

Bountiful harvest. Plant scientists are already using the rice genome sequence to improve productivity and disease resistance in rice.

CREDIT: MICHAEL FREEMAN/CORBIS

Whereas Monsanto and the public consortium took the traditional, orderly route of mapping the *japonica* genome and sequencing it piece by piece, the Syngenta and Beijing groups turned to the whole-genome shotgun method. They chopped up the entire genome into fragments, sequenced each fragment, and put the data in order with the help of powerful computers. Their successes "demonstrate how powerful a tool [the whole-genome shotgun] is," says Jeffrey Bennetzen, a plant geneticist at Purdue University in West Lafayette, Indiana. Syngenta did enough sequencing to cover the genome six times, yielding 99% coverage and accuracy. That still leaves thousands of gaps in the sequence coverage, however. The Beijing group achieved four times coverage, yielding a less complete sequence.

From their analysis, the Syngenta team estimates that the rice genome contains 32,000 to 50,000 genes, depending on how the genes are picked out. The Beijing team has picked up about 46,022 to 55,615 genes. As Bennetzen points out in a Perspective on page 60, the disparity between the two estimates probably largely reflects differences in the way the groups identified genes, and he expects that both estimates will shrink as more analysis is done. As expected, more than 80% of the *Arabidopsis* genes have counterparts in the rice genome, which should help researchers identify the functions of those shared genes.

From an analysis of coding and noncoding regions within the genes, the Beijing team concludes that each gene codes for just a single protein—unlike human genes, which often code for multiple proteins. Rice may get its diversity of proteins from an abundance of genes that acquired diverse functions after the whole genome was duplicated, an event that Goff and his colleagues estimate occurred about 45 million years ago.

The Beijing group has made all its data public, depositing them in GenBank. Already, some 350 researchers have used the data, says Yang. Syngenta is making its sequence available through its own Web site and on a CD-ROM, but it is also discussing sharing its data with IRGSP, says Syngenta's Briggs. "Details are still being negotiated," says RGP head Sasaki, "but it is likely that an agreement will be reached [that is] very similar to the agreement with Monsanto, allowing IRGSP members to use the Syngenta data to complete their work on the phase II draft." That would mean all the Syngenta data would end up in GenBank through the consortium's work. In the meantime, company researchers are collaborating with interested researchers; about 65 labs in 11 countries have made use of the information, says Goff.

What next

As news of these two drafts began circulating in the community, IRGSP was forced to rethink its goal of painstakingly closing almost all the gaps in the sequence before publishing a complete sequence. Japan's program, in particular, came "under a lot of pressure from the government to accelerate [its work]," says Joachim Messing, a molecular geneticist at Rutgers University's Waksman Institute of Microbiology in Piscataway, New Jersey. So in December 2001, IRGSP shifted gears and announced that instead of finishing the genome sequence on its original schedule, it would produce its own draft by December 2002. "We started with a certain road map, and then we had to adjust," says Robin Buell, who heads the rice sequencing effort at The Institute for Genomic Research in Rockville, Maryland.

IRGSP members have already placed the data for over 230 megabases, or more than half of the genome, in public databases, and three chromosomes are nearly finished. "I am sure we will complete our [draft] rice genome sequence by the end of 2002," says Sasaki. On average, each base will have been sequenced 10 times. This is more than enough to guarantee high accuracy, with each clone placed correctly and in the right orientation along the genome. Sequencers will then fill in remaining gaps as they finish the genome.

For many groups, the rough drafts are already providing good data. Fragmented though they are, these drafts capture 99% of the genes, notes Bennetzen. Both Monsanto and Syngenta say they are identifying genes that are important not only for rice but also for other cereals. Briggs says they have developed a microarray to study gene expression and are already moving into proteomics. "We're deep into the discovery of a lot of functional aspects of rice and other cereals," he says. Gerard Barry, a molecular biologist at Monsanto, confirms that his company, too, has used its sequence as a basic research tool to do a wide range of plant studies, to breed new varieties at an accelerated rate, and to look for genes involved in productivity and stress tolerance. And Cornell's McCouch reports, "Breeders are mining the [public] sequence data for improving traits such as grain quality and pathogen resistance."

The rough drafts are being used for more basic research as well. Susan Wessler, a molecular geneticist at the University of Georgia in Athens, is studying how plant species diverge, focusing on the role of transposable elements: blocks of DNA that have been copied from one location to another in the genome. Having the BGI data on *indica* "is absolutely fantastic," she says, as it allows comparisons of where transposable elements occur and don't occur between two closely related subspecies. "To my knowledge, it's the highest organism where there are two sequences of subspecies; it saved us literally a year of work," she says.

But for other groups, the drafts may not hold the answers they seek. Masahiro Yano, a rice geneticist at RGP, wants more precise data. His research seeks to link important agronomic traits, such as pest resistance and flowering time, to particular regions of chromosomes. Yano says that because the

IRGSP sequence data are accurately tied to maps of chromosomes, researchers can use computer programs to quickly home in on candidate genes. He has used this approach to find several genes that control flowering time in relation to length of day, and he hopes eventually to have a whole collection of genes so that researchers could control the flowering time of plants. This would allow breeders to take varieties with desirable traits and move them from northern to southern latitudes, or vice versa. "This is an ideal example of the promise of having the rice genome [sequence]," Yano says. He adds that for the rough drafts, where the sequence data are not tied to a map, "you can't do this isolation and identification so easily."

Such requirements make it imperative that the IRGSP's work be completed, argue Wing, Sasaki, and others. "I think Monsanto and Syngenta can get enough information out of the drafts to patent genes," explains Wing, "but we need to know more about the regulatory elements." But Wing and others are concerned that the public and the funding agencies will get caught up in the excitement over the rough drafts and that funds for finishing the job will dry up. In the United States, for example, there are no funds specifically set aside for finishing the rice genome sequence. Already, companies and researchers interested in maize are pressing to begin pilot sequencing of that genome. "It will be up to the community to decide" what's more important, says NSF's Dilworth. Failing to complete the genome sequencing would be a big mistake, says Wing, not just for basic research but for anyone interested in any of the cereals.

–**DENNIS NORMILE AND ELIZABETH PENNISI**

THE RICE GENOME

▶ PROFILE: BEIJING GENOMICS INSTITUTE

From Standing Start to Sequencing Superpower

Good timing and determination have helped geneticist Yang Huanming create an institute that has catapulted China into the front ranks of sequencing

BEIJING AND HANGZHOU—In August 1998, geneticist Yang Huanming led a skeptical crowd of scientists from around the world through a new, two-story brick building in the northern reaches of Beijing. As the scientists trooped through the empty building, their footsteps echoing off bare walls, Yang explained that it would soon become a world-class sequencing facility. He said that employees and sequencing machines were on the way, neglecting to mention that he didn't yet have the money for either. His colleagues were polite but dubious.

"The building had a nice double helix on the brick facade," recalls Maynard Olson, a geneticist at the University of Washington, Seattle. "But that was the only indication that this was a genome center as opposed to an empty warehouse. I really wondered if they could get the support to become an internationally competitive group."

Olson wonders no more. Today, visitors to the Beijing Genomics Institute (BGI) see 92 of the latest-model automated sequencing machines, four of the fastest supercomputers in China, and a staff of 500 that grows by a dozen or so every month. The sequencing center has moved from that tiny brick building to a spacious, modern industrial park and has spread to a second campus in the southern city of Hangzhou. And its science—including the shotgun sequencing of the *indica* rice genome reported on page 79—is certainly internationally competitive.

Olson says he always had confidence in the scientific capabilities of the group. His sequencing center has trained many BGI scientists, engineers, and technicians, and two of the four lead authors on the paper, Yu Jun and Wong Gane Ka-Shu, are on the staff of the University of Washington Genome Center in Seattle. But "it's pretty startling," Olson admits. "When you think of being a support center for a scientific program in a developing country, you don't expect them to become 10 times bigger than you are, in less than 4 years, and to start publishing papers in *Science*."

Rising son. Institute director Yang Huanming has made China a sequencing powerhouse using domestic computers from Dawning.

CREDIT: D. NORMILE

Such accomplishments no longer surprise fellow University of Washington geneticist Mary-Claire King. "The Beijing Genomics Institute would be a miracle," she says, "except that the BGI guys make genomic miracles routine."

Young and restless

The Ferrari-like acceleration from standing start to joining the global front-runners in genomic sequencing is a tale of timing, determination, and hustle. It also demonstrates Yang's ability to translate his vision into reality by tapping the increasingly diversified sources of support in a reform-minded China.

Yang, 50, is a spark plug of a man. The fact that he's considerably shorter than most of his staff would be obvious if he ever stood still. Likewise, his nonstop discourses jump from topic to topic. He sprinkles Chinese proverbs into his conversation, reciting them in Chinese and then looking around for a translator.

Born in Yueqing, Zhejiang Province, Yang earned his Ph.D. in genetics at the University of Copenhagen, Denmark. Over the next 6 years, he focused on mapping genes on the X chromosome during stints at the CNRS Immunology Center in Marseilles, France, Harvard Medical School in Boston, and the University of California, Los Angeles. In 1994 Yang returned to China with the idea of adapting to sequencing the large-scale, high-efficiency, low-cost techniques that have boosted the country's manufacturing capacity. "[Sequencing] is where a developing country can compete in big science," he says.

His target was the human genome sequencing effort that was already under way, and his intended vehicle was the Human Genome Center, a part of the Chinese Academy of Sciences' (CAS's) Institute of Genetics. But he and his colleagues realized that the academy's rules and traditions would prevent them from ramping up fast enough to join the rest of the world, and the center—