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—

the brick building the visiting scientists toured—never really got off the ground.

Instead, Yang and three colleagues who had worked with Olson at the University of Washington took advantage of new laws, and in spring 1999 they set up BGI as a private, nonprofit research organization. Seed money came from CAS, the Institute of Genetics, and even Yang's hometown municipal government, along with loans from employees, family, and friends. CAS also designated BGI as its Genomics and Bioinformatics Institute, although fewer than 10 members of BGI are actually CAS employees.

Team leaders. Deng Yajun (*right*) preaches speed and economy in sample preparation and sequencing; bioinformatics chief Wang Jun (*above*, at left) works with his mentor, Li Songgang, on shotgun assembly.

BGI bought its first batch of sequencing machines on an installment plan and trained its staff on *Thermoanaerobacter tengcongensis*, a thermophilic bacterium isolated from a hot spring in Tengcong, China. In September of that year, Yang made his pitch to be a global player at the 5th International Strategy Meeting on Human Genome Sequencing in Hinxton, U.K. There was only one question that stumped him: "Do you have the money?" "I lied," he now admits. "We didn't have the money, but I was sure we would get it."

Four months later he did. CAS agreed to fund three Chinese sequencing centers to tackle 1% of the human genome, and BGI received slightly more than half the total award. China completed its share of the draft on time and has recently closed the gaps and corrected the errors in the draft sequence.

Although the Western press barely mentioned China's participation, Chinese accounts emphasized the nation's role in this historic endeavor and its status as the only developing country in the global partnership. That participation helped Yang convince Hangzhou municipal officials to provide a rent-free building and enough money for BGI to set up and equip a sister center, the Hangzhou Genomics Institute. In return, city officials hope the institute will attract foreign high-tech investment.

The mood in the two labs is akin to that in a U.S. high school before the big game against its archrival. Posters in every room in Beijing remind employees that "Discovery can't wait!" and "Speed! Speed! Speed!" In a nod to the scientific task at hand, other posters proclaim "High throughput is everything!" The Hangzhou group prefers symbolic reminders of its mission, including a rusty hoe propped in a corner of the computer room "to remind us that we're data miners," says bioinformaticist Zhou Yan.

CREDITS: D. NORMILE

The youthful enthusiasm is no act. Excluding the dozen or so senior scientists, the average age of the 100 authors on the *Science* paper is in the mid-20s. Wang Jun, who leads 150 programmers and computer scientists in the bioinformatics department, entered Beijing University at 16 and is now 26. "This is a new field," he says, "so I don't think I'm too young for this job."

The head of the lab's 100-person sample preparation and sequencing group is 29-year-old Deng Yajun. A forensic investigator for the police department in her hometown of Xi'an, Deng began working part-time at BGI while studying for her Ph.D. in forensic medicine. "The policewoman," as her colleagues call her, quickly mastered the techniques and took charge of the group.

Yang likes to brag about BGI's parsimony. When BGI first started, it imported the 96-well plates used to hold the DNA samples during preparation for $2.54 apiece. Then Deng found a local glassmaker who could make them for just 36 cents. The same cost advantages apply to reagents used in sample preparation and to salaries. "The key to driving down costs is to scale up so we can negotiate better prices with suppliers," says Deng.

Shotgun success

Yang says his "addiction to sequencing" has helped the institute achieve that economy of scale. With work on the human genome sequence well under way, Yang landed two major sequencing jobs in 2000 that broadened the center's scope. In April 2000, he convinced CAS to support sequencing *indica* (*Science*, 5 May 2000, p. 795). Then in October, BGI reached an agreement with the Danish Pig Genome Consortium, a joint public-private endeavor, to sequence the pig (*Science*, 3 November 2000, p. 913).

For both projects, BGI chose a sequencing technique known as the whole-genome shotgun. Traditionally, sequencing has required mapping the genome, then developing a library of clones, or relatively short strings of DNA, tied to a known location, which are then sequenced. In the shotgun approach, the entire genome is broken into pieces and sequenced. Then powerful computers and sophisticated software sort the pieces into the proper order.

Shotgunning appealed to BGI officials because the maps and clones needed for the traditional method were not available for *indica* and would have taken years to prepare. Just as importantly, "we felt that mastering the shotgun technique would be the next step in our development," says Liu Bin, BGI's chief of research and collaborations.

The BGI scientists faced a big challenge in developing the computer expertise to pull off the whole-genome shotgun. Initially barred from acquiring U.S. supercomputers because of security-related export restrictions, BGI became one of the best customers of Dawning, a homegrown supercomputer maker. Particularly critical was the so-called assembler, which pieces together the data from the sequenced fragments. Whereas the pig project can use the human genome as something of a template, rice required its own assembler. "The assembler was our biggest worry," Yang says.

Rather than starting completely from scratch, BGI decided to modify the assembly program used in the public Human Genome Project, called phrap. A key step was developing a subroutine that would identify and temporarily mask repetitive strings, which make up 40% of the rice genome. By temporarily ignoring such repetitive sequences, the BGI assembler reduced the chance of making mistakes in stringing together sequence data and dramatically cut the computing time required. Li Songgang, a professor of bioinformatics at Beijing University and senior adviser to BGI, developed an algorithm to search for such strings, and the BGI group tested it against virtually all the sequence data for all organisms publicly available. The effort took nearly a year.

Once the group was convinced that it would work, BGI pulled out all the stops to beat the competition. In particular, Yang was taking aim at Syngenta Corp., which had announced that it had completed a draft of the rice genome sequence in early 2001 but had never made the data public. Starting last July, the sequencing team was split into two, 12-hour shifts. That allowed BGI to keep the machines running 24 hours a day for the 74 days needed to complete the actual sequencing. Then came the analysis team, working practically around the clock. "It was like going into battle," says bioinformatics head Wang.

Most of the Beijing staff lives in a housing development barely a kilometer down the road from the center. But to save commuting time, mattresses and sleeping bags were spread out in the halls. And some people didn't even make it that far. "I just slept in my chair," says Han Yujun. Ping-Pong was the only diversion from the data crunching. On 8 October the paper was sent off to *Science*.

The completion of the draft is certain to put BGI on the sequencing map. Olson calls it "a major accomplishment." Adds plant geneticist Jeff Bennetzen of Purdue University in West Lafayette, Indiana, "the Chinese showed how quickly you can do this if you take a modern approach."

Yang has drawn up a long list of organisms he thinks should be sequenced, and the institute has already dipped its toe into proteomics and drug discovery, including a project to isolate the active compounds in the herbs used in traditional Chinese medicine. "Diversification is a challenge all of the sequencing centers must face," says Olson.

So is finding money to take the next step. Chen Zhu, CAS vice president for life sciences, says BGI is free to compete for project grants but that "we think the big genome sequencing projects should depend on international cooperation." Yang isn't saying where he plans to get the institute's next round of funding. But after going from an empty building to a paper in *Science* in 40 months, Yang is confident that he hasn't run out of miracles.

–DENNIS NORMILE

With reporting by Ding Yimin and Elizabeth Pennisi. Ding Yimin writes for *China Features* in Beijing.

MATHEMATICS

Erdös's Hard-to-Win Prizes Still Draw Bounty Hunters

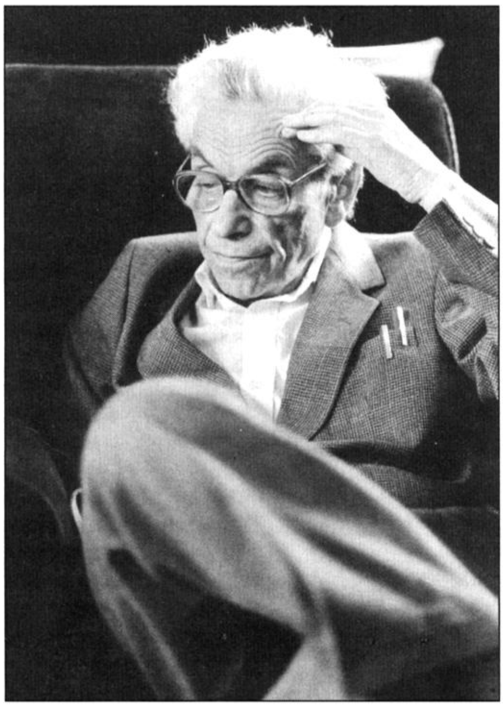
Half a decade after his death, a colorful mathematician continues to spur his colleagues with challenges—and checks

It takes more than death to keep a good mathematician down. Among many other things, Paul Erdös proved that. In life, the world's premier number theorist supported himself by wandering from colleague to colleague, freeloading off friends and strangers, and, in return, sharing his vast mathematical insight with all comers. His 6-decade frenzy of near-nonstop work resulted in more than 1500 papers that link him to almost every academic mathematician in the world. Erdös's death in 1996 has slowed, but not stopped, his publication rate: Over the past 5 years, journals have published some 62 new papers bearing his name. And he is still guiding the research of mathematicians with the problems he left behind.

Problems, problems, problems. Erdös hurled them out relentlessly, in lectures, papers, and informal talks: problems in number theory, logic, graph theory, geometry, combinatorics, and other disciplines. Every mathematician agrees on their importance. Yet nobody has listed them all; nobody even knows how many there are.

"I'd say it's in the small number of thousands," says Ronald Graham, a mathematician and computer scientist at the University of California (UC), San Diego, who managed many of Erdös's affairs in the last few years of Erdös's life. András Hajnal, a set theorist at Rutgers University in New Brunswick, New Jersey, says he can name at least 100 in set theory alone.

As remarkable as their variety are the incentives Erdös attached to them. Early in his career, Erdös began to offer small prizes for solutions to his problems. "They were $10, $25, $100. It was a way to stimulate people and calibrate how difficult he thought the problem was," Graham says. "Some prizes were larger: There is a famous $3000 one, and there's kind of a $10,000 one, but it's not well stated. There are hundreds and hundreds of such problems." Alert listeners jotted the figures down in lecture notes; compliant journal editors let Erdös publish them in his papers.



Puzzle master. Although he once called himself "a device for turning coffee into theorems," Paul Erdös was equally adept at posing problems.

CREDIT: GEORGE PAUL CSICSERY

During Erdös's lifetime, the mounting most-wanted list could reap unexpected windfalls for his colleagues. "Some years ago, he was visiting in Athens, Georgia. He was going back into town, and my colleague Helmut Maier was giving him a ride," says Carl Pomerance, a mathematician at Lucent Technologies' Bell Labs in Murray Hill, New Jersey. The conversation naturally turned to mathematics and to a theorem that Maier had just proven. " 'Maybe I offered a prize for that,' said Erdös, and instead of going to town, they went to the library," says Pomerance. Sure enough, Erdös had offered $100 for that particular problem in a math journal, and he paid up. "I said to Erdös, 'That's a pretty expensive taxi ride,' and he found that hilarious," adds Pomerance.

Collecting a bounty could be tricky. Erdös often changed his mind about a problem's value, stating one sum in a lecture and another in print, and he could alter the figure on a whim based on his own sense of aesthetics. "I solved a $250 problem, but I only got $50 because Erdös didn't like the proof," says Hajnal. Hajnal's proof, which had to do with partitioning the real numbers into different sets, hinged upon a logical trick rather than any special properties of the numbers. Erdös also abhorred proofs related to Gödel's incompleteness theorem, which states, in essence, that certain propositions are unprovable and unfalsifiable. As a result, there are statements that one can simply declare to