

THE SEQUENCE FACTORY

The bold ambitions of one institute could make China the world leader in genome sequencing. **David Cyranoski** asks if its science will survive the industrial ramp-up.

In 2006, Li Yingrui left Peking University for the BGI, China's premier genome-sequencing institute. Now, freckled and fresh-faced at 23 years old, he baulks at the way a senior BGI colleague characterized his college career — saying Li was wasting time playing video games and sleeping during class. “I didn't sleep in lectures,” Li says. “I just didn't go.”

He runs a team of 130 bioinformaticians, most no older than himself. His love of games has served him well when deciphering the flood of data spilling out of the BGI's sequencers every day. But “science is more satisfying” than video games, he says. “There's more passion.”

The people at the BGI — which stopped officially using the name Beijing Genomics Institute in 2007 after moving its headquarters to Shenzhen — brim with passion, and an ambition so naked that it unsettles some. In the past few years the institute has leapt to the forefront of genome sequencing with a bevy of papers in top-tier journals. Some recent achievements include the genomes of the cucumber¹, the giant panda², the first complete sequence of an ancient human³ and, in this issue of *Nature*⁴, the genomes of more than 1,000 species of gut bacteria, compiled from 577 billion base pairs of sequence data.

The mission, BGI staff say with an almost rehearsed uniformity, is to prove that genomics matters to ordinary people. “The whole institute feels this huge responsibility,” says Wang Jun, executive director of the BGI and a professor at the University of Copenhagen. The strategy is to sequence — well, pretty much anything that the BGI or its expanding list of collaborators wants to sequence. It has launched projects to tackle 10,000 microbial genomes and those of 1,000 plants and animals as part of an effort to create a genomic tree of life covering the major evolutionary branches. Important species, such as rice, will be sequenced 100 times over, and for humans there seems no limit to the number the institute would like sequenced.

To fulfil that mission, the BGI is transforming itself into a genomics factory, producing cheap, high-quality sequence with an army of young bioinformaticians and a growing arsenal



D. CYRANOSKI

The BGI's sequencing room, where thousands of projects will contribute to building a genomic tree of life.

of expensive equipment. In January, the BGI announced the purchase of 128 of the world's newest, fastest sequencers, the HiSeq 2000 from Illumina, each of which can produce 25 billion base pairs of sequence in a day. When all are running at full tilt, the BGI could theoretically sequence more than 10,000 human genomes in a year. This puts it on track to surpass the entire sequencing output of the United States, says David Wheeler, director of the Molecular Biology Computational Resource at Baylor College of Medicine in Houston, Texas. “It is clear there is a new map of the genomics world,” he says.

The charge that the BGI has reduced science to brute mechanization does little to ruffle feathers in Shenzhen. Wang himself quips that the BGI brings little intellectual capital into projects: “We are the muscle, we have no brain.” But such comments belie a quiet confidence, in everyone from the BGI's seasoned management to its youngest recruits, that they can make an impact not just to the balance of sequencing power but also in biology, medicine and agriculture. This will be a challenge given the significant loans taken out to expand capacity. Torn between scientific and financial goals, even its founder can't seem to

decide whether the BGI is a business or a non-profit research institute. Genome scientists around the world are watching to see how it will strike a balance. Edison Liu, director of the Genome Institute of Singapore and head of the Human Genome Organization warns: “If they are just a sequence-for-money operation, they will not be remembered.”

Getting far from the emperor

China was late to the genomics frenzy of the 1990s that led to the sequencing of the human genome. The fact that the country didn't miss out altogether is thanks largely to the BGI's determined, charismatic and sometimes abrasive leader Yang Huanming (‘Henry’). As the human genome project was nearing completion, Yang and a small group of sequencing advocates tried to get China involved. They found support from the Chinese Academy of Sciences (CAS), which secured a building and a start-up fund of 1 million renminbi (US\$150,000). Wang announced the establishment of the BGI on 9 September 1999, at nine seconds past the ninth minute of the ninth hour. Few moments portend more longevity in Chinese numerology, says Wang. That November, the government issued a grant of 3 million renminbi, of which the BGI got the lion's share, to support the sequencing of 1% of the human

“In Shenzhen, the mountains are high and the emperor is far away.”

genome. China was the only developing nation involved in the international project, and it finished its 30 million bases in less than a year.

Like those at other major sequencing centres at the time, Yang acquired a taste for big genome projects. While completing a somewhat underfunded 'scan' of the swine genome for Danish agencies in 2000, Yang says he decided to do something "more significant". The BGI launched a project to sequence the rice genome in 2001, using a grant of 60 million renminbi from the Hanzhou municipal government to buy 36 state-of-the-art sequencers. The BGI published the genome of the indica variety of rice in *Science* in 2002 (ref. 5), months before an international consortium published that of the japonica variety.

The BGI moved on to sequence the chicken⁶ and silkworm⁷ genomes. In 2003, it sequenced the corona virus⁸ that caused severe acute respiratory syndrome (SARS) and released a diagnostic kit that impressed Chinese President Hu Jintao. The BGI's reward was to be made part of the CAS, an honour that came with extra funding, but the academy turned out to have stipulations that didn't fit the BGI. CAS institutes are not supposed to have more than 150 scientists; the BGI had twice that and was looking to expand. Yang had to make some of his workforce official CAS staff and make special arrangements for others, stretching the CAS budget to the extreme. "No one was happy," he says.

The move to Shenzhen provided a release valve, luring the BGI in 2006 with 10 million renminbi in start-up fees and 20 million renminbi in annual grants. The city is a driving force in southern China's 'factory of the world', with many of its 12 million people producing the cheap clothing and electronics that helped to usher in China's economic miracle.

The BGI is at home in Shenzhen. Yang wants to sequence genomes at twice the speed and half the price of anyone else. And he was eager to slip away from some of the oversight in Beijing. Although he doesn't like talking about those with whom he's clashed, Yang likes to say that in Shenzhen, "the mountains are high and the emperor is far away".

The BGI gets big

With this breathing room, the BGI has grown to employ 1,500 people nationwide, more than two-thirds of them in Shenzhen, and this is expected to jump to 3,500 by the end of the year. With the investment in new sequencers, provided by a 10-billion-renminbi loan from the China Development Bank, the BGI's capacity will grow, but so will costs. Staff at

the BGI won't say how much they paid for the new sequencers, but the list price is about 3.4 million renminbi each. The purchase, which was announced on the same day the model launched, raised hackles among competing genomics centres. They accuse Illumina of making a secretive deal with the BGI while only granting others access to older models. Illumina denies such allegations, and says it has a trade-in programme for those who want to upgrade.

Of the machines, 100 will be installed in a new Hong Kong lab to facilitate international collaborations. But staff in Hong Kong cost more than the BGI is used to paying, and will be kept to a minimum (40–50 researchers). Reagents cost about 1 billion renminbi per year, and electricity for computers and cooling systems consumes another 9 million renminbi. Yang emphasizes that the loan will be paid back. But as the commodification of sequencing continues to push prices down, how the BGI will do this is an open question.

A BGI monopoly in providing sequencing services is far from assured. Aside from existing academic competitors, private ones using newer technology are starting up. Complete Genomics, based in Mountain View, California, which specializes in human genomes, expects to sequence 5,000 human genomes in 2010, starting in April. It has already logged more than 500 orders.

The BGI's solvency depends in part on scientists elsewhere paying to have microbe, plant and human genomes sequenced and resequenced faster and better than they could themselves. But like many sequencing centres, the BGI is looking to be more than a service provider. Maynard Olson, a genomics researcher at the University of Washington in Seattle who trained Wang and has close ties to the BGI, says it needs to be. "Outsourcing only works well when there is some scientific relationship between the parties. There are too many trade-offs during both the laboratory procedures and the low-level data analysis to commodify sequence data entirely."

Yang says he hopes that collaborators will pay half of the estimated costs of the genomes they want sequenced and then publish jointly, but for interesting projects he will cover 70% or even all of the cost if the collaborators lack funding.

For Eske Willerslev at the University of Copenhagen, it made sense. He collaborated with the BGI on the genome of a 4,000-year-old frozen Greenlander dubbed Inuk. Although his lab had the capacity to sequence up to about 50 billion bases in a week, he went for the BGI's technical expertise. "I have a lot of respect

"It is clear there is a new map of the genomics world."

MASS PRODUCTION

Over the past decade, the BGI has gone from sequencing millions of base pairs (Mbp) to billions (Gbp) in ever shorter amounts of time. Below are some of the institute's previous achievements.

RICE GENOME

466 Mbp

Coverage: -6x
Sequenced:
April 2000-
October 2001



SILKWORM GENOME

480 Mbp

Coverage: -6x
Sequenced:
June 2003-
September 2003



FIRST ASIAN GENOME

3 Gbp

Coverage: -36x
Sequenced:
January 2007-
October 2007



CUCUMBER GENOME

240 Mbp

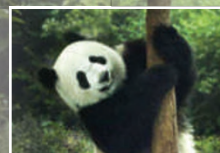
Coverage: -72x
Sequenced:
January 2007-
April 2009



PANDA GENOME

3 Gbp

Coverage: -73x
Sequenced:
March 2008-
October 2008



ICEMAN GENOME

3 Gbp

Coverage: -20x
Sequenced:
May 2009-
December 2009



for people like the folks at the BGI that really can run second-generation sequencing platforms to their perfection," he says. The ancient human genome was sequenced in two and a half months for roughly \$500,000, split evenly between Willerslev's funders and the BGI.

Willerslev says the BGI was integral not just to the sequencing but to the science. "The whole project was started because of an important scientific question agreed on by both Wang Jun and myself," he says.

For science or service

Proving its science goes beyond brute-force sequencing could be a challenge. The BGI's Luo Ruibang, also a student at the South China University of Technology in Guangzhou, turned 21 while at his last scientific meeting. He says he's had trouble convincing other scientists that, lacking doctoral training, he can do top-notch science. "A lot of the foreigners wonder if I'm really capable," he says. Luo and Li were co-first authors on a paper⁹ describing the discovery of large DNA segments in the Asian and African genomes that are absent in the Caucasian genome.

Li and his bosses are confident that this youth brigade can piece together and verify sequences. "It is a new field," says Wang. "There is not much experience anyway." But interpreting data and designing experiments are two different things, and BGI staff admit a dearth of knowledge in the latter. "We don't know much about biology," Li says. Liu says the BGI needs to overcome its biological blindspot, but he is supportive of its mission. "They are primarily sequencers, but smart ones with big guns," he says.

The panda genome was, in part, a way to show off those guns. As few biologists work on the animal, its genome is unlikely to lead to basic or applied breakthroughs. But it gave the BGI a chance to show the power of what many call next-generation sequencing, which produces shorter individual DNA reads — 100 base pairs or less, compared with the 1,000 base-pair fragments with previous technology — but at unprecedented speed. This increases the sequencing output thousands of times, but for an unfamiliar genome, it's difficult to assemble the finished product, says Wang. "Some thought we wouldn't be able to do it." There was also public interest and local government support because Jing Jing, the panda whose genome was sequenced, was the model for Beijing's 2008 Olympic mascot. "And pandas are cute," says Wang.

The BGI is also powering more biologically relevant projects coordinated by collaborators.



New faces of genomics: from left, Wang Jun, BGI executive director, Luo Ruibang and Li Yingrui.

For example, sequences of 40 silkworm strains, published last October, uncovered some 300 genes showing the history of breeding and domestication¹⁰. Sequencing was recently completed on the Tibetan antelope, whose ability to gallop at more than 4,500 metres above sea level might offer clues to adapting to high altitudes. Ge Ri-Li, a specialist in high-altitude medicine at Qinghai University in Xining plans to study antelope genes related to energy transport. Next he will sequence the genomes of Tibetan Chinese people, who suffer far less mountain sickness than the majority Han Chinese. "We want to know why," says Ge. "The final goal is to make humans more capable of adapting to high altitudes."

Research alone is not going to pay back the 10-billion renminbi bank loan. The BGI makes some income from collaborations, which account for 40% of the sequencing workload. Outsourced sequencing services for universities, breeding companies or pharmaceutical companies bring in higher margins and account for another 55% of the workload (the final 5% is the BGI's own projects). In 2009, the BGI pulled in 300 million renminbi in revenue. That is not enough, says BGI marketing director Hongsheng Liang. In 2010, Liang hopes to pull in 1.2 billion renminbi.

New income could come from proprietary rights to agricultural applications. The BGI, which owns more than 200 patents, has been attempting to do genomics-based breeding with foxtail millet in Hebei and has other agricultural projects in Laos. More cash could come from expansion of services overseas. Within three years, the institute plans to open offices in Copenhagen and San Francisco. The BGI may also charge for access to its Yanhuang database, a project launched in 2008 to sequence the genomes of 100 Chinese; BGI scientists say they

would like to expand this number into the thousands. Although according to Yang, it would be charging "at cost" — to cover computational expenses and maintenance, not for the data.

Wheeler says that the BGI will need constant innovation to keep up, and that means maintaining its high rate of collaboration. The United States has three large national centres that can test a broader range of technology, he says. "They constantly challenge one another to improve. They work cooperatively in large national sequencing projects, and critique one another to improve production and analytical methods." Single institutions that bank on a single technology may lose their edge when the technology goes out of date.

Yang recognizes the unpredictability of technological advances. Asked why he didn't stagger his investment in sequencers to take advantage of new technology as it appears, he says he'll just replace what he has when the time comes. He admits, however, that if that time comes too soon, he will be out of luck.

"There are plenty of risks, but I admire that," says Olson. By aggressively seeking collaborations and new technologies, the BGI's ambitious approach will no doubt continue to turn heads. "The bottom line is that the BGI is doing something exciting. It is a Chinese solution to the challenge of developing stronger science. Time will tell how well it works." ■

David Cyranoski is *Nature's* Asia Pacific correspondent.

- Huang, S. et al. *Nature Genet.* **41**, 1275–1281 (2009).
- Li, R. et al. *Nature* **463**, 311–317 (2010).
- Rasmussen, M. et al. *Nature* **463**, 757–762 (2010).
- Qin, J. et al. *Nature* **464**, 59–65 (2010).
- Yu, J. et al. *Science* **296**, 79–92 (2002).
- Hillier, L. V. et al. *Nature* **432**, 695–716 (2004).
- Xia, Q. et al. *Science* **306**, 1937–1940 (2004).
- Qin, E. et al. *Chinese Science Bulletin* **48**, 941–948 (2003).
- Li, R. et al. *Nature Biotechnol.* **28**, 57–63 (2010).
- Xia, Q. et al. *Science* **326**, 433–436 (2009).

See Editorial, page 7.

"It's the Chinese solution to developing stronger science."