

FEATURE

The international Human Genome Project: A milestone for life sciences and humanity

—The three stages and three major impacts of the HGP, and three contributions by China

Wei Dong¹, Boqing Qiang², Huanming Yang^{3,4,*}

¹ BGI-Beijing, Beijing 101318, China

² The State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China

³ BGI-Shenzhen, Shenzhen 518083, China

⁴ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: yanghm@genomics.cn

Received July 8, 2021; Revised July 28, 2021; Accepted August 9, 2021

The year 2021 marks the 20th anniversary of the official publication of the first draft of a human genome sequence generated by the Human Genome Project (HGP) consortium [1]. The whole world has further acknowledged the HGP's contributions to both life sciences and humanity amid combating the COVID-19 coronavirus.

In this review, we would like to summarize the history of the HGP by dividing its whole process into three stages with slight temporal overlap, *i.e.*, the 1st stage (1984–1990) of global debates and the initiation in the USA, the 2nd stage (1990–1999) of globalization, and the 3rd stage (1999–2013) of large-scale performance and completion, as well as its following sister projects.

We then wish to emphasize the HGP's contributions to three major impacts, including its role in fighting the present pandemic. Firstly, the HGP nurtures a new field of omics, which has been linking with various fields in life sciences. Secondly, the HGP fuels the development and improvement of genome sequencing, the core technology of omics. Thirdly, it cultivates a new culture of collaboration in science.

This article is dedicated to the feature in 20th anniversary of Human Genome (Eds. Michael Q. Zhang and Xuegong Zhang).

Historically, the HGP is the first international project opening its door to the whole world, especially to those developing countries including China. In return, China has made significant contributions to the three impacts or major achievements of the HGP and its following sister projects, which will also be described accordingly.

Similar to “a piece of news could be more impressive to mention the involvers”, the history of science might be more touching when describing HGP contributors. Therefore, this review will list many contributors in the HGP events to acknowledge them who deserve being remembered by the coming generations.

THREE STAGES OF THE HGP

The 1st stage: from a global debate to the initiation in the leading country (1984–1990)

The largest endeavor ever in history to decode life, the HGP is considered “the second revolution”, following the discovery of the DNA double helix [2]. The HGP is also publicly regarded as one of the three major projects that changed the world's trajectory in the 20th century, alongside the Manhattan Project and the Apollo Project.

As R. Gibbs, one of the leaders of the HGP, described, “The Human Genome Project changed everything” [3].

A global discussion was ignited immediately after the US Department of Energy (DOE) organized the first HGP meeting in 1984 [4]. The discussions or debates touched upon most natural sciences. They involved almost all countries in the world, both developed and developing, most agencies of the United Nations, such as its General Assembly and UNESCO. It attracted several generations of scientists, both established geneticists and young trainees or students, both under- and post-graduates, who were so enthusiastic that many of them later joined the major working force or even became leaders of the following sister projects. No doubt, it is the first time in history that a scientific research project becomes the focus of such global responses.

The discussions touched upon all the relevant issues related to science, bioindustry or other aspects of economy, with possible applications in medicines and other aspects. More importantly, social and ethical issues of genome sequencing triggered the widest and deepest ethics debate ever in every community, including clinicians and patients with genetic diseases.

In other words, four major issues were debated — how important it is, how difficult it is, how urgent it is, and how serious it is concerning the social and ethical challenges it might bring to the whole society.

The scale of a genome sequencing project became the subject of another furious debate, regarding which organism to begin with. A challenging question was the timing related to its scale. Should we begin with the genome of a bacterium or a yeast, a worm or a weed, a fly or a fish, a mouse or a primate? Then it led to another debate: could each be done independently in different countries or through an international collaboration?

The scientists, ethicists, entrepreneurs, and colleagues of the US have been leading the discussions, debates, and responses regarding the ethical issues of genetic sequencing. One of the great beginnings was to extend the relevant issues from ethics to ELSI by adding legal and social issues (later changed to social implications), all of which might not be given equal attention. J. Watson, a globally renowned Nobel Laureate who co-discovered the DNA double helix, organized a special ELSI committee composed of influential specialists, and proposed to assign 3%–5% of the total amount of funding towards ELSI studies, setting up a routine for those big projects afterward.

The 2nd stage: from one country to an internationalized consortium (1990–1999)

After almost 7 years of discussion and pilot research led

by its DOE and NIH, the US became the first country to initiate its HGP. The US Congress officially passed the bill on Oct. 1, 1990, with a total funding of US\$ 3 billion (later nicknamed the project of “one buck, one base”) and a 15-year completion plan (up to Oct. 2005). This 15-year completion plan covered three consecutive “Five-Year Plans”, which were revised and published periodically.

The first institution in the world specializing in genome research at the national level is the National Center for Human Genome Research (NCHGR), which was founded in 1989. It was renamed as the National Human Genome Research Institute (NHGRI) in 1997. The NHGRI has got three generations of Directors until now (J. Watson as the founding Director, F. Collins the second since 1993, and E. Green as the current one).

The first and second US “Five-Year Plan” concentrated on the construction of the short tandem repeats (STR) – based on genetic/linkage map and bacterial artificial chromosomes (BAC, used as vectors for human genome mapping and sequencing, one of the key technologies of the HGP) – backbone physical map, emphasizing the development/improvement of tools for sequencing and bioinformatics, as well as the ELSIs, especially the public education in which the US colleagues have done a great job.

The internationalization of the HGP was first proposed by J. Watson, documented in his letter to the governmental leaders of another developed country in 1990. The UK was the first country to promote a substantial international collaboration on the human genome sequencing. The most important contributor is M. Morgan, who proposed the historic Bermuda Meeting in 1996, joined by representatives from the US, UK, Japan, France, and Germany. The “Bermuda Rules” listed three principles, namely, “free access to, immediate release of, and no patenting on the human genome sequence” and were signed by all the participants led by J. Watson and other center-leaders. It laid the solid foundation of the HGP and later the “HGP Spirit”, and become one of the most important documents in the history of the natural sciences.

A. Patrinos of the US DOE further emphasized the significance of international collaboration for the HGP at the 3rd International Strategy Meeting for Human Genome Sequencing on May 19, 1999. In his opening address, he strongly expressed that “This (HGP) is still an international program and should be remain so”, which did encourage and lead to China’s participation in the HGP on Sept. 1, 1999, a noteworthy event in the HGP.

The 3rd stage: from global large-scale performance to the completion of the HGP (1999–2013) and its following sister projects

The human genome sequencing has worked in its full

swing since 1999 in dozens of major research centers in the world. The global efforts were humorously described as “sun-never-setting”, since there were researches working on the HGP from Japan to China, from Europe to East Coast then to West Coast of the US, accompanied with the improvement of the sequencers and newly developed software for assembly and annotation.

The US NHGRI, led by F. Collins, became the headquarter of the international HGP. The US chapter was conducted by a dozen “big” centers, including the Whitehead Center of the Massachusetts Institute of Technology (MIT), in Boston and Genome Sequencing Center of Washington University in St. Louis (led by B. Waterston and E. Lander, jointly contributed more than 30% to the international HGP), the DOE’s Joint Genome Institute (led by A. Patrinos, E. Rubin, and T. Hawkins, contributed around 10% to the international HGP), the Human Genome Sequencing Center of Baylor College (led by R. Gibbs, contributed ~8% to the international HGP), the University of Washington’s Genome Sequencing Center in Seattle (led by M. Olson, contributed ~1.5% to the international HGP), and several other important contributors. The US alone contributed almost 55% to the international HGP in all as the leading country. Its great contribution together with all those leaders as well as contributors from all contributing countries should forever be acknowledged and appreciated.

The UK contributed more than one third to the HGP. The UK chapter was completed by a single institution, The Sanger Institute (later renamed The Wellcome Sanger Research Institute or WSRI) led by J. Sulston, R. Durbin, and J. Roger. Together with its sister institute on the same campus, the European Bioinformatics Institute (EBI) led by E. Birney, also made significant contributions to the assembly and annotation of the sequence, together with a group led by D. Haussler and Jim Kent, one of his young students at University of California, Santa Cruz, in the US. Japan contributed around 7% to the HGP. The Japan chapter was led by Y. Sakaki, and done by three centers in the University of Tokyo, RIKEN, and Keio University, respectively.

France was one of the earliest countries to begin its own human genome research. Its CEPH (Centre du Polymorphisme Humain) led by J. Dausset, made great contributions to the STR-based genetic map and YAC (yeast artificial chromosome)-backboned physical map, proposal of “Free-Sharing” Principle, and the widely used human samples of the “CEPH Families”. The French chapter was mainly contributed by GenoScope led by J. Weissenbach, contributing close to 3% to the HGP.

Furthermore, Germany contributed around 2% to the

HGP. The German chapter was led by H. Lehrach and completed by three centers in IMB (Institute of Molecular Biology), MPIMG (MP Institute of Molecular Genetics), and GBF (Gesellschaft für Biotechnologische Forschung).

China made an approximate 1% contribution to the HGP as the newest member and the only developing country in the HGP consortium to date. The Chinese chapter was jointly completed by the South Center, led by Z. Chen (陈竺) and North Center, led by B. Q. Qiang (强伯勤) and Y. Shen (沈岩), of the National Genome Center, and the Human Genome Center, Institute of Genetics of the Chinese Academy of Sciences (CAS), and Beijing Genomics Institute (BGI).

In May of 2000, the HGP consortium decided to announce the first draft of the human genome sequence during the International Strategic Meeting on Human Genome Sequencing at the Cold Spring Harbor Laboratory (CSHL, USA). Based on the generally accepted standard at that time, the first draft was thought to represent at least 90% of the human genome with more than 5X coverage. Celera, a private company led by C. Venter. Incorporating the international HGP’s publicly available data into its own data developed under the “whole genome shotgun strategy”, Celera produced its own draft of the human genome sequence. A celebration, themed “Decoding the Book of Life, A Milestone for Humanity”, was staged in the capitals of the member countries of the HGP consortium (finally the US and UK held it together simultaneously through satellites) on June 26, 2000, which has been regarded as the first and most important event in the HGP.

The scientific paper on the draft of the human genome sequence was officially published in Nature on Feb.15, 2001 [1]. The completed genome sequence was considered to represent at least 99% of the human genome with more than 10X coverage based on the generally accepted standard. This completion was officially announced on Apr. 14, 2003, by a Joint Proclamation by the Heads of Government of Six Member Countries Regarding the Completion of Human Genome Sequence, and this announcement marked the successful completion of the HGP (Fig. 1)

The mission of the HGP is not limited to the human genome sequencing alone, because it also includes 8 genomes of model organisms (several completed in collaboration between some HGP teams and non-HGP). Among the 8 model organisms, there are two monocellular organisms (*E. coli* K-12 for prokaryote and *S. cerevisiae* for eukaryote models, approximately 5 Mb and 12 Mb in genome size, published in Sept. 1997 [5] and Oct. 1996 [6], respectively); a worm (nematode, *C. elegans*, approximately 100 Mb in size, published in Dec. 1998 [7]) and a fly (*D. melanogaster*,

Joint Proclamation by the Heads of Government of Six Countries Regarding the Completion of the Human Genome Sequence

April 14, 2003

We, the Heads of Government of the United States of America, the United Kingdom, Japan, France, Germany, and China, are proud to announce that scientists from our six countries have completed the essential sequence of three billion base pairs of DNA of the human genome, the molecular instruction book of human life.

Remarkable advances in genetic science and technology have been made in the five decades since the landmark discovery of the double-helix structure of DNA in April 1953. Now, in the very month and year of the 50th anniversary of that important discovery by Watson and Crick, the International Human Genome Sequencing Consortium has completed decoding all the chapters of the instruction book of human life. This information is now freely available to the world without constraints via public databases on the World Wide Web.

This genetic sequence provides us with the fundamental platform for understanding ourselves, from which revolutionary progress will be made in biomedical sciences and in the health and welfare of humankind. Thus, we take today an important step toward establishing a healthier future for all the peoples of the globe, for whom the human genome serves as a common inheritance.

We congratulate all the people who participated in this project on their creativity and dedication. Their outstanding work will be noted in the history of science and technology, and as well in the history of humankind, as a landmark achievement.

We encourage the world to celebrate the scientific achievement of completing the Human Genome Project, and we exhort the scientific and medical communities to rededicate themselves to the utilization of these new discoveries to reduce human suffering.

His Excellency
Jacques Chirac,
President of the French Republic.

The Honorable
George Bush,
President of the United States of America.

The Right Honorable
Tony Blair, M.P.,
Prime Minister of the United Kingdom.

His Excellency
Gerhard Schroeder,
Chancellor of the Federal Republic of Germany.

His Excellency
Junichiro Koizumi,
Prime Minister of Japan.

His Excellency
WEN Jiabao,
Premier of the State Council of the People's Republic of China.

Figure 1. Joint proclamation by the Heads of Government of six countries regarding the completion of human genome sequence to announce the successful completion of the HGP in 2003.

approximately 180 Mb in size, published in Mar., 2000 [8]) to represent multicellular invertebrates; a weed (*A. thaliana*, approximately 130 Mb, published in Dec. 2000 [9]) to represent higher plants; two fish, one that is smaller in body size but larger in genome size (the zebra fish, *B. rerio*, > 1 Gb in genome size, published in Apr. 2013 [10]); and another that is larger in body size but smaller in genome size (puffer fish, *F. rubripes*, approximately 390 Mb, but containing a number of protein-coding genes similar to that of mammals, published in Aug. 2002 [11]); and finally a mammal (mouse, *M. musculus*, approximately 2.5 Gb, published

in Feb. 2002 [12]).

First of all, the model organisms are primarily selected based on their significance in research and medicine, since several of them have been widely used in the laboratory as animal or plant models. Secondly, they each hold a representative position in the "Tree of Life". Thirdly, there are more knowledge about their genetics/genomics, making essential model organisms in scientific research. Thus, they play a significant role in phylogenomic and comparative genomic studies, such as when tracing homologous regions in evolution or assembling and annotating the human genome sequence.

It is important to keep in mind that the completion of the human genome sequence is not the end goal of the HGP consortium. Led by F. Collins, has initiated four subsequent sister projects, all of which are characterized by: (1) the major HGP centers form the main working force and can invite collaborators according to the project's mission; (2) improved sequencing and bioinformatic analysis are used as the major technology; (3) genomics-based omics is used as the major strategy; (4) the “needed by all, owned by all, done by all, and shared by all” of the “HGP spirit” are accepted as the unchanged principle.

The first sister project is the International HapMap Project, marking a leap “from a reference genome of a single ‘representative’ individual (actually it also contains the sequences of several other individuals for various reasons) to the human genome diversity with sequence polymorphic markers (SNPs, or single nucleotide polymorphisms) of the three major populations (*i.e.*, European, Asian, and African).

The HapMap Project began in 2002. The major centers from the US, UK, Japan, Nigeria, and China participated in the HapMap Project. Beginning with identified SNPs from the two homologous chromosomes obtained by the HGP, the HapMap Project aimed to extend to those with minor allele frequencies (MAF) equal to or higher than 5% in the three major populations: Europeans and Africans (each with 30 trios provided by CEPH, respectively); Asians (45 genetically unrelated individuals from Chinese and 44 from Japanese populations, respectively). The findings of the three phases of the HapMap Project were published in 2005 (Phase I, with more than 1 million SNPs [13]), 2007 (Phase II, with another 3 million SNPs), and 2010 (Phase III was extended to 11 populations and selected regional sequencing), respectively.

The International HapMap Consortium identified millions of SNPs with high MAF and high density, which have been used as the best genetic markers in scientific research and provided a solid foundation for the GWAS (genome wide association studies). GWAS is one of the major technologies used to link genotypes and phenotypes, especially effective when studying complex diseases or characteristics.

The second sister project is the ENCODE (ENCyclopedia of DNA elements) Project, marking the leap from general DNA sequencing to the identification of all DNA functional elements, especially those located in non-coding regions, such as promoters, enhancers, origin or termination sites of replications, transcription factor-binding sites, methylation or DNase hypersensitive sites, and numerous specific sites or all other characteristic regions with unknown functions. The ENCODE consortium was well organized into 35

working groups with members from about 80 institutions from 11 countries. The ENCODE Project was a brilliant example of the extension of the HGP collaboration culture. Its chief coordinator was E. Birney, the youngest leader in the HGP consortium.

The ENCODE Project was discussed in parallel with the HapMap Project. It was initially nicknamed “Two 1% Sequencing Projects” because its sequencing part was targeted at a ~30 Mb “gene-rich” region and a ~30 Mb “gene-desert” region. The ENCODE Project was formally initiated in 2003 and published nearly 60 highly informative papers in Jun. 2007 and Sep. 2010. China joined its discussion and initiation, but did not substantially contribute to it.

The third sister project is the 1000 Genomes Project (GIK), which was proposed and initiated by the UK and China, supported and joined by the US major centers. This project marked the leap from a single reference genome to complete genomes of many individuals from various major populations. Its pilot phase focused on the same three populations used for the HapMap Project (European, Asian, and African, around 299 samples in total). Its Phase I expanded to include North Americans (a total of 1094 samples) and its Phase II further expanded to five by incorporating individuals from the Middle East (a total of 2500 samples from 25 ethnic groups).

The fourth sister project is the International Cancer Genome Project (ICGP, or ICGC, its consortium), marking the leap from research on “normal” genomes to clinical studies. In Apr. of 2008, the ICGC announced its joint workforce of 47 research centers from 15 countries. Its coordination center has now moved to the UK and is led by A. Biankin. The present phase is subtitled ICGC-ARGO (Accelerating Research in Genomic Oncology) and is still ongoing. Its 4th coordination conference took place in Beijing on May 15–18, 2021.

THE THREE MAJOR IMPACTS OF THE HGP

About two decades have passed since the official completion of the HGP on Apr. 14, 2003. However, its achievements have generated significant impacts on life sciences and society in many respects.

The HGP nurtures a new field of omics

The first impact of the HGP is that it has nurtured a new field of science — “omics”. In other words, the HGP is the first vast and comprehensive practice of omics.

Omics has several layers of meanings. The term “genome” is originally derived from “genes”, preserving

its core meaning and extending it to encompass the totality of genes of an individual or species. The HGP has further extended its research from a single reference genome to thousands of human genomes, beginning with the initiation of G1K immediately afterward as a sister project. F. Collins, the chief coordinator and a major leader of the HGP, talked about sequencing millions of human beings in developed countries, indicating that genomics is aimed at studying all human genomes globally, in other words, “to sequence everybody in the world”.

Secondly, genomics has extended from studying only human genomes to also studying those of animals, plants, and microbes, reflecting its intended aim of studying all living organisms in the biosphere. This ambition has led to many international collaborations, such as the Earth BioGenome Project (EBP) led by H. Lewin, which aim to “sequence(ing) every (living) thing on the earth”.

Thirdly, it is no exaggeration that genomics has “omized” or “omicsized” most, if not all, fields of life sciences. “Where once there was the genome, now there are thousands of ‘omes’ [14].” “Today, we’ve gotten to the point where almost no biological phenomenon can escape “omicsization”, and within the next 25 years, omics will be the biggest, if not the only, game in town [15].”

Genomics, like all other fields in life sciences, has the same soul that “nothing in biology makes sense except in the light of evolution” [16]. The HGP also tells us that, based on comparative genomics or phylogenomics, together with what J. Watson said in 1953 during the discovery of the DNA double helix, “the precise sequence of the bases is the code which carries the genetic information.” [17]. Thereafter, we might say “the genome sequence is the ‘lab-logs’ of evolution in nature” of all species. Genomics has opened a new era for studies on evolution by genome sequencing and other relevant tools. These technologies are also used to study such topics as ancient DNA (aDNA) which might be a million years old to explore the core secrets of evolution and its origin.

Genomics also has two beliefs, that “life is of/in sequence, life is digital”. It has revolutionized our understanding of life. Together with the DNA double helix, it has further extended to all omics, characterized by a combination of big-data, machine/deep learning and AI essential for the creation of a new and more potent phase in the development of the life sciences and medicine.

Fourthly, we have realized, in practice, that no single “omics” could and should be conducted alone when studying any phenomenon of life. Thus, we have started to adopt the strategy of combining more than one

“omics” when trying to accomplish a single goal, such as incorporating RNAome/omics with transcriptomics, proteomics, metabolomics, glycome/omics, lipidome/omics.

The fight against the COVID-19 pandemic in late 2019 and onwards is a brilliant example of the importance of the concept and practice of the ome/omics. By sequencing the genome of the pathogen, we could efficiently identify the organism and use the sequence as a “golden standard” during the clinical diagnosis of both infected and asymptomatic patients, detection of possible variants, primer design for genetic engineering, and most importantly, vaccine development and future pandemic prediction and control.

It is no doubt that genomics is essential. But it is insufficient. In addition to studying the “humanomes (including both human genotypes and phenotypes)”, we must also study “pathogenomes”, the “interactomes” between both of them, the “ecologiomes” for all hosts and “intermediate transmitters”, as well as the “bio-responsomes” or “exposomes” of the changing environment (*i.e.*, temperature, humidity, wind, sunshine). In summary, no omics could be done alone.

The most evident characteristic of “omics” is BIG. A BIG science dealing with BIG data, on a BIG tech-platform, by a BIG team, through BIG collaboration. Also in other words, omics has “-lized” almost everything, such as sequentialization, digitalization, scalarization, industrialization, and globalization of life sciences. All could be seen from the whole process of the HGP. Even a small team can find its position and play its role within this BIG consortium.

The HGP fuels the development of new sequencing technologies

The HGP’s second impact was its push for the development of new technologies for the life sciences, particularly for DNAome/RNAome sequencing. The language of information science is comprised of two letters: 0s and 1s, while the language of life is comprised of four letters: ATCG. In a way, sequencing is the digitalization of life. F. Crick called this “sequentialization”, enabling us to read the “Book of Life”, though we are still not in a position to fully understand this “book”. More significantly, we have to realize that we still know so little about life, and sequencing is only one of the tools to help us begin to decode life, which is why “sequencing is not everything”, just like “nothing is everything”. It is essential to acknowledge and respect the fact that science includes all fields and branches, and that every effort has its own specific importance and value.

The significant improvement in sequencing technology

could be seen from the cost of sequencing a human genome with reasonably good quality. In the 1980s, the total cost for sequencing a single base was more than US\$10. In 1990, the US HGP had a total budget of US\$3 billion for the 3 billion bases of the human genome. At present of 2021, the commercially available price for sequencing a human genome is approximately US\$600, which is equal to “5 million bases, one buck” or 1/5,000,000 of the price three decades ago, setting up a record in the science history for any new technology development, *e.g.*, that Moore’s law.

In this review, we would acknowledge once again the irreplaceable contribution by our US colleagues. In 2003, the cost of genome sequencing was a staggering “US\$30 million, one genome” and F. Collins, former Director of the NHGRI proposed a “plan in two steps” based on his prediction of the potential applications of sequencing in laboratory research and the clinical setting. This plan aimed to lower the cost of sequencing to “US\$100,000, one genome” in five years, then further lowering it to “US\$1000, one genome” in the next 10 years. The whole plan was announced at the First Meeting of the International Conference on Genomics (ICG) in 2006 in Hangzhou, China. History has witnessed that this ambitious prediction has been realized far ahead of schedule.

Let us discuss the history of genomics. DNA sequencing was invented in the 1970s by F. Sanger [18] and W. Gilbert [19] for enzyme- and chemistry-sequencing. When we ourselves learned and did sequencing in the 1980s, the frequently available method was one of the first generation of radiolabeled “slab gel hand-operated Sanger sequencing, or SGHOSS”, which was very costly (approximately \$10 per base) and had low throughput. Thus, during that period, it was almost impossible for the HGP to come up with satisfying results within 15 years given the cost of all the sequencing work that had to be done. Moreover, a prestigious biologist had noted that “it is impossible for us to know the full genome sequence (base) of human. Even considering the rapid development of modern science, I can assure you that we will have to wait at least 300 years.” This point was not entirely without any reason at that time.

Sequencing, one of the three major impacts the HGP provided, will continue to provide the unique and solid foundation for life sciences. “Nothing could be done without the information of sequence” — a mantra often heard and echoed in the field of omics around the world, must be kept in proper perspective since it is related to the nature of omics, which is both concept/hypothesis-driven and tool-driven, aiming at discovering new things that require further explanation.

The HGP cultivates a new culture for collaboration

The HGP, the biggest collaborative project in the history of the natural sciences, has cultivated a culture of collaboration. The HGP is the first one that turns research by a single lab or a single country into an international consortium, at least in life sciences. It has set up a brilliant model which has been followed immediately by the four HGP sister projects as discussed above, and further followed by the International S.c2.0 Project on the design and synthesis of the first unicellular eukaryotic genome, joined by scientists from the USA, UK, Singapore, China, and others. The project has proceeded with full success, even while the field of synthetic biology or “genome writing” is still immature.

The culture of collaboration cultivated by the HGP could be summarized as the HGP Spirit, expressed in the simplest verbs, *i.e.*, “Owned by All, Done/Joined by All, and Shared by All”. This motto was proposed by the Chinese participants and firstly published in the news release on the completion of Chromosome 3 in Apr. of 2006, which were then endorsed and acknowledged by the HGP community.

The primary meaning at that time was that the human genome is the common heritage of mankind and should be “Owned by All”; the HGP is a common effort to achieve a shared goal so should be “Done/Joined by All” through vast international collaboration; surely the human genome sequence data should be freely “Shared by all”.

Presently, the HGP Spirit may be further enriched by adding another simple verb, making the motto become “Needed by All, Owned by All, Done/Joined by All, and Shared by All”. This can be explained as: We, as the human species of this planet, first have to realize that we need to face the same challenges, which is “Needed by All”; we have the same opportunity for development, *i.e.*, “Owned by All”; just like the HGP, this opportunity should be taken through vast international collaboration, *i.e.*, “Done/Joined by All”; the better future of mankind should be “Shared by All”.

We can also see that the HGP Spirit has been followed by numerous global collaborative projects. In addition to the ones mentioned above, several other examples for human being include:

The Global Alliance for Genomics & Health (GA4GH)

The International Efforts on GWAS

The Genotype-Tissue Expression (GTEx) Project

The Human Microbiome Project/The International Earth Microbiome Project

The Human Cell Atlas (HCA by single cell RNAome

sequencing)

The Human Protein Atlas

The Human Phenotype Ontology

The Human Knockout Project

The Human Induced Pluripotent Stem Cells (iPS) Initiative

The Human Brain Projects (worldwide, also many national ones)

And those for non-human, represented by the Earth BioGenome Project

The Global Genome Biodiversity Network (GGBN)

The Plant G10K Project

The Global Orphan Crop Consortium

The Global Crop Improvement Network (GCIN)

The 1001 Genomes Consortium on *Arabidopsis thaliana*

The Animal/Bird G1M Project

The Vertebrate Genome Project (VGP)

The Genomic Encyclopedia of Bacteria and Archaea (GEBA)

The Global Virome Project

The International Mouse Phenotyping Consortium

We also have to see the changing world with many unpredicted and complicated factors. The HGP Spirit or its culture of collaboration, has been constantly challenged in all respects all the time especially in the past years or presently. As scientists, we see any form of collaboration is based on mutual respect and mutual trust. As Chinese scientists, we should not let all our collaborator down. The Chinese participation in and contribution to the HGP, especially to the HGP Spirit, will be discussed below.

CHINA'S PARTICIPATION IN AND CONTRIBUTIONS TO THE HGP

China's participation in and contributions to the HGP is a noteworthy event in the history of the HGP and natural sciences.

China's pre-HGP efforts on human genome research and preparation for its participation in the HGP

The first meeting focusing on human genome research was proposed by Prof. C. C. Tan (谈家桢) of Fudan University and Prof. M. Wu (吴旻) of the Chinese Academy of Medical Sciences (CAMS) and organized by the Department of Life Sciences of the Natural Science Foundation of China (NSFC) in 1993.

This meeting led to NSFC immediately funding its first project on the human genome as one of its key

projects, termed "A study on the selected loci in the genome of Chinese ethnic groups and their genomic structures". The key leaders of this project include Prof. Z. Chen of Shanghai Second Hospital and Prof. B. Q. Qiang of CAMS, marking the beginning of China's pre-HGP efforts on human genome research. Two National Genome Centers, the South Center, headed by Prof. Z. Chen and Wei Huang (黄薇) in Shanghai, and the North Center, headed by Profs. B. Q. Qiang and Y. Shen in Beijing, were soon established afterwards.

It should be emphasized that relevant research on the human genome was already taking place in Fudan University, Hunan Medical University, and several institutes of CAS and CAMS. A series of achievements, such as those in human cytogenetics, X-chromosome mapping and library-making, were made, igniting the interest of China's scientific communities and their enthusiastic discussions on the HGP.

The Chinese scholars and students sent by the Chinese government to the United States and Europe also played important roles in China's genome research. The first Chinese scientist who brought the sequencing technology back to China is Prof. G. F. Hong (洪国藩). During our first visit to the UK's Sanger Institute, they told us to "go to Shanghai to learn from Dr. Hong, who helped improve the sequencing technology here and taught us how to do it." The fact is that after his return to China, Prof. Hong built the first sequencing laboratory in China and later led the Chinese Rice Genome Project as a part of the international efforts on the rice genome sequencing.

The Chinese students trained in the US, together with those in China, were fully supported by their supervisors and made irreplaceable contributions. The "Zhangjiajie Meeting" was organized by a newly established Committee of Young Geneticists of the Chinese Society of Genetics, supported by the GBI headed by Dr. J. Wang (汪建), and hosted by Prof. J. H. Xia (夏家辉) of Hunan University, in Zhangjiajie, Hunan Province in November of 1997. This was the first meeting officially dedicated to the HGP in China. The participants issued "A Letter to the Senior-Generation Chinese Geneticists", calling for China's human genome sequencing to join the internationally collaborative HGP.

Dr. J. Yu (于军), former Assistant Director of the Genome Sequencing Center at the University of Washington in Seattle (headed by Dr. M. Olson), was invited by Dr. J. Wang to the "Zhangjiajie Meeting". Dr. Yu inspired many with his talks on various occasions day and night regarding the updated version of the US HGP, as well as the sequencing, assembly, and annotation of genome sequencing technology. The meeting was joined by researchers from all over China,

such as Hunan, Shanghai, Beijing. The audience included many experts, such as F. C. He (贺福初), L. He (贺林), L. Yu (余龙), and J. Wang. Several of them were later elected as academicians of the CAS.

During the 8th International Genetics Conference in Beijing in August 1998, Profs. M. Olson and M-C. King from the University of Washington in Seattle, Prof. R. Waterson from the Washington University in St. Louis, both directors of major US HGP centers, were invited by CAS and its Genetics Institute, as well as the Chinese Society of Genetics to attend the opening ceremony of the Human Genome Center, Institute of Genetics, CAS on the campus of the Institute of Genetics. The ceremony was chaired by Prof. S. Y. Chen (陈受宜), Prof. L. H. Zhu (朱立煌), Prof. Z. H. Xu (许智宏, former Vice President of CAS), and Mr. G. H. Wang (王贵海, former Director of the Bureau of BioScience, CAS). This event attracted the attention of the international genetic community and was reported by Science on Aug. 20, 1998 [20].

Immediately after its establishment, the preparation for joining the international HGP began by systematically training a group of young researchers, including the young group leaders, Drs. W. Dong (董伟), X. Q. Zhang (张秀清), and S. N. Hu (胡松年). The process was overseen by Dr. J. Yu and Prof. M. Olson at the center. The trainees were quick learners and were appraised by their supervisors. It is they who successfully sequenced and assembled 4 BACs after returning to Beijing with the first ABI-377 sequencer. We would also like to express our gratitude towards all who supported us along the way. The China's application to join the international HGP consortium was submitted and publicized on July 7, 1999, in the name of "Human Genome Center, Institute of Genetics, Chinese Academy of Sciences", abbreviated as "Beijing Center".

Science-based discussions, debates, and other issues related to joining the HGP

First of all, both sides of the debates on "to join or not to join (the international HGP)", which began even before the application was publicized, were scientifically normal or reasonable and mainly around a scientific issue and fact-based.

We must admit that China was not strong in natural science research at the time. One side of the debates is concerned about China's reputation if we could not realize what was promised. It was a fact that we would be assigned a task that must be completed quantitatively and qualitatively, to meet an aggressive schedule, because the HGP was scheduled to be completed rather soon.

However, the majority of the China's scientific community was supportive of China's application in the HGP, believing that China should take the opportunity to improve "alongside the 'forest of nations' of the world" in this new field. Thus, participating in the HGP meant putting China's reputation and honor at stake. No doubt, it was a severe challenge.

However, the debates were fueled by a commercial company in the US which challenged the internationally collaborative HGP by publicly announcing that it would independently finish the human genome sequencing sooner and charge for data use fee and patent hundreds of "important genes". A question, which could still be difficult for us to answer today, was "if that company won, it is a waste of such a handsome money; if the internationally collaborative HGP won, then all the data would be freely available, thus forming an equally waste of "blood and sweat" of China. Anyway, it would be a "double loss" for China to join the international HGP. It was also true that the HGP might not be interested in China's participation when it was nearly close to its end and that "it would not be better to have China, not be worse without China, either." The reality is they were afraid that China would delay the whole HGP.

Dr. A. Patrinos, one of the HGP's pioneers and representative of the US DOE in the HGP consortium, strongly encouraged China's application. In his opening address at the 3rd International Strategic Meeting on Human Genome Sequencing (ISMHGS) on May 19, 1999 in CSHL, he emphasized the importance to keep the HGP as an international collaboration project by saying "This is still an international program and should remain so". We should not forget Prof. M. Olson who strongly supported China's participation in the HGP, had full trust on us, his Chinese students and young friends. Upon our request, he also called the major leaders to take into consideration of China's application. The door was really opening.

In response to our application, Prof. H. M. Yang, representative of the Beijing Center, was invited to defend China's application at the 5th ISMHGS on Aug. 31, 1999 at the Sanger Institute in the UK. Delegates from 15 HGP centers worldwide attended the event. Prof. Yang began his 5 minutes application address with: "I have just been asked by a colleague why I (as a stranger or a newcomer) am here. My answer is 'We are on your side by joining you'..." The final decision came after a long debate and was published in the news release by the International Human Genome Consortium on Sept. 1, stating that "China has become the latest contributor to the worldwide sequencing effort alongside France, Germany, Japan, the United Kingdom and the United States".

The Chinese admission was totally based on the

mutual trust before Chinese scientists proved their capability to complete the assigned task on schedule and not to delay the completion of the whole HGP in time, even most of the HGP leader met the Chinese representative for the first time. In its remarks, *Nature* said that “China’s scientific leaders overcame skepticism from some members of the HGP — and from many of their own researchers — to become the only developing country to take a role in sequencing the human genome [21].”

With all that being said, we would like to express our deepest and most sincere gratitude for trusting, accepting, and providing us with the opportunity “to demonstrate yourself (China) and to prove yourself!” That is what Mr. M. Morgan said when we asked him whether it would be a “double loss” for China to join the HGP several years later.

China’s contributions to the HGP

As the “only developing country to take a role” in the HGP, China does play an active role in the HGP.

“Collaboration, collaboration, and collaboration”

In Oct. of 1999, soon after China was admitted to the HGP, Drs. J. Sulston and M. Morgan invited a Chinese delegation for a bilateral meeting at the Sanger Center. The delegates were so impressed and touched by a big-size slogan at the Sanger Center’s entrance that reads, “Human genome, to get one free (by all) or to buy one (from those who would like to monopoly it)”. At the meeting, both China and the UK sides expressed a strong and firm commitment to “protect the human genome” for all.

After becoming a member of the International Bioethics Committee (IBC) of UNESCO, the Chinese representative submitted 6 subjects and/or program proposals to call IBC for discussions of “the most urgent issue in ethics now is to freely release and get access to human genome sequence data”. This continuous effort resulted in the UNESCO’s announcement on May 7, 2000 and was later officially written into the G7 Summit document and the United Nation’s Millennium Declaration. This contribution was widely and firmly applauded by the HGP consortium.

Realizing the importance of genome sequencing, especially for developing countries, Chinese participants in the HGP consortium proposed a brief expression of the HGP Spirit or the collaboration culture it cultivated and described above, *i.e.*, “Owned by All, Done by All, and Shared by All”. It has been already a banner for the following sister projects, such the HapMap and

ENCODE, G1K and ICGP, and still been followed by the international community on omics, and also in the fight against the COVID-19 pandemic. Just take the vaccine design as an example. “After the Chinese lab released the sequence of the virus in Jan. 2020, researchers around the world could begin researching the virus without needing a sample. That made it possible within 24 hours for the first vaccine design to get started!” [22]

Another example is the Chinese role in the S.c2.0 Project, which aimed to design and synthesize “the Second Version of the Yeast Genome” [23–26]. The initial innovation was made by Dr. J. Boeke at the New York University (NYU) Institute for Systems Genetics in the US. Chinese teams then proposed to turn a single laboratory research project into an international collaboration. The teams held their first international meeting in Beijing, China. This S.c 2.0 Project has achieved its first goal after successfully synthesizing the first unicellular eukaryotic genome in 2017 through the joint effort of USA, UK, Singapore and China, following the experiences of the HGP.

We, Chinese scientists, should raise this banner of vast collaboration even higher in every relevant research project for ever. It has been written into science textbook and into life science population reading materials, that collaboration is essential for natural sciences research.

Needless to say, the close collaboration between the three major centers and 12 collaborating partners in China should be applauded. For example, a few days before China’s Spring Festival, the HapMap Project was at the last stage. The South Center team co-led by Dr. W. Huang still accepted an unexpected urgent task and completed it in a few days on schedule. Instead of enjoying the whole holidays for the Spring Festival, they worked day and night, and finally they also completed HapMap assignment, gaining the full appreciation from the entire HapMap Consortium.

“Sequencing, sequencing, and sequencing”

Many of us might not be that happy to hear such a slogan, “sequencing, sequencing, sequencing,” but its concept was taken from F. Sanger, the “Father of Sequencing”, who called on “sequence, sequence, and sequence” [27]. It is true that, like any other techniques, “sequencing is NOT everything”, but “nothing could be done in biology without sequencing”. Thus, we have to keep it in mind that life science is both hypothesis and tool-driven, and strongly oppose pure technology determinism” in science. Sequencing could never be done alone, at least it could not be done without

bioinformatics, big data especially the phenotype data (phenomics), and artificial intelligence (AI) as well as other omics, if we agree with “scientific discovery, technology innovation, and bioindustry development” in life.

With that said, sequencing is one of the fundamental tools in the field of life sciences. China has become an active member of the international omics community, with more than 1% contribution to the HGP (China submitted approximate 64 Mb draft sequences and 38 Mb finished sequences), 10% contribution to the HapMap project, ~ 20% to the G1K project, and also as the second or third contributor to the ICGP. At the same time, we fully acknowledge that the US has been the first and the biggest contributors to the above “sister projects in addition to the HGP”.

After building the essential infrastructures for genome sequencing, China has become one of the biggest contributors in sequencing and annotating the genomes of other organisms. In addition to the big data generated by sequencing offers to decode the discoveries in life, China has also made numerous discoveries in various fields and species, such as rice [28], silkworm [29], and panda [30]. Although China is now a globally acknowledged player in the field of genomic sequencing, it is important to keep in mind that all these successful projects, without exception, were carried out through vast national and international collaboration.

“Learning, learning, and learning”

“If you try to innovate something good, you have to be a good learner.” Innovation is based on effective learning and Chinese scientists have proved their ability to learn and innovate throughout the working process.

The process to work on those big omics projects is the process for us to learn and to innovate. Just take four examples in four steps.

The 1st stage. Purchasing the right machines (sequencers)

It is not that easy to always make the right choice from many choices available, but Chinese scientists can make the right selection most of the time by considering various factors, such as price, efficiency, unit cost, potential for improvement, and durability. Such caution is taken because purchasing the wrong sequencer is a huge loss and could even bring down the entire sequencing center.

At the very beginning, we bought the most updated slab gel sequencers with non-radioactive labelings, such as the ABI-377 model, which was advantageous at the

time since it required inexpensive labor (relative to the price of the machine) force in China at that time. Later, when the first generation of capillary sequencers hit the market, we gave up the old machine for the MegaBACE 1000 due to its pronounced advantage of automation and higher throughput.

It is equally important to effectively and efficiently utilize sequencing machines. Through strict training or even “competition” and refined management, we made the longest “off-machine read length (OMRL)”. In return, the Wellcome Trust’ Genome Office, headed by M. Morgan, generously gifted us 34 pieces of this sequencer type as encouragement and affirmation. With that said, we would like to take this opportunity to express our deepest and most sincere gratitude for the gifts.

During the second wave of MPH (Massively, Parallel, High throughput) sequencing, it was even more difficult to select the ideal machine out of the countless available models, such as the “454”, “Illumina”, “Solexa”. For example, the Illumina had the revolutionary advantage of “naked” reactions and “dense” potential in further improved throughput, and had the reliable underlining mechanism of sequencing-by-synthesis (SBS). At the same time, it had the fatal disadvantage of short off-machine read length (OMRL).

The 2nd stage. Making software when unable to make hardware

A sequencer machine could not run well without a sufficient sequencer-providers’ software. Therefore, our first goal was to “turn biological repeats into mathematical repeats” when sequencing the rice genome using the “whole genome shotgun” strategy from 2001 to 2002. This approach was widely acknowledged as a crucial improvement in the practical application of the “whole genome shotgun” strategy.

Our second goal was the “*de novo* assembly” of short OMRL of the Illumina sequencer, which often shortened the OMRL from up to 630 bp to no longer than 36 bp. Our developed software was internationally evaluated and termed to be “the best of the best”. In addition, the genome of the panda was “the first reported mammalian genome” to be solely done by the Illumina sequencer [31].

The 3rd stage. Assembling new sequencers from scratch

We decided to make our own sequencers from scratch thanks to the suggestions, advice and support from our advisors in the US and other countries.

The first challenge is to select sequencer type, which

required us to consider many aspects, including, but not limited to, its sequencing mechanism, potential improvement, price and maintenance fees. After careful consideration, we purchased the Complete Genomics which again proved to be a right choice, since it also played an important role in combating the COVID-19 pandemic.

This review does not allow us to write more and longer. We do have to express our deepest gratitude to all our TEACHERS. It is true that we do have not disappointed our teachers and friends, who have all the reason to be proud of us and to further trust us; furthermore, we have to acknowledge that we would be students forever, as we still have much more to learn from our teachers and friends, and would love to be your more worthy collaborators.

We would like to conclude by saying that we believe that the ideal attitude towards a collaborative relationship should be 30% as competitors (no fair competition, no development), 30% as friends (mutual trust), and 40% as family members. Finally, it is important to keep in mind that science must be “Needed by All, Owned by All, Don/Joined by All, Shared by All”.

OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921
- Sharp, P. (2014) A speech at the annual meeting in 2014 of AAAS. Chicago
- Gibbs, R. A. (2020) The Human Genome Project changed everything. *Nat. Rev. Genet.*, 21, 575–576
- Cantor, C. R. Biotechnology and the Human Genome Project, Serving Science and Society in the New Millenium: DOE’s Biological and Environmental Research Program. Accessed: June 15, 2021
- Blattner, F. R., Plunkett, G. 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, 277, 1453–1462
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., *et al.* (1996) Life with 6000 genes. *Science*, 274, 546–567
- The *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282, 2012–2018
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*, 287, 2185–2195
- Federspiel, N. (2000) Deciphering a weed. Genomic sequencing of *Arabidopsis*. *Plant Physiol.*, 124, 1456–1459
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496, 498–503
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297, 1301–1310
- Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, 520–562
- International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, 437, 1299–1320
- Baker, M. (2013) Big biology: The omes puzzle. *Nature*, 494, 416–419
- Friend, S. (2011) Opinion: Thinking outside the genome. *The Scientist*, Accessed: June 15, 2021
- Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.*, 35, 125–129
- Watson, J. D. and Crick, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171, 964–967
- Sanger, F. and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.*, 94, 441–448
- Maxam, A. M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.*, 74, 560–564
- Normile, D., Li, H., Couzin, J., Lawler, A., Malakoff, D. (1998) China ramps up genome effort. *Science*, 281, 1121b
- Cyranoski, D. (2001) Chinese biology. A great leap forward. *Nature*, 410, 10–12
- Collins, F. (2020) Connecting the dots from the human genome project to the COVID-19 vaccine. *Forbes*, Accessed: June 15, 2021
- Xie, Z. X., Li, B. Z., Mitchell, L. A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B. X., Liu, H. M., *et al.* (2017) “Perfect” designer chromosome V and behavior of a ring derivative. *Science*, 355, 1046
- Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D.,

- Walker, R., Zhao, H., Chen, S., Liu, W., *et al.* (2017) Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science*, 355, eaaf4791
25. Wu, Y., Li, B. Z., Zhao, M., Mitchell, L. A., Xie, Z. X., Lin, Q. H., Wang, X., Xiao, W. H., Wang, Y., Zhou, X., *et al.* (2017) Bug mapping and fitness testing of chemically synthesized chromosome X. *Science*, 355, eaaf4706
26. Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J., *et al.* (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science*, 355, eaaf3981
27. Sanger, F. (1988) Sequences, sequences, and sequences. *Annu. Rev. Biochem.*, 57, 1–29
28. Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 296, 79–92
29. Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., *et al.* (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 306, 1937–1940
30. Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., *et al.* (2010) The sequence and *de novo* assembly of the giant panda genome. *Nature*, 463, 311–317
31. Worley, K. C. and Gibbs, R. A. (2010) Genetics: Decoding a national treasure. *Nature*, 463, 303–304