

# Genomic Research

To reveal an organism's genome sequence takes similarity in drawing a map. Resorting to such genomic maps would greatly help scientists understanding how genes work together properly and why diseases occur when things go wrong in the DNA. It radically changes many aspects of molecular biology and genetics. At the very least, the genome sequence will represent a valuable shortcut, helping scientists find genes much more easily and quickly, omitting the troublesome cloning and sequencing of genes of interests.

In the field of genomic research, CAS scientists have made their contributions in revealing the genomic maps of many organisms, such the humans, rice and locusts.

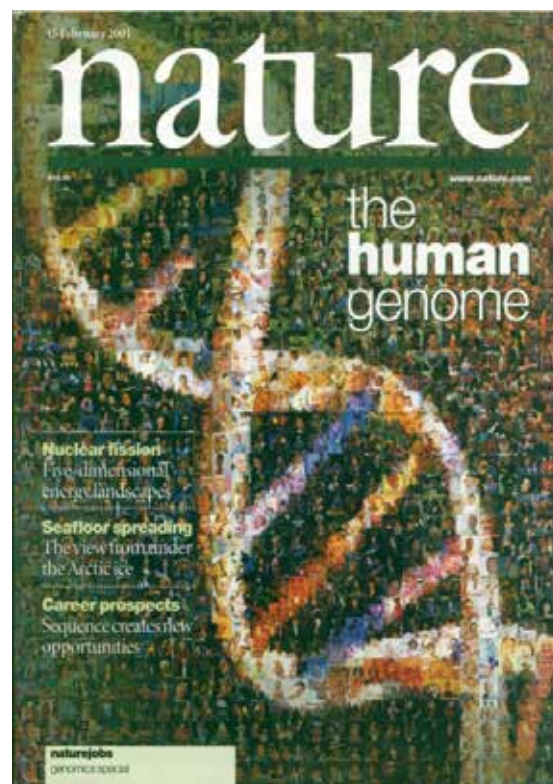
## Contribution to the Human Genome Project

The Human Genome Project (HGP) is the first and greatest endeavor so far to understand ourselves and all other forms of life through the vast international collaboration. It is widely acknowledged as one of the three most important projects in natural sciences of the 20th Century. It is also generally accepted by the scientific community as “the 2nd revolution” in life sciences, following “the 1st revolution” of discovering DNA double helix in 1953.

In July 1999, China participated in the Human Genome Project (HGP) and became the “latest contributor” following USA, UK, France, Germany, and Japan, as well as the only member from the developing world in the International Human Genome Sequencing Consortium (IHGSC).

China committed to sequencing, assembly, and analysis of a region on the tip of the short arm of chromosome 3, which accounts for about 1% of the entire human genome. The China's commitment to the HGP is thus called the “1% Project”, or the “Chinese Chapter of the Human Genome Sequence”.

The Chinese Human Genome Consortium (CHGC) committed to the HGP was composed of 15 teams from the Northern Center of National Human Genome Center (Beijing), the Southern Center of National Human Genome Center (Shanghai), the Human Genomics Center of the CAS Institute of Genetics and Beijing Genomics Institute (BGI), as well as Xi'an Jiaotong University, Southeast China University, and other institutions.



The 15 February 2001 issue of *Nature* highlights the completion of the “draft sequences” of the human genome. The data were made public in online databases, which have largely liberated molecular biologists from the need to clone and sequence genes involved in human health and disease. Despite their immediate usefulness, the draft sequences were not perfect. For example, the draft sequences omitted ~10% of the so-called ‘euchromatin’ — the gene-rich portion of the genome, which was fixed later in 2004 (*Nature* 2014, doi: 10.1038/nature03001).

Through the efforts of all teams, the CHGC submitted 64 Mb (a million of base pairs) raw data for the human genome “draft sequence” and 38 Mb data of “finished sequence” without even a single “gap” in the whole region.

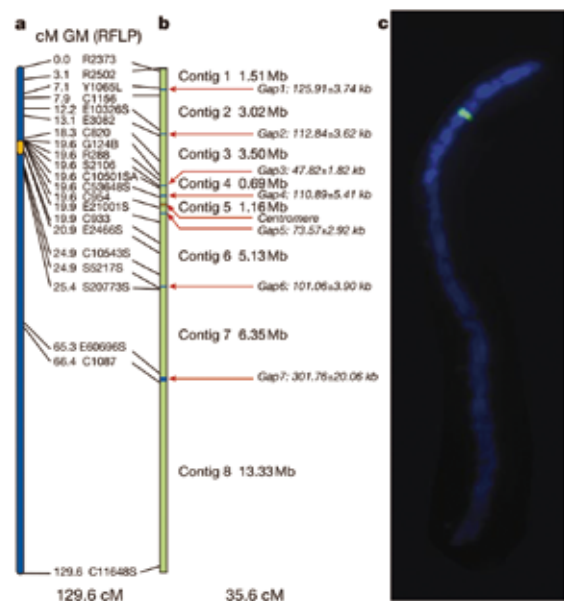
“China’s participation not only improves the international representation of the HGP and made it the first vast internationally collaborative project joined by both developed and developing countries in history, but also marks an important starting point for Chinese scientists to play their important roles on the international science stage as they are doing now,” stated YANG Huanming, a former member of the CHGC and dedicated to the HGP, in a recollection paper from the journal of *Protein Cell* in 2018.



China became one of the six contributing countries for the HGP at the 5th Strategic Meeting hosted by The International Human Genome Sequencing Consortium, August 31, 1999, Hinxton, UK.

## The Rice Genome

Later in November 2002, scientists at the CAS Institute of Genetics and Developmental Biology (IGDB) and Shanghai Institutes for Biological Sciences together with their collaborators released a completed sequence of the rice chromosome 4, with an accurate readout of



Maps of rice chromosome 4. a, Genetic map showing the positions of the centromere (yellow) and genetic markers. b, Physical map showing the locations of eight sequenced contigs (green) numbered from the top, seven remaining gaps (blue) indicated by red arrows, and the centromere (yellow) on the chromosome. c, A DAPI-stained pachytene spread of rice chromosome 4 showing heterochromatic (bright blue) and euchromatic (light blue) regions. The centromere (green) was detected by FISH with a centromere-specific probe. (Credit: *Nature*)



The December 2002 issue of *Nature* highlights the completed sequences on the rice chromosomes 1 and 4.

which bases sit at each position in rice genomic DNA.

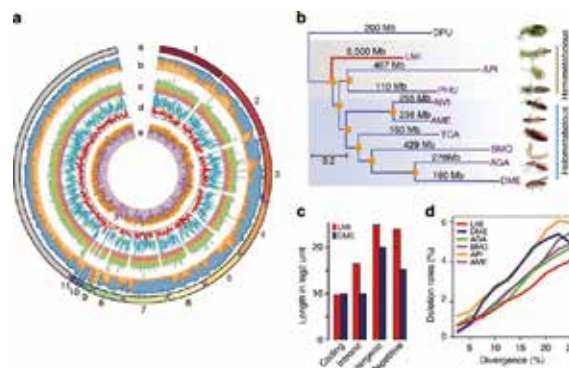
They found that 52% of the genes were not completely predicted by the draft sequence, which was published in *Science* earlier by another group of Chinese scientists (doi: 10.1126/science.1068037). This finding highlights the importance of a high-quality finished sequence. The results were published in the journal of *Nature*, entitled “Sequence and Analysis of Rice Chromosome 4” (doi: 10.1038/nature01183). Notably, they also found that there is little conservation in gene order between rice and *Arabidopsis*, a small flowering model plant. The completion of such fine sequence of rice genome greatly boosted our understanding of this principal crop that feeds over half of the population of the world.

## The Locust Genome

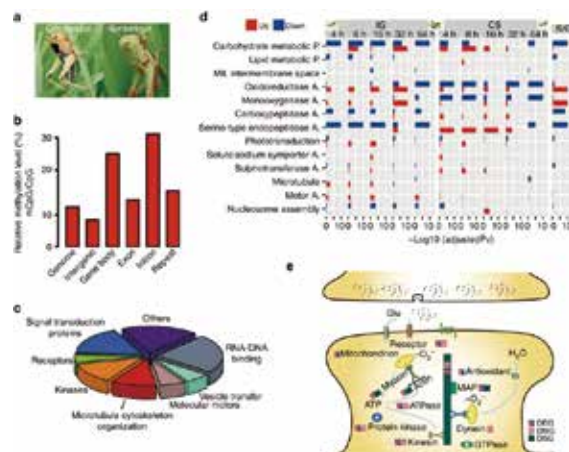
In 2014, the CAS Institute of Zoology successfully decoded the genome sequence of the locust (*Locusta migratoria*), which is the largest animal genome sequenced to date.

The researchers indicated that the large genome size of *L. migratoria*, up to 6.5 Gb, is likely due to transposable element proliferation combined with slow rates of loss for these elements (*Nature communications* 2014, doi: 10.1038/ncomms3957).

Through methylome and transcriptome analyses, the researchers revealed the complex regulatory mechanisms involved in microtubule dynamic-mediated synapse plasticity during the change of two different phases, the solitary phase and the gregarious phase. The solitary phase individual is relatively inactive and cryptically colored, but gregarious phase individual actively swarms and is conspicuously colored. They found a significant expansion of gene families associated with energy consumption and detoxification, consistent with the long-distance flight capacity and phytophagy of *L. migratoria*. They also reported hundreds of potential insecticide target genes, which hold great values for human interferences against locust outbreak that may bring disaster to the crops. The study won the 2017 Outstanding Scientific and Technological Achievement Award of the Chinese Academy of Sciences.



Locust genomic characterization and comparative analysis of insect genomes. (Credit: IOZ)



Transcriptome and methylome analysis of locust phase change between the solitary phase and the gregarious phase. (Credit: IOZ)