

## Sequencing, annotation, and comparative analysis of nine BACs of the giant panda (*Ailuropoda melanoleuca*)

ZHENG Yang

The BGI has carried out deep sequence exploration of nine BACs from the giant panda using traditional Sanger sequencing methods. The sequences and structural information of new genes and repeats were predicted, which make a large contribution to filling in gaps in the genomic study of the giant panda. In addition, the nine sequences, assembled by traditional phrap methods, will provide a validation resource for the accurate *de novo* assembly of whole genome shotgun sequencing reads of the giant panda generated by the Illumina GA sequencing technology. This very significant study is reported in Issue 53, No.1 of SCIENCE CHINA Life Sciences.

The giant panda (*Ailuropoda melanoleuca*), known as the “Chinese treasure”, is one of the most endangered species in the world. A survey revealed that only about 2500–3000 individuals remain in the wild. Its specialized diet, isolated habitat, and reproductive constraints have led to a perception that the species is at an “evolutionary dead end”. In recent years, strenuous efforts have been made to study and protect this animal. Considerable knowledge of its physiology, biochemistry, genetic diversity, and ecology has been obtained.

The MHC II genomic region and the mitochondrial sequence of the giant panda have been sequenced and annotated. A phylogenetic tree was generated using the mitochondrial sequences of the giant panda and other bear species, which suggested that the giant panda belongs to the family of Ursidae. Genome diversities of the giant panda have also been explored, with the data suggesting that the giant panda is in fact not a species at an evolutionary dead end. Conservation strategies should focus on the restoration and protection of wild habitats and the maintenance of the substantial regional genetic diversity.

In this work, we sequenced and assembled nine BACs of the giant panda, comprising a genomic region of 878 kb, which fills in gaps in the genomic study of the giant panda. Twelve protein coding genes were predicted, seven of which could be functionally annotated. About 27% of the BAC sequence was predicted to comprise repeats. A phylogenetic tree was constructed across five species, including the giant panda, human, dog, cat, and mouse, which reconfirms that, of the species analyzed, dog is the most related species to the giant panda.

Using the detailed sequence information, further studies should be carried out on interesting genes of the giant panda to obtain a deeper understanding of this species. Our data is a preview of the giant panda genome. The whole genome sequence of the giant panda has been released and an article has been recently published in *Nature*. The assembled sequences of these nine BACs were used as a validation resource for assessing the accuracy of the assembly of the whole genome sequencing reads based on new Illumina GA sequencing technology.

These studies greatly accelerate genomics studies of the giant panda, which supports the gathering of fundamental molecular biological information to guide the protection of the endangered giant panda.

Most authors are affiliated to the BGI (Beijing Genomics Institute), which is a front-runner in the field of genomics and is an influential sequencing and bioinformatics center.

Funding from the Shenzhen Municipal Government and the Yantian District Government supported this research.

**See the article:** Zheng Y, Cai J, Li J W, et al. Sequencing, annotation and comparative analysis of nine BACs of giant panda (*Ailuropoda melanoleuca*). *Sci China Life Sci*, 2010, 53: 107–111, doi: 10.1007/s11427-010-0001-z