

Yiming Bao

baoym@big.ac.cn



PROFESSIONAL EXPERIENCE

- Director of BIG Data Center, **Beijing Institute of Genomics, Chinese Academy of Sciences**, China, 2017 - Present
- Professor in “100-Talent” Program of CAS, **Beijing Institute of Genomics, Chinese Academy of Sciences**, China, 2017 - Present
- Staff Scientist, **National Center for Biotechnology Information (NCBI)/NLM/NIH**, USA, 2005 - 2017
- Viral Genome Scientist, **Computercraft Corporation (as a government contractor working at NCBI)**, USA, 2001-2005
- Postdoctoral Associate and Senior Research Associate, **Noble Foundation**, USA, 1994 - 2001
- Teaching and Research Assistant, **Peking University**, China, 1987-1991

EDUCATION

- PhD in Genetics, **John Innes Center (through University of East Anglia)**, UK, 1994
- BS in Biochemistry, **Peking University**, China, 1987

RESEARCH INTERESTS

- Bioinformatics
- Viral Genomics

PROJECTS & RESOURCES

- National Key Research and Development Project, Global Omics Data Sharing Initiative, Grant No. 2016YFE0206600, 2017-2020, leader
- The 13th Five-year Informatization Plan of Chinese Academy of Sciences, Grant No. XXH13505-05, 2017-2020, leader
- IUBS, Open Biodiversity and Health Big Data Initiative, 2017-2020, leader

HONORS & AWARDS

- 100-Talent Program, Chinese Academy of Sciences, 2017
- NIH Merit Award, 2006

ABSTRACT

BIG DATA CENTER AT BEIJING INSTITUTE OF GENOMICS AND THE BHBD INITIATIVE

Zhang Zhang, Wenming Zhao, Jingfa Xiao, Yiming Bao*

BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences,
Beijing 100101, China

*Corresponding author: baoym@big.ac.cn

Genome data are increasing dramatically as the result of new technologies. Sometimes, these data are not shared with the scientific communities because of the lack of centralized databases in China and other developing countries. Other times, data are required to be deposited into international databases such as NCBI, in order to obtain accession numbers needed for publication. This could be challenging for researchers in China and other countries because of the following reasons: 1. large data size; 2. slow data transfer due to limited international internet bandwidth; 3. language barrier and technical issues in communication. To alleviate these problems, several resources/databases were established at the BIG Data Center (BIGD, <http://bigd.big.ac.cn>), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences. One of such databases, the Genome Sequence Archive (GSA), is a data repository for archiving raw sequence reads, equivalent to SRA at NCBI. It supports data generated from a variety of sequencing platforms and provides data storing and sharing services free of charge for worldwide scientific communities. In addition to raw sequencing data, GSA also accommodates secondary analyzed files in acceptable formats. It has user-friendly web interfaces for easy data submission and provides technical support in both English and Chinese, which therefore will greatly promote genome data sharing among the scientific communities. The accession numbers provided by GSA are accepted by many journals. In addition, BIGD also hosts other resources, such as Gene Expression Nebulas, a data portal of gene expression profiles based entirely on RNA-Seq data; Genome Variation Map, a comprehensive collection of genome variations for featured species; Genome Warehouse, a centralized resource housing genome-scale data with particular focus on economically important animals and plants; Methylation Bank, an integrated database of whole-genome single-base resolution methylomes; and Science Wikis, a central access point for biological wikis developed for community annotations.

In response to the appeal "Open Data in a Big Data World" by the International Council for Science, Dr. Yongbiao Xue, director of BIG, recently proposed an initiative of Open Biodiversity and Health Big Data (BHBD) to the International Union of Biological Sciences (IUBS), of which he is an executive member. This proposal was approved by IUBS. Global sharing of BHBD is able to advance scientific research and promote the fair distribution of benefits throughout the

world, which yet cannot be accomplished without engagement of the entire global communities on the following three issues:

- To build the principles and mechanisms for global sharing of BHBD in accordance with laws and ethics of member countries.
- To develop a big data platform for BHBD integration, translation and sharing that is publicly accessible to worldwide communities.
- To promote the level of participation and influence of IUBS in global biological research.

The BHBD initiative (<http://bhbd-alliance.org/>) will be built based on BIGD resources.

Taken together, BIGD is dedicated to providing freely accessible data repositories and a variety of data resources in support of worldwide research activities. We welcome data submissions and comments/suggestions to our resources.