Home Page Text

**Revolutionizing data dissemination, organization, and use**

*Giga*DB contains discoverable, trackable, and citable data that have been assigned DOIs and are available for public download and use.

Search for dataset: {search bar: search by species, DOI or data type}

Browse by dataset {drop down menu}

{*GigaScience* logo}

For more information, see:

[www.gigasciencejournal.com](http://www.gigasciencejournal.com)

{BGI logo}

See more BGI data at <climb.genomics.cn>

**About**

*GigaScience* is a new integrated database and journal co-published in collaboration between [BGI](http://www.genomics.cn) [http://www.genomics.cn] and [BioMed Central](http://www.biomedcentral.com) [http://www.biomedcentral.com], to meet the needs of a new generation of biological and biomedical research as it enters the era of "big-data.

To achieve its goals, *GigaScience* has developed a novel publishing format that integrates manuscript publication with a database that will provide [DOI](http://www.datacite.org) [http://www.datacite.org] assignment to every dataset. Supporting the open-data movement, we require that all supporting data and source code be publically available in a suitable public repository and/or under a public domain [CC0 license](http://creativecommons.org/publicdomain/zero/1.0/) [http://creativecommons.org/publicdomain/zero/1.0/] in *Giga*DB.

*Giga*DB is the journal’s affiliated data base hosted by [BGI](http://www.genomics.cn) [http://www.genomics.cn] to increase the sharing and reuse of large biological datasets. Through our association with [DataCite](http://www.datacite.org) [www.datacite.org], we assign data DOIs to our datasets to make them more discoverable, trackable and citable. We also work with the [BioSharing](http://biosharing.org/) [www.biosharing.org] community and use [ISA-Tab](http://isa-tools.org/) [http://isa-tools.org/] tools to maintain the highest data and metadata standards.

To submit data to *Giga*DB, visit [*need to get link*] or contact us at: editorial@gigasciencejournal.com.

{DataCite logo} [www.datacite.org]

{BioSharing logo} [www.biosharing.org]

{ISA-Tab logo} [http://isa-tools.org/]

**Contact**

For more information or questions regarding submitting data to *Giga*DB, please contact us at: editorial@gigasciencejournal.com

**Terms of use**

**GigaScience Data Use Policies**

**GigaScience Database**

*GigaScience* and the BGI impose no restrictions on access to, or use of, the data provided. Data generated by the BGI and its collaborators is released into the public domain in accordance with community norms for scholarly communication. Data associated with *GigaScience* journal is specifically released under a creative commons [CC0 license](http://creativecommons.org/about/cc0). Such policies are only sustainable if scientific credit is generated for all parties involved, and the BGI is playing its part in developing a global research environment that rewards data sharing.

# BGI Data

# As one of the world’s largest Biological Data producers, the BGI’s goal is to maximize the use of its data by providing it to the research community in a timely manner. At the same time, BGI recognizes the need for researchers to be appropriately credited for their scientific contribution and investment in data generation. It is therefore expected that all researchers both honor agreements in line with the [Fort Lauderdale](http://www.genome.gov/10506537) and the [Toronto International Data Release Workshop](http://www.nature.com/nature/journal/v461/n7261/full/461168a.html) data sharing principles and appropriately acknowledge the contributions of others.

Accordingly, raw data such as individual sequence read traces are submitted to the relevant database as soon as they have exited our quality control pipelines. Whole genome sequence assemblies are released as soon as possible following appropriate quality analysis. Our repository contains draft versions of genome sequence assemblies, and we ask that you understand that these represent preliminary data, subject to omissions and errors. In addition, whole genome assemblies are likely to change upon the availability of new data, and our website will document new assembly versions as they are released.

In recognition of the extensive effort that underlies these projects, we ask that you appropriately acknowledge the use of any preliminary data. To aid researchers to find, access, and reuse data we have issued citable digital object identifiers for each dataset. Our recommended format for a data citation of each dataset is listed on each page. This recommendation is in accordance with the adopted guidelines by the genome sequencing community in a statement of principles for the distribution and use of large-scale sequencing data: [Community Resource Projects](http://www.wellcome.ac.uk/About-us/Publications/Reports/Biomedical-science/WTD003208.htm) and the resulting [NHGRI policy statement](http://www.genome.gov/page.cfm?pageID=10506537). If you have any questions regarding the use of this data, please contact us at web address: [editorial@gigasciencejournal.com](mailto:editorial@gigasciencejournal.com). In line with research norms we request that you ask us before publishing analyses of the sequences on a genome scale. We welcome collaborative interaction to provide the community with improved whole genome analyses and annotations.

**Disclaimer**

Some of the data provided from external sources may be subject to third-party constraints. Users are solely responsible for establishing the nature of and complying with any such intellectual property restrictions.

*GigaScience* and the BGI provide this data in good faith, but make no warranty, express or implied, nor assume any legal liability or responsibility for any purpose for which they are used.

**Adelie penguin**

**Image**

Available at [Image: http://eol.org/pages/1049602/overview]

*Text to Include under image*

Image source: [eol](http://www.eol.org/pages/1049602/overview) [http://eol.org/pages/1049602/overview]

**Introduction**

The Adelie penguin (*Pygoscelis adeliae*) is an iconic penguin of moderate stature and a tuxedo of black and white feathers. The penguins are only found in the Antarctic region and surrounding islands. Being very sensitive to climate change, and due to changes in their behavior based on minor shifts in climate, they are often used as a barometer of the Antarctic.

With its status as one of the adorable and cuddly flightless birds of Antarctica, they serve as a sticking point for conservationists, even though they are as of now low risk for endangerment. The sequence of the penguin can be of use in understanding the genetic underpinnings of its evolutionary traits and adaptation to its extreme environment; its unique system of feathers; its prowess as a diver; and its sensitivity to climate change. We hope that this genome data will further our understanding of one of the most remarkable creatures to waddle the planet Earth.

**Download**

[Genome assembly](ftp://ftp.genomics.org.cn/pub/G10K/Pygoscelis_adeliae/assembly/Pygoscelis_adeliae.scaf.fa.gz) [ftp://ftp.genomics.org.cn/pub/G10K/Pygoscelis\_adeliae/assembly/Pygoscelis\_adeliae.scaf.fa.gz]

**Citation**

Please cite this dataset in your references as:

Li, J; Zhang, G; Lambert, D; Wang, J (2011): Genomic data from Adelie penguin (*Pygoscelis adeliae*). GigaScience. [doi:10.5524/100006](http://dx.doi.org/10.5524/100006)   
<http://dx.doi.org/10.5524/100006>

{G10K logo} [http://www.genome10k.org/]

The Adelie penguin is part of the Genome 10K Project.

**Chinese Rhesus Macaque**

Image available at [Image: http://eol.org/pages/327960/overview]

*Text to Include under image*

Image source: [eol](http://eol.org/pages/327960/overview) [http://eol.org/pages/327960/overview]

**Introduction**

The Chinese rhesus macaque (*Macaca mulatta lasiota*) is a subspecies of rhesus macaques that mainly resides in western and central China. Due to their anatomical and physiological similarity with human beings, macaques are a common laboratory model. Also, as several macaques have been sequenced, such as the Indian rhesus macaque and the crab-eating macaque, examination of the CR genome offers interesting insights into the entire *Macaca* genus.

The DNA sample for data sequencing and analyses was obtained from a five-year old female Chinese rhesus macaque (CR) from southwestern China. The genome was sequenced on the IlluminaGAIIx platform, from which 142-Gb of high-quality sequence, representing 47-fold genome coverage for CR. The total size of the assembled CR genome was about 2.84 Gb, providing 47-fold on average. Scaffolds were assigned to the chromosomes according to the synteny displayed with the Indian rhesus macaque and human genome sequences. About 97% of the CR scaffolds could be placed onto chromosomes.

Additional information can be found at: [http://macaque.genomics.org.cn/](http://macaque.genomics.org.cn/" \t "_blank)

**Download**

Genome sequence

[CR.cns.all.fa.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.cns.all.fa.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.cns.all.fa.gz]

Gene sets

[CR.gff.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.gff.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.gff.gz]

CDS

[CR.cds.fa.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.cds.fa.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.cds.fa.gz]

Protein sequence

[CR.pep.fa.gz](http://climb.genomics.cn/Macaca_mulatta_lasiota) [We are getting the updated link]

Functional annotation data

[CR.ipr.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.ipr.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.ipr.gz]

[CR.kegg.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.kegg.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.kegg.gz]

[CR.wego.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.wego.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.wego.gz]

Genome depth

[CR.genome.depth.gz](ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.genome.depth.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CR/CR.genome.depth.gz]

**Citation**

Please cite this dataset in your references as:

Yan, G; Zhang, G; Fang, X; Zhang, Y; Li, C; Ling, F; Cooper, DN; Li, O; Li, Y; van Gool, AJ; Du, H; Chen, J; Chen, R; Zhang, P; Huang, Z; Thompson, JR; Meng, Y; Bai, Y; Wang, J; Zhuo, M; Wang, T; Huang, Y; Wei, L; Li, J; Wang, Z; Hu, H; Le, L; Stenson, PD; Li, B; Liu, X; Ball, EV; An, N; Huang, Q; Zhang, Y; Fan, W; Zhang, X; Li, Y; Wang, W; Katze, MG; Su, B; Nielsen, R; Yang, H; Wang, J; Wang, X; Wang, J (2011): Genomic data from the Chinese Rhesus Macaque (*Macaca mulatta lasiota*). GigaScience. [doi:10.5524/100002](http://dx.doi.org/10.5524/100002)   
<http://dx.doi.org/10.5524/100002>

{G10K logo} [http://www.genome10k.org/]

The Chinese Rhesus Macaque is part of the Genome 10K Project.

**Crab Eating Macaque**

[Image: http://eol.org/pages/327959/overview]

Image Source: [eol](http://www.eol.org/pages/327959" \t "_blank) [http://eol.org/pages/327959/overview]

**Introduction**

The crab-eating macaque (*Macaca fascicularis*), also known as the Java macaque or long-tailed macaque, is a species of primate located throughout Southeast Asia. Due to the frequent usage of the genus *Macaca* in scientific research, the sequence the crab-eating macaque furthers our understanding on how it differs from other macaque species, like the Chinese rhesus macaque and the Indian rhesus macaque. This is especially relevant considering the recent trend of using crab-eating macaque (CE) and Chinese rhesus macaques rather than the Indian rhesus macaque as laboratory models.

The DNA sample for genome sequencing and analyses was from a female CE that was a captive-bred descendent of a CE from Vietnam. The genome was sequenced on the IlluminaGAIIx platform, and we obtained 162-Gb of high-quality sequence, representing 54-fold coverage. The sequencing data were processed with Illumina custom computational pipelines. The genome was *de novo* assembled using SOAPdenovo program based on the de Bruijn graph algorithm methods. The total size of the assembled genome was about 2.85 Gb, providing 54-fold coverage on average. The scaffolds were assigned to the chromosomes according to the synteny displayed with the Indian rhesus macaque and human genome sequences. About 92% of the CE scaffolds could be placed onto chromosomes. For additional information see: <http://macaque.genomics.org.cn/>

**Download**

Genome sequence

[CE.cns.all.fa.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.cns.all.fa.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.cns.all.fa.gz]

Gene sets

[CE.gff.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.gff.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.gff.gz]

CDS

[CE.cds.fa.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.cds.fa.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.cds.fa.gz]

Protein sequence

[CE.pep.fa.gz](http://climb.genomics.cn/Macaca_fascicularis) [http://climb.genomics.cn/Macaca\_fascicularis]

Functional annotation data

[CE.ipr.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.ipr.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.ipr.gz]

[CE.kegg.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.kegg.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.kegg.gz]

[CE.wego.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.wego.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.wego.gz]

Transcriptome data

[Transcriptome](ftp://ftp.genomics.org.cn/pub/macaque/CE/transcriptome/) [ftp://ftp.genomics.org.cn/pub/macaque/CE/transcriptome/]

Genome depth

[CE.genome.depth.gz](ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.genome.depth.gz) [ftp://ftp.genomics.org.cn/pub/macaque/CE/CE.genome.depth.gz]

**Citation**

Please cite this dataset as:

Yan, G; Zhang, G; Fang, X; Zhang, Y; Li, C; Ling, F; Cooper, DN; Li, O; Li, Y; van Gool, AJ; Du, H; Chen, J; Chen, R; Zhang, P; Huang, Z; Thompson, JR; Meng, Y; Bai, Y; Wang, J; Zhuo, M; Wang, T; Huang, Y; Wei, L; Li, J; Wang, Z; Hu, H; Le, L; Stenson, PD; Li, B; Liu, X; Ball, EV; An, N; Huang, Q; Zhang, Y; Fan, W; Zhang, X; Li, Y; Wang, W; Katze, MG; Su, B; Nielsen, R; Yang, H; Wang, J; Wang, X; Wang, J (2011): Genomic data from the Crab Eating Macaque/Cynomolgus Monkey (*Macaca* *fascicularis*). GigaScience. [doi:10.5524/100003](http://dx.doi.org/10.5524/100003)   
<http://dx.doi.org/10.5524/100003>

{G10K logo} [http://www.genome10k.org/]

The Crab Eating Macaque is part of the Genome 10K Project.