Basic Medical Data sharing at BMICC

Heng Wang, Professor, Director

Institute of Basic Medical Sciences

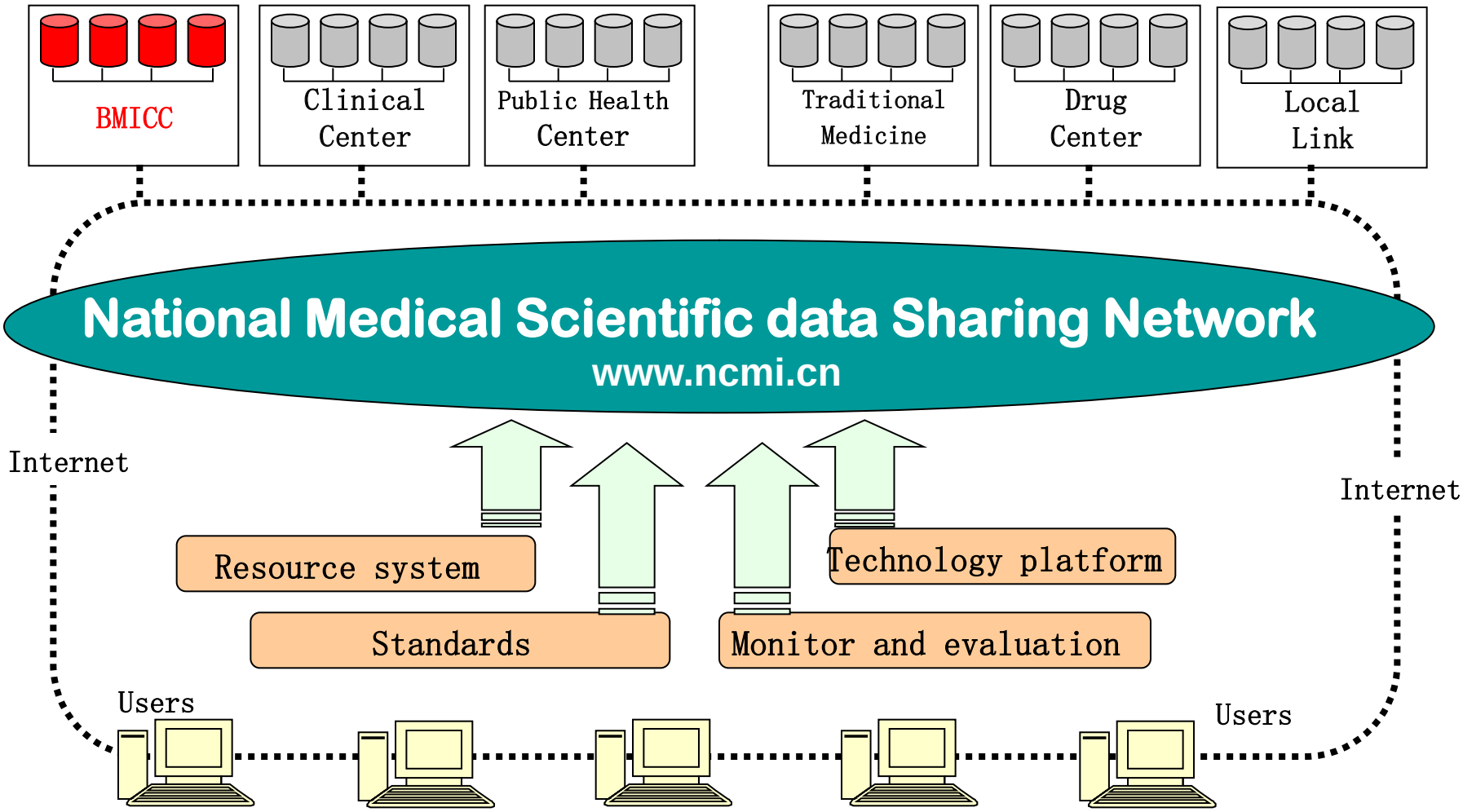
Chinese Academy of Medical Sciences

School of Basic Medicine

Peking Union Medical College

ABOUT BMICC

BioMedical Informatics Center of China


BMICC
Clinical Center
Public Health Center
Traditional Medicine
Drug Center
Local Link
National Medical Scientific data Sharing Network
www.ncmi.cn
Internet
Internet
Resource system
Technology platform
Standards
Monitor and evaluation
Users
Users

Review

National Medical Scientific Data Sharing Network
(2004-2009)

↓

National Scientific Data Sharing Network for Population and Health (2009)

National Biomedicine Database
(2001-2006)

↓

BioMedical Information Center of China (BMICC)
(2004 -)

Co-Units

12 Institutes/Universities, 4 Provinces,

- Institute of Basic Medical Sciences, CAMS
- Center for Bioinformatics in Peking University;
- Institute of Biophysics of CAS;
- Institute of Computing Technology of CAS ;
- Beijing Genomics Institute ;
- Beijing Genomics Institute at Shenzhen
- Bioinformatics Division Tsinghua University;
- Beijing Proteome Research Center;
- Model Animal Research Certer of Nanjing University
- Capital Institute of Pediatrics;
- Third Military Medical University ;
- Fuwai Hospital, CAMS

GOAL OF BMICC

- ★ To establish the technical platform for integrating and sharing the scientific data
- ★ To collect and integrate most valuable and available databases related to the study on health and diseases in China
- ★ To create a national scientific data sharing center for basic medical researchers and various users in different requirements

Resources of Data

Original

Requirements

Quality

Bmicc

Research Funding

National Science Fund of China (NSFC)
National Science Fund for Distinguished Young Schoolars, NSFC
Beijing Natural Sciences Foundation
The National Program for Key Basic Research Projects (973)
The National High Technology Research and Development Program (863)
The Research Fund for Doctoral Program of Higher Education
Project of Ministry of Science and Technology
International cooperation program from NIH, US
International Cooperation program between China and Europe
Chinese Medical Board (CMB) from America

Data from Laboratory

Data from Health Survey

Data from Health Survey

Administration Model

JOINY CONFERENCE

Authoritativeness

- Strategies
- Aim and assignment of the sub-centers
- Financial plan

ACADEMIC COMMITTEE

Academic activities

- Technical advisory
- Monitor and revise
- Evaluation

MANAGEMENT OFFICE

Implementation

- Reports to the Sharing Network
- Collect and integrate database
- Maintain the work on technique platform and network

RULES AND STANDARDS I

- ☆ Technical standards for data sharing
- ☆ The metadata standards
- ☆ The terminological standards in medicine
- ☆ The standards on the quality control for data searching

RULES AND STANDARDS II

- ☆ **The service guide**
- ☆ **A manual for developing technique framework**
- ☆ **Guide for data set execution.**
- ☆ **A rule for data backup manage**

Implementation & Progress


配电室
AMP
交换机

National Science & Technology Infrastructure Center
National Science Data Sharing Project

National Madical Scientific Data Sharing Network Biolicine information Center

Home Page | Database Search | Online Tools | MetaData Query | Standard Criterion | Data Center | About BMICC | Login

Original scientific data produced in China

Molecule --Individual--Population

24 hour service

Global Service (2versions)

Metadata Query

Metadata Query

Molecular Mechanisms Database

NcRNA(The database of all kinds of noncoding RNA)
NPInter(Functional interactions between noncoding RNAs and proteins Database)
ATID(Alternative Translational Initiation Database)
dbRES(dbRES: A web-oriente... annotated RNA Editing Site)
dbNEI(database for Neuro-En... Immune)
SPD(Secreted Protein Database)
SynDB(Synapse Database)
The Biology Database FTP Mirror Site
Yanhuang Database new!
Human Urinary Proteome Database new!
database of potential target genes for clinical diagnosis and immunotherapy of human carcinoma new!
natural sense-antisense transcripts database new!

Biomodel Organism Database

ChickVD(Chicken Variation Database)
_SilkDB(Silkworm Knowledgebase)
PigGIS(Pig Genomic Informatics System)
TFDB(Transcription Factor Databases)

Experiment Material Data Resource

China National Cell Resources Confederation new!
Phenotype Database for Genetically Engineered Mouse Disease Models new!

The aim of Sharing Scientific Data Program: Base on the public and communal resource, support by the research center, to build a data and service system with reasonable, global, and intellective frame. To Improve the public policy, rule management system. To Cultivate manager and intelligent with high personality and high ability of utilization information and public

About BMICC

...oRNA Identification Based on Sequence and Structure
MethCGI(Predict the methylation status of CpG islands in the human brain)
SubMito(Predict protein submitochondria locations from its primary sequence)
PhoScan
snap - a snap annotation platorm new!

...oject-specific databases
...ase new!
... and integrated search new!

Research Developments

New biomarker for heart fa..
Blood levels of resistin, a hormone ...dependently
...it..
...ntists have ...preferences f

Muscular protein bond -- strongest yet f..2009-07-21
Overfishing and evolution..2009-07-21
Slotted buses keep passengers cool..2009-07-21
Practice makes perfect -- motor memory p..2009-07-21
JCI table of contents: July 20, 2009..2009-07-21
Chasing tiny vehicles..2009-07-21
'Invisibility cloak' could protect ag...
more >>

Satisfaction investigate for BMICC

Are you satisfied with this website?

a. Satisfactory
b. Relative satisfactory
c. Dissatisfactory

Vote

Search

Search...

language/语言

English/ 中 文

Associate Of Bmicc

基础医学研究所
CBI Center of Bioinformatics
中国科学院 计算技术研究所
华大基因 BGI
中国科学院 生物物理研究所

Data Center

Biologic Medicine | Clinic Medicine
TCM | Public Health
Pharmacal Science | Local Notes

Related Links

Ministry of Science&Technology P.R China.
Ministry of Health P.R China.
National Science & Technology Infrastructure Center
...dical Scientific Data Sharing

Address:5 Dong Dan San Tiao BeiJing,100005,China E_mail:bmicc@pumc.edu.cn
Institute of Basic Medical Sciences Chinese Academy of Medical Sciences

Integrated and Shared

25 databases

- Population Survey Database (6)
- Molecular Mechanism Database (13)
- Biomodel Organism Database (4)
- Experiment Material Data Resource (2)

Sharing Database (1)

Database —— population, individual person	Sub-centers
(1) Physiological Reference Database of Chinese	IBMS
(2) Psychological Reference Database of Chinese	IBMS
(3) Sub-health database of Chinese	IBMS
(4) Maternal and Child Nutrition Reference Database	Capital Institute of Pediatrics
(5) A Multi-center Survey on Cardiovascular Diseases in Chinese	CAMS
(6) Visible Human Construction	Third Military Medical University

Sharing Database (2)

Database —— Molecular studies	Sub-centers
(7) NcRNA (The database of all kinds of noncoding RNA)	ICT CAS
(8) NPInter (Functional interactions between noncoding RNAs and proteins Database)	IB CAS
(9) ATID (Alternative Translational Initiation Database)	Tsinghua University
(10) dbRES (dbRES: A web-oriented database for annotated RNA Editing Site)	Tsinghua University

Sharing Database (3)

Database	Sub-centers
(11) dbNEI (database for Neuro-Endocrine-Immune)	Tsinghua University
(12) SPD (Secreted Protein Database)	Peking University
(13) SynDB (Synapse Database)	Peking University
(14) The Biology Database FTP Mirror Site	Peking University
(15) Human Urinary Proteome Database	IBMS
(16) Database of potential target genes for clinical diagnosis and immunotherapy of human carcinoma	ICT CAS
(17) Natural sense-antisense transcripts database	Peking University

Sharing Database (4)

Database	Sub-centers
(18) TFDB (Transcription Factor Databases)	Peking University
(19) Natural sense-antisense transcripts database	Peking University
(20) Database on antiCODE	IB CAS
(21) Liver Expression Profile	Beijing Proteome Research Center
Animal Modes	
(22) SilkDB (Silkworm Knowledgebase)	Beijing Genomics Institute
(23) PigGIS (Pig Genomic Informatics System)	Beijing Genomics Institute

Sharing Database (5)

Database	Sub centers
(23) ChickVD(Chicken Variation Database)	Beijing Genomics Institute
Experimental Materials	
(24) China National Cell Resources Confederation	IBMS
(25) Phenotype Database for Genetically Engineered Mouse Disease Models	Model Animal Research Certer of Nanjing University

Online Tools

- MiRAlign
- MethCGI
- SubMito
- PhoScan

User groups in China

- Scientists / specialists
- Health officials
- Students
- Public

Browse Record Distribution by Month

中国医学科学院欢迎您!
我的帐户 - 退出
- 页面设置

搜索...
系统管理 基础医学 博士 (个人)

平台管理 | 流量统计分析

本站访问日志统计分析:
www.bmicc.cn

摘要
按访问时间:
按月历史统计
按日期统计
按星期统计
每小时浏览次数
按访问者:
国家或地区
全部列出
主机
全部列出
最近访问日期
无法解析的IP地址
搜索引擎网站的机器人
全部列出
最近访问日期
浏览器统计:
每次访问所花时间
文件类别
存取次数
全部列出
入站处
出站处
操作系统
版本
无法得知
浏览器
版本
无法得知
反相链接:
来源网址
由那些搜索引擎转介
由那些其他网站转介
搜索
用以搜索的短语
用以搜索的关键词

最近更新: 2009年 08月 09日 13:55 立即刷新
报表日期: 八月 2009 OK

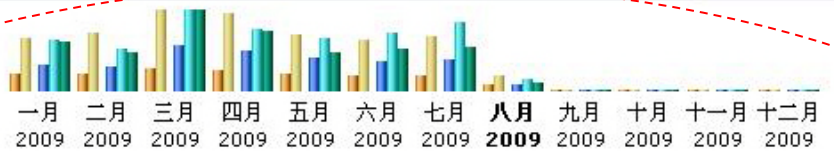
摘要

报表日期	2009年 08月 月报
首次访问日期	2009年 08月 01日 08:01
最近访问日期	2009年 08月 08日 22:45

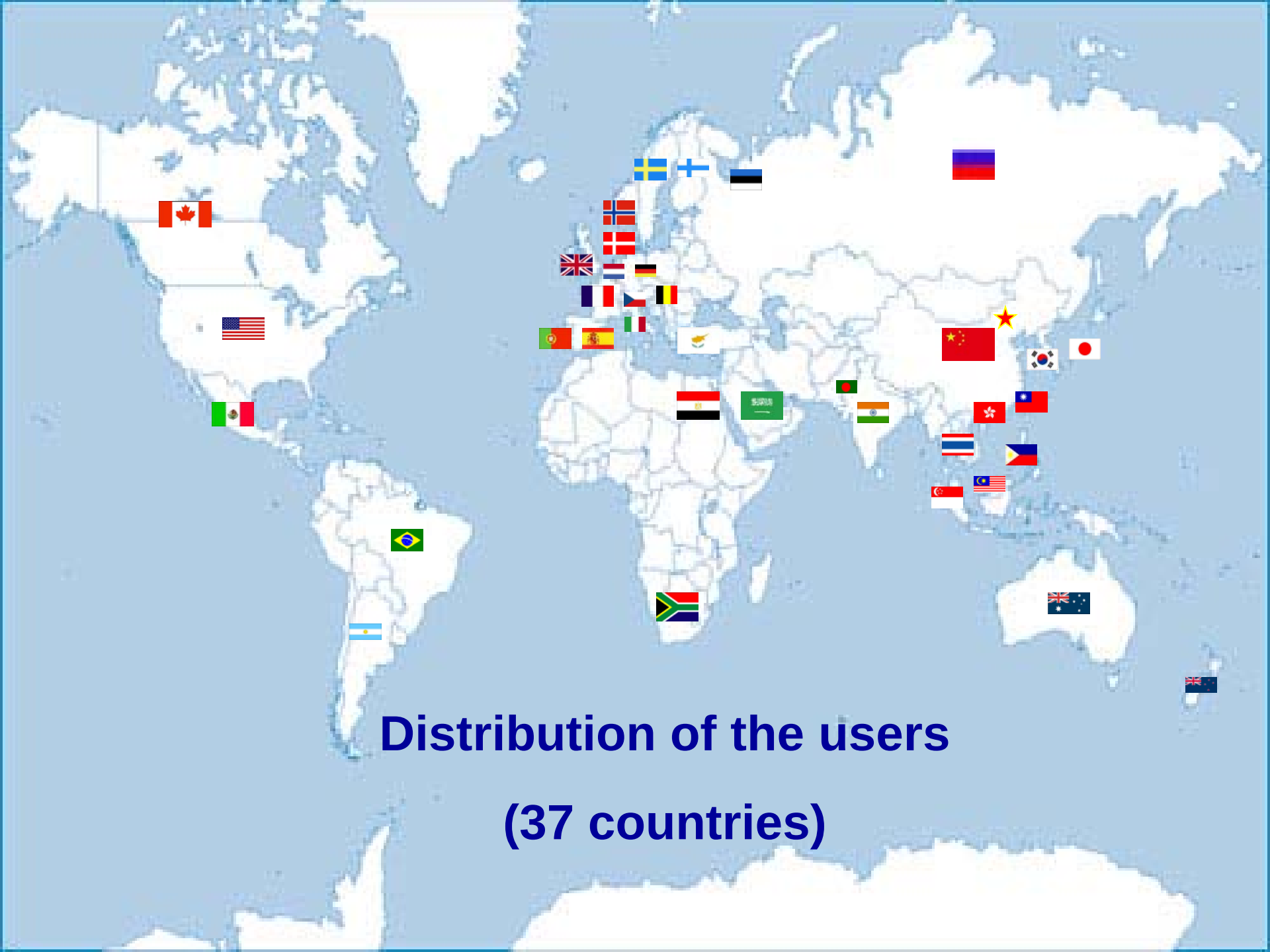
	访问者	访问人次	网页数	文件数	字节
浏览器流量 *	**1104**	**2373** (2.14 访问人次/访问者)	**9561** (4.02 网页数/访问)	**18248** (7.68 文件数/访问)	**40.25 M字节** (17.36 K字节/访问)
非浏览器流量 *			**1449**	**1583**	**605.76 K字节**

* 非浏览的流量包括搜索引擎机器人，蠕虫病毒产生的流量和非正常的HTTP相应

按月历史统计



月	访问者	访问人次	网页数	文件数	字节
一月 2009	2600	8470	39068	79041	228.25 M字节
二月 2009	2859	9270	37525	65922	184.57 M字节
三月 2009	3459	12912	69304	124290	378.58 M字节
四月 2009	3365	12354	61174	94977	282.14 M字节
五月 2009	2712	9134	51341	81072	180.87 M字节
六月 2009	2495	8224	46652	88967	196.04 M字节
七月 2009	2571	8824	49217	107648	206.98 M字节
八月 2009	1104	2373	9561	18248	40.25 M字节
九月 2009	0	0	0	0	0



Distribution of the users

(37 countries)

VISITERS from all of the world

(Jan. 2008 --- July 2009)

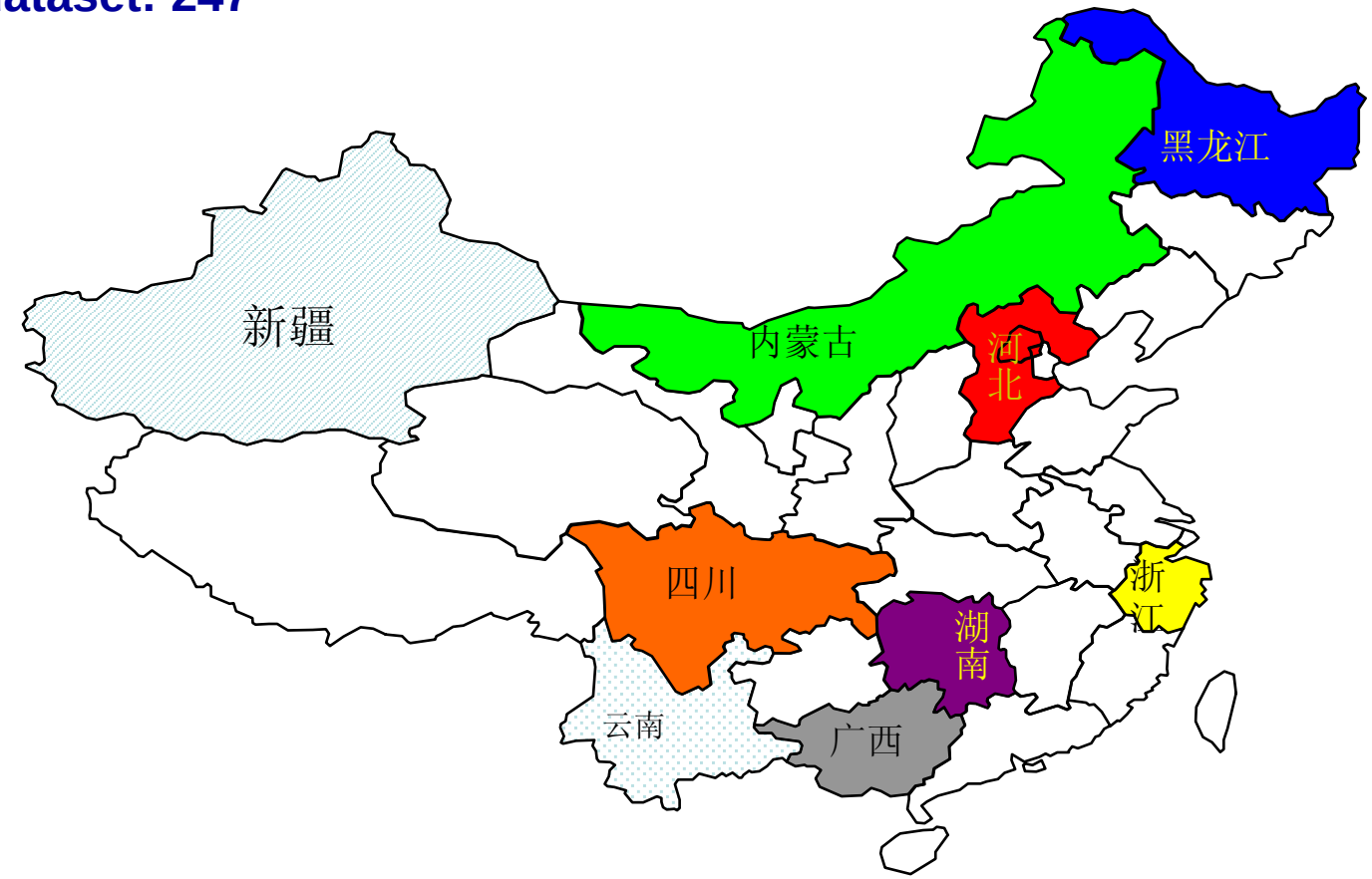
Countries	person times
United States	97322
China	36688
Australia	30567
Germany	2936
Japan	737
Great Britain	423
South Korea	351

Data Sharing Between CHINA & USA from BMICC

Chinese physiological reference dataset

Physiological Reference of Health Population in China

Population: 130,000
Provinces: 9
Races: 8
Variables in dataset: 247

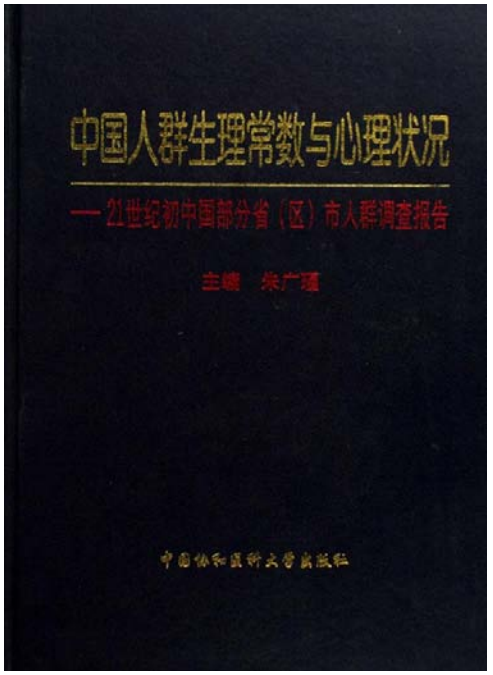


Data profile of Chinese physiological reference

The categories of variables surveyed

- Characteristics of demography
- Anthropometric measures
- Blood cell counts
- Blood biochemical variables
- Urinalysis
- Electrocardiogram ECG
- Immunological analysis
- variables of circulation system
- variables of respiratory system

Published Data
for 30,000 population



Continued survey & Data expanding

- **Continue survey**
- **More genome data of individual person**
- **Proteomics data**
- **Government support (the twelve five plan)**

Yan Huang Database

Yan Huang database

(Beijing Genomics Institute at Shenzhen)

- The genome, named as YH, is a very start of YanHuang Project, which aims to sequence 100 Chinese individuals in 3 years.
- To illustrated the personal genome data in a MapView, which is powered by GBrowse.
- A new module was developed to browse large-scale short reads alignment. This module enabled users track detailed divergences between consensus and sequencing reads.

Bmicc

- Efforts on designing the YH database are helpful attempts to organize and present personal genome data, which is a useful resource for genomic and medical researches.
- As the third published personal genome, YH diploid genome accelerates the discovery of disease gene and mutation in Asian population.

Bmicc

Mapview
Blast
Download
Craig Venter's Genome
James Watson's Genome

Home | Mapview | BLAST | Download | Help

search Chr 01 YHSNP ID YHSNP0144025 Search

Introduction

On October 11th, 2007, Beijing Genomics Institute at Shenzhen (BGI-Shenzhen) announced the completion of first diploid genome sequence of a Han Chinese, a representative of Asian population. The genome, named as YH, is a very start of YanHuang Project, which aims to sequence 100 Chinese individuals in 3 years.

We set up this 'YH database' to present the entire DNA sequence assembled based on 3.3 billion reads (117.7Gbp raw data) generated by Illumina Genome Analyzer. In total of 102.9Gbp nucleotides were mapped onto the NCBI human reference genome (Build 36) by self-developed software SOAP (Short Oligonucleotide Alignment Program), and 3.07 million SNPs were identified.

We illustrated the personal genome data in a MapView, which is powered by GBrowse. A new module was developed to browse large-scale short reads alignment. This module enabled users track detailed divergences between consensus and sequencing reads. In total of 53,643 HGMD recorders were used to screen YH SNPs to retrieve phenotype related information,to superficially explain the donor's genome. Blast service to align query sequences against YH genome consensus was also provided.

Data Statistics		
Nucleotide	Total	117.7Gbp
	Map to genome	102.9Gbp
	Coverage of genome	99.97%
Polymorphism	SNP	3.07M
	Indel	135262
	Structural Variation	2682

Our efforts on designing the YH database are helpful attempts to organize and present personal genome data, which is a useful resource for genomic and medical researches. As the third published personal genome, YH diploid genome accelerates the discovery of disease gene and mutation in Asian population. Companying with other personal genome projects, this endeavor will achieve fundamental goals for establishing personal medicine.

What's new?

Nature News: China launches large-scale human sequencing initiative

Data Statistics		
Nucleotide	Total	117.7Gbp
	Map to genome	102.9Gbp
	Coverage of genome	99.97%
Polymorphism	SNP	3.07M
	Indel	135262
	Structural Variation	2682

ARTICLES

The diploid genome sequence of an Asian individual

Jun Wang[1,2,3,4]*, Wei Wang[1,3]*, Ruiqiang Li[1,3,4]*, Yingrui
Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,7], Bi
Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,7], Y
Ines Hellmann[9], Michael Inouye[8], John Pool[9], Xin Yi[1,7], Jing
Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1],
Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,7], Qibin Li[1,7], Hongmei Zh
Jianguo Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,7], Quan Che
Shuang Yang[1], Fang Chen[1,7], Li Li[1], Ke Zhou[1], Hongkun Z
Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristians
Richard Durbin[8], Lars Bolund[1,11], Xiuqing Zhang[1,6], Songg

**Here we present the first diploid genome sequence of an Asia
coverage using massively parallel sequencing technology. We
genome to 99.97% coverage, and guided by the reference ge
high-quality consensus sequence for 92% of the Asian individ
single-nucleotide polymorphisms (SNPs) inside this region, of
analysis showed that SNP identification had high accuracy an
assembly. We also carried out heterozygote phasing and hapl
(Chinese and Japanese, respectively), sequence comparison w
C. Venter), and structural variation identification. These variati
sequence data and analyses demonstrate the potential usefuln
genomics.**

The completion of a highly refined, encyclopaedic human genome sequence[1,2] was a major scientific development. Such reference sequences have accelerated human genetic analyses and contributed to advances in biomedical research. Given the growth of information on genetic risk factors, researchers are developing new tools and analyses for deciphering the genetic composition of a single person to refine medical intervention at a level tailored to the individual. The announcements that J. Craig Venter and James D. Watson have had their genomes sequenced[3,4], along with the announcement of the Personal Genome Project[5], highlight the growth of personal genomics.

Using a massively parallel DNA sequencing method, we have generated the first diploid genome sequence of a Han Chinese individual, a representative of an East Asian population that accounts for nearly 30% of the human population. The consensus sequence of the donor, assembled as pseudo-chromosomes, serves as one of the first sequences available from a non-European population and adds to the small number of publicly available individual genome sequences. This sequence and the analyses herein provide an initial step towards attaining information on population and individual genetic variation, and, given the use and analysis of next-generation sequencing

[1]Beijing Genomics Institute at Shenzhen, Shenzhen 518000, China. [2]Genome Research Insti
Center for Genomics and Bioinformatics, Beijing 101300, China. [4]Department of Biochemistr
[5]College of Life Sciences, Peking University, Beijing 100871, China. [6]Beijing Genomics Institute, Beijing Institute of Genomics of Chinese Academy of Sciences, Beijing 101300, China.
[7]The Graduate University of Chinese Academy of Sciences, Beijing 100062, China. [8]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [9]Departments of Integrative Biology and Statistics, University of California, Berkeley, California 94720, USA. [10]Department of Biological Sciences and Department of Medicine, University of Alberta, Edmonton AB, T6G 2E9, Canada. [11]Insitute of Human Genetics, University of Aarhus, Aarhus DK-8000, Denmark.
*These authors contributed equally to this work.

Maternal and Child Nutrition Reference Database

Maternal and Child Nutrition Reference Database

- **Population**

  Children and women in 14 provinces, China

- **Information**

  Incidence of low birth weight

  Incidence of anaemia in reproductive aged women

  Prevalence of Vitamin A deficiency in child

- **Self-evaluation**

  Fetus weight

  Anaemia in reproductive aged women

  Vitamin A deficiency in child

快速导航
项目首页
低出生体重
育龄妇女贫血
儿童维生素A缺乏
健康自测

首页 | 低出生体重 | 育龄妇女贫血 | 儿童维生素A缺乏 | 健康自测

ITEM INTROS 项目背景介绍

1990年世界儿童问题首脑会议和国务院颁发的“九十年代中国儿童发展规划纲要”提出了一系列2000年战略目标，其中卫生保健目标占大部分。国务院妇女儿童工作委员会“九十年代中期中国儿童发展状况报告”表明卫生保健绝大部分目标已达到或超过中期目标的要求，中国妇女儿童的健康状况有了明显提高。但低出生体重儿发生率，育龄妇女缺铁性贫血患病率，儿童维生素A缺乏患病率三项指标尚缺乏全国性资料。在联合国儿童基金会的资助下，1998-2000年卫生部委托首都儿科研究所牵头，组织全国14个省、自治区、直辖市对以上反映妇女儿童营养状况的重要指标进行全国范围的调查研究。

低出生体重儿发生率是反应孕期工作质量、孕妇、胎儿及新生儿营养状况的重要指标，也是社会发展的重要指标。低出生体重儿死亡也是我国5岁以下儿童死亡的第三位死因，占我国5岁以下儿童死亡的10% 以上。监测资料表明，由于感染性疾病的死亡近几年明显下降，而低出生体重儿死亡率下降不明显，低出生体重儿死亡占5岁以下儿童死亡构成比和重要性呈上升趋势。然而至今我国尚无确切的90年代全国低出生体重发生率资料，1995年全国年报资料低出生体重发生率2.02% ，低于世界所有国家；国内有几次城市住院低出生体重调查发生率4-6% 左右，不能反映全国情况；1992年国家统计局进行中国儿童基本情况调查，低出生体重发生率，广东14.3% ，四川18.9% ，甘肃26.3% ，广西33.8% ，尤其海南高达53% ，可能大大超过实际情况。因此需要获得准确、可靠的全国低出生体重发生率资料。

育龄妇女缺铁性贫血和儿童维生素A缺乏患病率是分别反映妇女和儿童营养状况的重要指标。妊娠妇女贫血不仅影响母亲，也影响胎儿的健康成长，而儿童维生素A缺乏可导致儿童呼吸道感染和腹泻发病率、死亡率的增加，此外也是导致儿童失明的重要原因，但以上患病情况国内均只有局部地区的资料，也需要进行全国性专门调查。

本数据库的数据资料可作为实施《中国儿童发展纲要（2001-2010年）》和《中国妇女发展纲要（2001-2010年）》的基础数据。

THE CRUCIAL QUESTIONS ABOUT THE INTERNATIONAL DATA SHARING FOR BIOMEDICINE

- Standard for data collection and management between US & China
- Standard for the construction of the data sharing platform
- Strategies and mechanism of the data sharing system
- A professional team fully supported by the government is necessary for the program.

Challenge

- ★ Consciousness on sharing among researchers
- ★ Requirement for specialists on cross-sciences (biomedical and computer science)
- ★ Financial support on maintaining the network for a long term development

Plan for the next step

- ★ Continuing to search new data resource
- ★ To improve performance of technique platform
- ★ To develop the conjunction among datasets
- ★ International communication & cooperation

.....

Next steps ?

- **It may require close collaboration on point-to-point between us to make the co-data work effectively.**
- **It is needed to discuss how to design and implement science data transfers system and present a framework of data transfers linking.**
- **Complete the understanding of network performance & transfer rates**
- **Need to identify user requirements & set the routing to allow use of the co-data links.**
- **It is necessary to set up the linking mirrors in two sites for widening the user base.**
- **Considering to provide extensive international connectivity to other world regions**


Work team

THANK YOU