中国NCBI之梦

China's NCBI – a dream?

基因组所大数据中心国际顾问委员会第1届会议

2017年1月17日

中科院北京基因组研究所

罗静初

北京大学生命科学学院

北京大学生物信息中心

luojc@pku.edu.cn

Outline

- The significance of bioinformatics centers for biomedical research and development
- The bioinformatics service and resource at CBI
- Prof Hao's proposal to set up China's NCBI
- A survey project to set up China's NCBI
- BIGD – a good chance and a big challenge

The discovery of the double helix

Nature 171, 737-734 (1953) (C) Macmillan Publishers Ltd.

Molecular structure of Nucleic Acids

WATSON, J. D. & CRICK, F. H. C.

Medical Research Council Unit for the Study of Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

Figure 1

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

The paradigm shift in biology research

We must hook our individual computers to the worldwide network that gives us access to daily changes in the database and also makes immediate our communication with each other. The programs that display and analyse the materials for us must be improved …

Towards a paradigm shift in biology

Walter Gilbert, *Nature*, 1991

The birth of biology databases

- The 1st biological database Protein Data Bank (PDB) was started in 1971.
- The two DNA sequence databases EMBL and GenBank were founded in early 1980’.
- The two protein sequence databases PIR and Swiss-Prot were built in middle 1980’.

核酸序列数据库

- 1980年，欧洲分子生物学实验室核酸序列数据库EMBL正式宣告诞生，1982年6月，EMBL第1版正式对外发布。
- 1982年，Walter Goad等竭力推动，美国国家健康研究院NIH、科学研究基金会NSF、能源部DOE和国防部DOD等共同资助，创建了核酸序列数据库GenBank，由洛斯阿拉莫斯国家实验室负责运行。1987-1992年，由InteliGentics公司负责分发。
- 1992年起，美国国家生物技术信息中心NCBI接管GenBank，负责核酸序列收集、储存、管理、注释、分发，并开发基于网络浏览器的数据库检索系统。
- 1986年，日本国立遗传研究所NIG建立日本DNA数据库DDBJ，并和GenBank、EMBL共同成立国际核酸序列数据库协会INSDC。

蛋白质序列数据库

- 1984年，美国国家生物医学基金会NBRF Winona Barker等基于Margaret Dayhoff收集的数据，建立了蛋白质序列数据库PIR。
- 1986年，瑞士日内瓦大学Amos Bairoch创建了蛋白质序列数据库Swiss-Prot，该数据库具有大量注释信息和交叉链接。
- 1995年，欧洲生物信息学研究所Rolf Apweiler创建了蛋白质序列数据库TrEMBL，收集从EMBL翻译得到的蛋白质序列。
- 2003年，Swiss-Prot、TrEMBL和PIR合并，建立了国际蛋白质知识库UniProt，统一收集、管理、注释、发布蛋白质序列数据。

美国国家生物技术信息中心NCBI

- 1988年11月，由已故参议员Claude Peper提议成立。位于华盛顿北郊马里兰州，隶属NIH下的NLM。成立初期仅8名工作人员，现已增加到500多名。
- 运用最新的计算机和信息技术，创建方便实用的生物信息存储和分析系统，开发先进的生物信息处理方法，整合国际公共数据库资源，为生物医学领域提供内容丰富、更新及时的生物信息资源。
- David Lipman任NCBI主任，2003年当选为美国科学院院士，2004年获ISCB颁发的Senior Accomplishment Award。2009年应邀参加在北京举行的亚太地区生物信息学大会，作关于流感病毒起源和演化的报告。2013年获白宫Open Science奖。

About NCBI

www.ncbi.nlm.nih.gov

Our Mission

NCBI's contribution to the NIH mission of 'uncovering new knowledge'

Organizational Structure

The role of the branches within NCBI and the Board of Scientific Counselors.

Programs & Activities

NCBI's resources for genomic, genetic, and biomedical data

Researchers at NCBI

The basic research program conducted by our intramural investigators

Contact us

More questions? Write to us. We are here to help.

Learn more about our site

We offer webinars, courses, tutorials, help documentation and more...

Follow Us

NCBI News

July 27th NCBI Minute: Important Changes to NCBI Web Protocols

Wednesday, July 13, 2016

Sequence Viewer 3.15 is now available

Wednesday, July 13, 2016

July 20th NCBI Minute: Important Changes Coming to Sequence Databases

Tuesday, July 12, 2016

Conserved Domain Database (CDD) version 3.15 now available online and via FTP

Tuesday, July 12, 2016

RefSeq release 77 is now available

Thursday, July 7, 2016

Mouse and zebrafish genome annotations updated

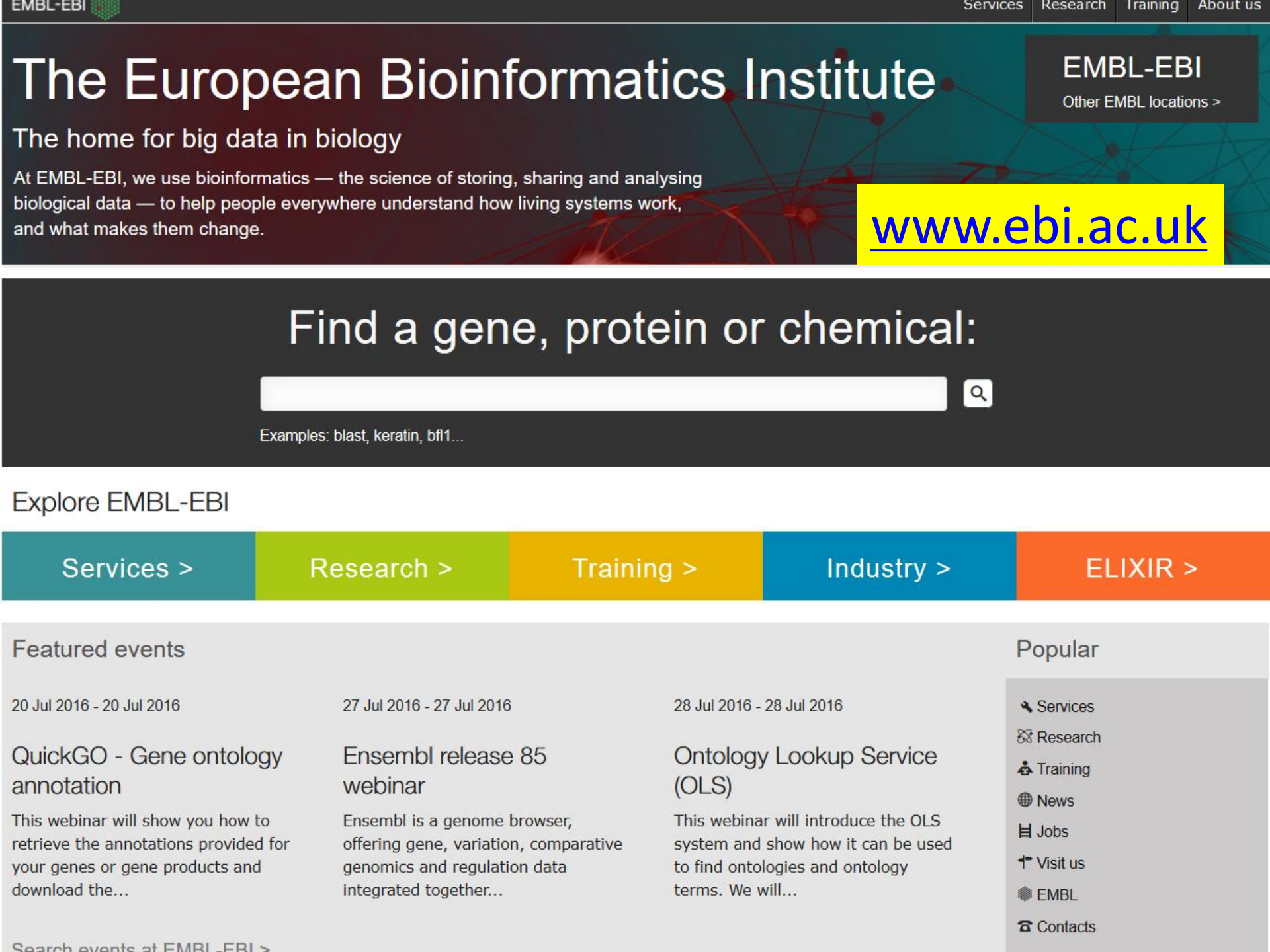
Wednesday, July 6, 2016

NCBI Databases

- PubMed – Biomedical literatures
- PMC – PubMed Central
- Bookshelf – Free online books
- GenBank – Nucleic acid sequences
- RefSeq – Reference sequences (DNA, RNA, Protein)
- CDD – Conserved Domain Database
- SRA – NGS sequence read archive
- Genome – Genomic sequences and annotations
- UniGene – Unique RNA transcripts and ESTs
- SNP – Single nucleotide polymorphism
- Taxonomy – Classification of biological species
- PubChem – Small molecules and drug compounds
- Flu – Influenza virus resources

欧洲生物信息学研究所EBI

- 成立于1994年，坐落在英国剑桥南Wellcome基因组园区。欧洲分子生物学实验室EMBL下属单位，研究人员主要来自英国、德国、法国等欧洲各国。
- 仅次于NCBI的国际生物信息中心，为欧洲各国和世界各地用户提供生物信息资源服务，包括核酸序列数据库EMBL、蛋白质序列数据库UniProt和基因组数据库Ensembl库等。
- 第一任主任为剑桥大学果蝇遗传学家Michael Ashburner，Graham Cameran任副主任。2003年，Janet Thornton接任EBI主任，2011年，Rolf Apweiler (蛋白组学) 和Ewan Birney (基因组学) 任副主任；2015年起任共同主任。


EMBL-EBI
Services
Research
Training
About us
The European Bioinformatics Institute
The home for big data in biology
At EMBL-EBI, we use bioinformatics — the science of storing, sharing and analysing biological data — to help people everywhere understand how living systems work, and what makes them change.
EMBL-EBI
Other EMBL locations >
www.ebi.ac.uk
Find a gene, protein or chemical:
Examples: blast, keratin, bfl1...
Explore EMBL-EBI
Services >
Research >
Training >
Industry >
ELIXIR >
Featured events
20 Jul 2016 - 20 Jul 2016
QuickGO - Gene ontology annotation
This webinar will show you how to retrieve the annotations provided for your genes or gene products and download the...
27 Jul 2016 - 27 Jul 2016
Ensembl release 85 webinar
Ensembl is a genome browser, offering gene, variation, comparative genomics and regulation data integrated together...
28 Jul 2016 - 28 Jul 2016
Ontology Lookup Service (OLS)
This webinar will introduce the OLS system and show how it can be used to find ontologies and ontology terms. We will...
Search events at EMBL-EBI >
Popular
Services
Research
Training
News
Jobs
Visit us
EMBL
Contacts

EBI Databases

- ENA – European Nucleotide archive
- Ensembl – Genomic sequences and annotations
- Expression Atlas – Differential and Baseline Expression
- Array Express – NGS functional genomics experiments
- DGVa – Database of Genomic Variants archive
- TreeFam – Database of animal gene trees
- Rfam – Database of non-coding RNA families
- UniProt – Database of protein sequences
- InterPro – Classification of proteins families and domains
- Pfam – Collection of protein families and domains
- Pride – Proteome identification database
- IntAct – Database of molecular interaction
- PDBe – Macromolecular 3D structures
- PDBeChem – Chemical Components in the PDB

www.uniprot.org

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

Basket 7

UniProtKB

UniProt Knowledgebase

Swiss-Prot (551,705)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (65,378,749)

Automatically annotated and not reviewed.

Records that await full manual annotation.

UniRef

The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

UniParc

UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

Proteomes

A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

Supporting data

Literature citations

Taxonomy

Subcellular locations

Cross-ref. databases

Diseases

XXX

Keywords

News

Forthcoming changes
Planned changes for UniProt

UniProt release 2016_07
(Bacterial) immigration under control

UniProt release 2016_06
Strength through unity | Removal of the cross-references to NextBio | Change of URIs for neXtProt

News archive

Getting started

Text search
Our basic text search allows you to search all the resources available

BLAST
Find regions of similarity between your sequences

Sequence alignments
Align two or more protein sequences using the Clustal Omega program

Retrieve/ID mapping
This tool merges the "Retrieve" and "ID Mapping" tools

UniProt data

Download latest release
Get the UniProt data

Statistics
View Swiss-Prot and TrEMBL statistics

How to cite us
The UniProt Consortium

Submit your data
Submit your sequences and annotation updates

SPARQL
Query UniProt data using a SQL like graph query language

Protein spotlight

On Releasing Tension
June 2016

Like life, cells are subject to continuous change. Nothing in the vicinity of a cell remains still - unless death has interrupted its course. And the same goes for the inside of each cell. All sorts of molecules are being shuttled from one part to another, after having been created or on their way to being degraded. The cell membrane is also a very dynamic and supple structure, with molecules wandering through it constantly...

ExPASy
Bioinformatics Resource Portal

www.expasy.org

Swiss Institute of Bioinformatics

Home About

Query all databases | search | help

- Visual Guidance
- Categories
  - proteomics
    - protein sequences and identification
    - Proteomics experiment
    - Function analysis
    - Sequence sites, features and motifs
    - Protein modifications
    - protein structure
    - Protein interactions
    - similarity search/alignment
  - Genomics
  - Structure analysis
  - Systems biology
  - Evolutionary biology
  - Population genetics
  - Transcriptomics
  - Biophysics
  - Imaging
  - IT infrastructure
  - Medicinal chemistry
  - glycomics
- Resources A..Z
- Links/Documentation

SIB resources
External resources - *(No support from the ExPASy Team)*

Databases

- UniProtKB • functional information on proteins • [more]
- UniProtKB/Swiss-Prot • protein sequence database • [more]
- STRING • protein-protein interactions • [more]
- SWISS-MODEL Repository • protein structure homology models • [more]
- PROSITE • protein domains and families • [more]
- ViralZone • portal to viral UniProtKB entries • [more]
- neXtProt • human proteins • [more]

- EMBnet services • bioinformatics tools, databases and courses • [more]
- ENZYME • enzyme nomenclature • [more]
- GlyTouCan • international glycan structure repository • [more]
- GPSDB • gene and protein synonyms • [more]
- HAMAP • UniProtKB family classification and annotation • [more]
- MatrixDB • protein-glycosaminoglycan interactions • [more]
- MetaNetX • Metabolic Network Repository & Analysis • [more]
- MIAPEGelDB • MIAPE document edition • [more]
- MyHits • protein domains database and tools • [more]
- PaxDb • protein abundance database • [more]
- Prolune • Popular science articles (in French) • [more]
- Protein Model Portal • structural information for a protein • [more]
- Protein Spotlight • Informally written reviews on proteins • [more]
- Rhea • expert curated resource of biochemical reactions • [more]
- SugarBind • pathogen sugar-binding • [more]
- SWISS-2DPAGE • proteins on 2-D and SDS PAGE maps • [more]
- SwissBioIsostere • bioIsosteres for small molecules • [more]
- SwissLipids • knowledge resource for lipid biology • [more]
- SwissPalm • database of S-palmitoylation events • [more]
- SwissSidechain • non-natural amino-acid sidechains • [more]
- SwissVar • variants in UniProtKB entries • [more]

Tools

- SWISS-MODEL Workspace • structure homology-modeling • [more]
- SwissDock • protein ligand docking server • [more]

- 2ZIP • Prediction of leucine zipper domains • [more]
- 3of5 • find user-defined patterns in protein sequences • [more]
- AACompIdent • protein identification by aa composition • [more]
- AACompSim • amino acid composition comparison • [more]
- Agadir • Prediction of the helical content of peptides • [more]
- ALF • simulation of genome evolution • [more]
- Alignment tools • Four tools for multiple alignments • [more]
- AllAll • protein sequences comparisons • [more]
- APSSP • Advanced Protein Secondary Structure Prediction • [more]
- Ascalaph • Molecular modeling software • [more]
- big-PI • predict GPI modification sites • [more]
- Biochemical Pathways • Biochemical Pathways • [more]
- BLAST • sequence similarity search • [more]
- BLAST (UniProt) • BLAST search on the UniProt web site • [more]
- BLAST - NCBI • Biological sequence similarity search • [more]
- BLAST - PBIL • BLAST search on protein sequence databases • [more]
- Blast2Fasta • Blast to Fasta conversion • [more]
- boxshade • MSA pretty printer • [more]
- CFSSP • Protein secondary structure prediction • [more]
- ChloroP • chloroplast transit peptides & cleavage sites • [more]
- Click2Drug • Directory of computational drug design tools • [more]
- ClustalO (UniProt) • Align two or more protein sequences • [more]
- ClustalW • Multiple sequence alignment • [more]
- ClustalW - PBIL • Multiple sequence alignment program • [more]
- ClustalW2 • Multiple sequence alignment program • [more]
- Coiled-Coils prediction • Prediction of coiled coils regions • [more]
- COILS • Prediction of Coiled Coil Regions in Proteins • [more]

RCSB PDB PROTEIN DATA BANK

An Information Portal to 120642 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | E

PDB-101 | Worldwide PDB | EMDataBank | NDB Nucleic Acid Database | StructuralBiology Knowledgebase | Worldwide Protein Data Bank Foundation

www.rcsb.org

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

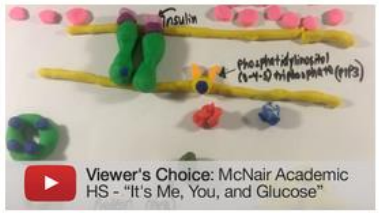
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Video Challenge Awards

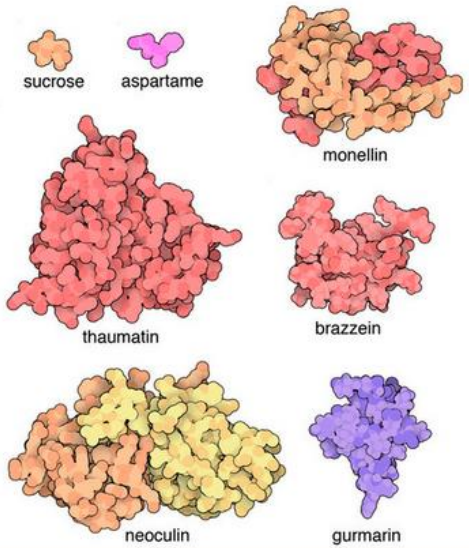
More Info

1st Place: West Windsor Plainsboro South HS - "Sweet Signals"



Viewer's Choice: McNair Academic HS - "It's Me, You, and Glucose"

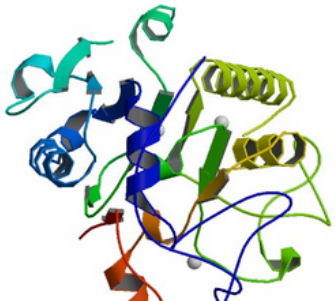
July Molecule of the Month



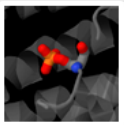
Monellin

Latest Entries

As of Tuesday Jul 19

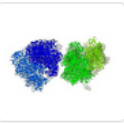


Features & Highlights



Explore Protein Modifications

Browse, search, and visualize protein modifications in the PDB. » 07/05



Biological Assembly files for large structures

Download biological assembly files for large structure in PDBx/mmCIF format. » 07/05

News

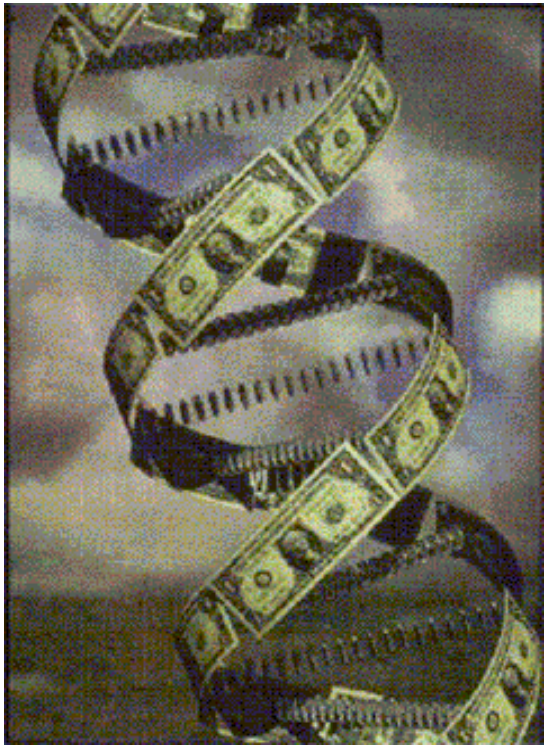
Publications

Meet the RCSB PDB at ACA

Visit us at booth #46, and learn about new rcsb.org tools, a resource for exploring Irving Geis' molecular images, and a data dictionary for archiving integrative/hybrid models. » 07/19

Summer Newsletter Published » 07/12

The human genome sequence



- IHGS, Initial sequencing and analysis of the human genome, *Nature*, 2001.
- Venter, *et al*. The Sequence of the Human Genome. *Science*, 2001.

1999年自然科学基金委21世纪核心科学论坛纪要

以人类及其它各物种基因组核酸、蛋白质等生物大分子数据为主要研究对象，以系统生物学为主要研究思路，以计算生物学为主要研究方法，以数理科学、信息科学和计算机科学为主要研究手段，以计算机网络为主要研究环境，以计算机软件为主要研究工具，构建各种类型的专用、专门、专业数据库，研究开发面向生物学家的新一代计算机软件，对浩如烟海的原始数据进行存储、管理、注释、加工，使之成为具有明确生物意义的生物信息，并通过对生物信息的查询、搜索、比较、分析，从中获取基因编码、基因调控、核酸和蛋白质结构功能及其相互关系等理性知识。在大量信息和知识的基础上，探索生命起源、生物进化以及细胞、器官和个体的发生、发育、病变、衰亡等生命科学中重大问题，搞清它们的基本规律和时空联系，建立“生物学周期表”。

第一届中国生物信息学大会

Life Sciences in
The Internet Times
9-10 Apr 2001
Beijing, China

First Chinese
Bioinformatics
Conference
11-13 Apr 2001
Beijing, China

Center for Bioinformatics at PKU

Supported by the Ministry of Education, we started the center of bioinformatics (CBI) at Peking University in 1997, and joined the European Molecular Biology Network (EMBnet). Our primary goal was to provide bioinformatics resource, service and training to domestic users.

Mirrors of GDB, RGD and ExPASy, etc, were set up, database query (SRS) and search (BLAST) platforms were installed locally. The EBI FTP server was cloned to provide better service for local users.

::北大生物信息中心首页:: - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://www.cbi.pku.edu.cn/chinese/ 转到

北京大学 生物信息中心

中文版 English Version

2003年09月28日

国外站点

- NCBI
- EBI
- APBionet

More>>

国内站点

- 上海生物信息中心
- 华大基因组信息中心
- 天大生物信息中心
- 计算所生物信息中心

More>>

友情链接

最新消息

学术报告：Biodiversity and molecular evolution: a personal account,(29 Sep, 2003)
欢迎订阅国家自然科学基金委员会提供的Nature期刊, (September, 2003)
EMBL Nucleotide Sequence Database: Release Notes 76, (September, 2003)
EMBL Nucleotide Sequence Database: User Manual, (September, 2003)

more>>

SRS

SRS（Sequence Retrieval System）是由欧洲生物信息研究所（EBI）开发的以WWW界面运行的数据库检索及导航系统， 是目前生物信息界应用最为广泛的数据库系统。CBI目前已经可以提供74种数据库的搜索，其中 比较常用的EMBL和GDB等已做到每日更新。

BLAST

BLAST（Basic Local Alignment System）， 数据库查询工具，是最常用的比较核酸和蛋白质同源性比较工具，本中心采用的是NCBI的BLAST2.0网络版。可以直接通过网页提交查询。如果网速不够快的话，还可以选择用E-mail接收BLAST结果。

本站建立了多个世界著名生物信息网站的镜像，从而使得教育网用户可以

ExPASy Molecular Biology Server - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹 媒体
地址(D) http://cn.expasy.org/ 转到

ExPASy Molecular Biology Server

The ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References).

[Announcements] [Job opening] [Mirror Sites]

Databases

- Swiss-Prot and TrEMBL - Protein knowledgebase
- PROSITE - Protein families and domains
- SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
- ENZYME - Enzyme nomenclature
- SWISS-3DIMAGE - 3D images of proteins and other biological macromolecules

Tools and software packages

- Proteomics and sequence analysis tools
  - Proteomics [PeptIdent, PeptideMass, ...]
  - DNA -> Protein [Translate]
  - Similarity searches [BLAST]
  - Pattern and profile searches [ScanProsite]

rat genome database - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://rgd.cbi.pku.edu.cn/ 转到

RGD RAT GENOME DATABASE 中国 China mirror site

SEARCH DATABASE FOR GO

(use * for wildcard search)

MAPS GENES ESTS QTLS SSLPS SEQUENCES STRAINS REFERENCES HELI

"The Rat Genome Database (RGD) curates and integrates rat genetic and genomic data and provides access to this data to support research using the rat as a genetic model for the study of human disease".
Related links: About RGD, Article describing RGD.

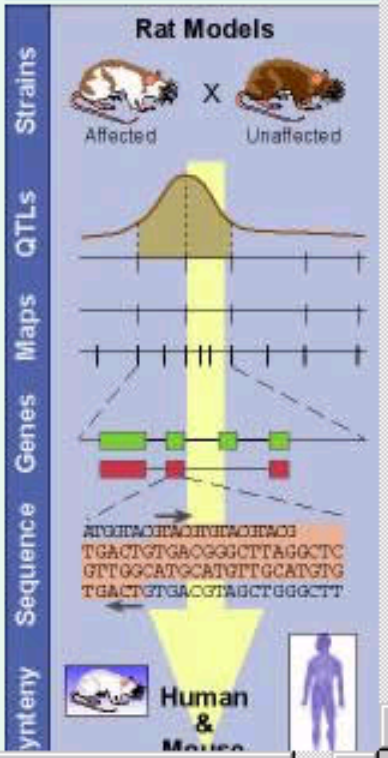
RGD Highlights

- **1450 new QTL and gene annotations in RGD** NEW
  There are 1450 new QTL and gene annotations in RGD, curated from the literature. In addition the annotations have references associated with them, linked to the abstract in RGD.

  On the QTL reports, the annotations include more detailed information about the QTL, the associated genes, mapping data, and disease.

  On Gene reports there are many new annotation fields, including biological process, molecular function, regulation, mutant phenotype, disease, expression, pathway, domains, and drugs.

How do I ... use RGD?



Colibri Web Server - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

地址(D) http://genolist.mirror.edu.cn/Colibri/ 转到

Colibri

© Copyright Institut Pasteur, 1999

Help | Mail

Search

Reset

Gene name ?

Synonym

Region 20 kb

Location ?

From

To

kb

Chromosome

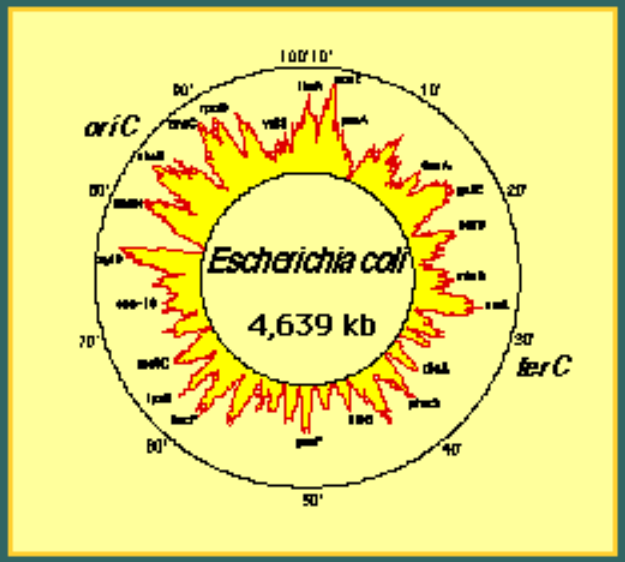
Free text ?

Welcome to the Colibri World-Wide Web Server

Mirror of Pastear Institute Colibri

All data, programs and the database structure of this mirror are Copyright of Pasteur Institute

Data Release R5.1 - EcoGene 11 (May 4, 1999)
WWW server v2.5 (May 4, 1999)



This image is clickable if your client supports HTML 3.0

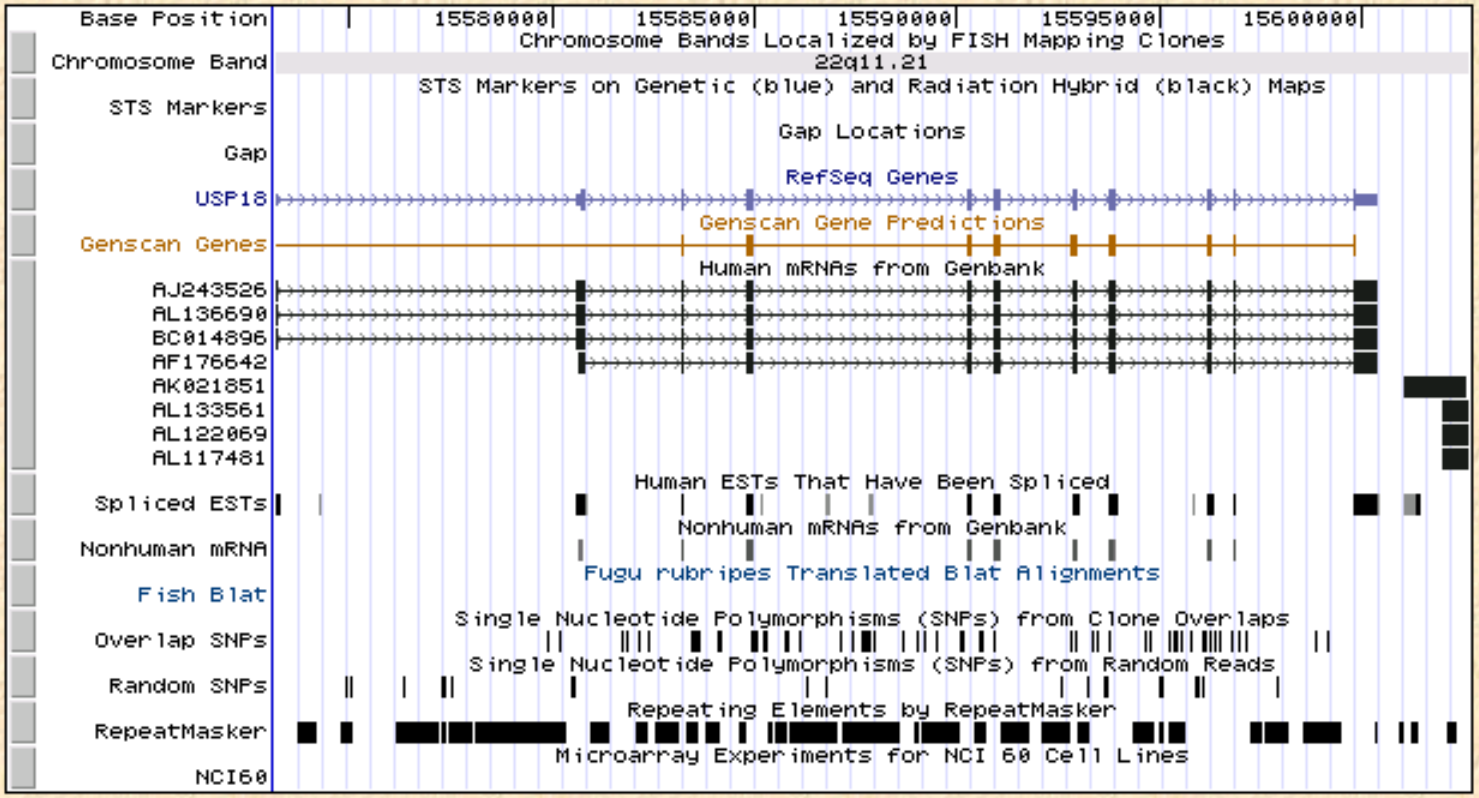
Visit the following links to learn more about this server:

UCSC Genome Browser v15 - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://goldenpath.cbi.pku.edu.cn/cgi-bin/hgTracks?hgsid=1087&position=chr22%3A15573184-15601
转到

UCSC Genome Browser on Jun. 28, 2002 Freeze

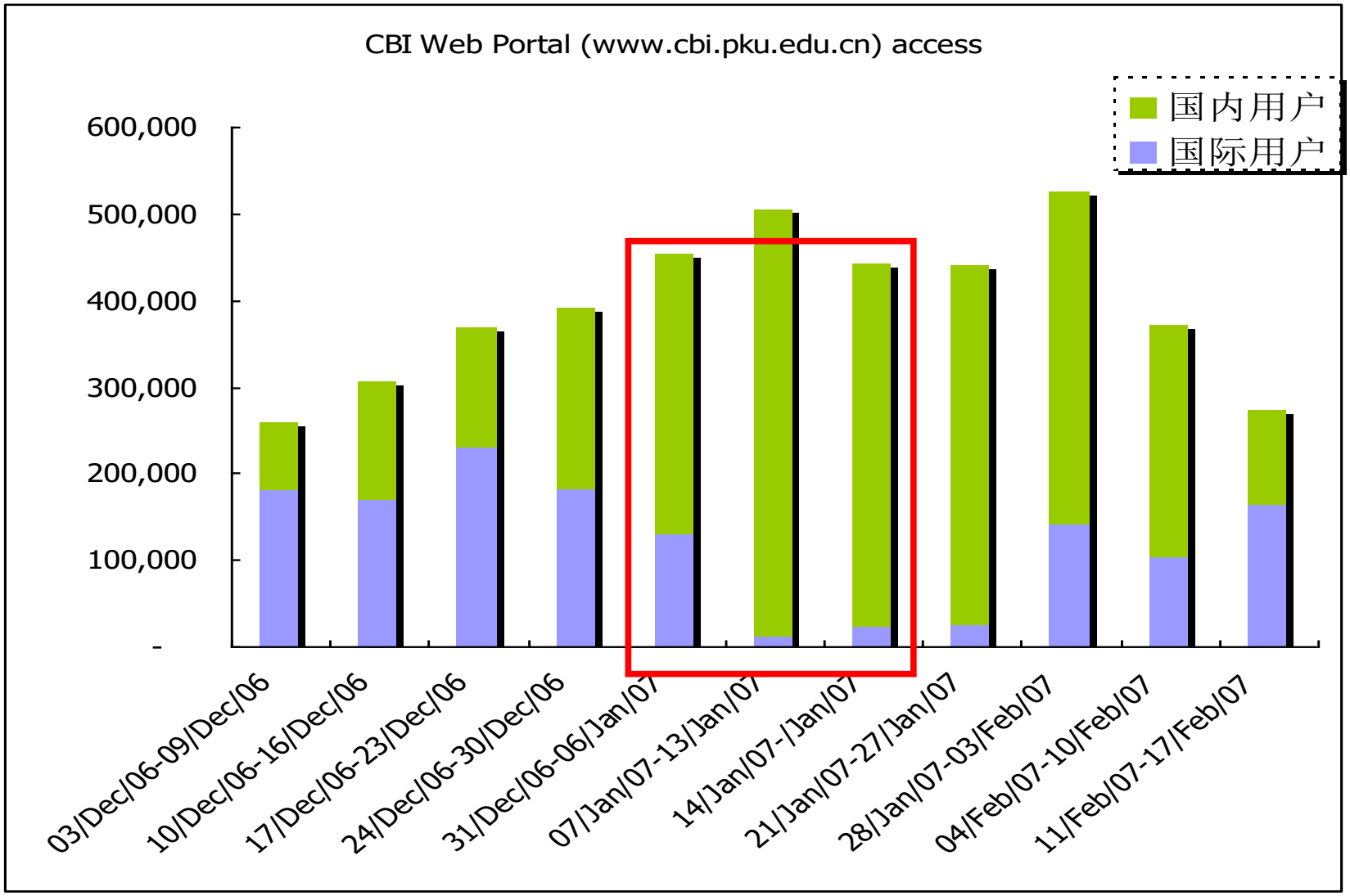
move <<< << < > >> >>> zoom in 1.5x 3x 10x zoom out 1.5x 3x 10x

position chr22:15573184-15602640 size 29457, image width 610 jump



Click on a feature for details. Click on

More CBI network access due to the earthquake



http://planttfdb.cbi.pku.edu.cn/

PlantTFDB

Plant Transcription Factor Database v2.0

Center for Bioinformatics, Peking University, China

Previous version

Home | Blast | Search | Download | WebService | Help | About | Links

Search (eg: SPL2)

Browse by Species

Arabidopsis lyrata	*Arabidopsis thaliana*	*Arachis hypogaea*
Artemisia annua	*Brachypodium distachyon*	*Brassica napus*
Brassica rapa	*Carica papaya*	*Chlamydomonas reinhardtii*
Chlorella sp. NC64A	*Citrus sinensis*	*Coccomyxa sp. C-169*
Cucumis sativus	*Glycine max*	*Gossypium hirsutum*
Helianthus annuus	*Hordeum vulgare*	*Lotus japonicus*
Malus x domestica	*Manihot esculenta*	*Medicago truncatula*
Micromonas pusilla CCMP1545	*Micromonas sp. RCC299*	*Mimulus guttatus*
Nicotiana tabacum	*Oryza sativa subsp. indica*	*Oryza sativa subsp. japonica*
Ostreococcus lucimarinus CCE9901	*Ostreococcus sp. RCC809*	*Ostreococcus tauri*
Panicum virgatum	*Physcomitrella patens subsp. patens*	*Picea glauca*
Picea sitchensis	*Pinus taeda*	*Populus trichocarpa*
Prunus persica	*Raphanus sativus*	*Ricinus communis*
Saccharum officinarum	*Selaginella moellendorffii*	*Solanum lycopersicum*
Solanum tuberosum	*Sorghum bicolor*	*Theobroma cacao*
Triticum aestivum	*Vigna unguiculata*	*Vitis vinifera*
Volvox carteri	*Zea mays*	

Browse by Family

AP2 (716)	ARF (646)	ARR-B (323)	B3 (1505)	BBR/BPC (218)	BES1 (247)

Secreted Protein Database - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹 媒体
地址(D) http://fir.cbi.pku.edu.cn:8080/index.jsp

Secreted Protein Database

hello, guest

REGISTER name: _____ password: _____ GO

- Home
- PipeLine
- Query
- Download
- GuestBook
- About Us

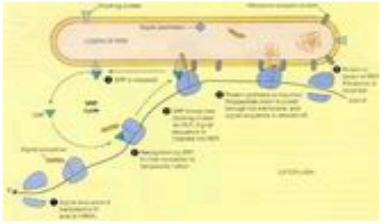
:: News :::::

Just 2.5% of DNA turns mice into men

Mice and men share about 97.5 per cent of their working DNA, just one per cent ...

CBI 北大生物信息中心

The 1999 Nobel Prize in Physiology or Medicine has been awarded to Gunter Blobel for the discovery that "proteins have intrinsic signals that govern their transport and localization in the cell." The first such signal to be discovered was the secretory signal peptide.



We want to point out, secreted protein is not a functional defination, like the kinases or transcription factors. It is a kind of defination based on protein post-translational processing and sorting and localization. So according to this principle, all proteins, which are synthsized on ER, then go through post-translational processing and sorting and localization steps, are clustered into secreted protein group.

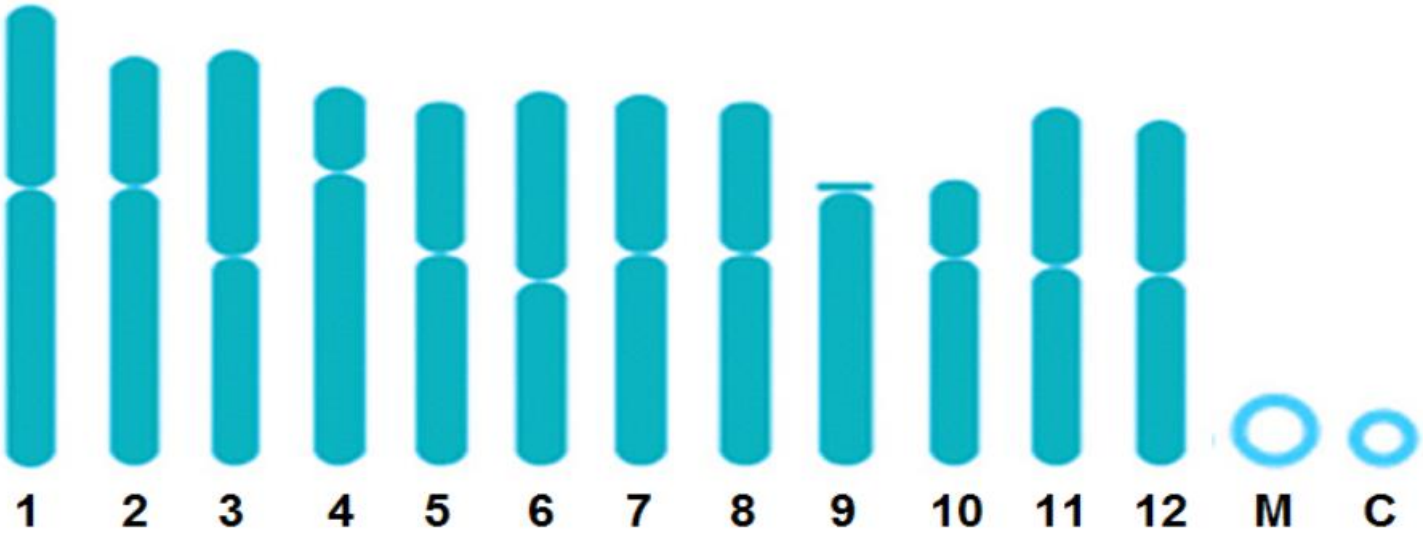
Most of human protein factors involved in cell signaling pathway, blood coagulation, immmune defence, and carcinogenesis, as well as digestive enzymes, components of the extracellular matrix belong to secreted proteins. many of the factors are therapeutic agents, or have important biological regulatory functions. Therefore they can be potential targets in the drug discovery. Now in the world, more than half of biotechnology drugs belong to secreted protein, which may prevent and cure tumor, heart disease and blood vessel, etc.

完毕 Internet

http://ricemap.cbi.pku.edu.cn/

Rice-Map

Home | Genome Browser | Rice Mart | Help | Track Desc | Sign In | Register Now | Download



>> Browse *japonica* genome

Input format: (1)chr:start-end, like *chr2:175000-196000*. (2)chr:start+length, like *chr2:175000+21000*. (3)gene ID, like *LOC_Os08g39890*. (4)function description, like *SBP-box gene family member*.

chr8:25265540-25278060 GO

>> Browse *indica* genome

Input format: (1)chr:start-end, like *chr2:175000-196000*. (2)chr:start+length, like *chr2:175000+21000*. (3)gene ID, like *BGIOSIBCE028446*. (4)function description, like *TBC domain*.

chr8:26921516-26932966 GO

Database Retrieve System SRS

http://srs.pku.edu.cn/srs5/ - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

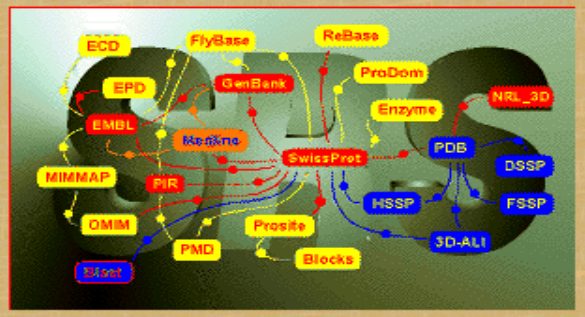
地址(D) http://srs.pku.edu.cn/srs5/ 转到

Sequence Retrieval System

Network Browser for Databanks in Molecular Biology

The EMBL database has been updated to Ver. 62, EMBL daily updates are in EMBLNEW and EMBLNEWEST

PDB contains entries to 31 Dec 1999, PDBNEW contains entries from 1 Jan 2000

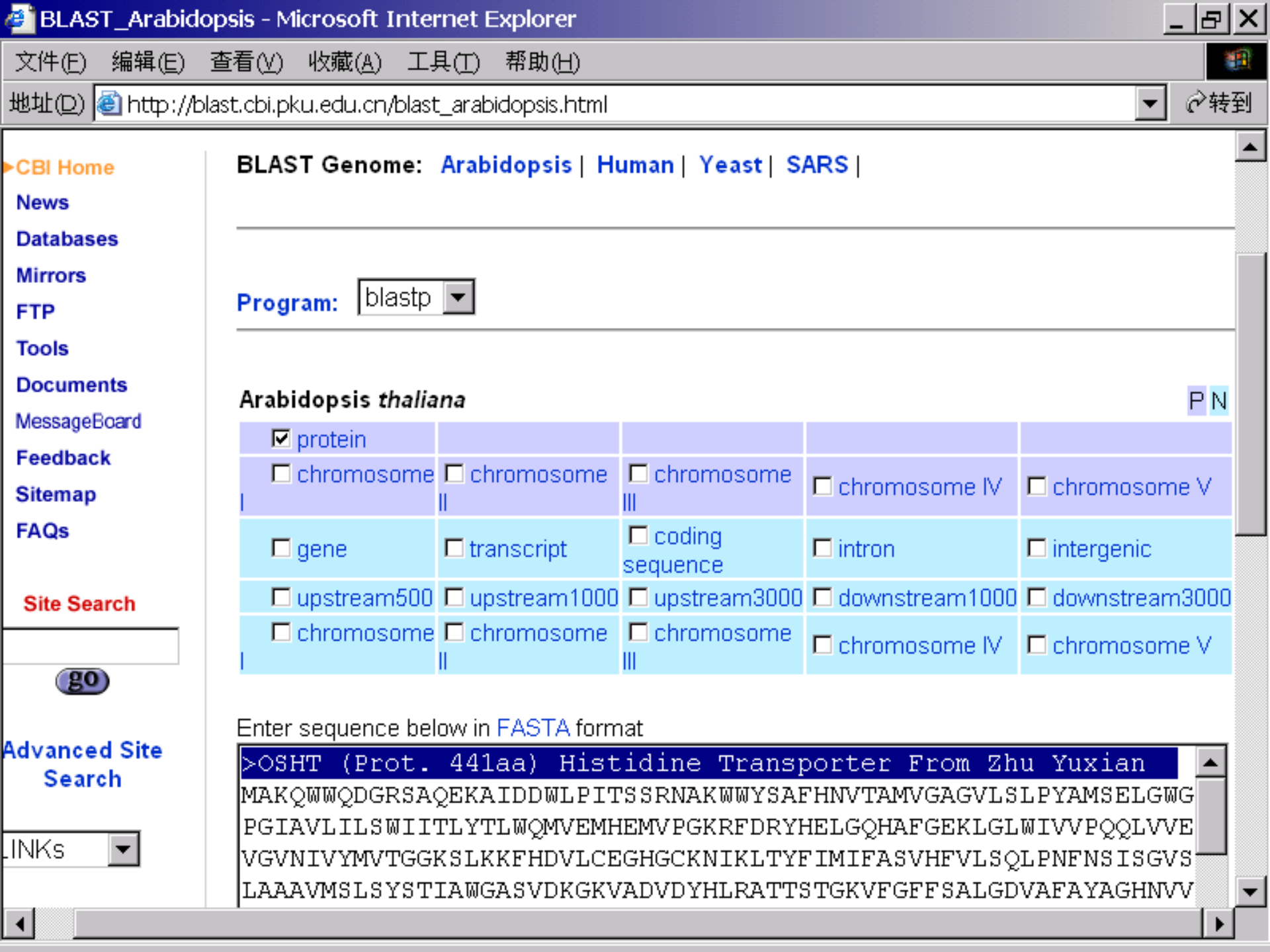


Start Start a new SRS session

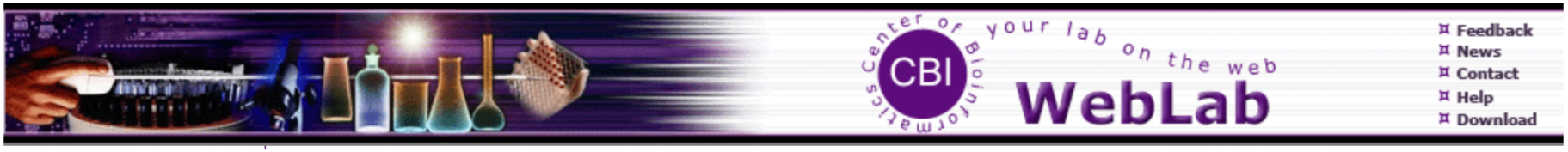
The SRS Manual

List of all SRS5 serverss

The SRS newsgroup


BLAST_Arabidopsis - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://blast.cbi.pku.edu.cn/blast_arabidopsis.html
转到
CBI Home
News
Databases
Mirrors
FTP
Tools
Documents
MessageBoard
Feedback
Sitemap
FAQs
Site Search
go
Advanced Site Search
LINKs
BLAST Genome: Arabidopsis | Human | Yeast | SARS |
Program: blastp
Arabidopsis thaliana
P N
protein
chromosome I
chromosome II
chromosome III
chromosome IV
chromosome V
gene
transcript
coding sequence
intron
intergenic
upstream500
upstream1000
upstream3000
downstream1000
downstream3000
chromosome I
chromosome II
chromosome III
chromosome IV
chromosome V
Enter sequence below in FASTA format
>OSHT (Prot. 441aa) Histidine Transporter From Zhu Yuxian
MAKQWWQDGRSAQEKAIDDWLPITSSRNAKWWYSAFHNVTAMVGAGVLSLPYAMSELGWG
PGIAVLILSWIITLYTLWQMVEMHEMVPGKRFDRYHELGQHAFGEKLGLWIVVPQQLVVE
VGVNIVYMVTGGKSLKKFHDVLCEGHGCKNIKLTYFIMIFASVHFVLSQLPNFNSISGVS
LAAAVMSLSYSTIAWGASVDKGKVADVDYHLRATTSTGKVFGFFSALGDVAFAYAGHNVV

Sequence analysis platform



Service

- Program
- Protocol
- Macro
- Utility
- Resource

User Space

- My Data
- My Literature
- My MetaPackage
- My Toolbox
- History

Account

- Login
- Try Out
- Registration

Welcome to WebLab

WebLab is a multifunctional bioinformatics analysis platform integrating diversified tools with unified, user-friendly web interface. However, WebLab is not a mere bioinformatics toolbox, we also offer powerful data management function, group strategy and knowledge sharing mechanism, which will bring considerable advance of efficiency for both wet bench and in silico scientists working in biomedicine community.

What's New in WebLab

- BioMart module is updated. Instead of Java API, we use REST style web service API to communicate with MartService. Users can now do sequence retrieval through the updated BioMart module. (2009-04-02)
- The tag system of WebLab has been updated. Now "my data", "my literature" and "my metapackage" use one common rather than three stand-alone tag systems. (2009-03-24)
- WebLab will be under maintenance from 14:00 to 22:00 on 2009.03.24 GMT +8. (2009-03-24)
- Video tutorials are online. (2009-01-06)
- Literature space supports citation manipulation. (2008-12-18)

Get Service in WebLab

Get service quickly through keyword [ ] search for service description .

How to Cite

Liu, X., Wu, J., Wang, J., Liu, X., Zhao, S., Li, Z., Kong, L., Gu, X., Luo, J. and Gao, G. (2009) WebLab: a data-centric, knowledge-sharing bioinformatic platform. Nucleic Acids Res. (Full Text)

Please Note

- According to our testing, WebLab can work well under , and . (detailed browser test result)
- This project is supported by National High-Tech (863) Programme.

http://weblab.cbi.pku.edu.cn/

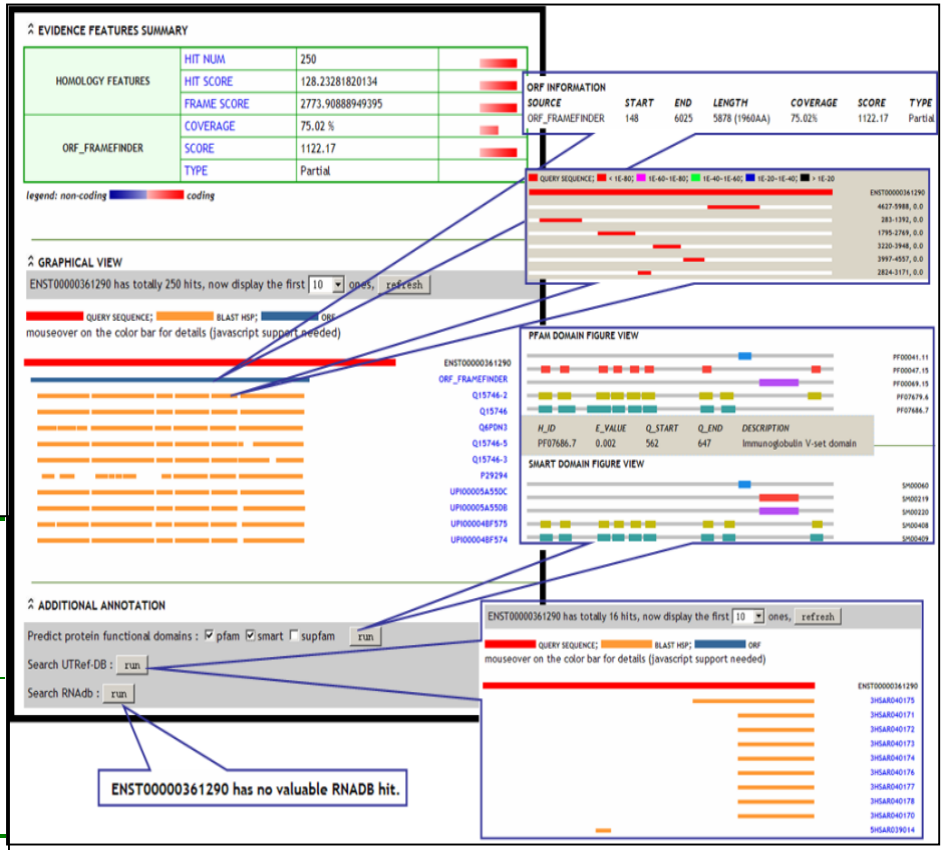
http://cpc.cbi.pku.edu.cn/

$$S_i = \underset{j}{mean}\{-\log_{10} E_{ij}\}, i \in [0,1,2]$$

$$\text{HIT SCORE} = \underset{i\in\{0,1,2\}}{mean}\{S_i\} = \frac{\sum_{i=0}^{2} Si}{3}$$

$$\text{FRAME SCORE} = \underset{i\in\{0,1,2\}}{variance}\{S_i\} = \frac{\sum_{i=0}^{2}(S_i - \bar{S})^2}{2}$$

Dataset	Dataset Type	Dataset Size[a]	Accuracy		Time (in minutes)	
			CPC	CONC	CPC	CONC
Rfam	noncoding	30,770	98.62%	97.12%	3,513	46,376
RNADB	noncoding	3,996	91.50%	85.44%	598	7,322
Embl cds	coding	121,914	99.08%	98.70%	69,116	826,210 [b]



International Collaboration

Bernd Jagla, 2002
Ernst Kretschmann, EBI, 2002
Martin Senger, EBI, 2002
Torsten Crass, 2002
Andreas Zell, Tuebingen,Germany, 2002
Yan P Yuan, EMBL, 202
Michael Poindiger, ANGIS, Australia, 2002
Robert Giegerich, Bielefeld, Germany, 2002
Burkhard Morgenstern, Bielefeld, Germany, 2002
Ralf Hofestaedt, Bielefeld, Germany, 2002
Anping Zeng, GBF, Germany, 2002
Ueng-Cheng Yang, YMU, Taiwan, 2002
Cathy Wu, PIR, 2002
Antoine Danchin, HK Pasteur Centre, 2001, 2002
Steve Bryant, NCBI, 2002
Mike Sternberg, IC, UK, 2002
Keith Dunker, Washington Univ., 2002
Roderic Guigo, Barcelona, Spain, 2002
David Sankoff, Montreal Canada, 2002
Peter Tonellato, WMC, USA, 2001, 2002
Rob Harper, EBI, 2002
James Bonfield, Cambridge, UK, 2002
Lisa Mullan, HGMP, 2002
Tim carver, HGMP, UK, 2002
Janet Kelso, SANBI, South Africa, 2002
Johann Visagie, egenetics, South Africa, 2002
Victor Jorgneel, SIB, Switzerland, 1999,2002

2001, 1999, 1998, Jack Leunissen CMBI, Holland
2001, 2000, 1999, Tim Littlejohn ANGIS, Australia
2001, Rolf Apweiler, EBI
2002, 2000, 1999, Edgar Wingender, GBF, Germany
2000, Martin Ebeling, Roche
2000, Luciano, Milanasi, Milan, Italy
2000, Win Hide, SANBI, South Africa
2000, Jamie Cotticia, GDB, Canada
2000, Voker Brendel, Iowa, US
2002, Weimin Zhu, GDB, Canada
1999, 1998, Frank Wright, CropNet, UK
1999, Peter Rice, Sanger Center/LION, UK
1999, Sandor Pongor, ICGEB, Italy
1999, Bruno Gatea, ANGIS, Australia
1999, Amos Bairoch, ExPASy, Switzerland
1999, 1998, David Judge, Cambridge, UK
1998, Thure Etzold, EBI, LION, UK
1998, Alan Bleasby, HGMP, UK

Shanghai biomedical big data center

DataBases

Genomics
- Node
- Gene
- Genome

Proteomics
- PhoSigNet
- SysPTM
- dbDEPC

Transcriptomics
- EST
- Plant Trans DB
- WIPA

Epigenomics
- Seqspider

Phenotype / Disease
- HBV-DIAP
- SysFinder

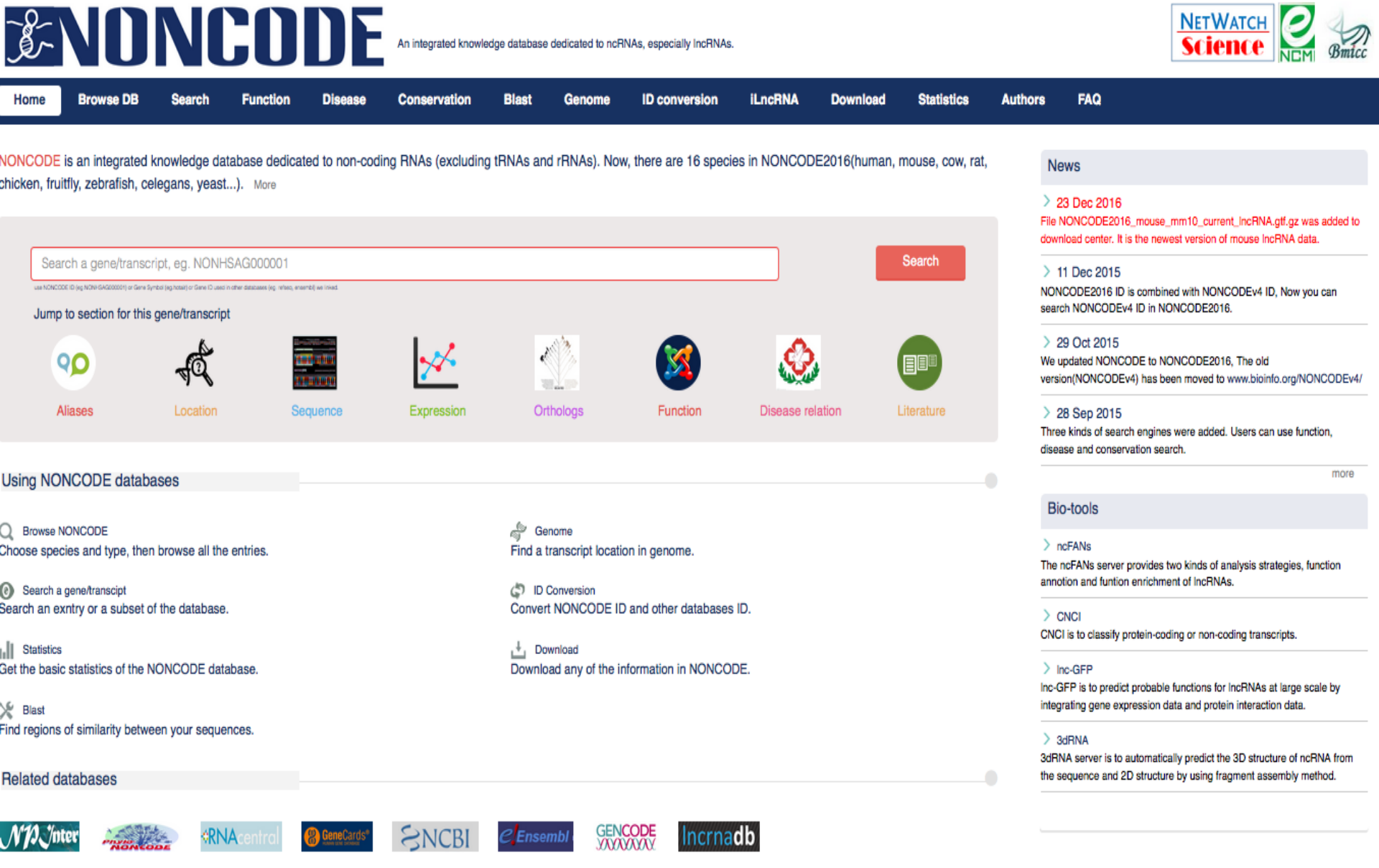
Variation
- DRUMS
- CanProVar

Tools

Alignment

Assembly

The Noncode database at CAS



Protein modification server at HUST

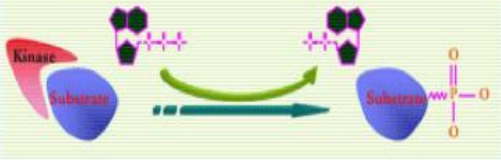
GPS . Group-based Prediction System

Version 3.0

The CUCKOO Workgroup

THE CUCKOO WORKGROUP

HUST



HOME | WEB SERVER | DOWNLOAD | CITATION | USER GUIDE | LINKS | CONTACT

PRODUCTS OF CUCKOO

+ PTMs Predictor

+ Tools

+ Databases

0310450

Last update: Sep. 10th, 2014

❋ GPS INTRODUCTION:

Computational prediction of phosphorylation sites with their cognate protein kinases (PKs) is greatly helpful for further experimental design. Although ~10 online predictors were developed, the PK classification and control of false positive rate (FPR) were not well addressed. Here we adopted a well-established rule to classify PKs into a hierarchical structure with four levels. Also, we developed a simple approach to estimate the theoretically maximal FPRs. Then **GPS 2.0** (**Group-based Prediction System**, ver 2.0) software was implemented in JAVA and could predict kinase-specific phosphorylation sites for **408** human PKs in hierarchy. As an application, we performed a large-scale prediction of >13,000 mammalian phosphorylation sites with high performances. In addition, we also provided a proteome-wide prediction of Aurora-B specific substrates including protein-protein interaction information. As the first stand-alone software for computational phosphorylation, GPS 2.0 will be an excellent tool for further experimental consideration and construction of phosphorylation networks.

Recently, we released **GPS 3.0** with novel **Peptide Selection** and **Weight Training** methods to improve the prediction performance and robustness. Additionally, more than **6,000** phosphorylation sites were used for training and could predict kinase-specific phosphorylation sites for **464** human PKs in hierarchy.

The Z-curve Database of Genomes

Home Browse Plotting Z-curve Gene recognition Identifying oriC Genome segmentation Special issue Search About

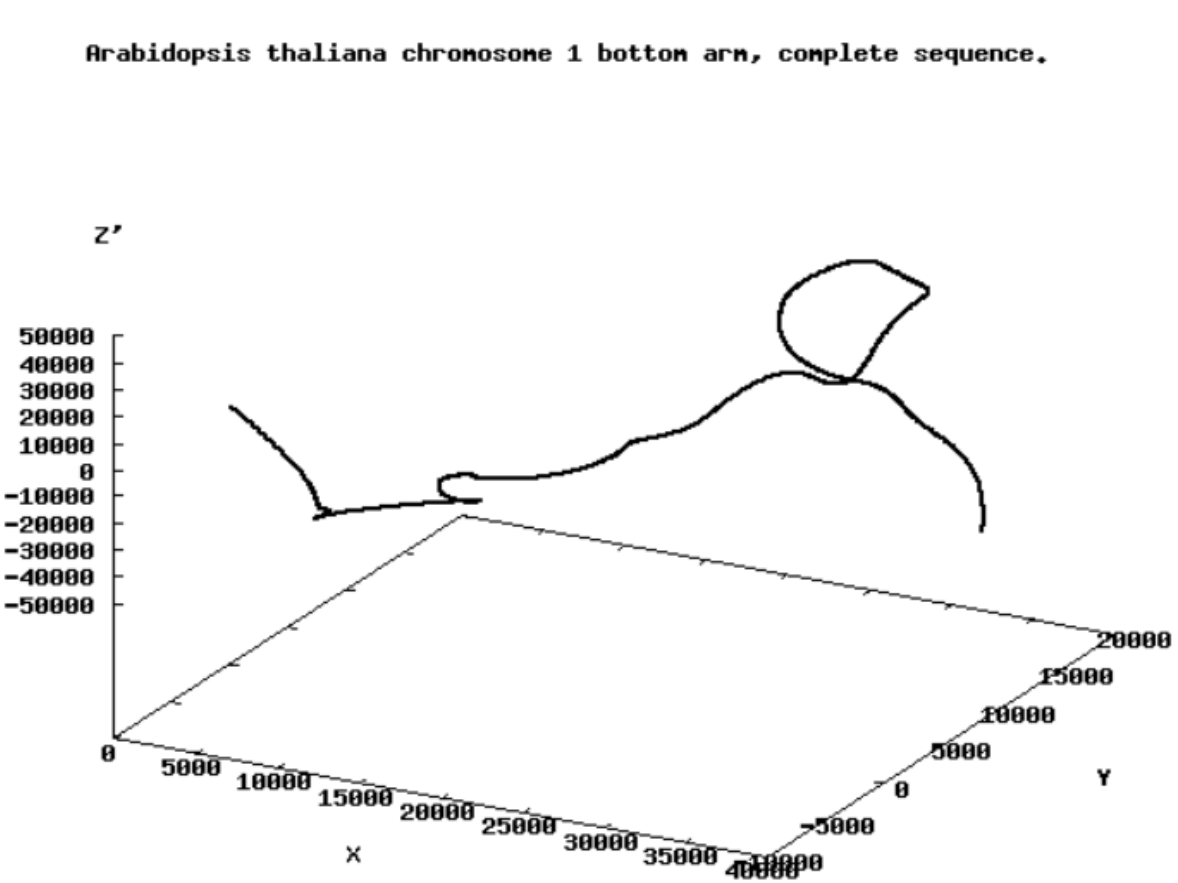
» Home » Information of Eukaryota

http://tubic.tju.edu.cn/zcurve/

the Z curve

3D X Y Z Z' XY SRS segments

AC	AE005173
DE	Arabidopsis thaliana chromosome 1 bottom arm
SQ	Length: 14668883 bp; 4744006 A; 4718470 T; 2608943 G; 2597461 C; 3 N.



天津大学张春霆院士提出了Z曲线 方法

$X_n = (A_n+G_n) - (C_n+T_n)$

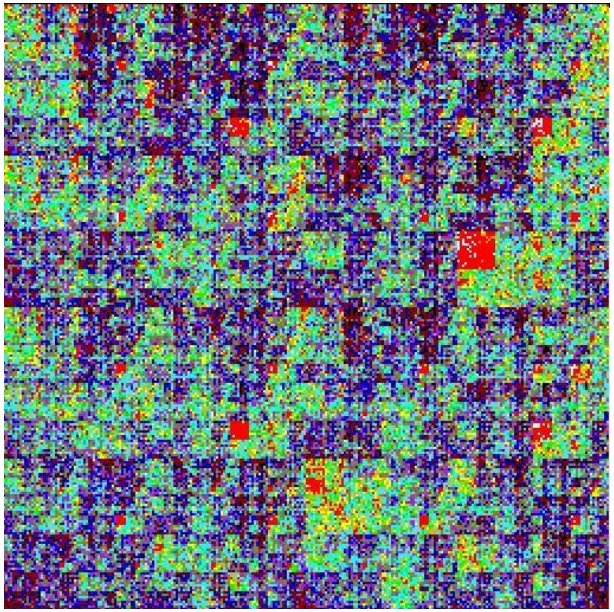
$Y_n = (A_n+C_n) - (G_n+T_n)$,

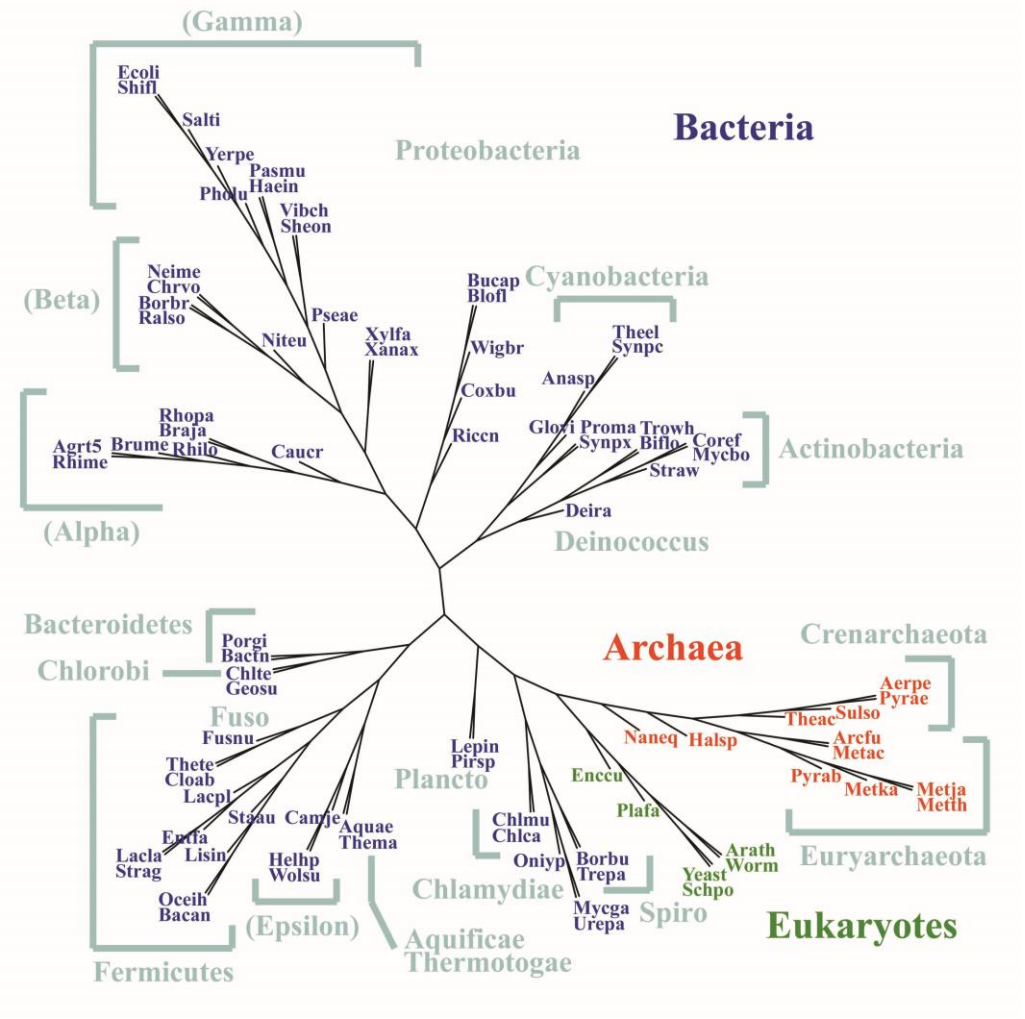
$Z_n = (A_n+T_n) - (C_n+G_n)$

（n = 0, 1, 2, …, N）

The CVTree method for bacteria taxonomy

复旦大学郝柏林院士提出了不用序列比对的组分矢量(Composition Vector)方法，用于细菌基因组比较分类





http://www.itp.ac.cn/~hao/

郝柏林院士建议组建“国家生物医学信息中心”

- 1999年6月，郝柏林院士写了“建议尽快组建国家级的生物医学信息中心”的院士建议。9月27日李岚清副总理做了批示。科技部委托中国生物工程开发中心组织多次论证会、评审会。由于种种原因，至今未果。
- 863“十五”计划拨款支持“国家生物信息基地建设”，由北京大学主持，上海生物技术信息中心和北京基因组研究所参加。
- 863“十一五”计划拨款支持“基于网格的生物信息平台建设”，由北京大学主持，上海生物技术信息中心、北京基因组研究所、军事医学科学院和哈尔滨工业大学参加。

郝柏林院士2009年6月博文

博文

国家生物医学信息中心搁浅十年! 精选

已有 11108 次阅读 2009-6-10 03:48 | 个人分类:科学史料 | 系统分类:科研笔记 | 关键词:生物医学信息中心,生物信息学,科学技术领导体制

"国家生物医学信息中心"搁浅十年!

郝柏林

随着当代科学进入基因组时代，生物和医学成为产生数据量最大的科学技术领域。在上个世纪80年代，美国、日本和联合起来的欧洲国家先后建立了国家级的生物信息中心。这三个中心实行合作，同步更新，但又各有特色。中国对数据库贡献甚少，无偿享用生物数据的情况很难保持长期稳定，国内也存在着自有数据共享的各种障碍，将来数据产量增大以后，更面临着全国性的服务协调。

1999年6月10日我写了一份"院士建议"，题为"建议尽快组建国家级的生物医学信息中心"（附件一），它后来又被国家自然科学基金委印发在一期简报上。9月27日李岚卿副总理在简报上做了批示。科技部组织了不止一个专家委员会、起草过"招标指南"，许多人忙碌了一番，事情也开始走样（附件二），继而搁浅。2000年9月19日我写信给科技部秘书长林泉，问"国家生物医学信息中心筹建工作为何搁浅？"希望"在李岚卿副总理批示一年之际，有点实质性进展"（附件三）。然而情况依旧。2003年1月1日我写信给科技部部长徐冠华，题目是"国家生物

郝柏林

加为好友 给我留言

打个招呼 发送消息

扫一扫，分享此博文

作者的精选博文 全部

- 纪念黄克孙先生
- 不常在教室上课的北大生物系
- 从若干自然规律看某些社会现
- 国民素质教育的关键是人民公
- 圣菲研究所与复杂性研究

“建立国家生物医学信息中心”调研项目

- 2013年1月24日，“建立国家生物医学信息中心”调研项目启动会议在医学科学院基础医学研究所召开。
- 2013年9月25日， 项目组第四次调研会在北京蛋白组中心召开。
- 2013年12月17-19日，项目组5名成员组成代表团访问NCBI，与David Lipman主任等展开了深入的讨论。
- 2015年4月21日，项目结题会在中科院生物物理所召开，调研结果形成的文件，已上交国家有关部门。

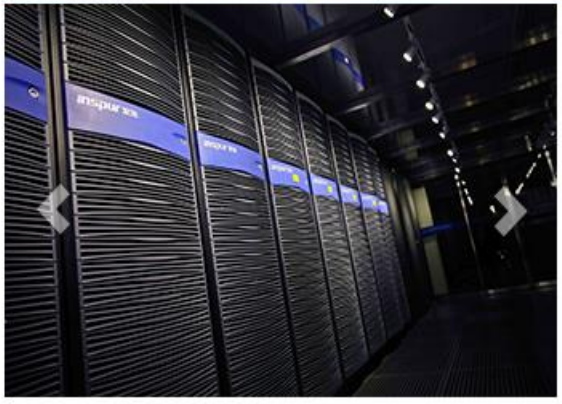
BIGD: Big challenge

BIGD Databases Tools Publications About Us English 中文

BIG Data Center in Beijing Institute of Genomics

The BIG Data Center advances life & health sciences by providing open access to a variety of resources, with the aim to translate big data into big discoveries and support activities in both academia and industry.

Deposition → Integration → Translation



Databases

Genome Sequence Archive Raw Sequence Reads	Genome Warehouse Genes, Genomes and Sequences	Gene Expression Nebulas RNA-Seq Expression Profiles
Genome Variation Map Sequencing and Array Variants	Methylation Bank DNA & RNA Methylomes	Science Wiki Wikis for Community Annotations

See the list of all database resources

News

- CloudPhylo: Spark-based Phylogeny Reconstruction Tool Scalable for Big Data Analysis. (2016-11-1)
- Progress on molecular mechnism studies of rice heterosis. (2016-10-26)
- LncRNAWiki joins RNAcentral. (2016-10-18)
- The 5th Young Bioinformatics PI Workshop will be held in Beijing, China. (2016-7-13)
- The 2016 Genomics Frontiers Symposium will be held in Shenyang, China. (2016-7-13)

Featured Resources

Databases

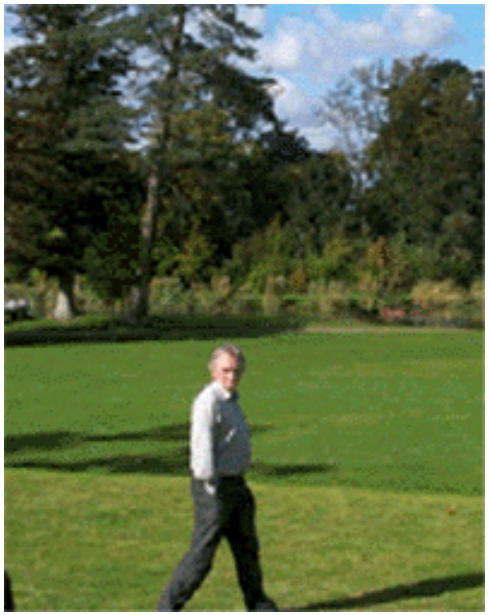
- Database Commons
- GSA
- IC4R

An email message from Alan Bleasby

Last, but not least …

I don't think we can get a Nobel prize by what we are doing, but the Nobel prize winners know what we are doing for.

So,

I will go to my death with a smile. ☺

- Alan Bleasby