生物信息学之我见

My View on Bioinformatics

北京大学香港青年才俊英才培训班（第二期）

2016年7月26日 北京大学电教304

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

http://abc.cbi.pku.edu.cn/talk/myview.pdf

The Definition of Bioinformatics

The term was apparently coined by B. Hesper and P. Hogeweg of the University of Utrecht Theoretical Biology/Bioinformatics Group in the mid 1970s.

The branch of science concerned with information and information flow in biological systems, esp. the use of computational methods in genetics and genomics.

1976 - *Acta Biotheoretica* 25 3 We publish papers on the philosophy of biology, biomathematics and bioinformatics.

1987 - *Science* 4 Sept. 1108/3 A new research program in bioinformatics. This is intended to bring together research in computing science, structural biology, and molecular genetics.

Medical Subject Heading index database still uses computational biology for bioinformatics.

Wikipedia lists more than 10 special areas that involving in bioinformatics.

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”。

以核酸蛋白质等生物大分子为主要研究对象

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

以信息、数理、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

对序列数据进行存储、管理、注释、加工

对各种数据库进行查询、搜索、比较、分析

构建各种类型的专用数据库信息系统

研究开发面向生物学家的新一代计算机软件

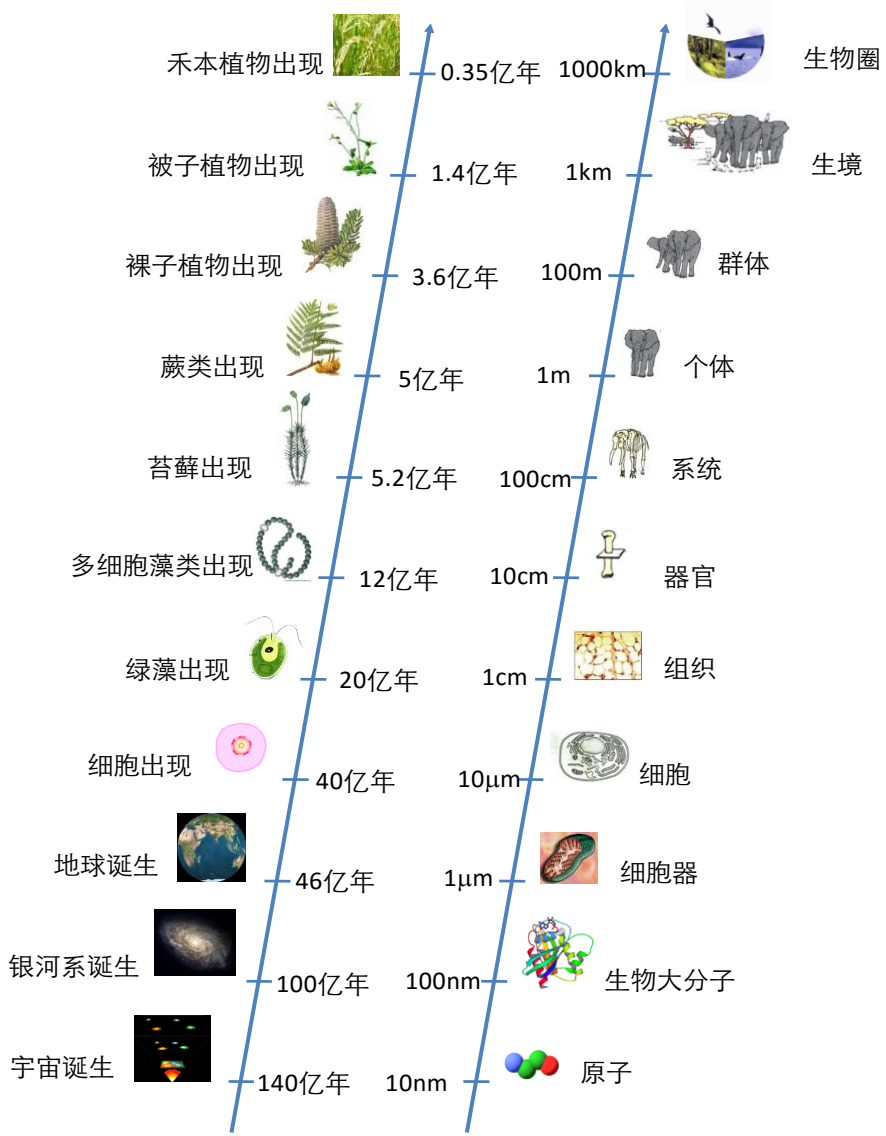
利用数理统计、模式识别、动态规划、密码解读、语意解析、信令传递、神经网络、遗传算法以及隐马氏模型等各种方法

对序列、结构数据进行定性和定量分析，从中获取基因编码、基因调控、序列-结构-功能关系等理性知识

阐明细胞、器官和个体的发生、发育、病变、衰亡的基本规律和时空联系

探索生命起源、生物进化、生命本质等重大理论问题，最终建立“生物学周期表”

生物学中的时间和空间尺度

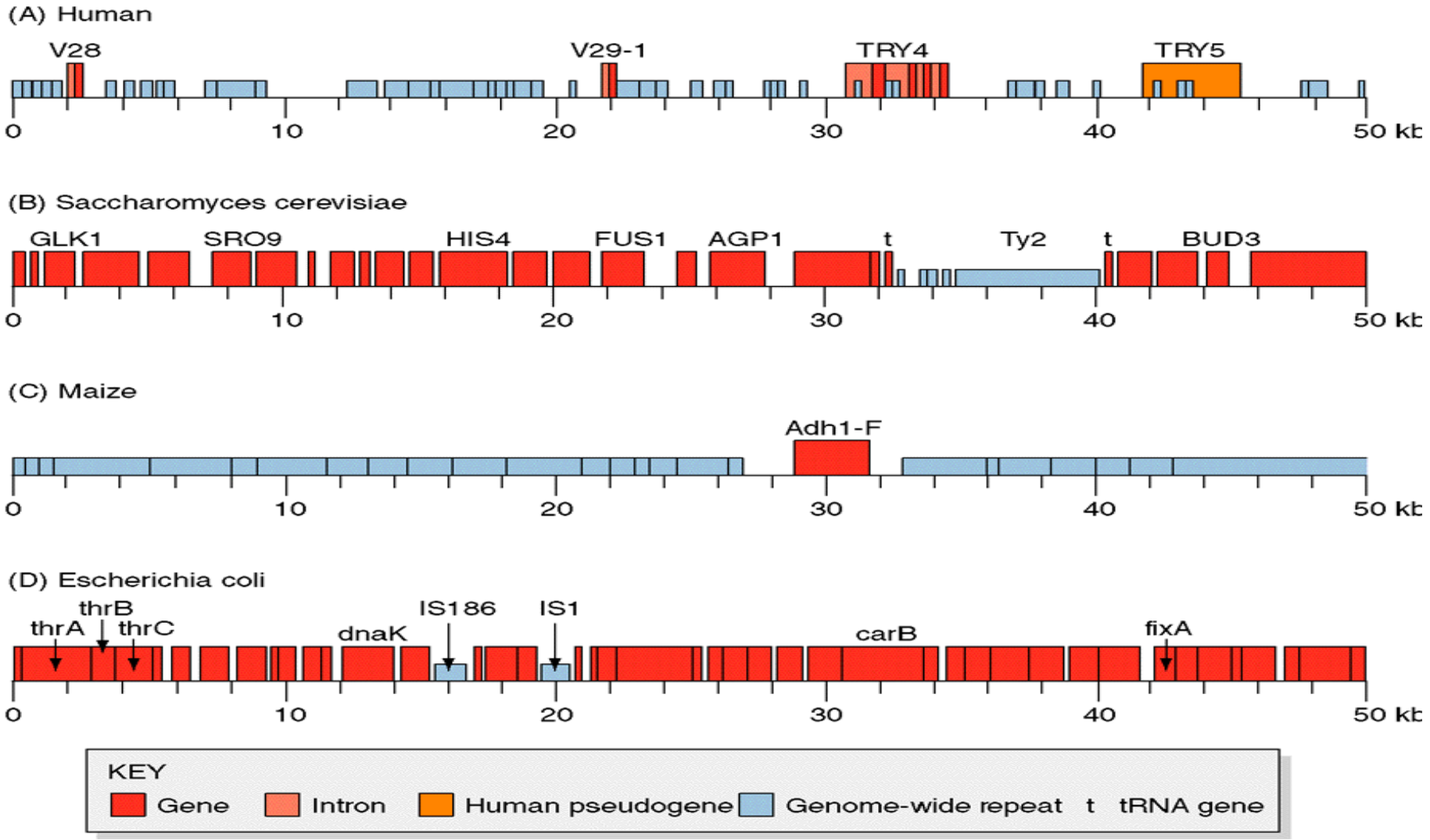


宇宙大爆炸	140亿年
地球诞生	46 亿年
单细胞生物	40亿年
多细胞生物	12亿年
寒武纪大爆发	5亿年
被子植物出现	1.4亿年
恐龙灭绝	6千万年
人类诞生	5百万年

生境 — 千米
群体 — 米
个体 — 厘米
细胞 — 微米
分子 — 纳米

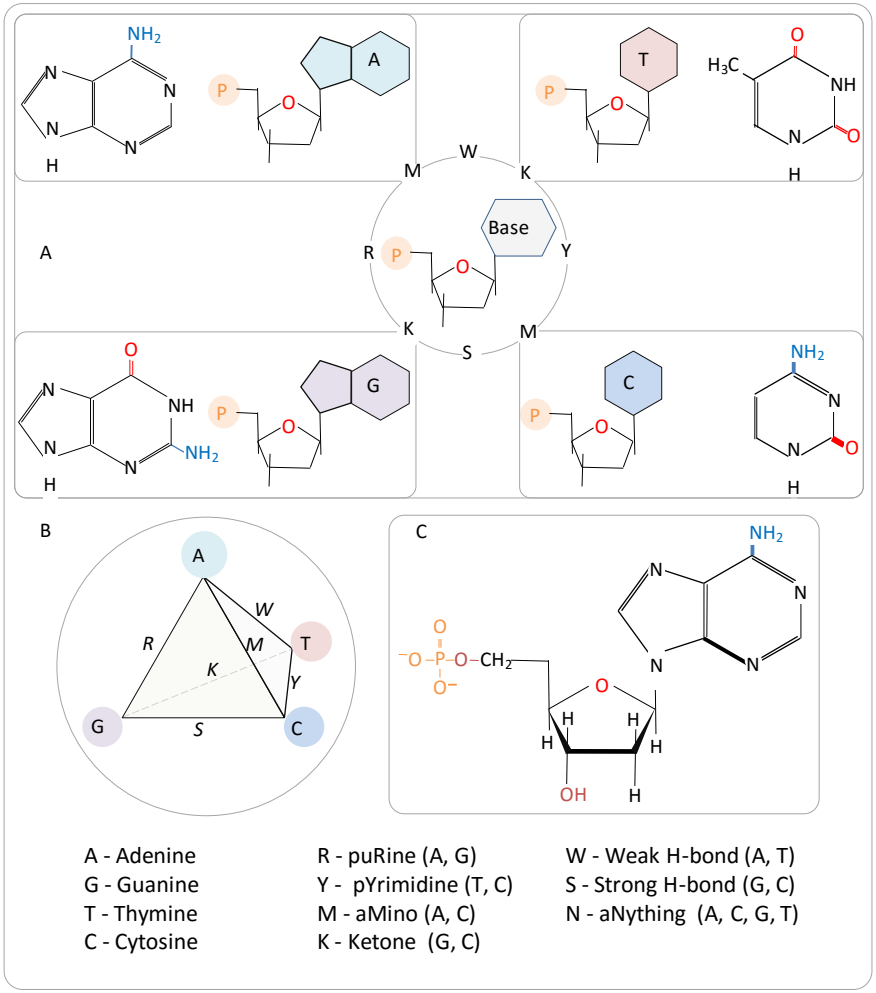
遗传物质和外界环境共同决定了生物多样性

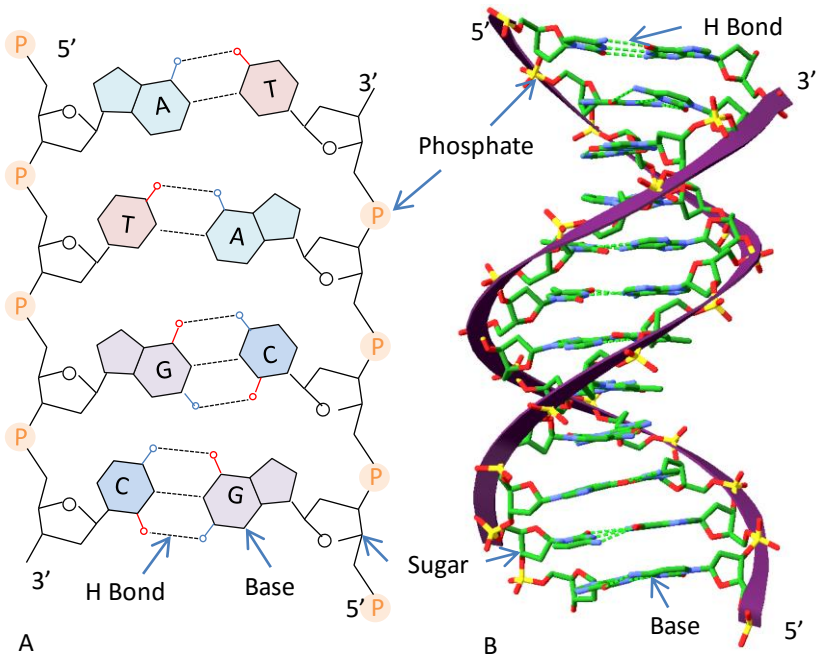
Genome composition among different organisms



from TA Brown: Genomes

DNA双螺旋由四种核苷酸配对形成

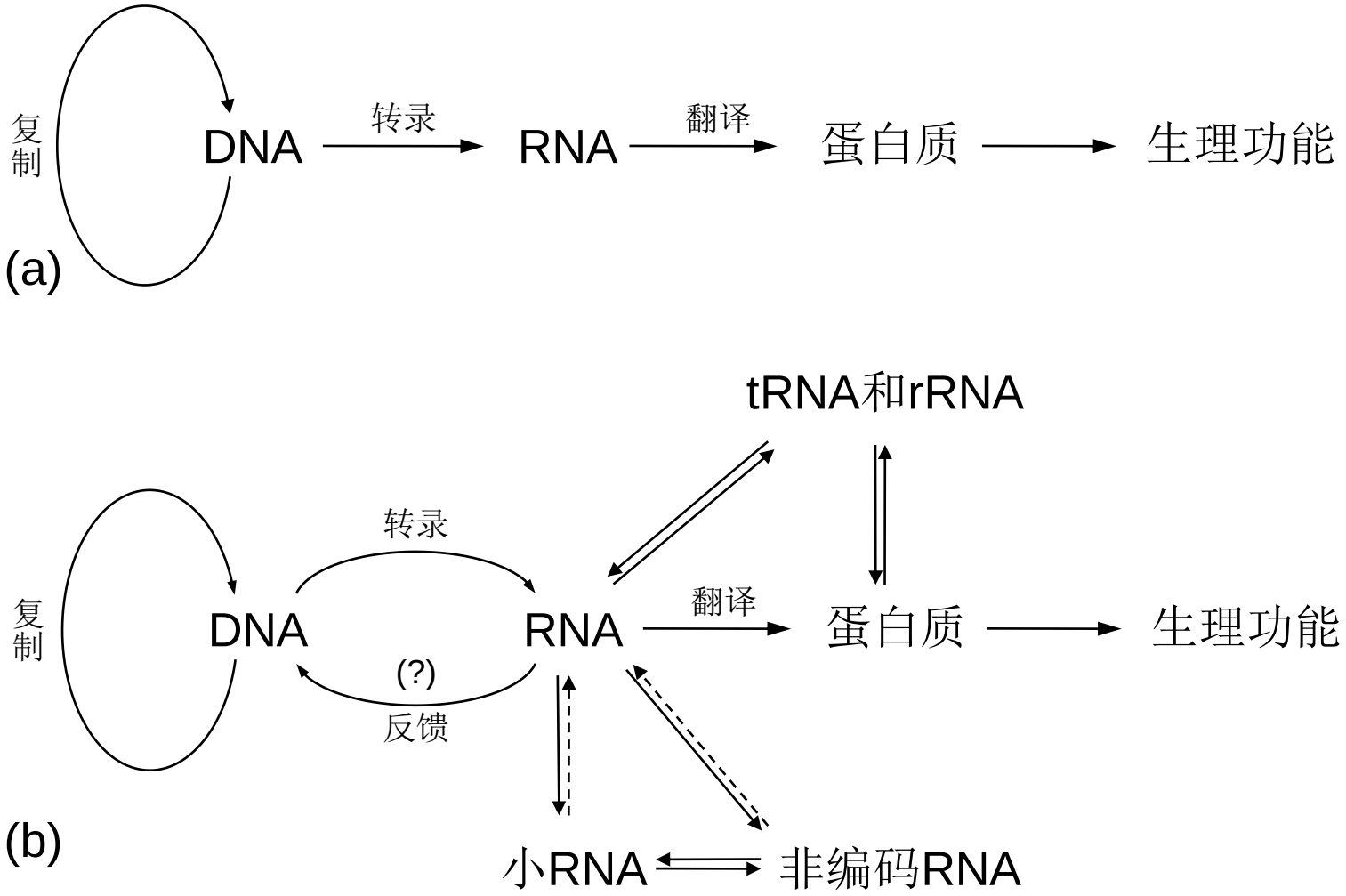




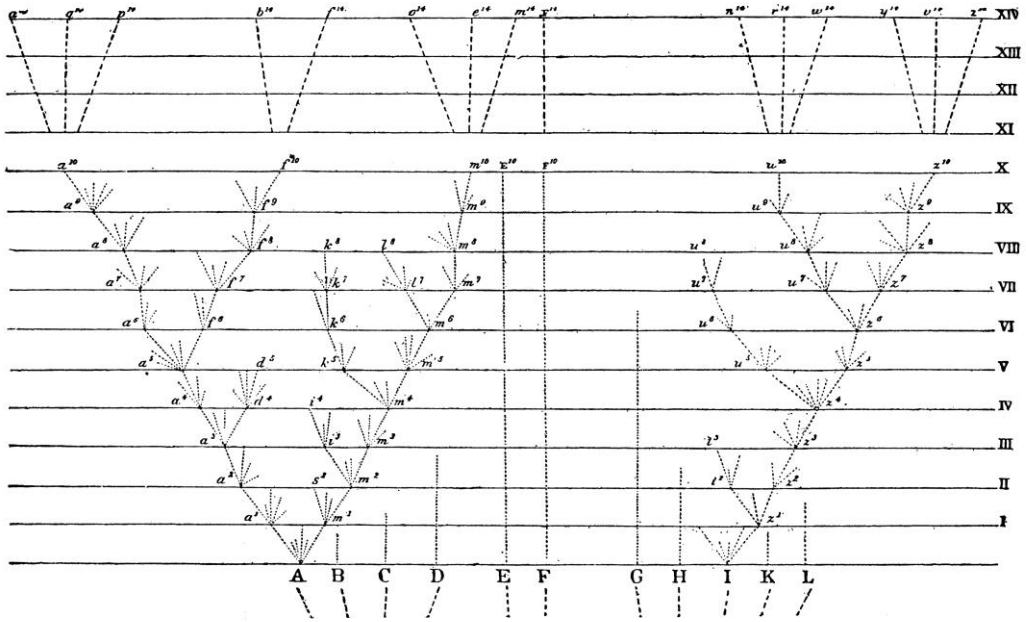
The DNA double helix is formed by base paring between A and T, C and G

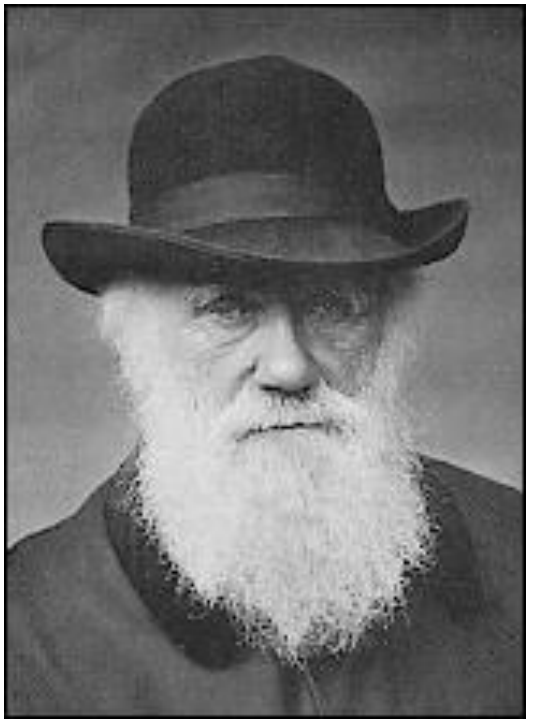
Luo JC
2008-08-08

分子生物学中心法则（Central Dogma）



1859年达尔文提出“物竞天择”学说

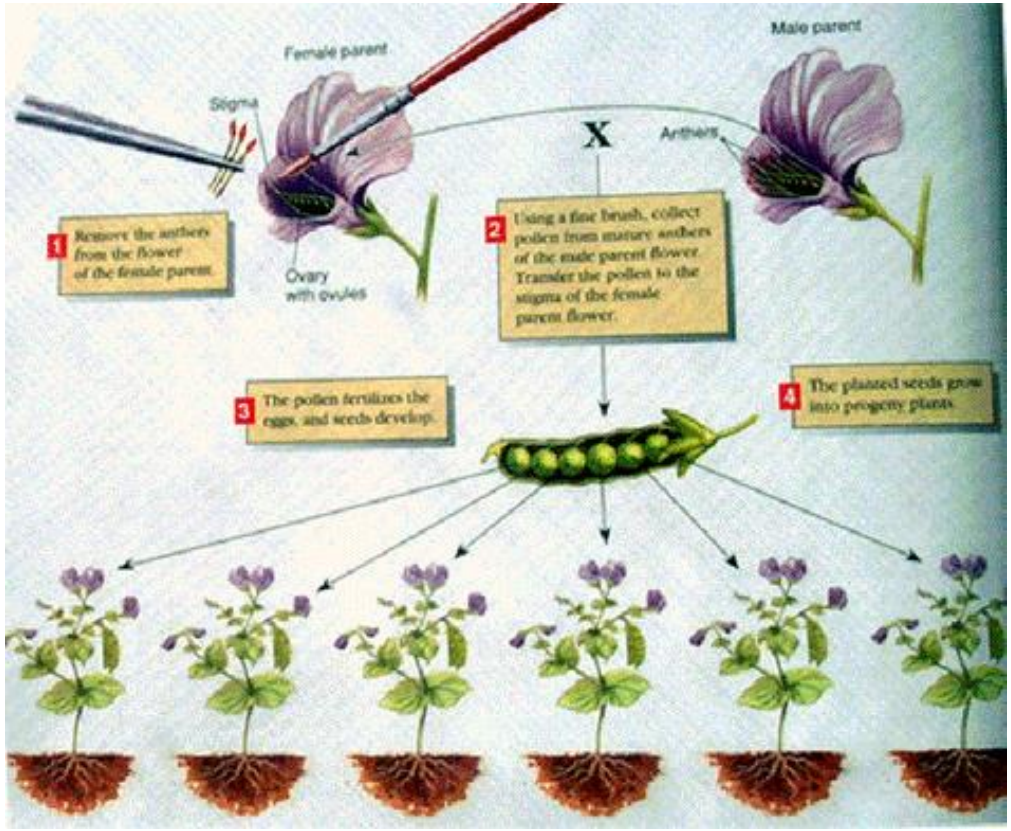




Charles Dawin and the species tree in his book “on The Origin of Species - by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life”, publisheed on 4 Nov 1859.

– from Zhong Yang , 2007

1865年孟德尔奠定遗传学基础



Mendel and his genetic experiment using garden pea as a model system – from The Science of Genetics, A Atherly, 1999

1926年摩尔根发表“基因论”

Morgan and his fly room at Columbia University

- 证明了基因是染色体的一部分
- 证明了孟德尔定律，发现了伴性遗传
- 测定了果蝇中基因的相对位置和距离，绘制了果蝇染色体图谱
- 发现了缺失、重复、倒位和易位等染色体畸变现象

1953年DNA双螺旋发现

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

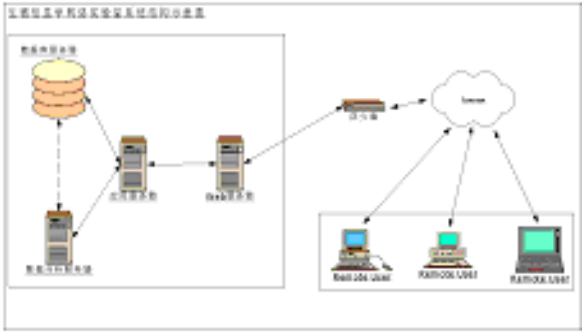
This figure is purely diagrammatic. The two ribbons symbolize the two phophate-sugar chains, and the horizonal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

1991年Gilbert 提出生物学研究新模式

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*, Vol 349, 10 Jan 1991, p99



2000年2月2日北大燕北园小区联网

On 2 Feb 2000, 300 family PCs at a resident area of Peking University connected to the Internet via fibre cables; on 12 Feb 2001, 2000 dorms of undergraduate students of Peking University connected to the Internet via fibre cables.

02022000

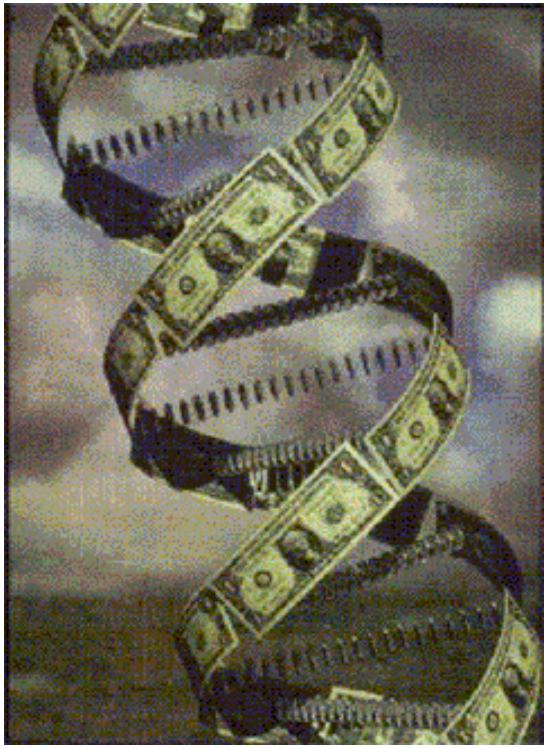
01011000

AGAGGAAA

BioMedX

Bio-Informatics

2001年2月人类基因组计划草图完成



International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature*, **409**:860-921, 15 Feb 2001.

J. Craig Venter, *et al*. The Sequence of the Human Genome. *Science*, **291**:1304-1351, 16 Feb 2001.

2016精准医疗与未来医学大会

一、精准医学与未来大健康产业论坛

议题一：组学领域研究新进展
议题二：生命组学与精准医学
议题三：未来大健康产业发展模式
议题四：精准医疗领域投资风口

二、精准诊断前沿进展高峰论坛

议题一：出生缺陷防控
议题二：微生物快速检测
议题三：基因测序技术
议题四：肿瘤早期精准筛查技术
议题五：医疗大数据的采集与管理

三、精准医疗临床应用高峰论坛

议题一：国内测序行业政策大观
议题二：临床应用端案例分享
议题三：精准医疗市场空间展望

主办: 千人杂志, Engineering Information Institute The Gordon Life Science Institute
承办: 千人智库
时间： 2016年9月20-22日
地点： 西安

The Birth of Bioinformatics Databases

- The 1st biological database Protein Data Bank (PDB) was started in 1977.
- 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

- 1979年，美国科学基金会NSF召集会议，对建立核酸序列数据库达成共识。
- 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，创建核酸序列数据库GenBank，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和Robert Ledley获NIH资助，建立了蛋白质序列数据库PIR；1988年，NBRF和德国、日本联合成立了国际蛋白质序列数据库。
- 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。UniProt包括UniProtKB/UniRef/UniParc三部分和Swiss-Prot/TrEMBL两个子库。

The Great Men in Bioinformatics

- **Needleman** – Saul Needleman and Christian Wunsch proposed an algorithm for global sequence alignment in 1970 [PMID 5420325]
- **Waterman** – Temple Smith and Michael Waterman developed an algorithm for local sequence alignment in 1981 [PMID 7264238]
- **Lipman** – The BLAST database sequence similarity search tool developed and maintained by a team at NCBI headed by David Lipman is widely used.
- **Fasman** – Peter Chou and Gerald Fasman developed a method for the prediction of protein secondary structures.
- **Dayhoff** – Margaret Dayhoff delivered the first scoring matrix PAM250 for protein sequence alignment in the 1980'.
- **Henikoff** – Steven Henikoff and Jorja Henikoff built the BLOSUM scoring matrices in 1992 [PMID 1438297].

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

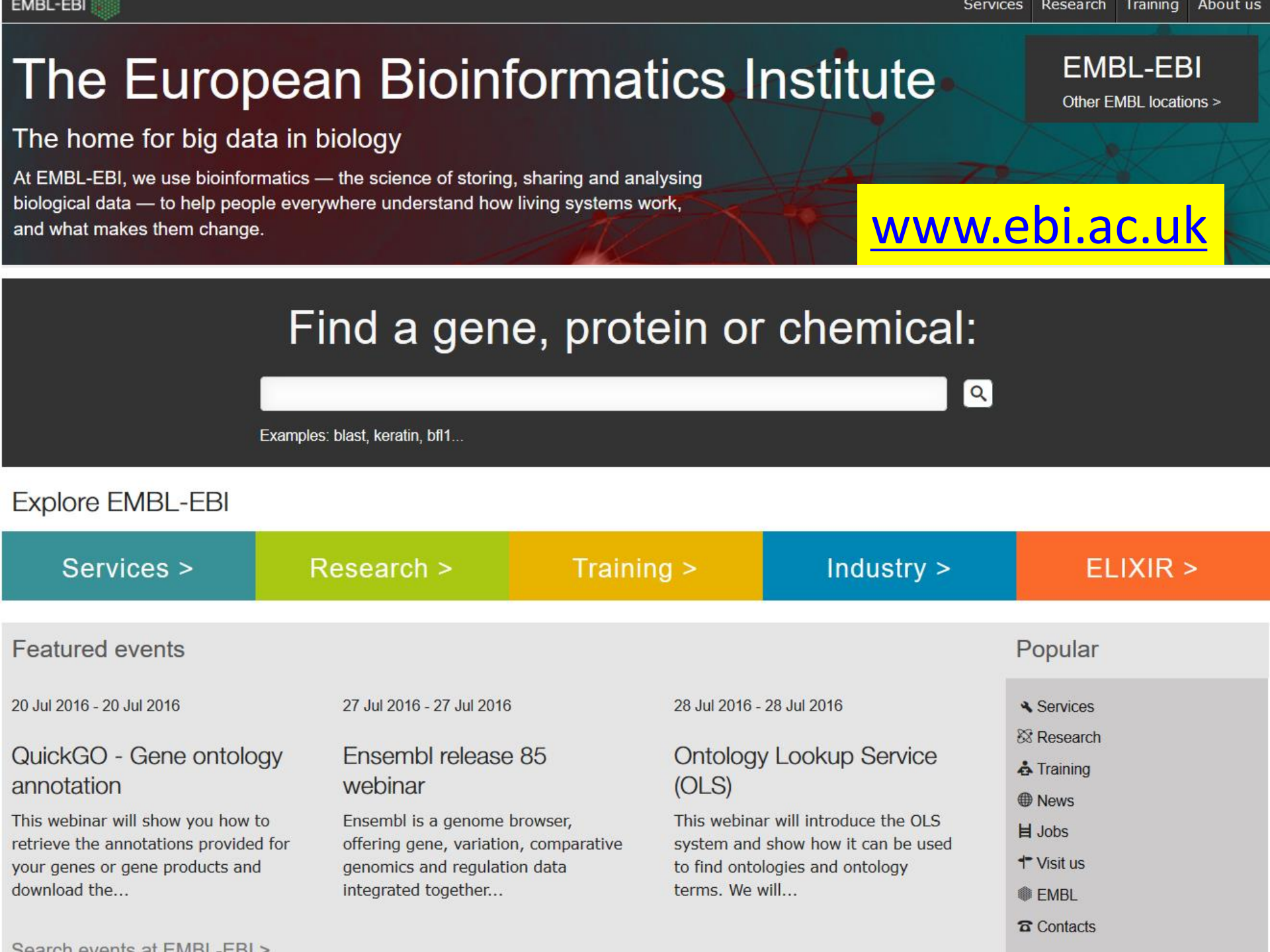
Wednesday, July 6, 2016

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。

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


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

欧洲生物信息学研究所EBI

- 成立于1994年，坐落在英国剑桥南部12英里Wellcome基金会基因组园区内。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、法国等西欧各国。
- 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，并从事生物信息研究开发。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由EBI负责管理发布。
- 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，著名英国生物信息学家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，Rolf Apweile和Ewan Birney任共同主任。

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

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]

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

YouTube

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...

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.

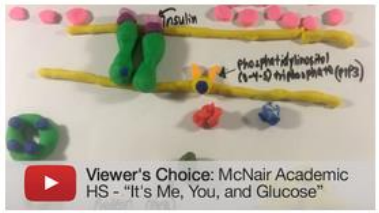
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

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

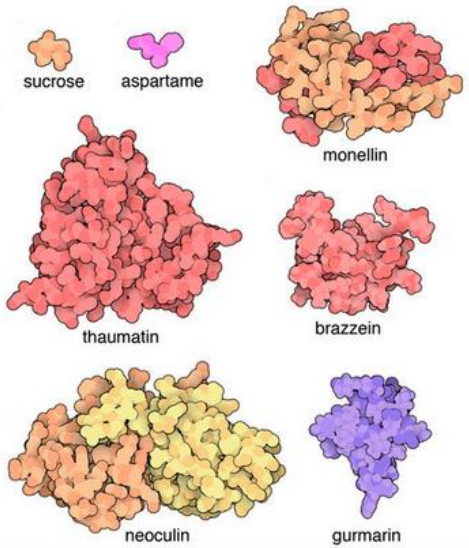
More Info

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



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

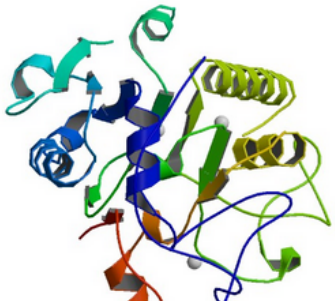
July **Molecule of the Month**



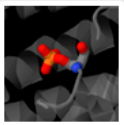
Monellin

Latest Entries

As of Tuesday Jul 19

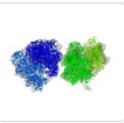


Features & Highlights



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*

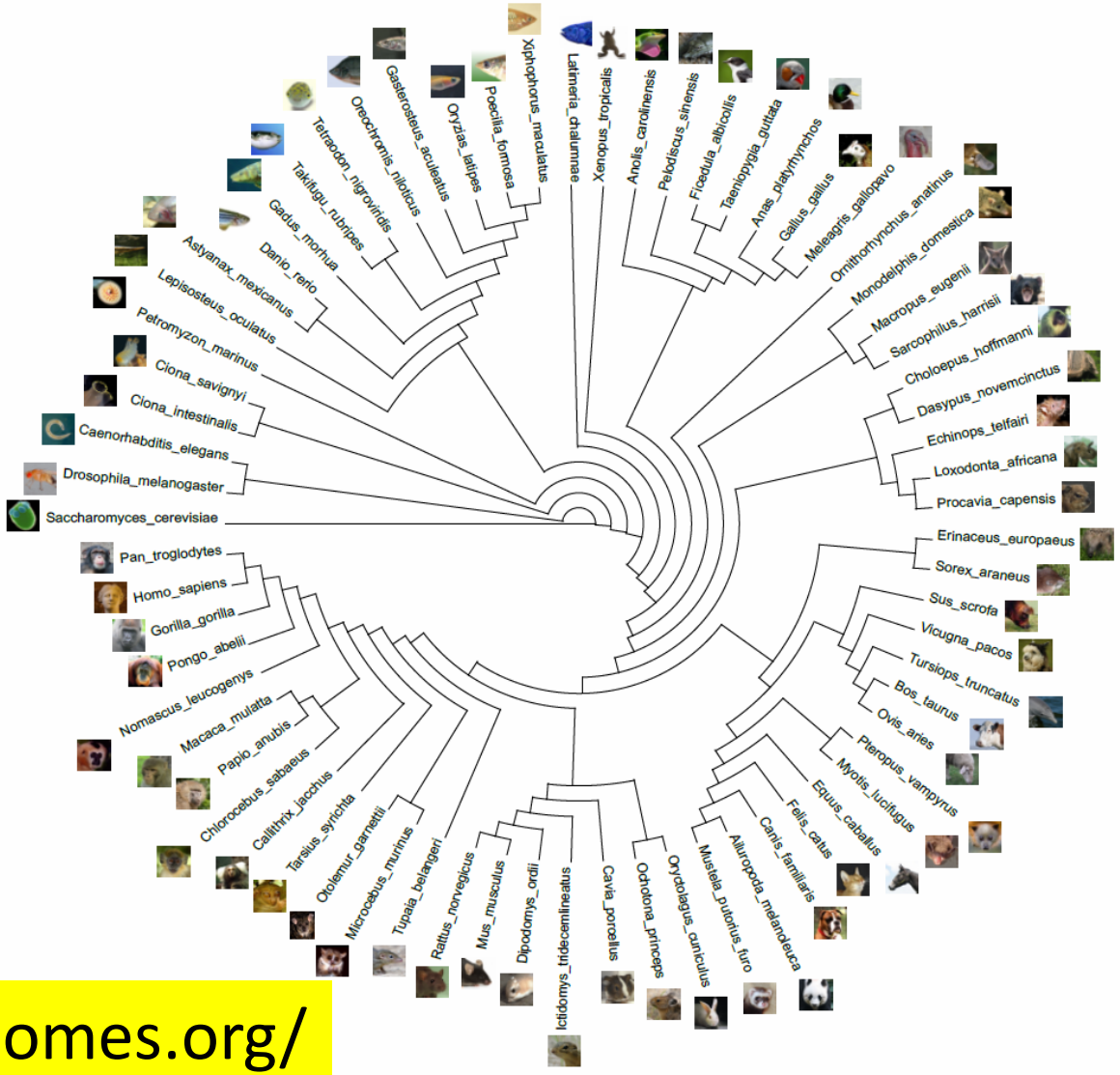
wwPDB News: Announcement: Map Volume Deposition

Ensembl基因组数据库中已完成测序动物

No of genomes

Vertabrates: 62
Other Metazoa: 54
Plants: 43
Fungi: 670
Bacteria: 47233

(Data collected in 25 July 2016)



http://ensemblgenomes.org/

1999年郝柏林院士建议组建国家生物医学信息中心

- 1999年4月，郝柏林院士以学生身份，全程参加了北京大学生物信息中心举办的为期一周的生物信息培训班。6月10日，**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，登载在国家自然科学基金委简报上。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。

2001年第一届中国生物信息学大会召开

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

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

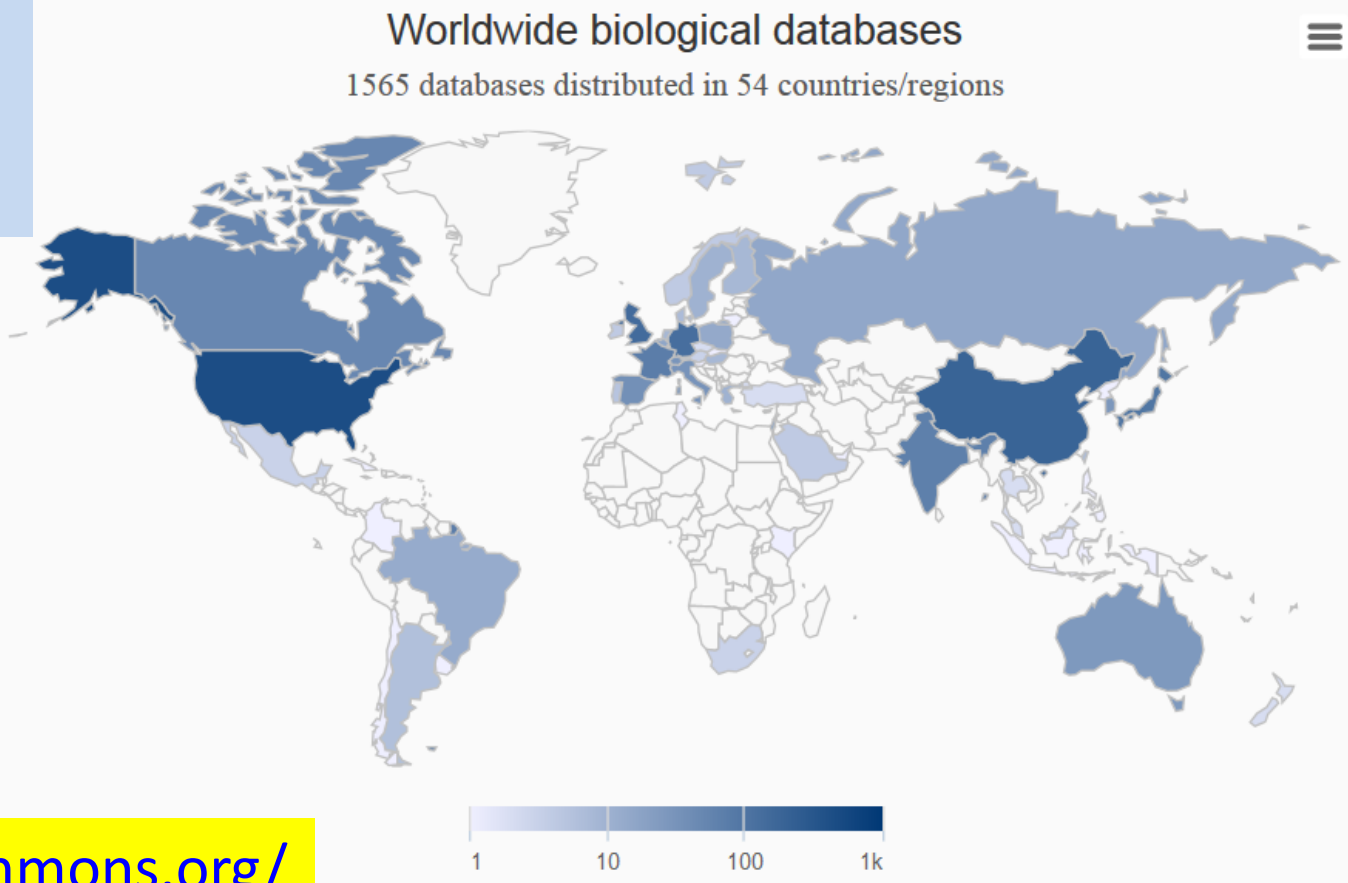
Database Commons a catalogue of biological databases

Home | Browse | Search | Statistics | Submit | Help | Database name

1565 databases
413 organisms
54 countries/regions
14 categories

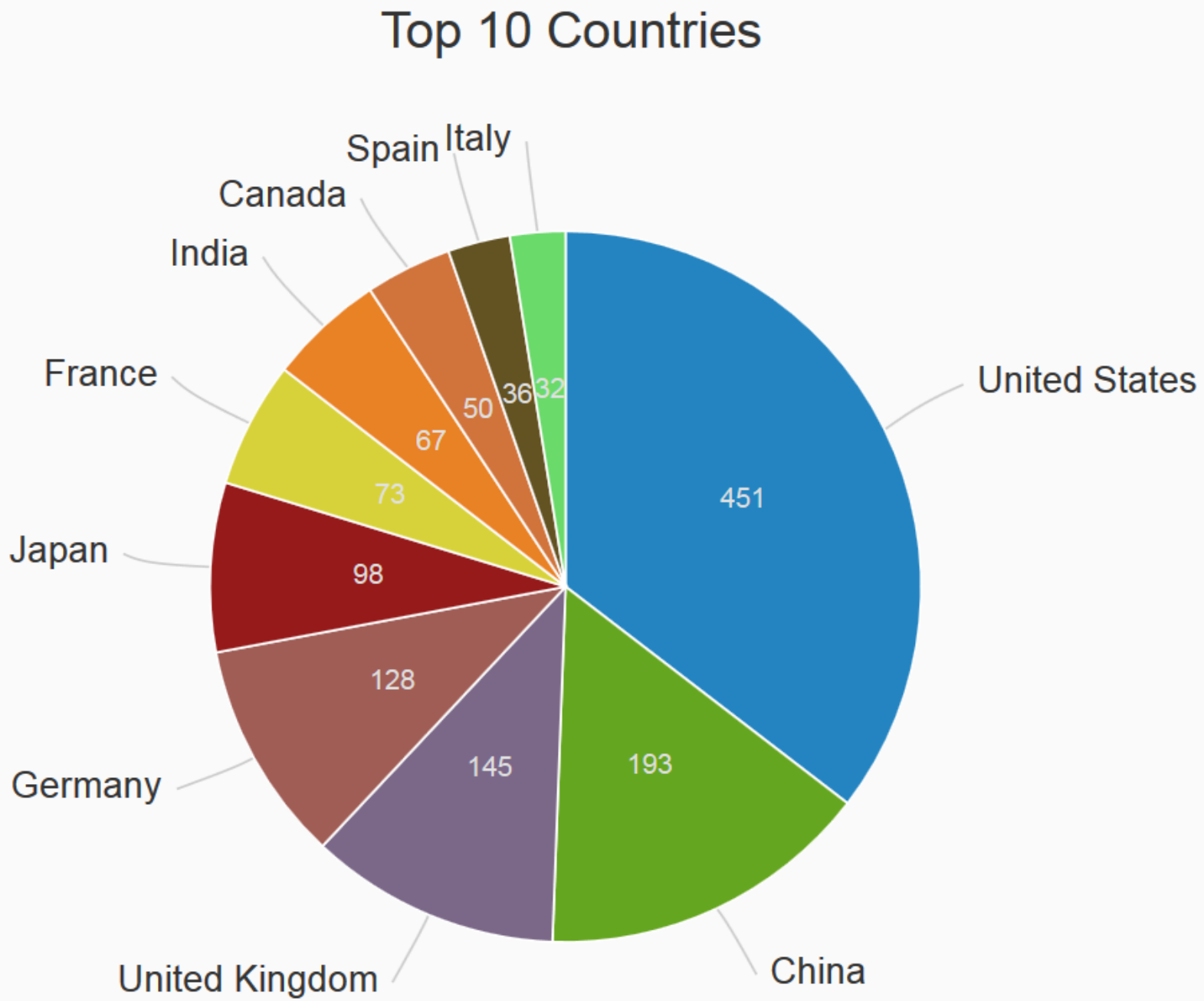
News & updates

- Light-weight visualizations of worldwide databases
- More than 1000 databases incorporated
- Latest additions:
  - Epitome
  - TED
  - SoyGD
  - RMD
  - Plant MPSS databases



http://databasecommons.org/

Zhang Lab, Beijing Institute of Genomics, Chinese Academy of Sciences
中科院基因组所章张课题组


Top 10 Countries
United States
451
China
193
United Kingdom
145
Germany
128
Japan
98
France
73
India
67
Canada
50
Spain
36
Italy
32

我国生物信息数据库论文

- 总计416篇，国际刊物389篇，国内刊物27篇
- NAR数据库专辑133篇，Database专刊54篇，其它刊物229篇
- 2011-2015五年内257篇，2014和2015两年内150篇

Databases built in China

- NAR Database papers from China
  - The PubMed list of database papers published on NAR by authors from China.
- Database papers from China
  - The PubMed list of papers published on the journal Database by authors from China.
- Other database papers from China
  - The PubMed list of database papers published on other journals by authors from China.

http://abc.cbi.pku.edu.cn/databases.php

The Z-curve Database of Genomes

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

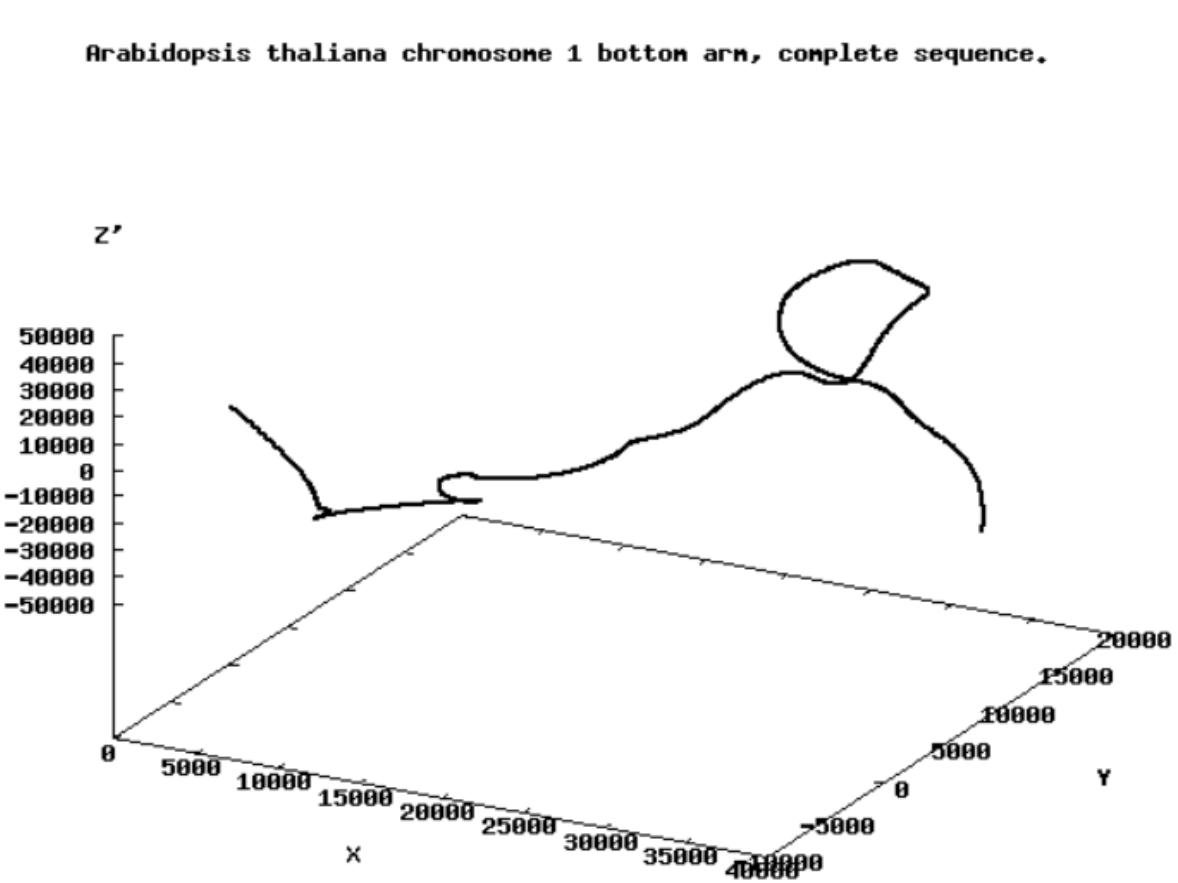
» Home » Information of Eukaryota

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

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

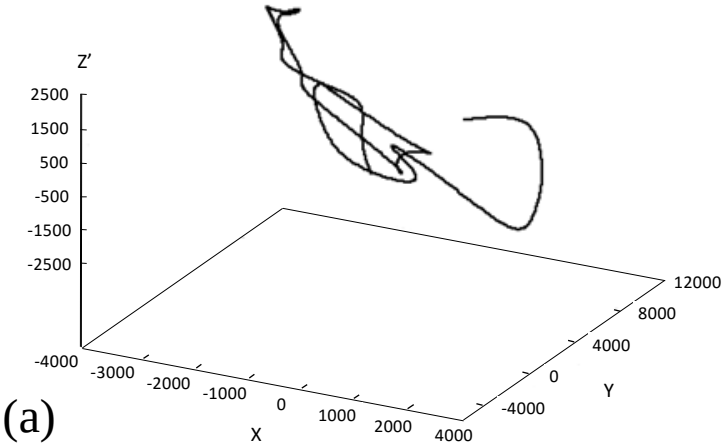
$Y_n = (A_n+C_n) - (G_n+T_n)$,

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

（n = 0, 1, 2, …, N）

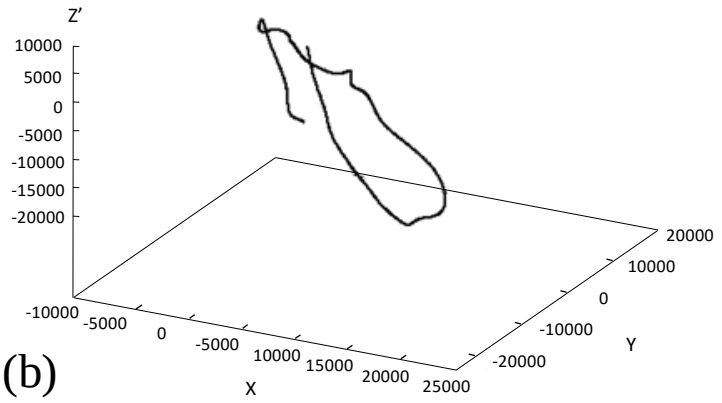
Z-Curve Presentation of bacterial genomes

Haemophilus influenzae Rd KW20



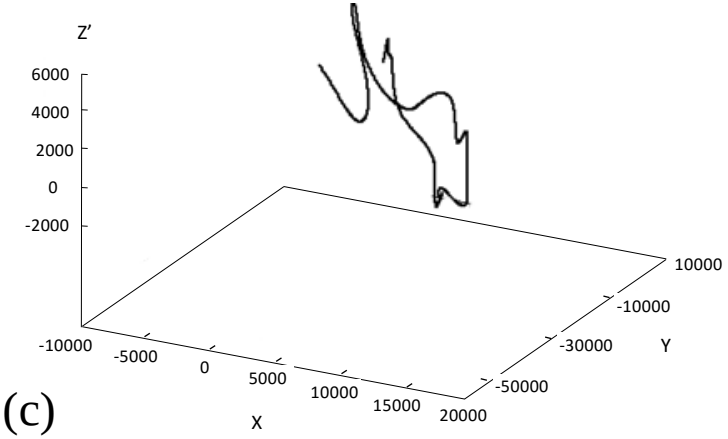
(a)

Escherichia coli K12 MG1655



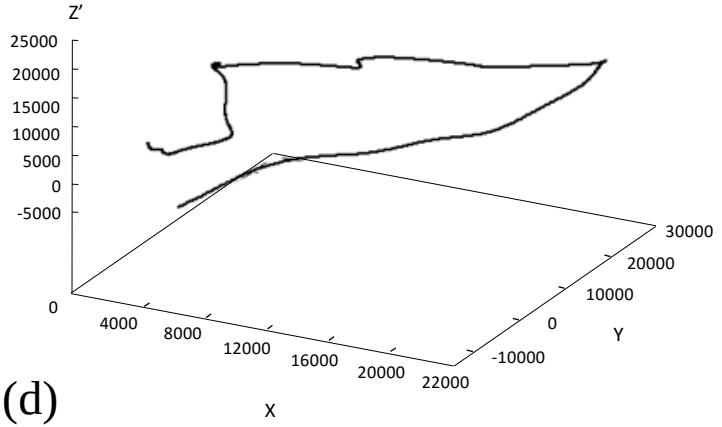
(b)

Mycobacterium tuberculosis CDC1551



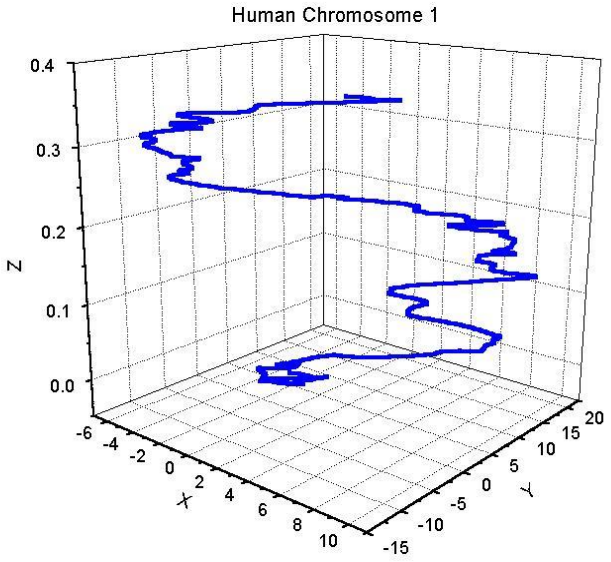
(c)

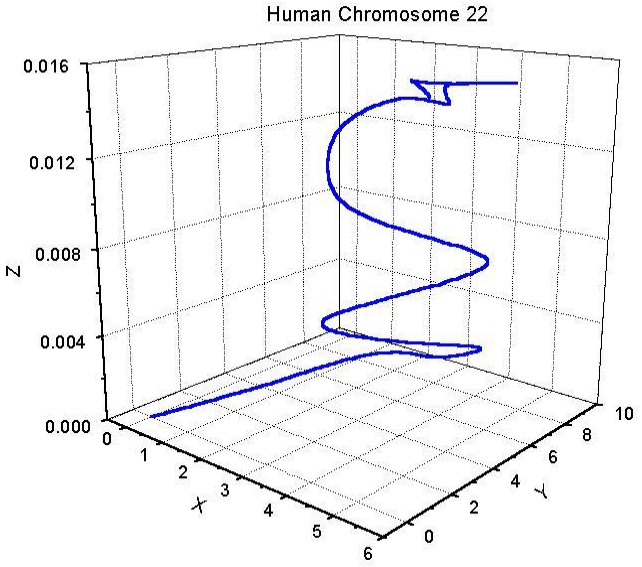
Methanococcoides burtonii DSM 6242

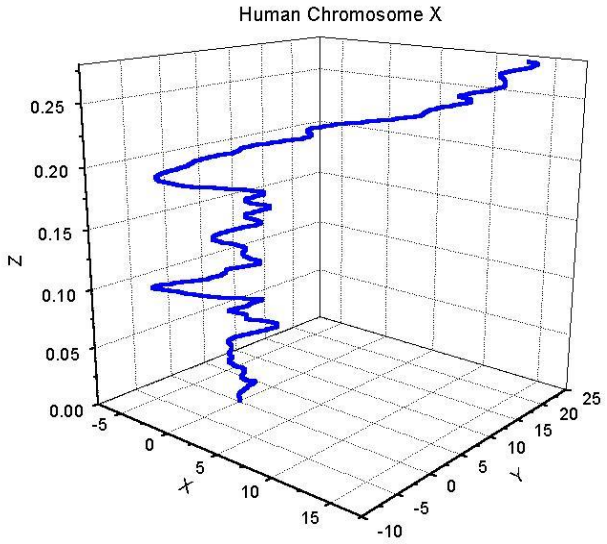


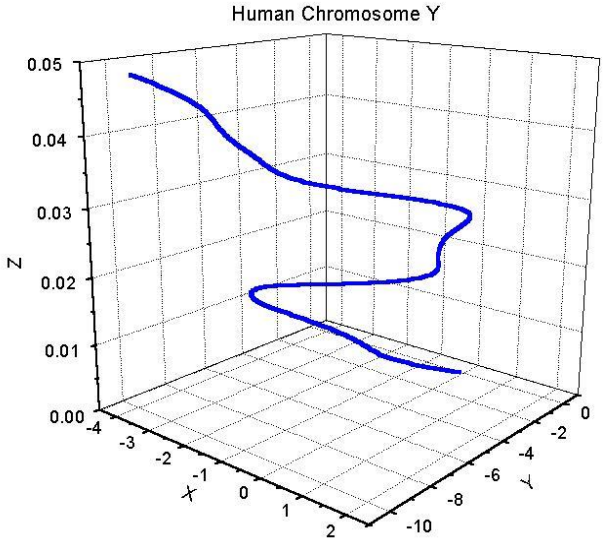
(d)

Z-Curve Presentation of Human Chromosomes



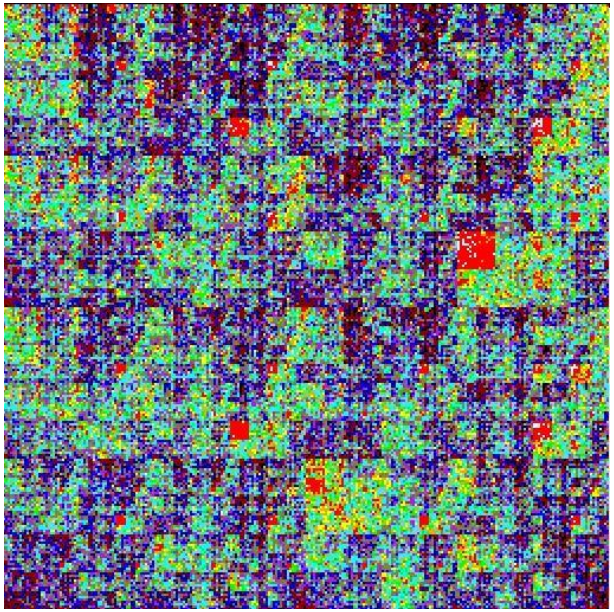


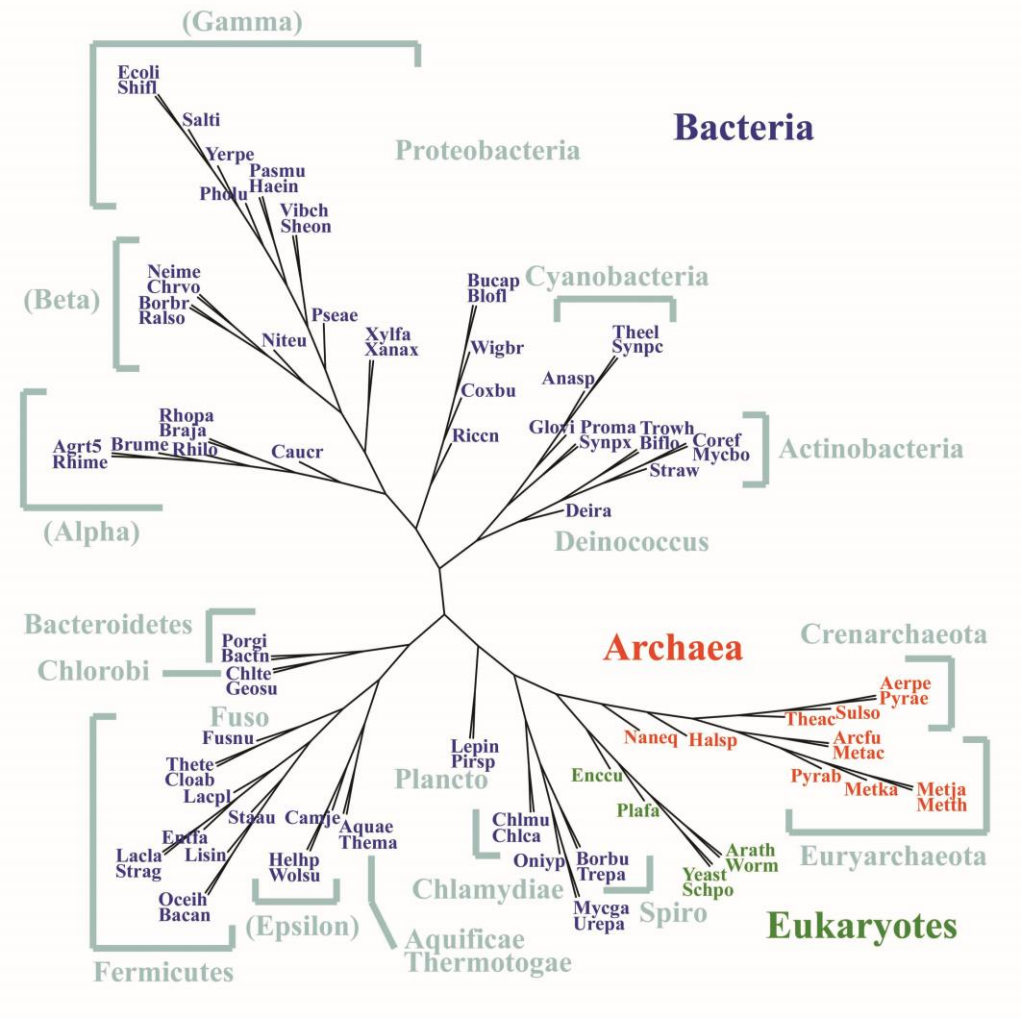




郝柏林院士提出的CVTree方法

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/

CVTree3: Composition Vector Tree Version 3

>>> Previous Version

Description:

CVTree constructs whole-genome based phylogenetic trees without sequence alignment by using a Composition Vector (CV) approach. It was first developed to infer evolutionary relatedness of microbial organisms and then successfully applied to viruses, chloroplasts, and fungi. CVTree3 makes comparison with taxonomy and reports tree-branch monophyleticity from domain to species. Please read the *Online User's Manual* for details.

Reference:

Ji Qi, Bin Wang, Bailin Hao (2004) Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach, J Mol Evol, 58: 1 –11

Guanghong Zuo, Bailin Hao (2015) CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy, Genomics Proteomics & Bioinformatics, being submitted.

Load/Create Project:

Enter Project Number to reload a previously completed project for checking or changing parameters and re-run. A blank input creates a new project. A project will be kept for 7 days after the last run.

Load/Create Project Example

Announcement:

CVTree3 web server is under final testing. *Feedback and criticism* welcome!

http://tlife.fudan.edu.cn/archaea/cvtree/cvtree3/cvtree.html

1996年北京大学生物信息中心成立

Supported by the Ministry of Education, we started the center of bioinformatics (CBI) at Peking University in 1996, and joined the European Molecular Biology Network (EMBnet). Our primary goal was to provide bioinformatics resource, service and education to domestic users, mainly in the molecular biology community.

Mirrors of GDB, RGD and ExPASy 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.

1998年CBI举办第一届生物信息培训班

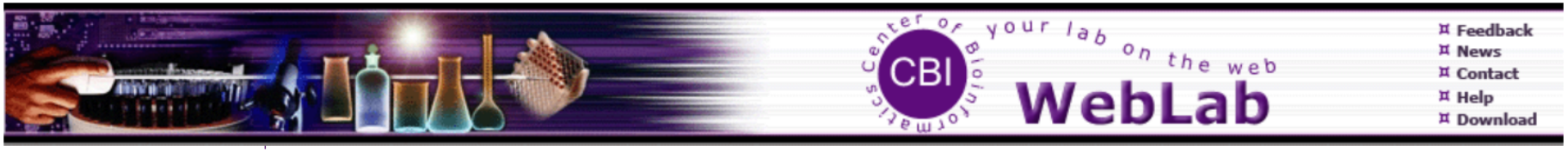
The first user training course was sponsored by European Molecular Biology Network (EMBnet) and organised by EMBnet China node. Peter Rice, Alan Bleasby, Jack Leunissen, Frank Wright and David Judge came to teach the course. A video show was displayed at the 10th EMBNet meeting in Hinxton in October 1998.

More than 10 training courses, workshops, seminars and conferences were organised during the past 15 years. Peter Rice released the 1st EMBOS program SeqRet at the 1999 course.

北京大学生物信息中心部分数据库

简称	内容	网址
PlantTFDB	植物转录因子数据库	http://planttfdb.cbi.pku.edu.cn/
LSD	植物叶片衰老数据库	http://psd.cbi.pku.edu.cn/
AHD	拟南芥激素及相关基因数据库	http://ahd.cbi.pku.edu.cn/
SeedGeneDB	种子发育相关基因数据库	http://sgdb.cbi.pku.edu.cn/
SPD	哺乳动物分泌蛋白数据库	http://spd.cbi.pku.edu.cn/
AutismKB	孤独症相关基因数据库	http://autismkb.cbi.pku.edu.cn/
HomeoDB	同源异形框相关基因数据库	http://homeodb.cbi.pku.edu.cn/

2001年开发生物信息网上实验室



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

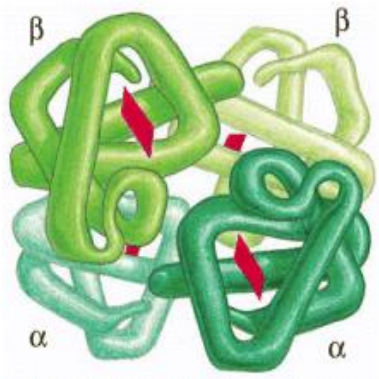
Liu, X., Wu, J., Wang, J., Liu, X., Zhao, S., Li, Z., Kong, L., Gu, X., Luo, J. and Gao, G. (2009) WebLab: a data-centric, knowledge-sharing bioinformatic platform. Nucleic Acids Res. (Full Text)

Please Note

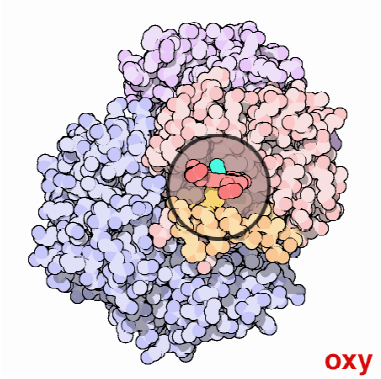
- According to our testing, WebLab can work well under , and . (detailed browser test result)
- This project is supported by National High-Tech (863) Programme.

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

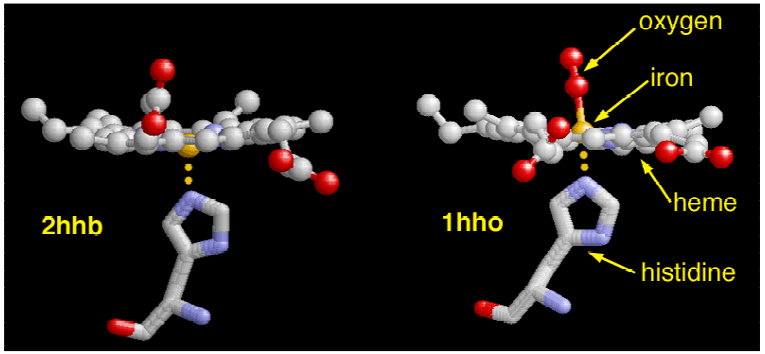
斑头雁血红蛋白序列/结构/功能分析



血红蛋白四个亚基



血色素携带氧分子



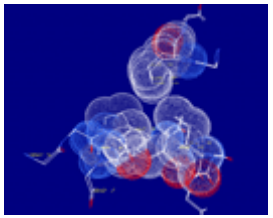
脱氧和含氧两种状态下的血色素分子

```
  1 VLSAADKTNVKGVFSKIGGHAEEYGAETLERMFTAYPQTKTYFPHFDLQH    50
    ||||||||||||||||||.||||||||||||||||||||||||||||||||
  1 VLSAADKTNVKGVFSKISGHAEEYGAETLERMFTAYPQTKTYFPHFDLQH    50

 51 GSAQIKAHGKKVAAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKF   100
    ||||||||||||.|||||||||||||||||||||||||||||||||||||||
 51 GSAQIKAHGKKVVAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKF   100

101 LGHCFLVVVAIHHPSALTPEVHASLDKFLCAVGTVLTAKYR    141
    ||||||||||||||||||.|||||||||||||||||||||
101 LGHCFLVVVAIHHPSALTAEVHASLDKFLCAVGTVLTAKYR    141
```

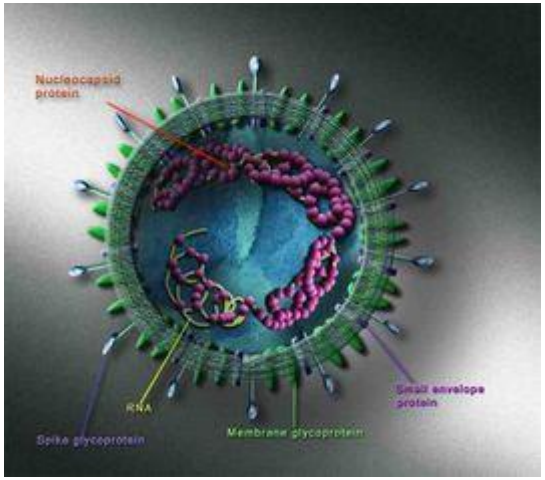
	α119	β55
斑头雁	Ala	Leu
安第斯雁	Pro	Ser
灰雁	Pro	Leu



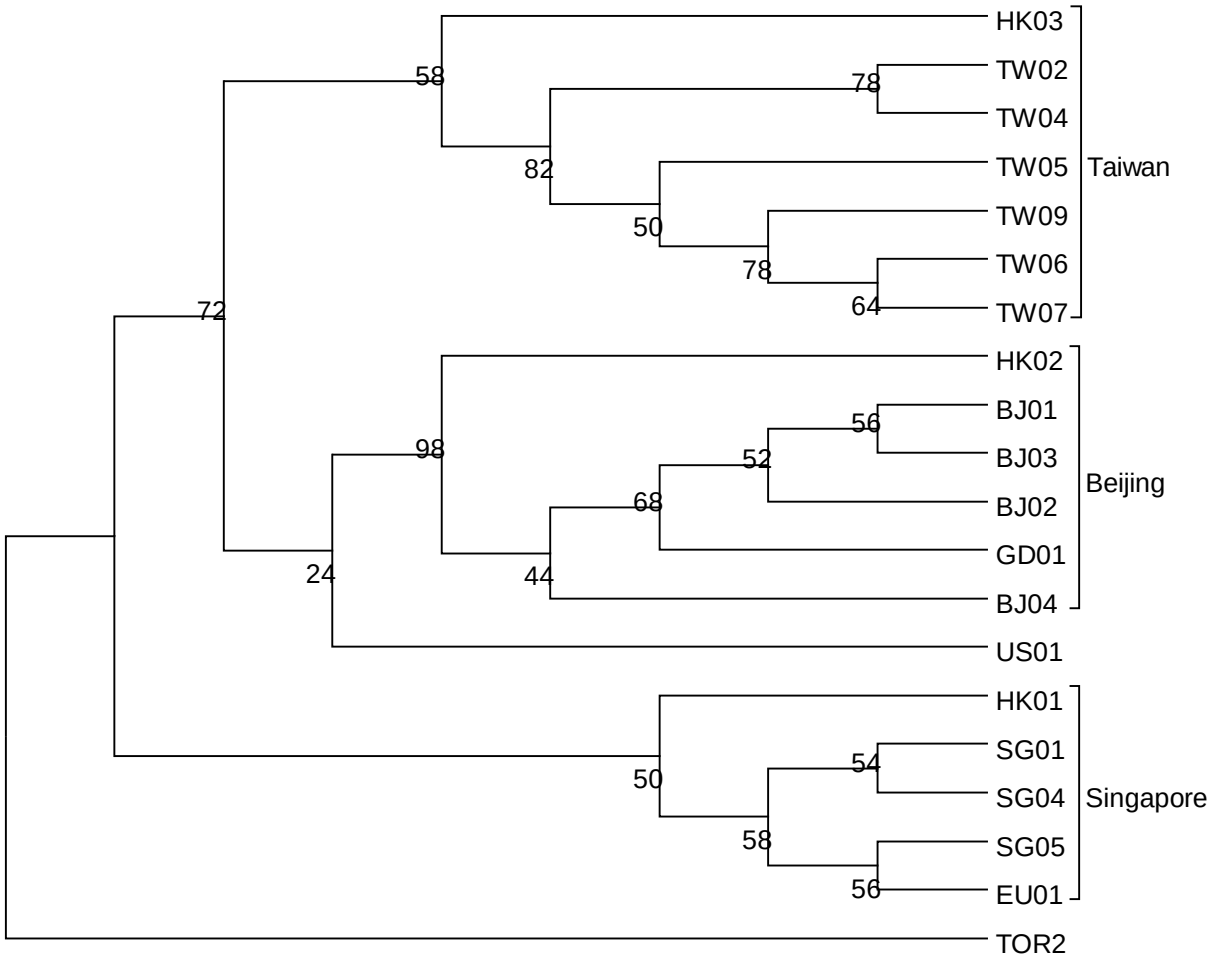


SARS病毒基因组系统发生树

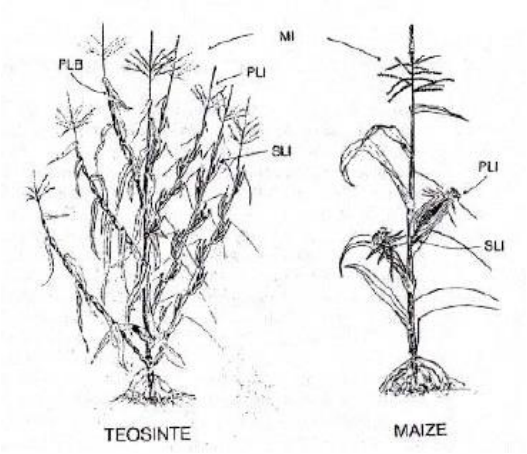


In 2003, the SARS epidemic caused by the coronavirus was a devastating flood to the Hong Kong economy. Bioinformatics approaches can be used to trace the origin and spread of the virus.

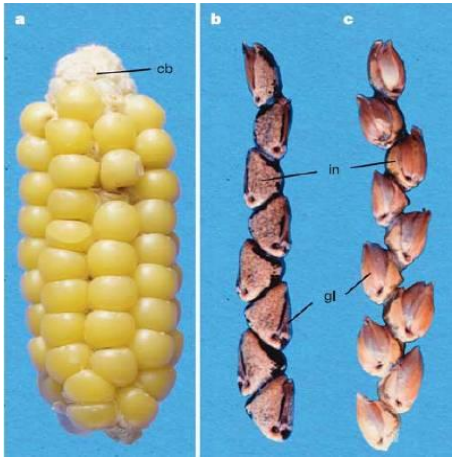


The Neighbor-Joining phylogenetic tree of 20 SARS virus genomes generated by MEGA6. A total of 30 sites with different nucleotide sequence were used. The evolutionary distances were computed using the p-distance method and the bootstrap values are shown next to the branches. Abbreviation: HK – Hong kong, TW – Taiwan, BJ – Beijing, GD – Guangdong, US – The United States, SG – Singapore, EU – Europe, TOR – Toronto.

基因突变与农作物性状相关



Doebley *et al.*(1997) *Nature* 386:485



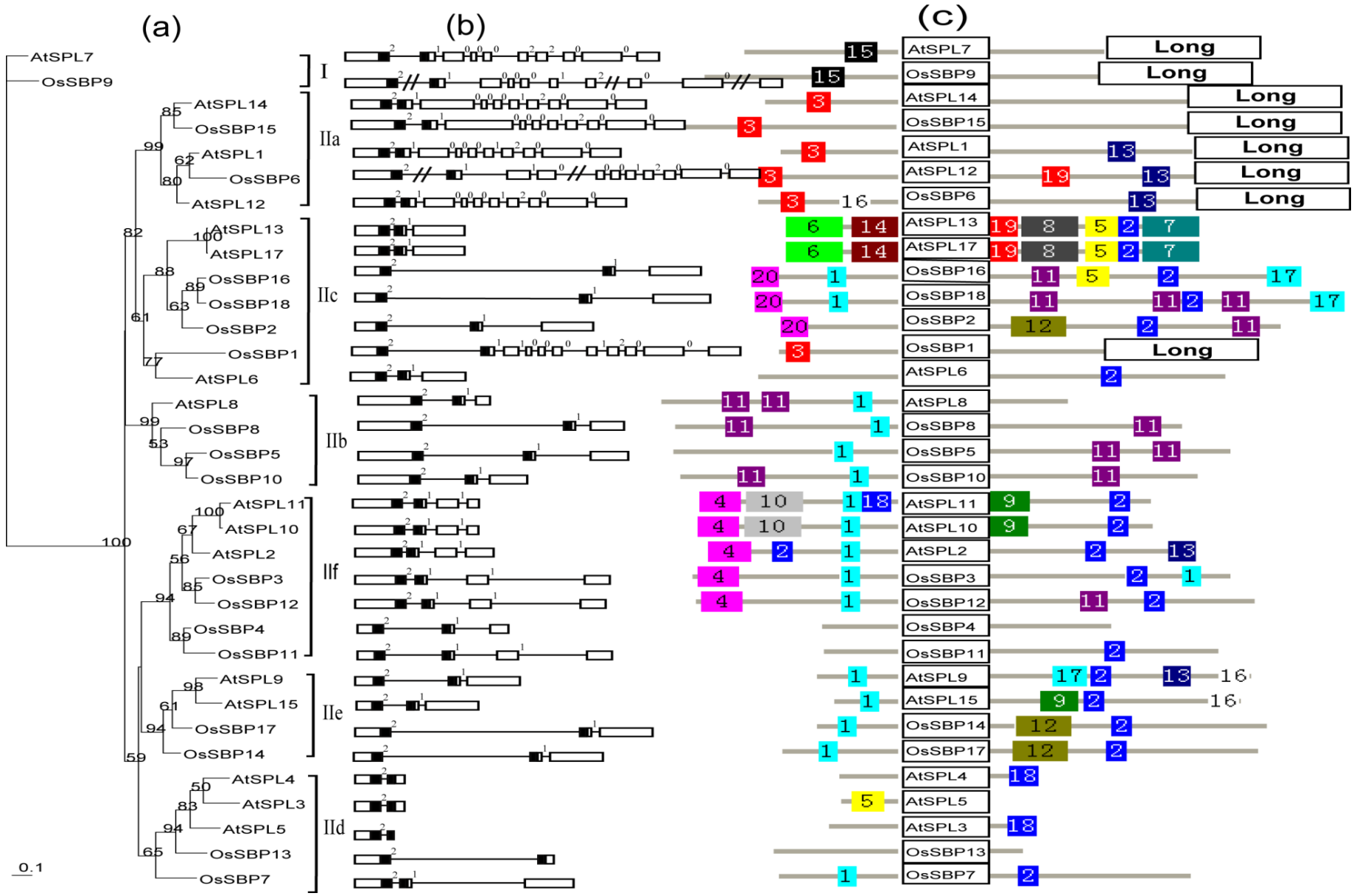
Wang *et al.* (2005) *Nature*, 436:714

Manning *et al* (2006) Nature Genetics, 38:948

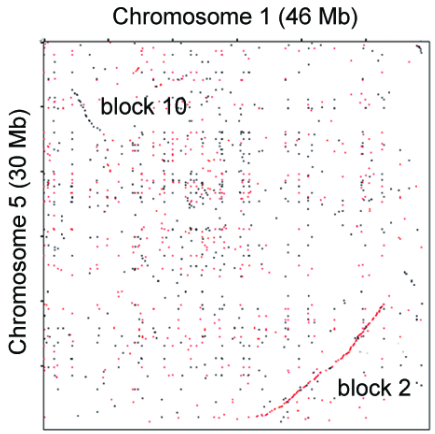
Kinishi *et al* (2006) Science, 312:1392
Li *et al* (2006) *Science*, 312:1936

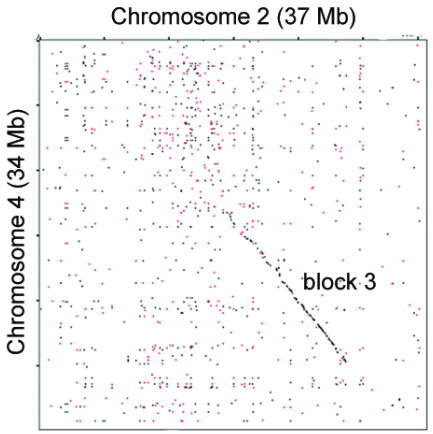
Jiao *et al* (2010) Nature Genetics, 42:541
Miura et al (2010) Nature Genetics, 42:545

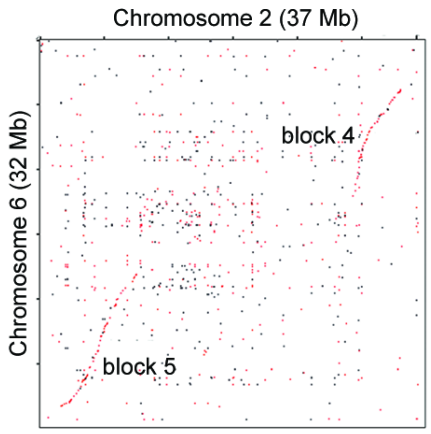
拟南芥和水稻SBP转录因子分析

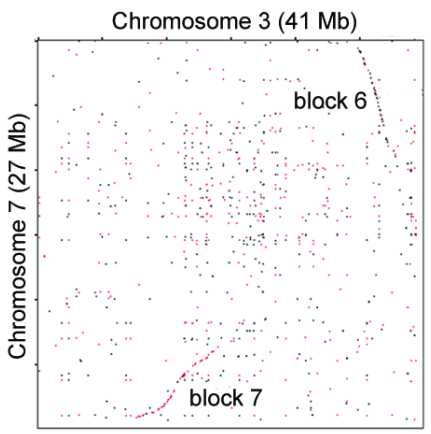


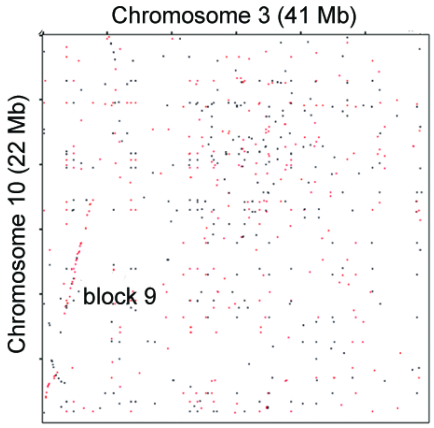
水稻基因组重复序列

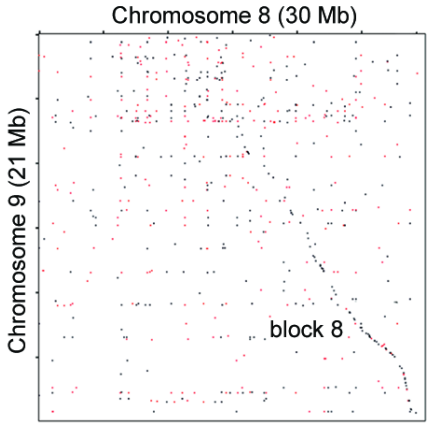


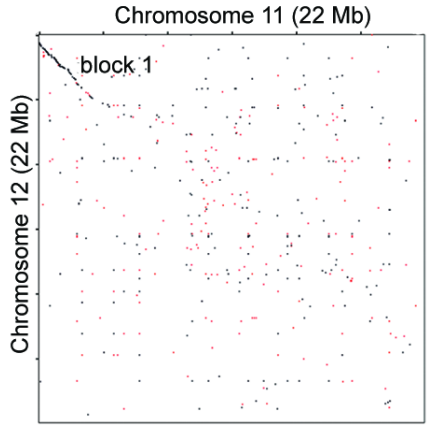


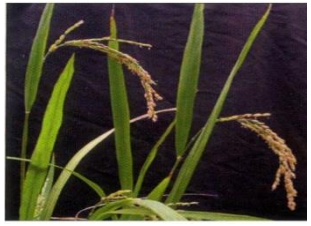












- 最大同源基因对间隔 500 Kb
- 共找到377个重复片段
- 76% 基因分布在重复区
- 最大重复片段含296个基因
- 最大重复片段长度为14M

英国帝国理工大学生化实验室大楼

New biological lab of Imperial College at London

Do your experiment, at least once a year!

- Sydney Brenner

Half day on the Web, save you half month in the lab!

- Alan Bleasby

英国生物信息学家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