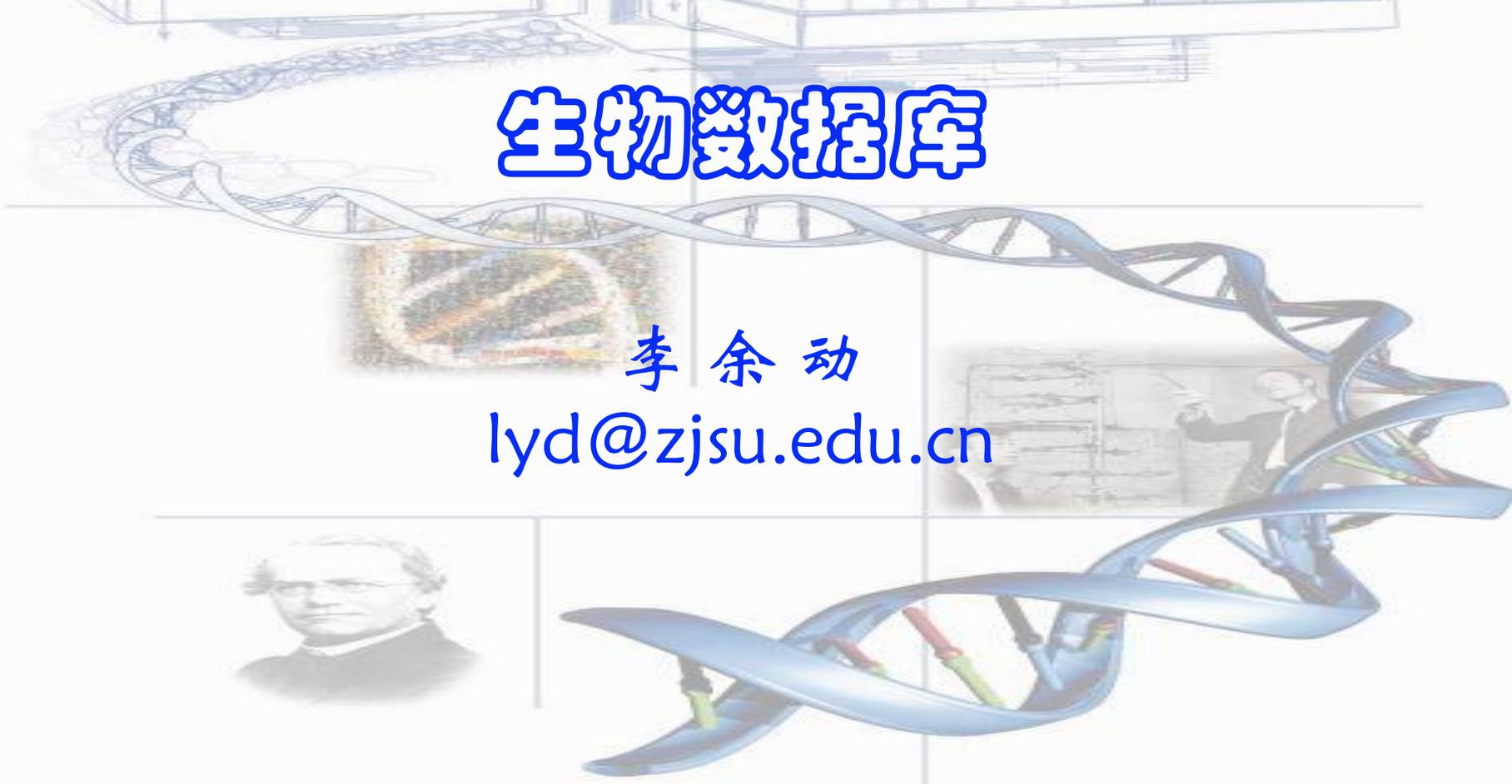


生物数据库



李余动

lyd@zjsu.edu.cn



内容提要 (Outline)

生物数据库简介

NCBI数据库: GenBank, RefSeq, ...

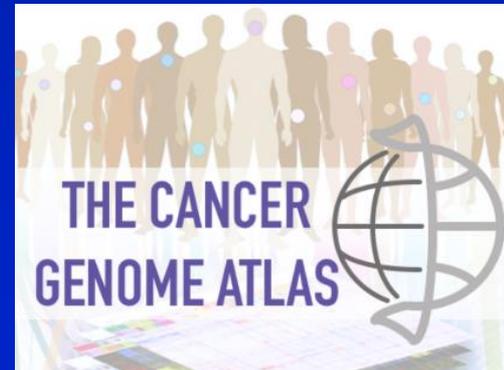
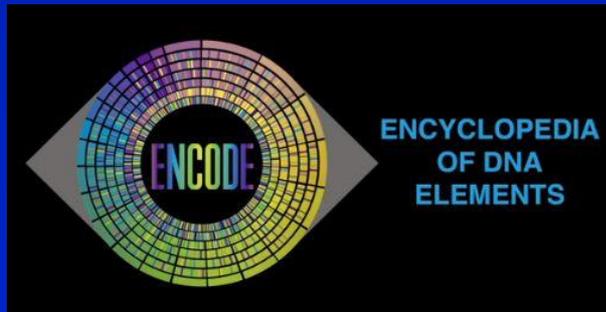
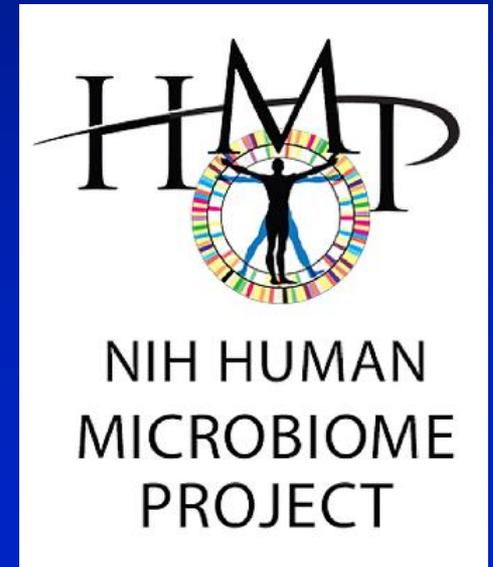
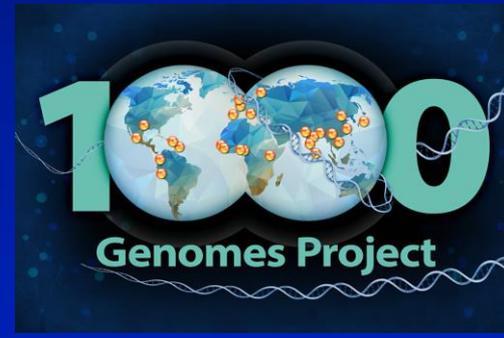
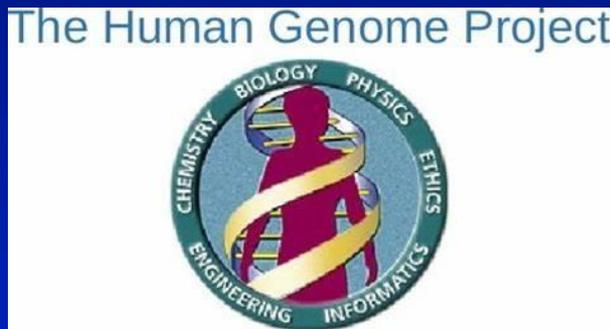
Entrez查询系统

常用数据格式: FASTA, GenBank

如何查找文献 (PubMed)?

生命科学进入大数据时代

- 测序技术快速发展产生海量的序列数据
- 为有效利用这些数据，大量生物数据库已经被开发用来存储、分析与维护这些数据



基因组计划

生物数据库

- 核酸序列数据库
 - GenBank、ENA等
- 蛋白质序列与结构数据库
 - UniProt、PDB等
- 基因组数据库
 - Ensembl、UCSC
- 其它专用数据库
 - GEO、KEGG、PubMed等。

NAR Database Summary Paper Category List

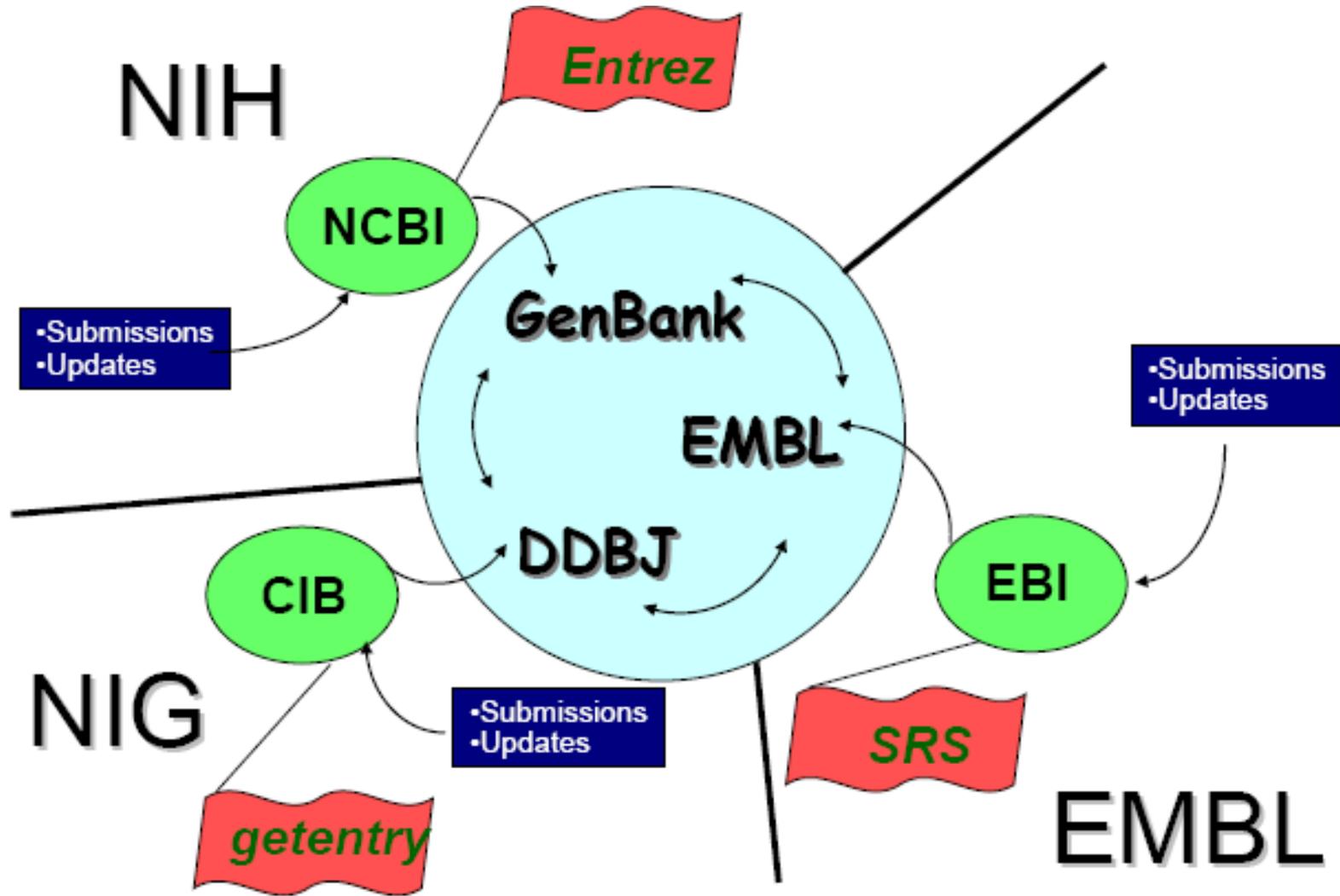
- Nucleotide Sequence Databases
 - RNA sequence databases
- Protein sequence databases
- Structure Databases
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression Databases
- Proteomics Resources
- Other Molecular Biology Databases
- Organelle databases
- Plant databases
- Immunological databases
- Cell biology

<https://www.oxfordjournals.org/nar/database/c/>

核酸数据库

- The Nucleotide Sequence Database
 - GenBank (美国核酸数据库)
 - EMBL(欧洲核苷酸档案库)
 - DDBJ (日本DNA数据库)
- 三大核酸数据库中的数据基本一致，仅
在数据格式上有所差别，对于特定的查
询，三个数据库的查询结果一样。

There are three major public DNA databases



The underlying raw DNA sequences are identical

National Center for Biotechnology Information (NCBI)

NCBI创建于1988年，主要任务：

1.开发数据库；

2.开发基因数据分析工具；

3.发布生物医学信息等

National Institutes of Health

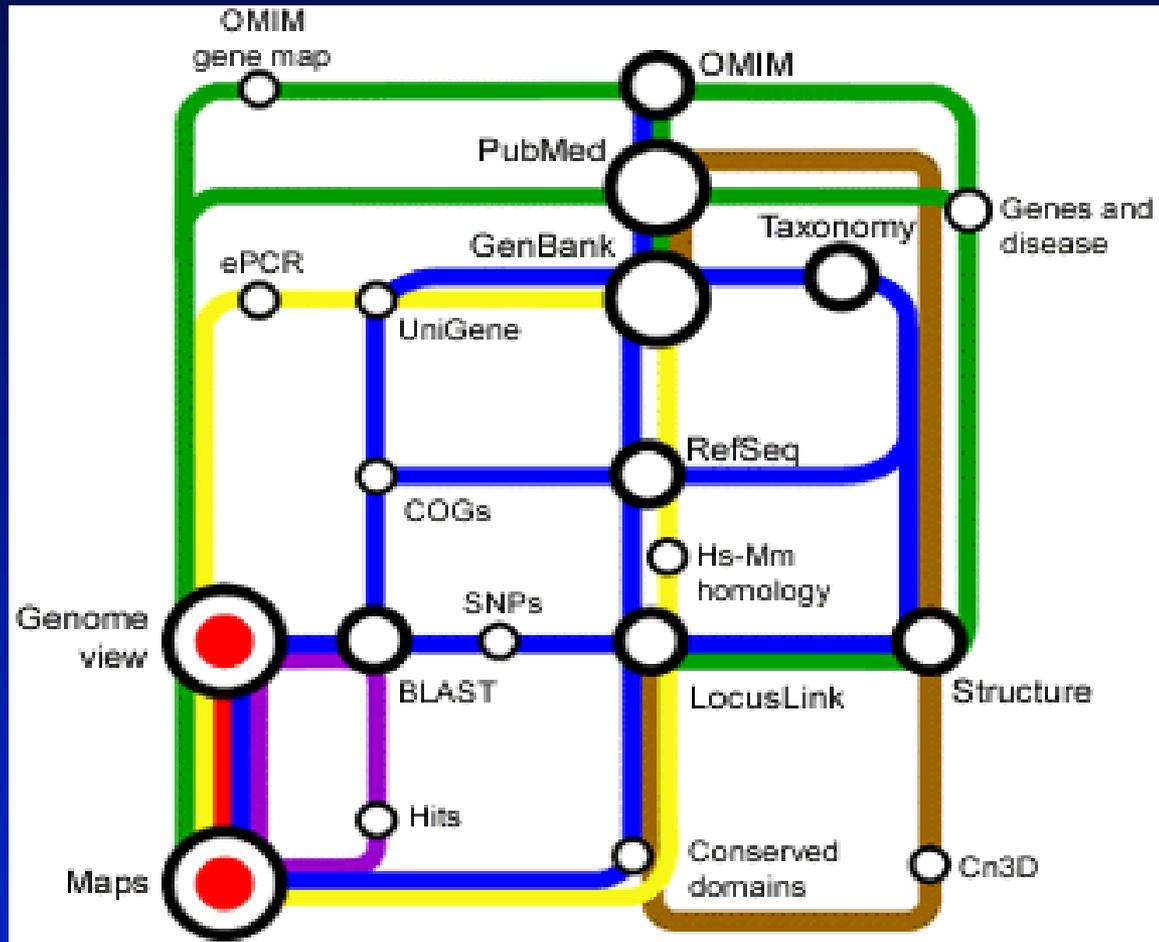
National Library of Medicine

www.ncbi.nlm.nih.gov

The screenshot shows the NCBI website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' menus, and a search bar. The main content area is divided into several sections:

- Resources:** A vertical menu on the left lists various categories such as 'NCBI Home', 'All Resources (A-Z)', 'Literature', 'DNA & RNA', 'Proteins', 'Sequence Analysis', 'Genes & Expression', 'Genomes', 'Maps & Markers', 'Domains & Structures', 'Genetics & Medicine', 'Taxonomy', 'Data & Software', 'Training & Tutorials', 'Homology', 'Small Molecules', and 'Variation'.
- PubMed Central:** A featured banner with the text: 'Free Full Text. Over 1,500,000 articles from over 450 journals. Linked to PubMed and fully searchable.'
- Notice: Upcoming Systems Maintenance:** A yellow banner dated 10 Nov 2009, stating: 'NCBI servers will undergo maintenance beginning November 13 at 3:00 PM and lasting until November 14 at 8:00 PM (EST). During this time, NCBI services, including BLAST and the Entrez search and retrieval systems, will be operational but may be intermittently slow. Data submission systems, including Bankit, GenBank FTP submissions, SequinMacroSend, and the SRA, Trace and GEO submission systems, will be unavailable. Please contact NCBI with concerns: info@ncbi.nlm.nih.gov.'
- How To...:** A section with a list of tasks: 'Obtain the full text of an article', 'Retrieve all sequences for an organism or taxon', 'Find a homolog for a gene in another organism', 'Find genes associated with a phenotype or disease', 'Design PCR primers and check them for specificity', 'Find the function of a gene or gene product', and 'Find syntenic regions between the genomes of two organisms'.
- Help Us Improve:** A section with a link to 'Send us feedback on the new home page and guide pages design'.
- NLM/NCBI H1N1 Flu Resources:** A section featuring a 'FLU.GOV' widget with the text 'Know what to do about the flu.' and a 'VISIT FLU.GOV' button. Below the widget is a list of links: 'Newest H1N1 influenza sequences', 'Submit flu sequences to GenBank', 'Latest H1N1 citations in PubMed', 'MedlinePlus (consumer health information)', and 'Enviro-Health links'.
- Popular Resources:** A section with a list of links: 'PubMed', 'PubMed Central', 'Bookshelf', 'BLAST', and 'Gene'.

Tour of NCBI

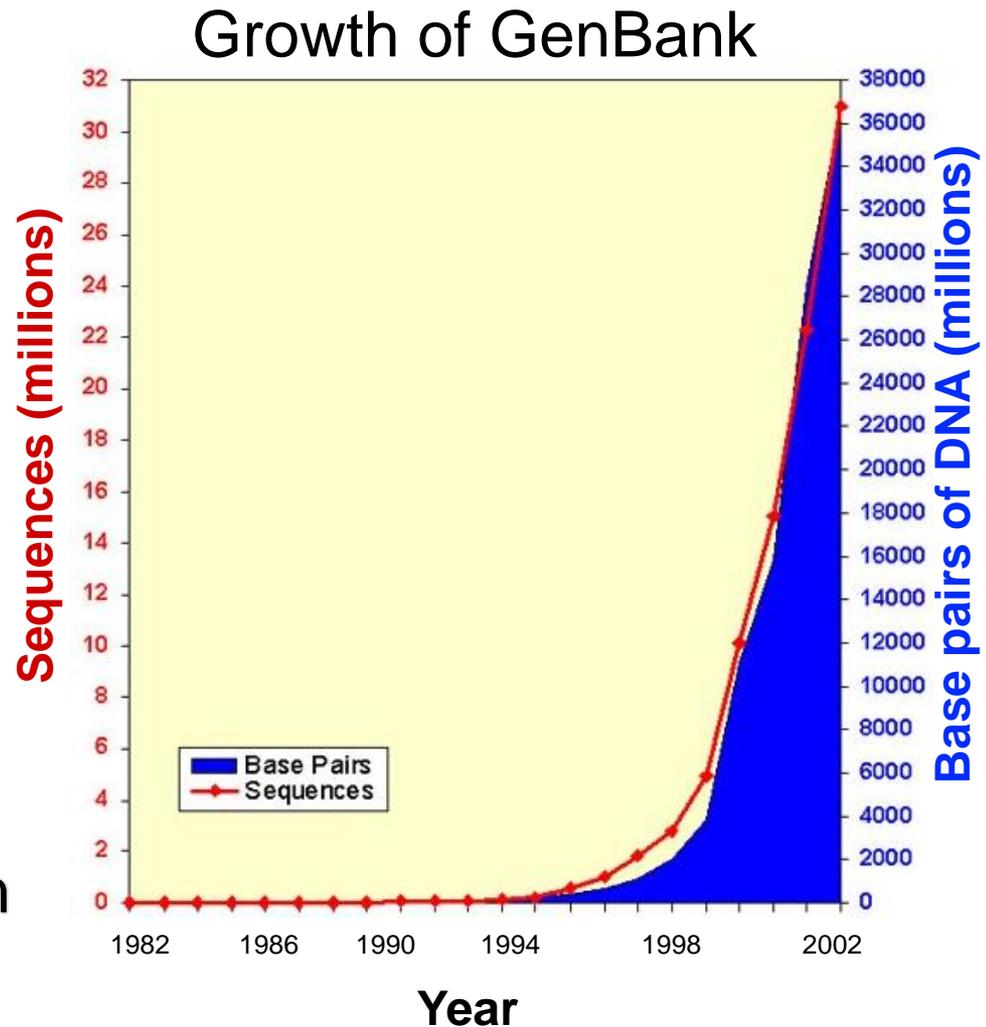


Key	
Green	Literature
Blue	Sequence
Yellow	Mapping
Purple	BLAST
Red	Genome View
Brown	Structure

GenBank: 基因序列数据库

- GenBank数据库是生物信息最基本的数据库
 - 核酸
 - 蛋白质
- GenBank数据库主要包括两个部分：
 - 原始序列(Sequence)
 - 生物学的注释信息

Bankit & SEQUIN submission



新型冠状病毒基因组 (参考菌株Wuhan-Hu-1)

Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

LOCUS NC_045512 29903 bp ss-RNA linear VRL 28-JAN-2020

DEFINITION Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome.

ACCESSION NC_045512

VERSION NC_045512.2

DBLINK BioProject: [PRJNA485481](#)

KEYWORDS RefSeq.

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Nidovirales; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29903)

AUTHORS Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Hu, Y., Song, Z.-G., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C. and Zhang, Y.-Z.

TITLE A novel coronavirus associated with a respiratory disease in Wuhan of Hubei province, China

JOURNAL Unpublished https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2/

RefSeq编号

基因组长度
病毒类型

BLAST分析

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

SARS Coronavirus Resource

Retrieve, view, and download SARS coronavirus genomic and protein sequences.

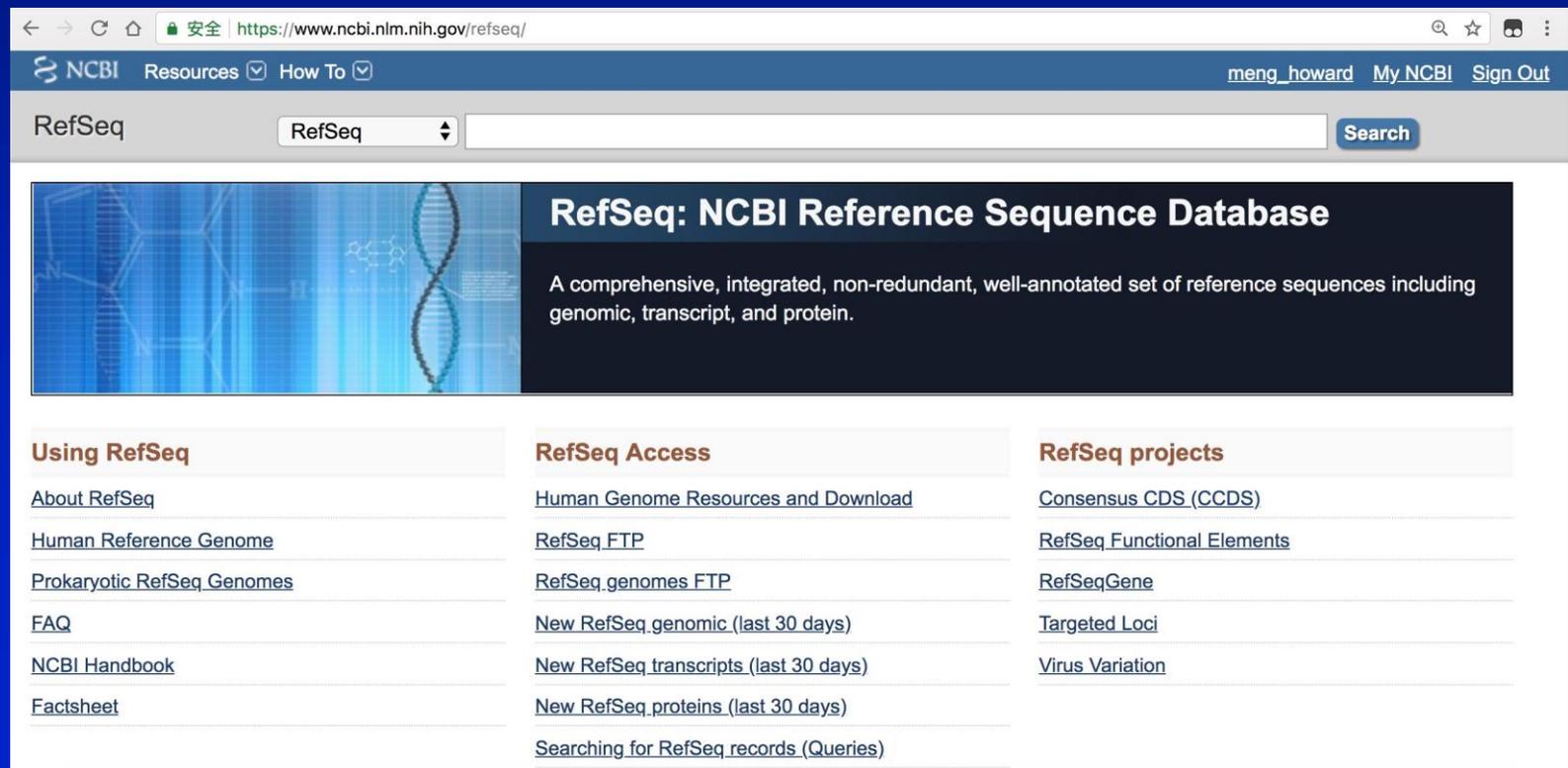
Related information

Assembly

BioProject

RefSeq: Reference Sequence Database

参考序列数据库(RefSeq)提供高质量的，无冗余的，完整的序列信息；包括基因组DNA、转录RNA以及蛋白质序列信息。



The screenshot shows the RefSeq website homepage. At the top, there is a navigation bar with the NCBI logo, "Resources", and "How To" menus. The main header area includes a search bar with "RefSeq" selected in a dropdown menu and a "Search" button. Below the header is a large banner with a blue background featuring a DNA double helix and a network diagram. The banner text reads: "RefSeq: NCBI Reference Sequence Database" and "A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein." Below the banner are three columns of links:

- Using RefSeq**
 - [About RefSeq](#)
 - [Human Reference Genome](#)
 - [Prokaryotic RefSeq Genomes](#)
 - [FAQ](#)
 - [NCBI Handbook](#)
 - [Factsheet](#)
- RefSeq Access**
 - [Human Genome Resources and Download](#)
 - [RefSeq FTP](#)
 - [RefSeq genomes FTP](#)
 - [New RefSeq genomic \(last 30 days\)](#)
 - [New RefSeq transcripts \(last 30 days\)](#)
 - [New RefSeq proteins \(last 30 days\)](#)
 - [Searching for RefSeq records \(Queries\)](#)
- RefSeq projects**
 - [Consensus CDS \(CCDS\)](#)
 - [RefSeq Functional Elements](#)
 - [RefSeqGene](#)
 - [Targeted Loci](#)
 - [Virus Variation](#)

GenBank VS. RefSeq

- RefSeq源于GenBank，因为GenBank综合了他人提交的数据，并未经处理；
- 而 RefSeq则是经过NCBI的努力，整合各方面的结果(GenBank+文献检索+算法预测)。
- RefSeq中的序列信息采用与GenBank一样的格式（GBK）格式。

What is an accession number?

An accession number (检索号) is label that used **to identify a sequence**. It is a string of letters and/or numbers that corresponds to a molecular sequence.

Examples (all for retinol-binding protein, RBP4):

X02775	GenBank genomic DNA sequence	DNA
NT_030059	Genomic contig	
Rs7079946	dbSNP (single nucleotide polymorphism)	
N91759.1	An expressed sequence tag (1 of 170)	RNA
NM_006744	RefSeq DNA sequence (from a transcript)	
NP_007635	RefSeq protein	protein
AAC02945	GenBank protein	
Q28369	SwissProt protein	
1KT7	Protein Data Bank structure record	

NCBI's RefSeq: best representative sequences

RefSeq序列具备独特的特点，便于我们加以识别：RefSeq序列的accession number是由字母前缀+下划线(_) +数字组成的，下划线是refseq序列独一无二的特征，并且RefSeq序列还会有NCBI成员提供的comment信息。

RefSeq identifiers include the following formats:

Complete genome	NC_#####
Complete chromosome	NC_#####
Genomic contig	NT_#####
mRNA (DNA format)	NM_##### e.g. NM_006744
Protein	NP_##### e.g. NP_006735
Protein(predicted)	XP_#####

Types of Databases

原始数据

加工后数据

Primary databases

(一级数据库)

- Raw nucleotide & amino acid data is stored.
- Annotated database info

GenBank

Secondary databases

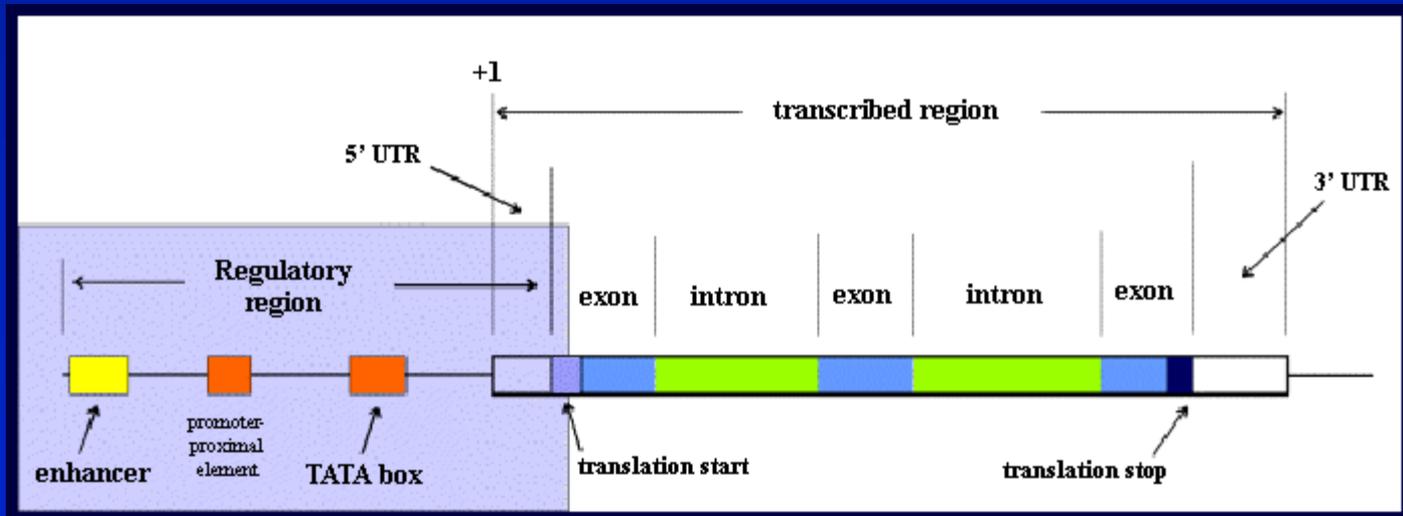
(二级数据库)

- Select and combine data from different primary databases

RefSeq

NCBI Gene

- <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>
- 序列从Refseq数据库中得到;
- 详尽的注释信息, 包括基因在基因组的定位, 基因名称、蛋白质名称, 基因结构等。



如何查找连续的mRNA、cDNA序列？

- Human Insulin (INS)

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. [Explain](#)

Genomic

NG_007114.1 RefSeqGene

Range	4986..6416
Download	GenBank , FASTA , Sequence Viewer (Graphics)

mRNA and Protein(s)

NM_000207.2 → NP_000198.1 insulin preproprotein

[See identical proteins and their annotated locations for NP_000198.1](#)

Status: REVIEWED

Description	Transcript Variant: This variant (1) represents the shortest variant. All variants encode the same protein.
Source sequence(s)	BC005255_BM510748
Consensus CDS	CCDS7729.1
UniProtKB/TrEMBL	I3WAC9
UniProtKB/Swiss-Prot	P01308

Conserved Domains (1) [summary](#)

cd04367	IIGF_insulin_like; IIGF_like family, insulin_like subgroup, specific to vertebrates. Members include a number of peptides including insulin and insulin-like growth factors I and II, which play a variety of roles in controlling processes such as metabolism, growth and ...
Location:26 → 110	

内容提要 (Outline)

NCBI数据库: GenBank, RefSeq, ...

Entrez查询系统

常用数据格式: FASTA, GenBank

如何查找文献 (PubMed)?

Entrez is a search and retrieval system that integrates NCBI databases

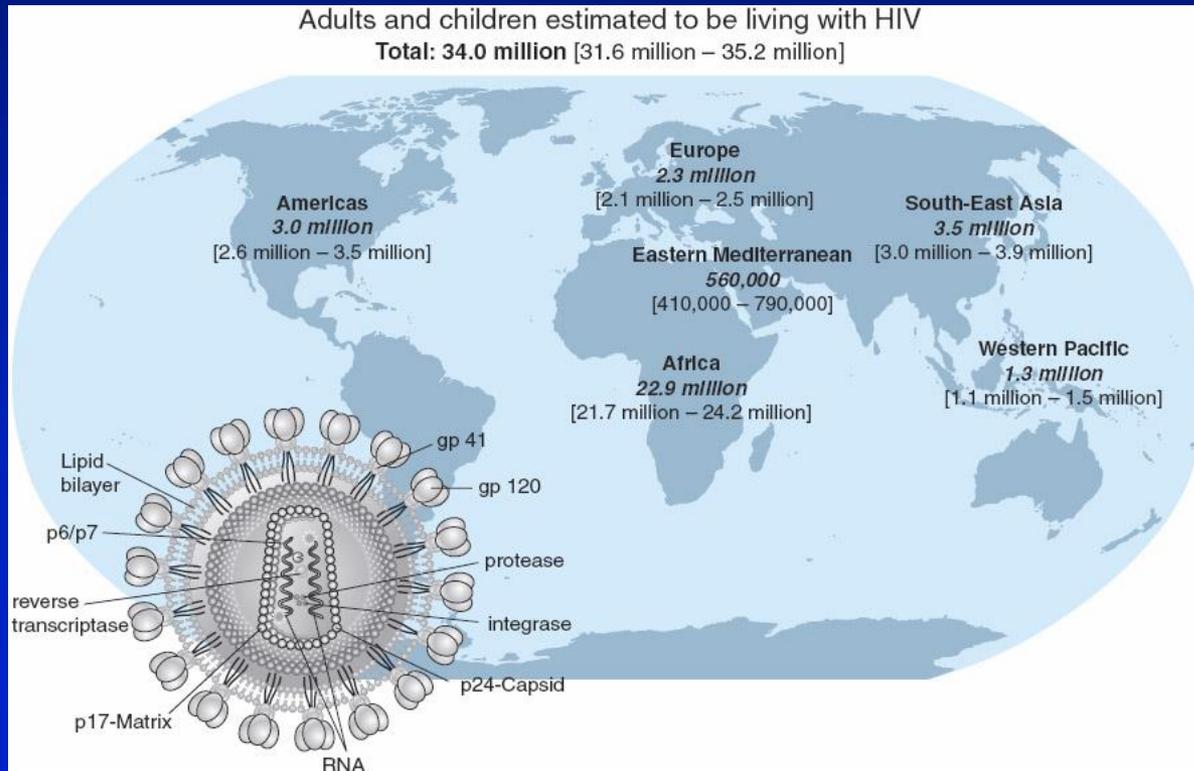


Entrez integrates...

- the scientific literature;
- DNA and protein sequence databases;
- 3D protein structure data;
- population study data sets;
- assemblies of complete genomes

查询序列的例子: HIV-1 *pol*

在NCBI的主页，通过Entrez查询: hiv-1 pol



查询 HIV-1 pol: 结果有>130,000 条目

NCBI  Entrez, The Life Sciences Search Engine

SEARCH SITE MAP PubMed All Databases Human Genome GenBank Map Viewer BL

Search across databases [Help](#)

 - Result counts displayed in gray indicate one or more terms not found

2178  PubMed: biomedical literature citations and abstracts	65  Books: online books
4367  PubMed Central: free, full text journal articles	7  OMIM: online Mendelian Inheritance in Man
13  Site Search: NCBI web and FTP sites	none  OMIA: online Mendelian Inheritance in Animals

138108  Nucleotide: Core subset of nucleotide sequence records	none  dbGaP: genotype and phenotype
none  EST: Expressed Sequence Tag records	none  UniGene: gene-oriented clusters of transcript sequences
none  GSS: Genome Survey Sequence records	1  CDD: conserved protein domain database
132690  Protein: sequence database	1492  3D Domains: domains from Entrez Structure
36  Genome: whole genome sequences	2  UniSTS: markers and mapping data
311  Structure: three-dimensional macromolecular structures	1182  PopSet: population study data sets
none  Taxonomy: organisms in GenBank	16  GEO Profiles: expression and molecular abundance profiles
none  SNP: single nucleotide polymorphism	none  GEO DataSets: experimental sets of GEO data
246  Gene: gene-centered information	none  Cancer Chromosomes: cytogenetic databases
none  SRA: Short Read Archive	11  PubChem BioAssay: bioactivity screens of chemical substances
25  BioSystems: Pathways and systems of interacting molecules	none  PubChem Compound: unique small molecule chemical structures
1  HomoloGene: eukaryotic homology groups	175  PubChem Substance: deposited chemical substance records

减少查询结果的方法

- 由于直接查询 hiv-1 pol得到大量核苷酸或蛋白质序列 (如果只查“hiv-1”会得到>300,000更多记录)
- 但可以通过以下两步减少命中记录:
 - 指定某个物种(organism), e.g. hiv-1[organism]
 - 限制(limit)查询结果只为数据库(Genome/RefSeq)的条目!

The screenshot shows the NCBI Entrez Genome search interface. At the top, there is a navigation bar with 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', 'Structure', 'OMIM', 'PMC', 'Journals', and 'Books'. The search bar contains 'Genome' in the dropdown and 'hiv-1[organism]' in the input field. Below the search bar are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The 'Display' dropdown is set to 'Summary', 'Show' is set to '20', and there is a 'Send to' dropdown. A 'All: 1' button is visible. The search results show a single entry: '1: NC_001802 Human immunodeficiency virus 1, complete genome'. The entry details include 'ssRNA; linear; Length: 9,181 nt', 'Replicon Type: viral genome', and 'Created: 1998/01/22'. A 'Recent activity' box at the bottom right shows 'Your browsing activity is empty.'

在“genome”库中查询 HIV-1 pol: 结果为 1 条目

NCBI Resources How To

Genome Genome hiv-1 pol
[Save search](#) [Limits](#) [Advanced](#)

[Display Settings](#) Overview [Send to](#)

[Organism Overview](#) ; [Genome Project Report](#) ; [Genome Annotation Report](#)

Human immunodeficiency virus 1

Human immunodeficiency virus 1 overview

Lineage: [Viruses\[3735\]](#); [Retro-transcribing viruses\[118\]](#); [Retroviridae\[55\]](#); [Orthoretrovirinae\[46\]](#); [Lentivirus\[11\]](#); [Primate lentivirus group\[5\]](#); [Human immunodeficiency virus 1\[1\]](#)

Genome Sequencing Projects

Chromosomes [1] Scaffolds or contigs [0] SRA or Traces [1]

Organism	BioProject	Status	Chrs	Size (Kb)	GC%	Gene	Protein	Nbrs
Human immunodeficiency virus 1	PRJNA15478	●	1	9.18	42.1	9	9	1,803
Human immunodeficiency virus 1	PRJNA198138	●	-	-	-	-	-	-

Genome Region

[Go to nucleotide](#) [Graphics](#) [FASTA](#) [GenBank](#)

Tools

1. [Genome Resource](#)

Publications

1. A genotypic test for HIV-1 tropism combining Sanger sequencing with ultradeep sequencing predicts virologic response in treatment-experienced patients. Kagan RM, et al. PLoS One 2012
2. Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. Martoglio B, et al. EMBO J 1997 Nov 17

内容提要 (Outline)

NCBI数据库: GenBank, RefSeq, ...

Entrez查询系统

常用数据格式: FASTA, GenBank

如何查找文献 (PubMed)?

FASTA format:

versatile, compact with **one header line** followed by a string of nucleotides or amino acids in the single letter code

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure

Search Protein for Go

Display FASTA Show 5 Send to

Range: from begin to end Refresh

1: [NP_000509](#). Reports beta globin [Homo...[gi:4504349]

```
>gi|4504349|ref|NP_000509.1| beta globin [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVVMGNPKVKAHGKKVLG
AFSDGLAHL DNLKGT FATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAAYQKVVAGVAN
ALAHKYH
```

- **标题行**，以“>”开头，记录序列的相关注释信息，如序列名称(ID)，序列来源等；
- **序列行**，从标题行的下一行开始，记录序列的具体字符串，字符必须为标准IUB/IUPAC核苷酸或氨基酸符号。

GenBank文件格式—Header

NCBI Entrez Protein

My NCBI
[\[Sign In\]](#) [\[Register\]](#)

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for

Limits Preview/Index History Clipboard Details

Display GenF **Accession number** Range: from CDD

1: [NP_477382](#). Reports Bub1 CG7838-PA [Drosophila melanogaster] [gi:17137586] [BLink](#), [Conserved Domains](#), [Links](#)

[Comment](#) [Features](#) [Sequence](#)

LOCUS NP_477382 1460 aa linear INV 12-OCT-2006

DEFINITION Bub1 CG7838-PA [Drosophila melanogaster].

ACCESSION NP_477382

VERSION NP_477382.1 GI:17137586

DBSOURCE REFSEQ: accession [NM_058034.3](#)

KEYWORDS .

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM [Drosophila melanogaster](#)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (residues 1 to 1460)

AUTHORS [Quattrone, A.](#), [Bergman, C.M.](#), and [Anxolabehere, D.](#)

TITLE [Conservation annotation of the Drosophila melanogaster genome](#)

JOURNAL *PLoS Comput. Biol.* 1 (2), e22 (2005)

PUBMED [16110336](#)

REFERENCE 2 (residues 1 to 1460)

AUTHORS [Hoshino, D.](#), [Smith, C.D.](#), [Gardner, J.W.](#), and [Gambetta, A.P.](#)

序列长度

数据类型

Definition: 标题

版本号

GI number

GenBank的数据类型

1. PRI - primate sequences
2. ROD - rodent sequences
3. MAM - other mammalian sequences
4. VRT - other vertebrate sequences
5. INV - invertebrate sequences
6. PLN - plant, fungal, and algal sequences
7. BCT - bacterial sequences
8. VRL - viral sequences
9. PHG - bacteriophage sequences
10. SYN - synthetic sequences
11. UNA - unannotated sequences
12. EST - EST sequences (expressed sequence tags)
13. PAT - patent sequences
14. STS - STS sequences (sequence tagged sites)
15. GSS - GSS sequences (genome survey sequences)
16. HTG - HTGS sequences (high throughput genomic sequences)
17. HTC - HTC sequences (high throughput cDNA sequences)
18. ENV - Environmental sampling sequences
19. CON - Constructed sequences

GenBank格式—FEATURES

特性关键词

source

CDS

mRNA

gene

序列位置与限定词

序列位置

限定词

GenBank格式—FEATURES

序列来源生物的简称
(或分子类型)

蛋白编码序列

基因

整个核苷酸序列

FEATURES	Location/Qualifiers
source	1..5028 /organism="Saccharomyces cerevisiae" /mol_type="genomic DNA" /db_xref="taxon:4932" /chromosome="IX"
<u>mRNA</u>	<1..>206 /product="TCP1-beta"
<u>CDS</u>	<1..206 /codon_start=3 /product="TCP1-beta" /protein_id="AAA98665.1" /translation="SSIYNGISTSGLDLNNGTIADMRQLGIVESYKLRVSSSEA AEVLLRVDNIIRARPRPTANRQHM"
<u>gene</u>	<687..>3158 /gene="AXL2"
<u>mRNA</u>	<687..>3158 /gene="AXL2"
<u>CDS</u>	<687..>3158 /gene="AXL2" /note="plasma membrane glycoprotein" /codon_start=1 /product="Axl2p" /protein_id="AAA98666.1" /translation="MTQLQISLLLTATISLLHLVWATPYEAVPIGKQYPPVARVNESF TFQISNDTYKSSVDKTAQITYNCFDLPSWLSFDSSSRFTSGEPSSDLLSDANTTLVFN LTSDDRLNLTISQTELESTIPGSLP"
ORIGIN	1 gatcctccat atacaacggt atctccacct caggtttaga tctcaacaac ggaaccattg 61 ccgacatgag acagtttagt atcgtcgaga gttacaagct aaaacgagca gtagtcagct 121 ctgcatctga agccgctgaa gttctactaa gggtggataa catcatccgt gcaagaccaa

特性关键词——source

Source通常包含序列来源生物的简称，有些时候也包含分子类型。第一个限定词organism表示物种分类，描述限定词的内容放在等号后面的引号中（如= “Saccharomyces cerevisiae”）。

```
FEATURES          Location/Qualifiers
  source          1..5028
                  /organism="Saccharomyces cerevisiae"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:4932"
                  /chromosome="IX"
```

特性关键词——CDS

- CDS是一个重要的基因特性，表示**蛋白编码序列**。
- 当CDS表示一部分蛋白质序列时，核苷酸数字范围将会以“<”开头或以“>”结尾。
 - 由符号“<”开头表示在该符号之前的蛋白质序列并不在此CDS序列中，如第一个CDS范围是<1..206，
 - 同样末尾的“>”符号表示CDS序列不包括蛋白质序列末尾部分。

CDS

```
<1..206  
/codon_start=3  
/product="TCP1-beta"  
/protein_id="AAA98665.1"  
/translation="SSIYNGISTSGLDLNNGTIADMRQLGIVESYKLRVAVSSASEA  
AEVLLRVDNIIIRARPRTANRQHM"
```

特性关键词——gene

- 图中基因“AXL2”位置在687..3158，即在此核苷酸序列中的687-3158区域。
- 另一个基因“REV7”位置在complement（3300..4037），在数字范围前的“complement”表示这段DNA序列的编码链在互补链（complementary strand）上，即编码核苷酸序列是ORIGIN下面显示序列的反向互补链序列，注意方向是反向的。

REV7		AXL2	
4035	4037	687	689
5'	3'	5'	3'
cat	(显示的序列)	atg	(显示的序列)
3'	5'	3'	5'
gta	(反向互补序列)	tac	(反向互补序列)
Met	(编码氨基酸)	Met	(编码氨基酸)

内容提要 (Outline)

NCBI数据库: GenBank, RefSeq, ...

Entrez查询系统

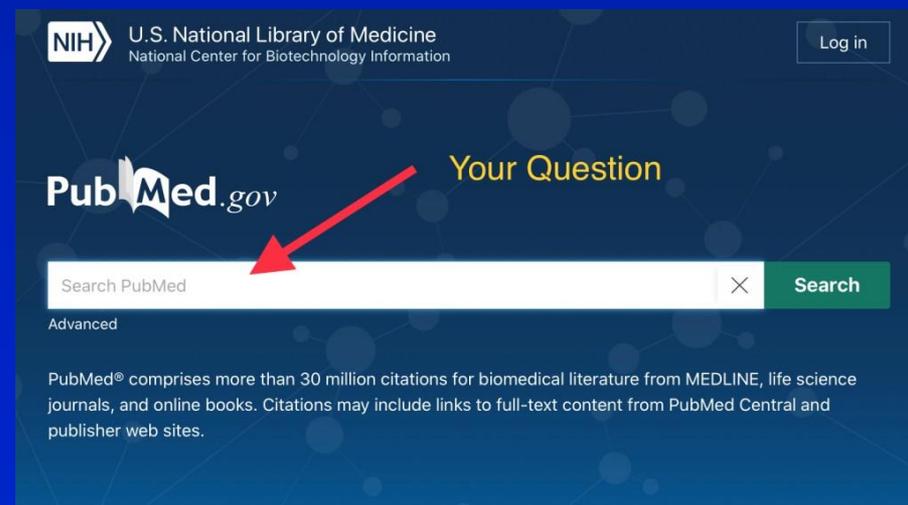
常用数据格式: FASTA, GenBank

如何查找文献 (PubMed)?

Pubmed简介

- PubMed 是NCBI的MEDLINE数据库门户。
- MEDLINE是当前国际上最权威的生物医学文献数据库，收录来自世界发行的超过4,600个杂志的**文献摘要**，包括 >19 million记录，最早可以追溯到1950s.
- PubMed Central (PMC)是论文**全文**数据库。根据NIH Public Access Policy（2008年执行），获得NIH资助的项目发表论文必须在12个月内上传到PMC。

通过PubMed查找文献



使用 “Advanced search” 可限制不同条件 如作者(author), 年份(year), etc.

NCBI Resources How To My NCBI | Sign In

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed
hbb

RSS Save search **Advanced search** Help

Search Clear



Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to:

Are you looking for gene information?

HBB hemoglobin, beta [Homo sapiens]

▶ **hbb** in [Homo sapiens](#) | [Mus musculus](#) | [Rattus norvegicus](#) | [All 20 Gene records](#)

Filter your results:

All (440)

[Review \(10\)](#)

[Free Full Text \(114\)](#)

[Manage Filters](#)

Results: 1 to 20 of 440

<< First < Prev Page 1 Next > Last >>

- [Control of fetal hemoglobin: new insights emerging from genomics and clinical implications.](#)

Thein SL, Menzel S, Lathrop M, Garner C.
Hum Mol Genet. 2009 Oct 15;18(R2):R216-23.
PMID: 19808799 [PubMed - in process]
[Related articles](#)

- [Multiplex ligation-dependent probe amplification screening of isolated increased HbF levels revealed three cases of novel rearrangements/deletions in the beta-globin gene cluster.](#)

Lee ST, Yoo EH, Kim JY, Kim JW, Ki CS.
Br J Haematol. 2009 Oct 5. [Epub ahead of print]
PMID: 19807730 [PubMed - as supplied by publisher]
[Related articles](#)

Also try:

- ▶ **hbb** gene
- ▶ **hbb** mutation

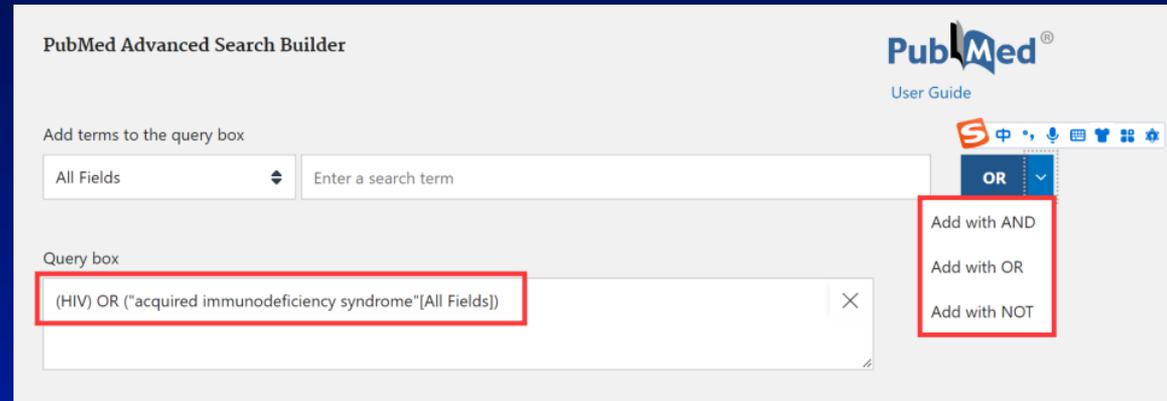
Titles with your search terms

- ▶ Comparison of the mismatch-specific endonuclease mett [BMC Biotechnol. 2008]
- ▶ Role of hyoscine N-butyl bromide (**HBB**, buscopan) as labo [Indian J Med Sci. 2008]
- ▶ [Family screening for **HBB***S gene and detection of new [Rev Saude Publica. 2008]

» See more...

高级查询(Advanced Search)

组合检索



- 使用布尔(boolean) 运算符查询 (大写AND, OR, NOT)

SARS-CoV-2 NOT SARS-CoV-1

默认AND:

genome thornton science

PubMed 查询技巧

1. Searching Pubmed using fields:

如要跟踪Nature的一篇文章，就搜“Nature[journal]”，同理，如对肿瘤感兴趣的话，可试试“Tumor[Keyword]”，等等
例如：

genome[TIAB] thornton[AU] science[TA]

- TIAB: Title/Abstract
- AU: Author
- TA: Journal
- DP: Publication Date
- AD: Affiliation

<https://pubmed.ncbi.nlm.nih.gov/help/>

2.结果过滤

可利用结果页面左边的过滤选项可快速筛选文献：

- Article types->Review, 只看综述文献；
- Publication dates->5 years, 近五年文献等。

The screenshot displays a search results interface with several filter sections on the left and a list of search results on the right.

RESULTS BY YEAR: A bar chart showing the number of results per year from 1945 to 2024. The number of results increases significantly starting around 2000.

TEXT AVAILABILITY:

- Abstract
- Free full text
- Full text

ARTICLE ATTRIBUTE:

- Associated data

ARTICLE TYPE:

- Books and Documents
- Clinical Trial
- Meta-Analysis
- Randomized Controlled Trial
- Review
- Systematic Review

PUBLICATION DATE:

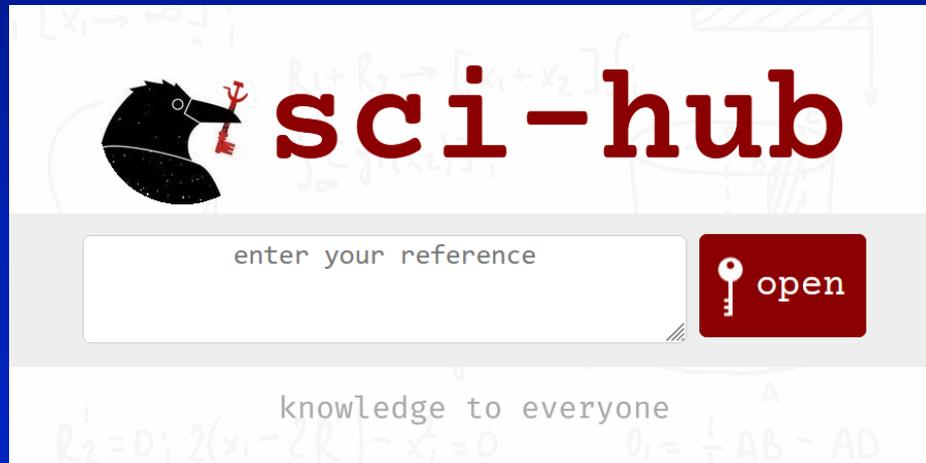
- 1 year
- 5 years

Search Results:

- Human immunodeficiency virus and acquired immunodeficiency syndrome: correlation but not causation.**
1
Cite: Duesberg PH.
Share: Proc Natl Acad Sci U S A. 1989 Feb;86(3):755-64. doi: 10.1073/pnas.86.3.755.
PMID: 2644642 [Free PMC article.](#) [Review.](#)
AIDS is an **acquired immunodeficiency syndrome** defined by a severe depletion of T cells and over 20 conventional degenerative and neoplastic diseases. ... (viii) It is highly improbable that within the last few years two viruses (**HIV-1** and **HIV-2**) ...
- The discovery of HIV as the cause of AIDS.**
2
Cite: Gallo RC, Montagnier L.
Share: N Engl J Med. 2003 Dec 11;349(24):2283-5. doi: 10.1056/NEJMp038194.
PMID: 14668451 [Free article.](#) [No abstract available.](#)
- Antiretroviral therapy.**
3
Cite: Thiers BH.
Share: J Am Acad Dermatol. 1989 Sep;21(3 Pt 1):443-54. doi: 10.1016/s0190-9622(89)70206-1.
PMID: 2550525 [Review.](#)
Research into the biology of human **immunodeficiency virus type 1 (HIV-1)**, the causative agent of the **acquired immunodeficiency syndrome**, has yielded valuable information about the replicative cycle of the virus. ... Although many drugs have been ...
- [AIDS: acquired immunodeficiency syndrome].**
4
Cite: Sener B, Ustaçelebi S.
Share: Mikrobiyol Bul. 1989 Oct;23(4):379-87.
PMID: 2488947 [Review.](#) [Turkish.](#)
In this article the molecular biology of **HIV** (Human **Immunodeficiency Virus**) which is the causative agent of **acquired immune deficiency syndrome** and the pathogenesis, epidemiology and clinical picture of the **syndrome**, the current therapeutic appr ...
- Strengths and weaknesses in methods for identifying the causative agent(s) of**

3. 获得全文方法

- 不是免费的文献可以通过图书馆或SCI-HUB 下载pdf文件:



<http://sci-hub.shop>

作业

- 2020年初，我国科学家首先完成新冠病毒基因组测序，对全球防疫作出重大贡献。
 - 在NCBI核酸数据库上查找SARS-CoV-2参考病毒株(Wuhan-Hu-1)的基因组序列，并观察病毒基因注释信息；
 - 在PubMed检索最早报道Wuhan-Hu-1基因组的Nature论文(2020)。