



苏州大学

一步一步教你使用 NCBI

查找 DNA、mRNA、cDNA、Protein、promoter、引物设计、BLAST 序列比对等

作者: urbest

2007-8-1

苏州大学生命科学学院

最近看到很多战友在论坛上询问如何查询基因序列、如何进行引物设计、如何使用 BLAST 进行序列比对……，这些问题在 NCBI 上都可以方便的找到答案。现在我就结合我自己使用 NCBI 的一些经历（经验）跟大家交流一下 NCBI 的使用。希望大家都能发表自己的使用心得，让我们共同进步！

我分以下几个部分说一下 NCBI 的使用：

Part one 如何查找基因序列、mRNA、Promoter

Part two 如何查找连续的 mRNA、cDNA、蛋白序列

Part three 运用 STS 查找已经公布的引物序列

Part four 如何运用 BLAST 进行序列比对、检验引物特异性

特别感谢本版版主，将这个帖子置顶！

从发帖到现在，很多战友对该帖给与了积极的关注，在此向给我投票的（以及想给我投票却暂时不能投票的）各位战友表示真诚的感谢，谢谢各位战友！

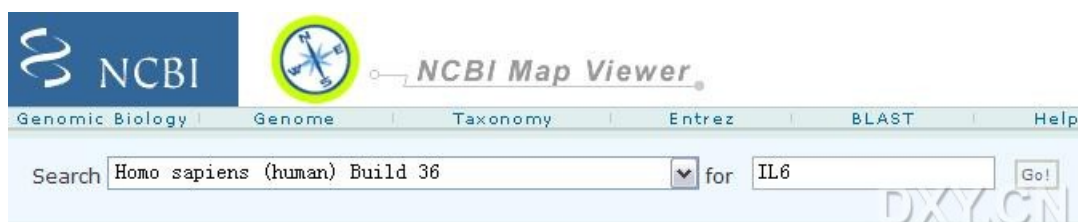
请大家对以下我发表的内容提出自己的意见。关于 NCBI 其他方面的使用也请水平较高的战友给予补充

First of all，还是让我们从查找基因序列开始。

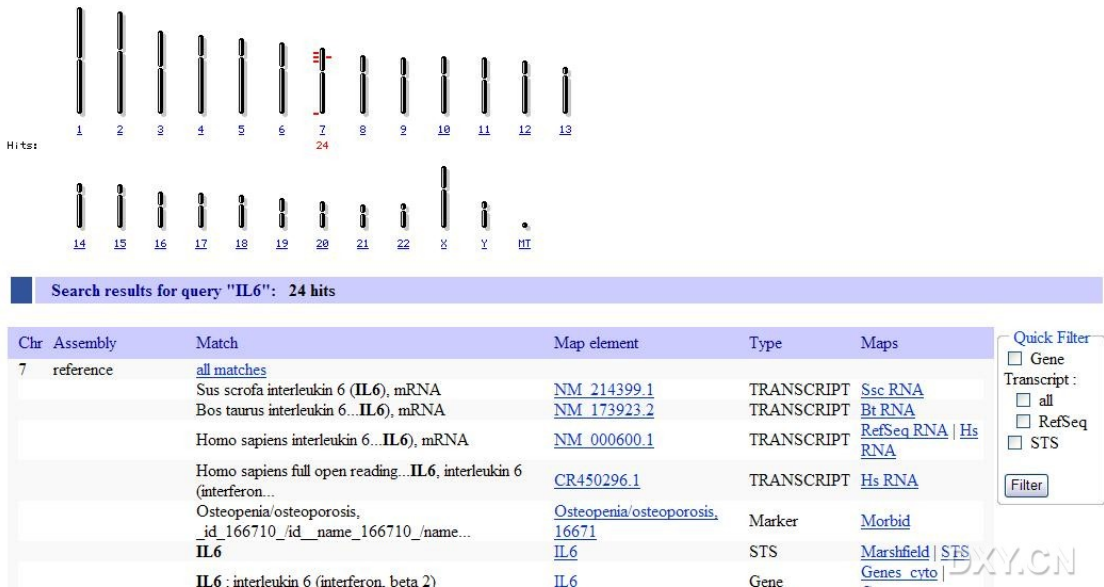
第一部分 利用 Map viewer 查找基因序列、mRNA 序列、启动子（Promoter）

下面以人的 IL6（白细胞介素 6）为例讲述一下具体的操作步骤

1. 打开 Map viewer 页面，网址为：<http://www.ncbi.nlm.nih.gov/mapview/index.html>
在 search 的下拉菜单里选择物种，for 后面填写你的目的基因。操作完毕如图所示：

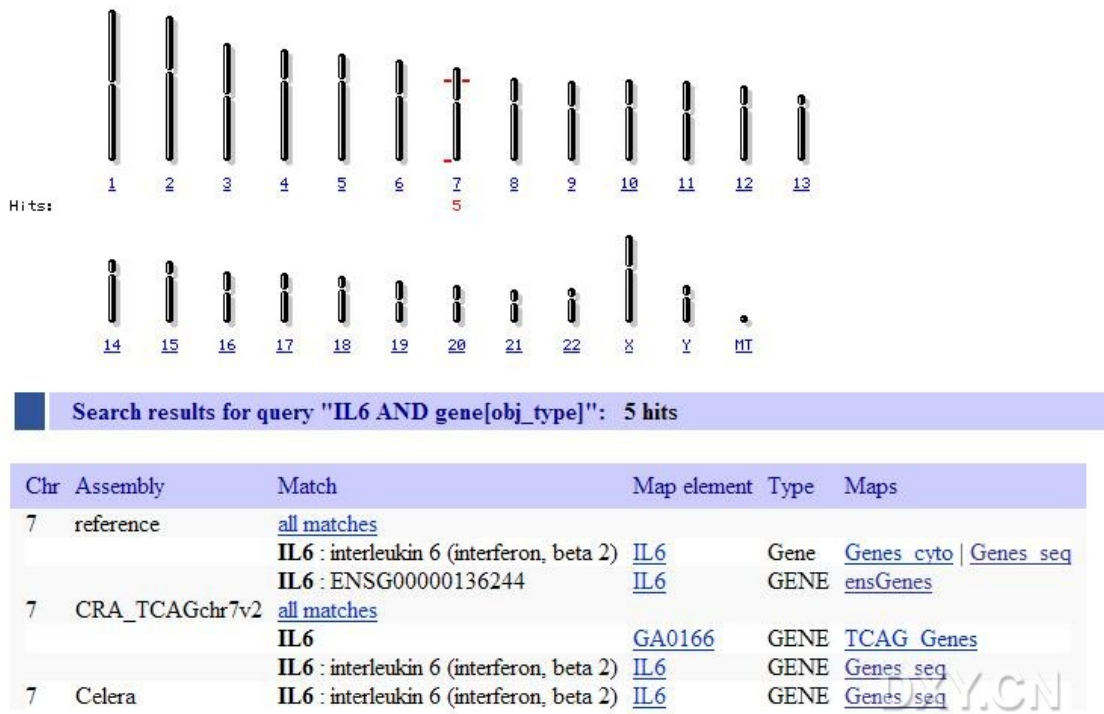


2. 点击“GO”出现如下页面：



3. 在步骤二图示的右下角有一个 Quick Filter, 下面是让你选择的几个复选框, 在 Gene 前面的小方框里打勾, 然后点击 Filter. 出现下图:

说明一下: 1、染色体的红色区域即为你的目的基因所处位置。2、下面参考序列给出了三个, 是不同的部门做出来的, 经我验证, 序列有微小的差异, 但总体来说基本相同。尽管你分别点击后, 序列代码、序列代码等有所差异, 但碱基基本一致, 不影响大家研究分析序列。现在普遍采用的是最上面的那个序列, 这一条是世界范围的生物科学家用计算机合成的一个序列。我也推荐大家使用这个序列。



4. 点击上述三条序列第一条序列（即 reference）对应的“Genes seq”，出现新的页面，页面下方为：

Summary of Maps:

Map 1: Genes On Sequence

[Table View](#)

Region Displayed: 22,729,700-22,741,700 bp [Download/View Sequence/Evidence](#)

Total Genes On Chromosome: 1452 [\[11 not localized\]](#)

Genes Labeled: 1 Total Genes in Region: 1

DXY.CN

5. 点击上图出现的“Download/View Sequence/Evidence”，即下载查看序列等功能，结果如图所示：

[Homo sapiens](#) (Build 36.2)

Region to retrieve (in chromosome coordinates):

Chromosome: Strand:

from: adjust by:

to: adjust by:

Sequence Format:

This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand	
NT_007819.16	22252181	22264171	+	Display Save to Disk View Evidence ModelMaker

先对上面这张图做点简要的说明，在 Sequence Format（序列输出格式）后面是一个下拉式选择菜单，默认的为 FASTA 格式，还有一个是 GenBank 格式。我推荐大家选择 GenBank 格式，因为这个格式提供了很多该基因的信息，而 FASTA 格式只有基因序列。

6. 在 Sequence Format 后选择 GenBank，然后点击下面的 Display，目的基因的相关信息和序列就出现在眼前了。点击后如图所示（网页较大，只抓取一小部分以作示范）：

1: NT_007819. Reports Homo sapiens chro...[gi:89024890]					
Comment Features Sequence					
LOCUS	NT_007819	11991 bp	DNA	linear	CON 29-AUG-2006
DEFINITION	Homo sapiens chromosome 7 genomic contig, reference assembly.				
ACCESSION	NT_007819 REGION: 22252181..22264171				
VERSION	NT_007819.16 GI:89024890				
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (sites)				
AUTHORS	Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A and Enright AJ.				
TITLE	miRBase: microRNA sequences, targets and gene nomenclature				
JOURNAL	Nucleic Acids Res. 34 (DATABASE ISSUE), D140-D144 (2006)				
PUBMED	16381832				
REFERENCE	2 (bases 1 to 11991)				
AUTHORS	International Human Genome Sequencing Consortium.				
TITLE	Finishing the euchromatic sequence of the human genome				
JOURNAL	Nature 431 (7011), 931-945 (2004)				
PUBMED	15496913				
REFERENCE	3 (sites)				
AUTHORS	Griffiths-Jones S.				
TITLE	The microRNA Registry				
JOURNAL	Nucleic Acids Res. 32 (DATABASE ISSUE), D109-D111 (2004)				
PUBMED	14681370				
COMMENT	GENOME ANNOTATION REFSEQ : Features on this sequence have been produced for build 36 version 2 of the NCBI's genome annotation [see documentation]. On or before Mar 1, 2006 this sequence version replaced				

在上述打开的网页中，你可以看到基因长度，基因序列，以及这个基因是如何被报道出来的等各种信息。

你会看到：[mRNA join\(3598..3678, 3841..4031, 5090..5203, 5911..6057, 7803..8394\)](#) 这代表了从基因的 3598 位开始就是转录区了，即我们常说的 mRNA 片断，由于内含子的存在，所以 mRNA 在 DNA 序列上分成了几段。

[CDS join\(3660..3678, 3841..4031, 5090..5203, 5911..6057, 7803..7970\)](#)

CDS 代表编码序列，即蛋白编码区是从 3660 开始的 (ATG)，由于剪接作用所以 CDS 区也是不连续的。

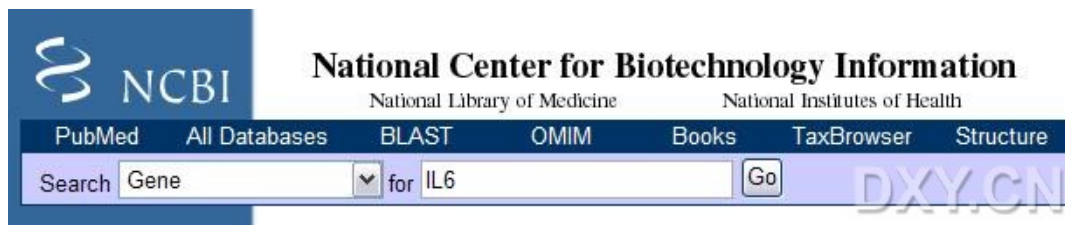
说到这里，可能很多朋友都已经明白了 promoter 即启动子区域在哪里了。但我还是再唠叨几句：转录起始位点前面是基因的调控区，启动子区没有明显的位置定义，大家也只是猜测它的大体位置，如果你要研究 promoter 区的话，建议你选择转录起始位点前的 2000 个碱基进行研究，一般默认的是这样。当然你如果觉得长度太长不好研究的话，也可以只研究-1000 到 0 这一千个碱基，因为一般情况下，启动子区的变异都在这个区域内。

这样大家就可以找到自己的目的基因序列和启动子了，这种方法可能使用的人不是很多，但我个人比较喜欢，因为它最大的优点是可以找到启动子区域和其他调控区域。希望大家可以发帖交流，让我们把 NCBI 用的更好！

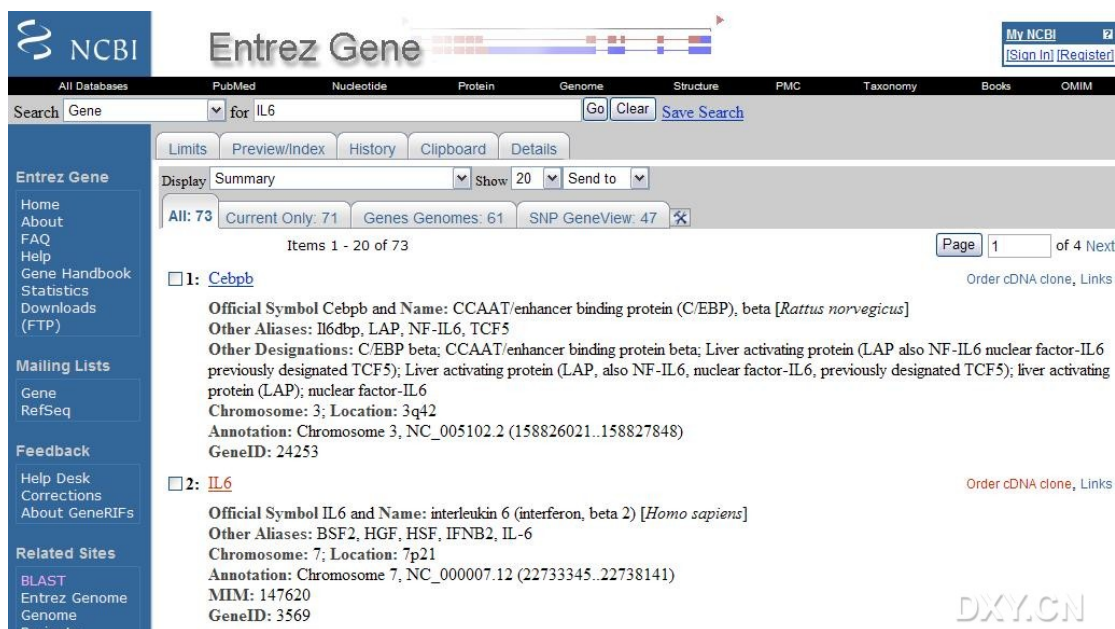
第二部分 如何查找连续的 mRNA、cDNA、蛋白序列（依然以人类的 IL6 为例）

1. 进入 NCBI 主页: <http://www.ncbi.nlm.nih.gov/>

在 search 后面选择 Gene，在 for 后面填写需要查找的基因的名字。如图所示：



点击“Go”，出现以下界面：



出现了很多基因序列，在每个序列的右边还有“Order cDNA clone”的链接，这些序列中有些序列是跟你的目的基因同名的，有些是别名（Other Aliases）与你的目的基因一致，根据每个序列的介绍认真选择你的目的基因。上图中我需要的 IL6 是标号为 2 的序列。

2.1 查找 cDNA 序列

2.1.1 点击 Order cDNA clone，出现目的页面如图所示：

NCBI ? Clone Registry ? Order Clones

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy

Search All Databases [Go]

Order cDNA Clones for Homo sapiens gene IL6
IL6: interleukin 6 (interferon, beta 2)

The following clone(s) can be purchased through any of the IMAGE distributors. Some of them have provided dire

Clone: **MGC:9215 (IMAGE:3884652)**
 Clone Sequence: [BC015511.1](#)
 Vector: pCMV-SPORT6
 Corresponding RefSeq mRNA: [NM_000600.1](#)

from American Type Culture Collection
 from RZPD German Resource Center for Genome Research
 from Geneservice Ltd
 from Harvard Institute of Proteomics
 from Open Biosystems

DX.Y.CN

2.1.2 点击 Clone Sequence 后面的链接即可得到 cDNA 序列。点击后如图所示（只抓取其中一部分）：

NCBI Nucleotide

PubMed Nucleotide Protein Genome Structure P

Search Nucleotide for [Go] [Clear]

Limits Preview/Index History Clip

Display GenBank Show 5 Send to Hide: ☐ sequence ☐ all but gene, CDS and mRNA fe

Range: from begin to end ☐ Reverse complemented strand Features: ☐ SNP ☒ MGC

1: BC015511.1 Reports Homo sapiens inte...[gi:15930148]

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

LOCUS BC015511 1127 bp mRNA linear PRI 15-JUL-2006

DEFINITION Homo sapiens interleukin 6 (interferon, beta 2), mRNA (cDNA clone MGC:9215 IMAGE:3884652), complete cds.

ACCESSION BC015511

VERSION BC015511.1 GI:15930148

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
 Catarrhini; Hominidae; Homo.

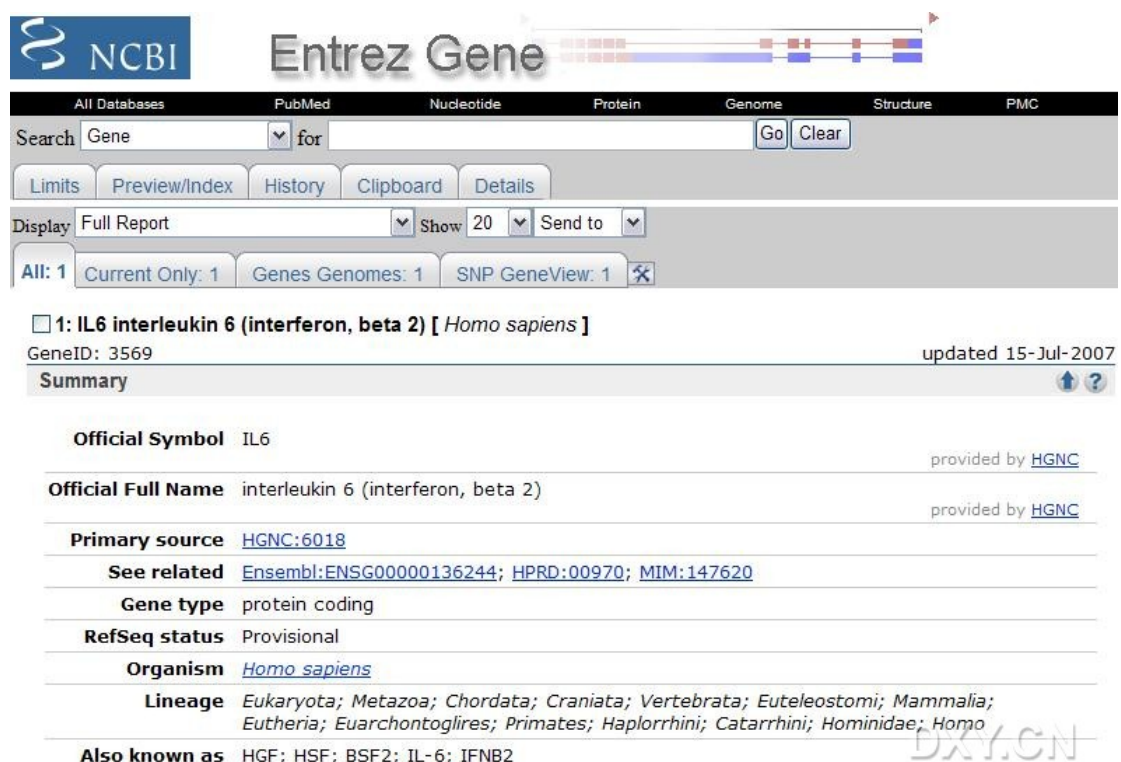
REFERENCE 1 (bases 1 to 1127)

DX.Y.CN

2.2 查找 mRNA、蛋白序列

回到步骤 1 点击“Go”之后出现的页面，点击目的基因的名字，出现以下页面（只抓取

相关部分):



Entrez Gene

Search Gene for [Go] [Clear]

Display Full Report Show 20 Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

☐ 1: IL6 interleukin 6 (interferon, beta 2) [Homo sapiens]

GeneID: 3569 updated 15-Jul-2007

Summary

Official Symbol IL6 provided by [HGNC](#)

Official Full Name interleukin 6 (interferon, beta 2) provided by [HGNC](#)

Primary source [HGNC:6018](#)

See related [Ensembl:ENSG00000136244](#); [HPRD:00970](#); [MIM:147620](#)

Gene type protein coding

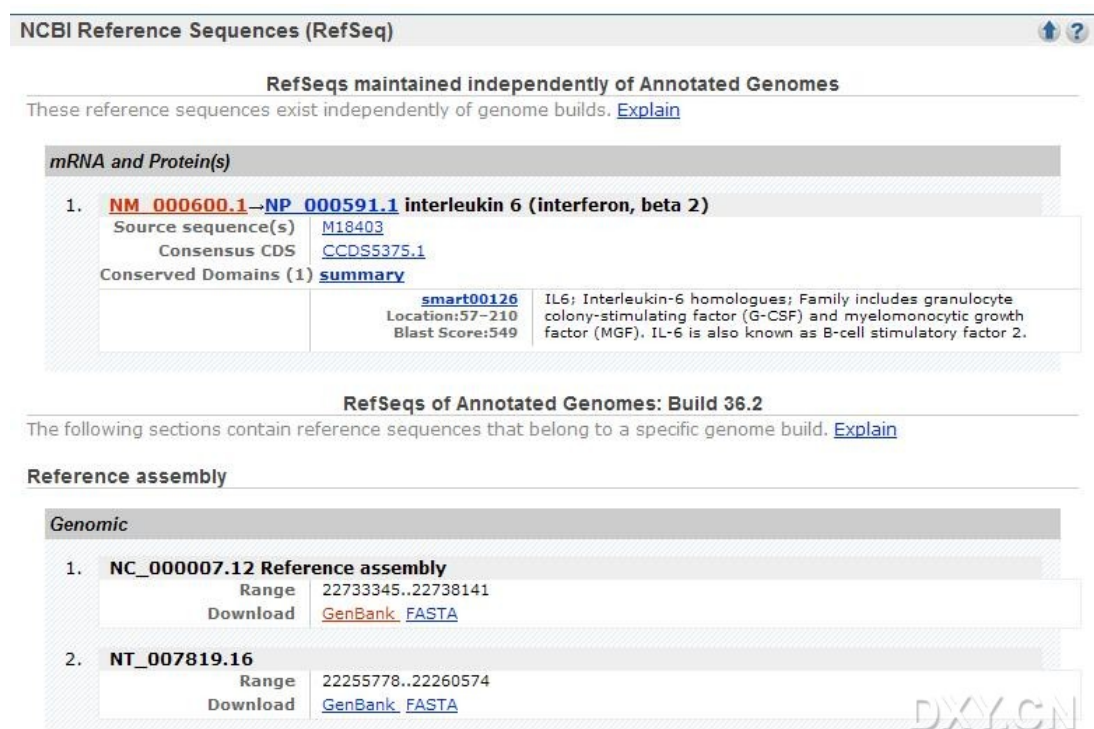
RefSeq status Provisional

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as HGF; HSF; BSF2; IL-6; IFNB2

页面的下半部分，即可以获取 mRNA 和蛋白序列的部分：



NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

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

mRNA and Protein(s)

1. **NM_000600.1--NP_000591.1 interleukin 6 (interferon, beta 2)**

Source sequence(s) [M18403](#)

Consensus CDS [CCDS5375.1](#)

Conserved Domains (1) [summary](#)

[smart00126](#) Location:57-210 Blast Score:549

IL6; Interleukin-6 homologues; Family includes granulocyte colony-stimulating factor (G-CSF) and myelomonocytic growth factor (MGF). IL-6 is also known as B-cell stimulatory factor 2.

RefSeqs of Annotated Genomes: Build 36.2

The following sections contain reference sequences that belong to a specific genome build. [Explain](#)

Reference assembly

Genomic

1. **NC_000007.12 Reference assembly**

Range 22733345..22738141

Download [GenBank](#) [FASTA](#)

2. **NT_007819.16**

Range 22255778..22260574

Download [GenBank](#) [FASTA](#)

找到“NCBI Reference Sequences (RefSeq)”，它分为几个板块，第一个“mRNA and Protein”区可以让我们找到连续的编码 mRNA 序列和蛋白序列。在 mRNA and Protein

下面有两个序列代码（中间划有一个箭头），这代表了 mRNA 序列和蛋白序列。分别点击就可以得到相应的序列页面。点击后如图所示，mRNA 序列：

```

ORIGIN
      1 ttctgcccctc gagcccaccg ggaacgaaag agaagctcta tctcgccctcc aggagcccag
     61 ctatgaactc cttctccaca agcgcccttcg gtccagttgc cttctccctg gggctgctcc
    121 tgggtgttgc tgctgccttc cctgccccag taccgccagg agaagattcc aaagatgtag
    181 ccgccccaca cagacagcca ctcacctctt cagaacgaat tgacaaacaa attcgggtaca
    241 tctctgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa
    301 gcagcaaaga ggcaactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg
    361 gatgcttcca atctggattc aatgaggaga cttgcctggt gaaaatcatc actggtcttt
    421 tggagtttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag
    481 ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaaga
    541 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc
    601 aggcacagaa ccagtggctg caggacatga caactcatct cattctgctg agctttaagg
    661 agttcctgca gtccagcctg agggctcttc ggcaaagtga gcatgggcac ctcagattgt
    721 tgttgttaat gggcattcct tcttctggtc agaaacctgt ccaactggga cagaacttat
    781 gttgttctct atggagaact aaaagtatga gcgttaggac actattttta ttatttttaa
    841 tttattaata tttaaatatg tgaagctgag ttaatttatg taagtcatat ttatattttt
    901 aagaagtacc acttgaaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta
    961 tgcagtttga atatcctttg tttcagagcc agatcatttc ttggaaagtg taggcttacc
   1021 tcaaataaat ggctaactta tacatatattt taaagaaata tttatatatt atttatataa
   1081 tgtataaatg gttttttata caataaatgg catttttaaa aattc
//
  
```

蛋白序列如下：

```

ORIGIN
      1 mnsfstsafig pvafslglll vlpaafpapv ppgedskdva aphrqpltss eridkqiryi
     61 ldgisalrke tcnksnmces skealaennl nlpkmaekdg cfqsgfneet clvkiitgll
    121 efeyyleylq nrfesseeqa ravqmqstkv l iqflqkkakn ldaittpdpt tnaslltklq
    181 aqnqwlqdm t hllilrsfke flqsslrall qm
//
  
```

NCBI Reference Sequences (RefSeq) 的第二个板块是 Reference assembly，它下面显示的是 Genomic，点击 Genomic 下面 Reference assembly 对应的 Genbank 或 FASTA 即可出现编码的 DNA 序列（注意：只是编码序列，其中包括内含子，但一般没有 5' 非编码区）。这一步就不做贴图演示了吧，呵呵。

这样我们就可以找到基因的 cDNA 序列、连续的编码 mRNA 序列、蛋白序列以及含有内含子的编码 DNA 序列了。相信这些操作对很多战友还是有用的。

如果大家有更好的方法，欢迎发帖交流！

友情提示：在 NCBI 里打开的每一个页面都会给我们提供大量的信息，大家不妨好好看看，可能会有令我们惊喜的收获！

最后唠叨一句：最近我实验比较忙，只能在深夜发帖，可能要过几天再发第三部分 [Part three 运用 STS 查找已经公布的引物序列]，希望“期待下集”的朋友可以理解。

第三部分 运用 STS 查找已经公布的引物序列

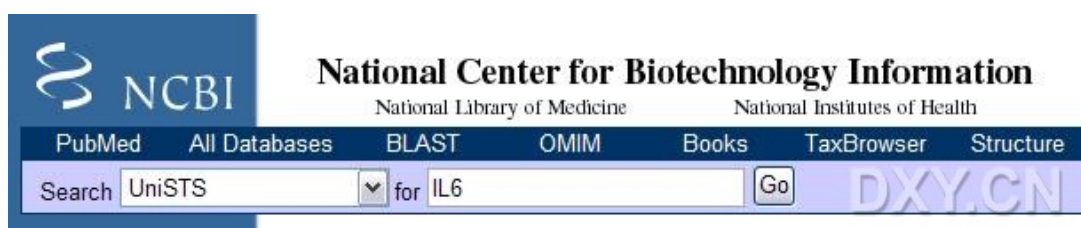
STS, 序列标签位点 (Sequence Tagged Site): 一段短的 DNA 序列 (200—500 个碱基对), 这种序列在染色体上只出现一次, 其位置和碱基顺序都是已知的。在 PCR 反应中可以检测处 STS 来, STS 适宜于作为人类基因组的一种地标, 据此可以判定 DNA 的方向和特定序列的相对位置。

以上内容基本是 STS 的定义, 我主张活学活用, 下面就介绍一下我个人用 STS 数据库查找引物的一点经验。

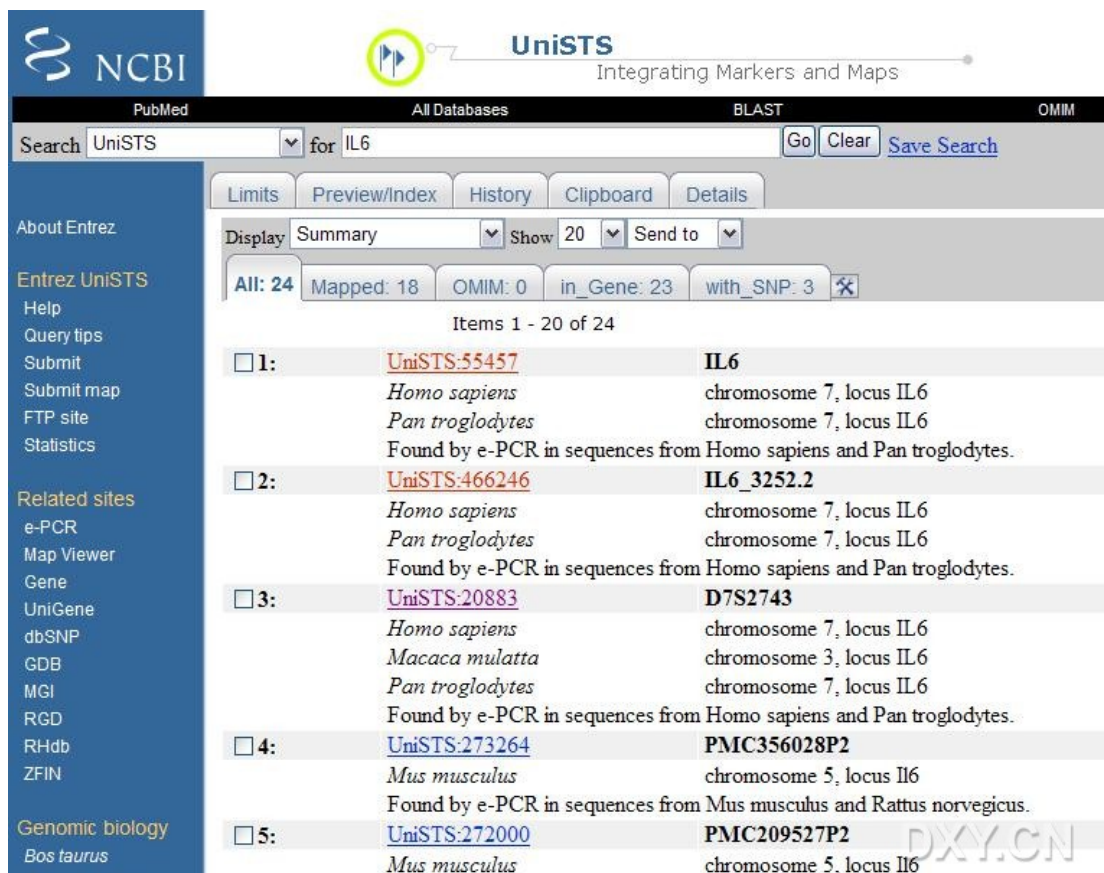
还是使用人的 IL6 基因为例, 呵呵

1. 打开 NCBI 主页, 在 Search 后面的下拉菜单选择 UniSTS, 在 FOR 后面填写目的基因。

操作完毕如图所示:



点击 GO 以后出现以下页面,



这是你会发现 NCBI 又提供了很多序列, 下面我们还是要初步筛选我们需要的序列。

2. 根据物种、目的阴物所在染色体的位置等选择相应序列（可能不只一个），点击。
下面以点击第一个进入的画面为例。

The screenshot displays the NCBI UniSTS database interface. At the top, the NCBI logo and UniSTS title are visible, along with navigation links for PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. A search bar contains 'UniSTS' and a 'Go' button. The left sidebar lists various tools and resources under 'Entrez UniSTS' and 'Related sites'. The main content area shows the UniSTS entry for IL6, including its location on chromosome 7 for both *Homo sapiens* and *Pan troglodytes*, and a note about its discovery via e-PCR. Below this is a 'Primer Information' section with details on forward and reverse primers, PCR product size, and GenBank accession number. Further down, the 'Homo sapiens' section lists the gene name (IL6) and its alternative name (sWSS1282). The 'Cross References' section provides links to GeneID, Symbol, Description, and Position.

NCBI UniSTS Integrating Markers and Maps

PubMed Entrez BLAST OMIM Taxonomy Structure

Search UniSTS for Go

Entrez UniSTS
Help
Query tips
Submit
Submit map
FTP site
Statistics

Related sites
e-PCR
Map Viewer
Gene
UniGene
dbSNP
GeneMap'99
RHdb
GDB
MGD
ZFIN

Genomic biology
Bos taurus
Canis familiaris
Danio rerio

UniSTS:55457 Links

IL6
Homo sapiens chromosome 7, locus IL6
Pan troglodytes chromosome 7, locus IL6
Found by e-PCR in sequences from *Homo sapiens* and *Pan troglodytes*.

Primer Information

Forward primer: GGTGGCAAAAAGGAGTCACA
Reverse primer: TAGAATTTAGCGTTCCAGTT
PCR product size: 247 (bp), *Homo sapiens*
GenBank Accession: G31676

Homo sapiens

Name: IL6
Also known as: sWSS1282

Cross References

Gene GeneID: 3569
Symbol: IL6
Description: interleukin 6 (interferon, beta 2)
Position: 7p21

你会发现这个页面直接就给出了引物序列，PCR 之后的片段长度也是给了的（247bp）。
下面还有很多相关的信息……

3. 点击 GeneBank Accession 后面的代码，进入下一个页面。


```
Primer A: GGTGGCAAAAAGGAGTCACA
Primer B: TAGAATTTAGCGTTCAGTT
STS size: 247
PCR Profile:
    Presoak:      0 degrees C for 0.00 minute(s)
    Denaturation: 92 degrees C for 1.00 minute(s)
    Annealing:    60 degrees C for 2.00 minute(s)
    Polymerization: 72 degrees C for 2.00 minute(s)
    PCR Cycles:   35
    Thermal Cycler: PerkinElmer TC
Protocol:
    Template:     30-100 ng
    Primer:       each 1 uM
    dNTPs:        each 200 uM
    Taq Polymerase: 0.05 units/ul
    Total Vol:    5 ul

Buffer:
    MgCl2:        2.5 mM
    KCl:          50 mM
    Tris-HCl:     10 mM
    pH:           8.3

    This STS has been incorporated into the NHGRI chromosome 7
    physical map, but was developed by another investigator. See
    GenBank record: Y00081 For additional information about the NHGRI
    chromosome 7 mapping project, see
    http://www.nhgri.nih.gov/DIR/CTB/CHR7. Also see Genomics
    11:548-64 (1991) [MUID=92128937].
```

啊！前后引物都呈现在眼前了，还有反应体系和反应条件！其中 Primer A 是前引物序列，Primer B 则是后引物序列，并且给出了他们在 DNA 序列中的位置。有兴趣的朋友可以在序列中找一下，是可以找到的，不过要注意，PCR 是双链扩增，在序列中可以直接找到的是 Primer A 的原序列 和 Primer B 的互补序列。

在步骤二里面我只点开了一个序列，继续打开其他的可能还会有对自己有用的引物，不过这要你自己慢慢发掘了。

这种寻找引物的方法有点投机取巧的味道，实用程度不是很高，但如果这里面恰好有你想 P 的片段的话，恭喜你，这些引物都是很成熟的引物，可以直接拿过来使用了。

如果想寻找引物，大家可以查阅相关论文，已经报道的引物我们为什么不用呢？！既省时间，可靠性又强。

如果这两种方法都不能找到你需要的引物的话，那就自己设计吧，建议使用 Primer 5 和 Oligo。引物设计的详细内容我在这里就不多说了，推荐两个帖子给大家看一下，第一个是本版版主 liuzeyi2002 发起的，内容很丰富，很值得学习，另一个则是我发的。

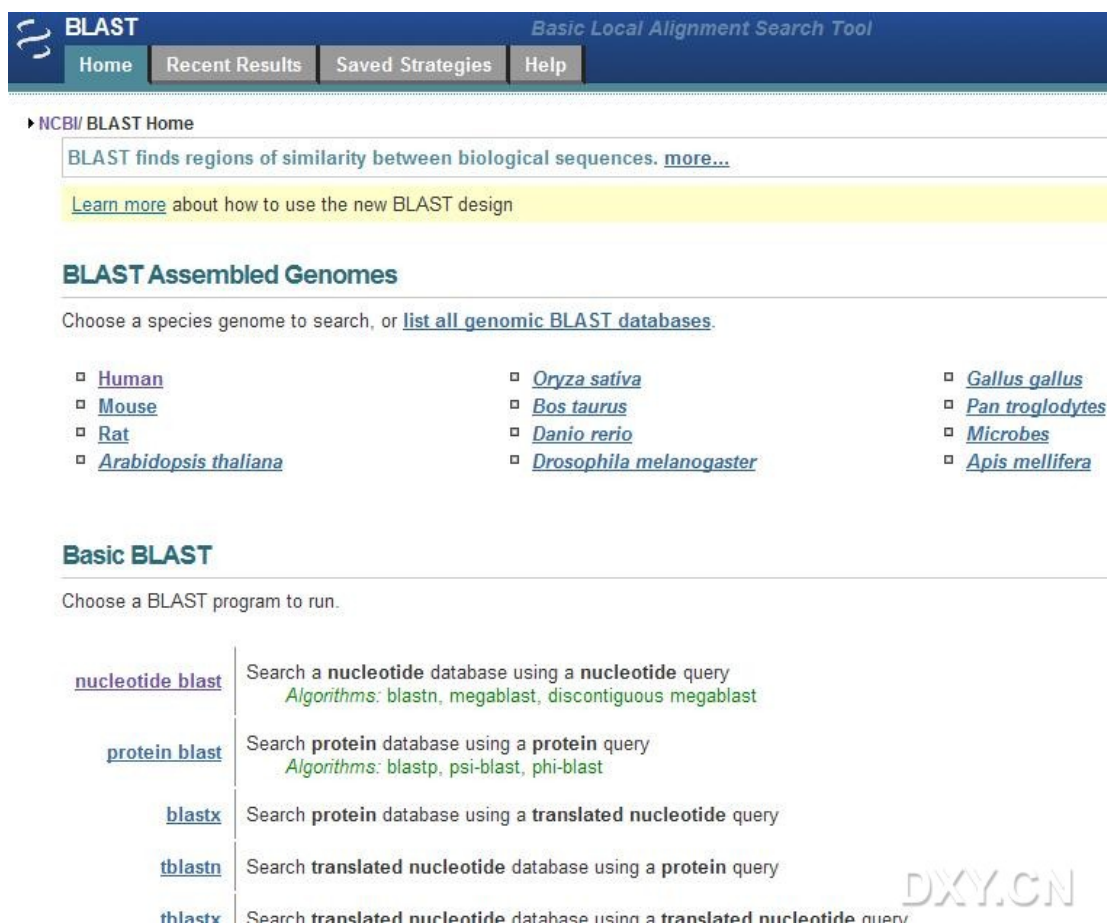
<http://www.dxy.cn/bbs/post/view?bid=64&id=9517792&sty=1&tpg=1&age=0>

<http://www.dxy.cn/bbs/post/view?bid=67&id=9523263&sty=1&tpg=1&age=0>

第四部分 如何运用 BLAST 进行序列比对、检验引物特异性

提到序列比对，绝大多数战友都会想到 BLAST，但 BLAST 的使用确实又是一个很大的难题，因为他的功能比较强悍，里面涉及到的知识比较多，而且比对结束后输出的结果参数（指标）又很多。如果把 BLAST 的使用详细的都讲出来，我想我发帖发到明天也发不完，更何况我自己也不是完全懂得 BLAST 的使用。所以我在这里也就“画龙点睛”——以比对核酸序列为例来给大家介绍一下 BLAST 的使用，也算是 BLAST 的入门课程吧。请看帖的战友好好体会，如果你用心看，在看帖完毕之后 BLAST 的基本使用（包括其他序列的比对）应该没有问题了。

1. 打开 BLAST 页面，<http://www.ncbi.nlm.nih.gov/BLAST/> 打开后如图所示：



对上面这个页面进行一下必要的介绍：

BLAST 的这个页面主体部分（左面）包括了三部分：BLAST Assembled Genomes、Basic BLAST、Specialized BLAST。相信大家可以看懂这三个短语的意思，我就不多说了；我要说的是，可以认为这是三种序列比对的方法，或者说是 BLAST 的三条途径。

第一部分 BLAST Assembled Genomes 就是让你选择你要比对的物种，点击相应物种之后

即可进入比对页面。

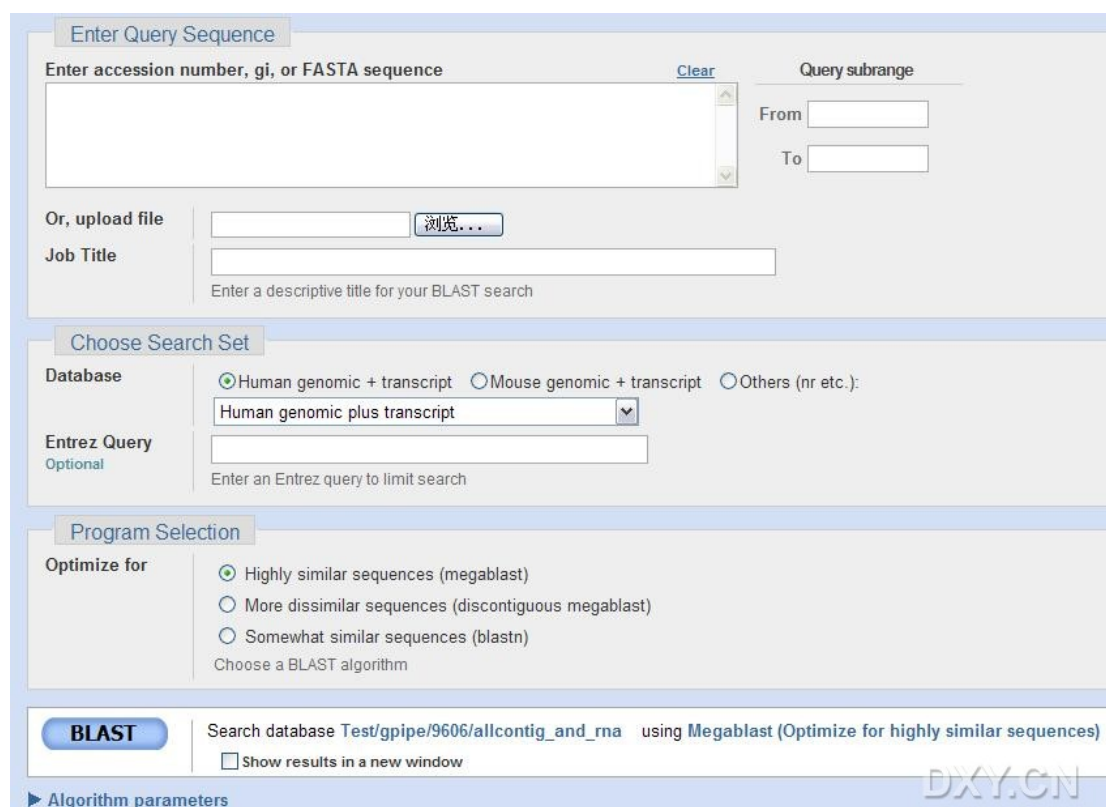
第二部分 Basic BLAST 包含了 5 个常用的 BLAST，每一个都附有简短的介绍。

第三部分 Specialized BLAST 是一些特殊目的的 BLAST，如 IgBLAST、SNP 等等，这个时候你就需要在 Specialized BLAST 部分做出适当的选择了。

总之，这是一个导航页面，它的目的是让你根据自己的比对目的选择相应的 BLAST 途径。

下面以最基本的核酸序列比对来谈一下 BLAST 的使用，期间我也会含沙射影的说一下其他序列比对的方法。

2. 点击 Basic BLAST 部分的 nucleotide blast 链接到一个新的页面。打开后如图所示：



介绍一下上述页面：

Enter Query Sequence 部分是让我们输入序列的，你可以直接把序列粘贴进去，也可以上传序列，还可以选择你要比对的序列的范围（留空就代表要比对你要输入的整个序列）。Job Title 部分还可以为本次工作命一个名字。

Choose Search Set 部分是让我们选择要与目的序列比对的物种或序列种类（genome DNA、mRNA 等等）。如果是人或老鼠的话，就可以直接选择了如果是其他物种就要选择“others”了，这时候网页会主动跳出一个下拉对话框和一个输入式对话框，你可以分别选

择和输入要跟你的序列比对的序列种类和物种。下面的 Entrez Query 可以对比对结果进行适当的限制。

Program Selection 部分其实是让我们选择本次比对的精确度，种内种间等等。

在 BLAST 按钮下面有一个 “Algorithm parameters”，这是参数设置选项，一般用户使用不到此项，所以它比较隐蔽，点击，原网页下方即可增加了 Algorithm parameters 的内容。大部分战友都用不到更改这里的选项，我也不多说了，有兴趣的朋友可以自己研究一下。

3. 依次填写上述网页必须部分，点击 BLAST 按钮后，出现如下界面（只截取其中一部分）：

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
NT_007819.16	Homo sapiens chromosome 7 genomic contig, reference assembly	42.1	74.3	100%	0.005	100%	
NW_923240.1	Homo sapiens chromosome 7 genomic contig, alternate assembly (ba	42.1	74.3	100%	0.005	100%	
NT_079592.2	Homo sapiens chromosome 7 genomic contig, alternate assembly (ba	42.1	74.3	100%	0.005	100%	
NT_009237.17	Homo sapiens chromosome 11 genomic contig, reference assembly	36.2	36.2	85%	0.31	100%	
NW_925006.1	Homo sapiens chromosome 11 genomic contig, alternate assembly (b	36.2	36.2	85%	0.31	100%	
NT_016354.18	Homo sapiens chromosome 4 genomic contig, reference assembly	34.2	66.4	90%	1.2	100%	
NW_922162.1	Homo sapiens chromosome 4 genomic contig, alternate assembly (ba	34.2	34.2	80%	1.2	100%	
NT_010194.16	Homo sapiens chromosome 15 genomic contig, reference assembly	32.2	62.4	76%	4.8	100%	
NT_008705.15	Homo sapiens chromosome 10 genomic contig, reference assembly	32.2	32.2	76%	4.8	100%	
NT_008046.15	Homo sapiens chromosome 8 genomic contig, reference assembly	32.2	32.2	76%	4.8	100%	
NT_025741.14	Homo sapiens chromosome 6 genomic contig, reference assembly	32.2	62.4	76%	4.8	100%	
NT_007592.14	Homo sapiens chromosome 6 genomic contig, reference assembly	32.2	32.2	76%	4.8	100%	

出现的这个结果页面信息含量非常大，如果我们用心观察，还是可以发现其中的一些主要指标的。列举上图也是为了给大家展示一下这些评价标准。其中 Description 部分推荐大家详细看一下，另外说一下 “E value” 这个指标与其他指标不同，它的数值越小相似程度越高，其他几个（如 Total score）都是数值越高相似度越高。

在这个图示的表格下方就是具体的相似性的核酸序列了，还配合着各种参数的得分。

好了，各位亲爱的战友，我的 BLAST 就发到这里为止了，更具体的东西有待大家一起去努力研究。伴随着 BLAST 的终结，我的 “一步一步教你使用 NCBI” 也要暂时告一段落了，很高兴自己发第一个帖子时说的话今天终于做到了。以后如果我有新的 NCBI 使用方法的话，我还会添加到这里来，但我想这一阵子是不会接着发了，呵呵。

真心希望各位战友在这里一起交流自己使用 NCBI 的一些技巧，正如丁香园的宗旨一样——“我为人人，人人为我”，让我们互相学习、共同进步，最后再一次祝愿大家试验顺利！

作者：urbest

编排：lzfist

苏州大学生命科学学院

本文章已在丁香园 (<http://www.dxy.cn>) 网站发表