

参考基因组及注释文件下载 - 小xuo生 - 博客园

 [cnblogs.com/jessepeng/p/9681749.html](https://www.cnblogs.com/jessepeng/p/9681749.html)

参考基因组及注释文件下载

从三大核酸数据库NCBI、Ensembl、UCSC 下载参考序列及注释文件

0.人类基因组版本对应关系

NCBI	Ensembl	UCSC
GRCh36	release_52	hg18
GRCh37	release_59/61/64/68/69/75	hg19
GRCh38	release_76/77/78/80/81/82	hg38

1.NCBI

人类基因组

GRCh38下载（默认）：

ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/

GRCh37下载：

ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/ARCHIVE/BUILD.37.1/

ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/ARCHIVE/BUILD.37.2/

ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/ARCHIVE/BUILD.37.3/

GRCh36及其他版本下载：

ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/ARCHIVE/

常用的GRCh37基因组大小压缩约900Mb，解压后约3G。人类的注释文件一直在更新，默认为GFF3格式。gff和gtf格式文件关系及相互转换见：<https://www.jianshu.com/p/48b5a0972301>

也可以通过NCBI的genome数据库下载，默认是GRCh38，若要下载其他版本，直接检索关键词。

如输入GRCh37或hg19：

Genome Genome GRCh37
[Create alert](#) [Limits](#) [Advanced](#)

GENOME ASSEMBLY

[GRCh37.p13](#)

Homo sapiens (human)
Genome Reference Consortium (June 2013)
RefSeq assembly: GCF_000001405.25
[PubMed \(27\)](#)

[Genome Browser](#) [BLAST](#) [Download](#) [Genome Resources](#)

参考序列和GFF文件均可从此处下载，其他物种类似。

2.Ensembl

同NCBI一样，可通过网页检索下载，也可通过ftp直接下载。





(1) 官网下载：

e!Ensembl ASSEMBLY [BLAST/BLAT](#) [BioMart](#) [Tools](#) [Downloads](#) [Help & Documentation](#) [Blog](#) [Mirrors](#)

Search: All species for [Go](#)
e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [coronary heart disease](#)

Browse a Genome
The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Popular genomes

 Human GRCh38 p5	 Human GRCh37
 Mouse GRCm38 p4	 Zebrafish GRCz10

★ [Log in to customize this list](#)


All genomes


-- Select a species --

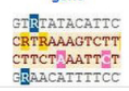
[View full list of all Ensembl species](#)


Other species are available in [Ensembl Pre/4](#) and [EnsemblGenomes](#)


Still using Human GRCh37?
[Go to GRCh37](#)

Variant Effect Predictor


Gene expression in different tissues


Find SNPs and other variants for my gene


Retrieve gene sequence


Compare genes across species


[Use my own data in Ensembl](#) [ENCODE data in Ensembl](#)

What's New in Ensembl

- [Update to Ensembl-Havana](#)
- [Mouse: update to Ensembl](#)
- [Chicken and pig dbSNP 14](#)
- [Filtering Variants by MAF](#)
- [Advanced Filtering and Cc](#)

[Full details](#) | [All web u](#)

Latest blog posts

- 16 Dec 2015: [Ensembl Gen](#)
- 14 Dec 2015: [New image ex](#)
- 08 Dec 2015: [Ensembl 83 h](#)

[Recent tweets from @ensembl](#)



BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog



Human (GRCh38.p12) ▼

Search Human (*Homo sapiens*)

Search all categories ▼

Search Human...

Go

e.g. BRCA2 or 17:63992802-64038237 or rs1333049 or osteoarthritis

Genome assembly: GRCh38.p12 (GCA_000001405.27)

 More information and statistics

 Download DNA sequence (FASTA) 

 Convert your data to GRCh38 coordinates

 Display your data in Ensembl

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▼

Go



View karyotype



Example region

或者通过进入download下载。

微生物或原生生物的下载，如幽门螺杆菌：

Find a Genome

The main Ensembl site focuses on vertebrate genomes - [scroll down](#) for links to our sister sites covering invertebrates, plants, bacteria, etc.

Species tree

[View the full Ensembl species tree.](#)


Available genomes

Note: to find out which species were in previous releases, please see the [table of assemblies](#)

Show All ▼ entries

Show/hide columns

Helicobacter pylori



Common name	Scientific name	Taxon ID	Ensembl Assembly	Accession	Genebuild Method	Variation database	Regulation database	Pre assembly
No matching records found								

[Credits page for species images](#)

Other Metazoa

Additional metazoan genomes (initially insect vectors and nematodes) are available from [EnsemblMetazoa](#)

Plants and Fungi


Plant and fungal genomes can be found at [EnsemblPlants](#) and [EnsemblFungi](#)

Protists, Bacteria and Archaea


Unicellular eukaryotic and prokaryotic genomes can be found at [EnsemblProtists](#) and [EnsemblBacteria](#) respectively.

或者直接从这里进入：<http://bacteria.ensembl.org/index.html>

<http://bacteria.ensembl.org/species.html>



HMMER | BLAST | Tools | Downloads | Documentation | Website help

 Search Ensembl Bacteria...

Find a Species

Show 25 ▼ entries

Filter Helicobacter pylori

Species	Assembly	Taxonomy ID	Serotype	Publications	Present taxonomically
Helicobacter pylori	ASM25882v1	210			N
Helicobacter pylori 2017	ASM19231v1	985081			N
Helicobacter pylori 2018	ASM19233v1	985080			N
Helicobacter pylori 26695	ASM852v1	85962			N
Helicobacter pylori 26695-1	ASM82698v1	1391726			N
Helicobacter pylori 26695 (GCA_000307795)	ASM30779v1	85962			N
Helicobacter pylori 35A	ASM17893v2	585535			N

(2) ftp下载:

ftp://ftp.ensembl.org/pub/release-75/fasta/homo_sapiens/

更改release后的数字下载相应的版本, 包括dna、cdna、cds等序列信息, release-75是目前最新的hg19版本。

注释文件下载(默认gtf, 大部分比对软件输入格式):

ftp://ftp.ensembl.org/pub/release-75/gtf/homo_sapiens/

3.UCSC

参考序列下载很简单(尤其是人)

进入官网: <http://hgdownload.cse.ucsc.edu/downloads.html>, 下载对应的各个版本

也可进入<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz>, 修改中间数字下载不同版本

但是注释文件下载稍微有点麻烦, 需要设置一系列参数来生成:

<http://genome.ucsc.edu/cgi-bin/hgTables>

Select the following options:

clade: Mammal

genome: Human

assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Predictions

track: UCSC Genes

table: knownGene

region: Select "genome" for the entire genome.

output format: GTF - gene transfer format

output file: enter a file name to save your results to a file, or leave blank to display results in the browser

Click 'get output'.

Reference:

<http://www.novogene.com/tech/suppor/gene-calss/comprehensive/228.html>

<http://www.bio-info-trainee.com/tag/ensembl>

分类: Transcriptomics

posted @ 2018-09-20 16:44 小xuo生 阅读(1871) 评论() 编辑 收藏