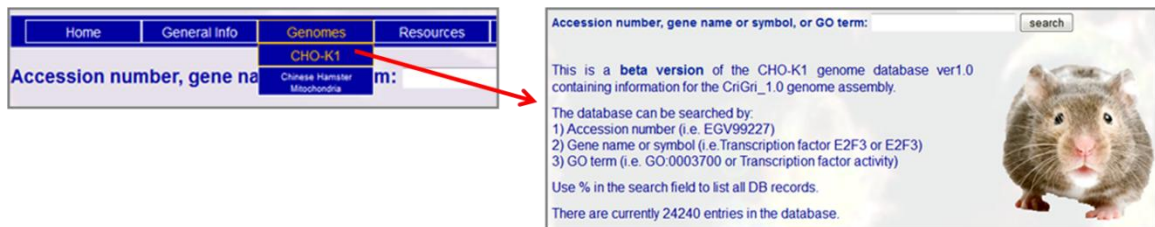


## Tutorial 3 – Searching the Chinese Hamster Genome Database

There are separate search pages for the CHO-K1 genome and the Chinese hamster mitochondrial genome. Both can be accessed from the **Genomes** menu at the top of the page. The CHO-K1 genome can also be searched directly from the homepage using the search box located at the top of the homepage.

### Searching the CHO-K1 Genome

- 1) Select the **CHO-K1** genome from the **Genomes** menu.



Accession number, gene name or symbol, or GO term:

This is a **beta version** of the CHO-K1 genome database ver1.0 containing information for the CriGri\_1.0 genome assembly.

The database can be searched by:

- 1) Accession number (i.e. EGV99227)
- 2) Gene name or symbol (i.e. Transcription factor E2F3 or E2F3)
- 3) GO term (i.e. GO:0003700 or Transcription factor activity)

Use % in the search field to list all DB records.

There are currently 24240 entries in the database.

- 2) Type a keyword into the search box at the top of the page.

Typing % in the search box will list all 24,240 entries currently in the database.

The CHO-K1 genome database can be searched by

**Accession number** (i.e. EGV99227)

The database can be searched using the GenBank WGS protein accession IDs. For the CHO-K1 WGS project, these accession numbers are EGVXXXXX or EGWXXXXX.

These accession numbers can also be found in the NCBI protein database by restricting the search to “cricetulus griseus” and the locus tag “I79.”

Found 1 search results for: EGV99227

No.	Gene	Accession	Parent Accession	Locus Tag	Product Qualifier
1	E2f3	EGV99227	JH000056	I79_002249	Transcription factor E2F3

**Gene name or symbol** (i.e. Transcription factor E2F3 or E2F3)

The database can be searched using the protein names assigned to gene products during annotation of the WGS project. The database can also be searched by gene symbols. Gene symbols were assigned based on annotation of homologous proteins.

Found 1 search results for: transcription factor E2F3

No.	Gene	Accession	Parent Accession	Locus Tag	Product Qualifier
1	E2f3	EGV99227	JH000056	I79_002249	Transcription factor E2F3

**Gene ontology (GO) term** (i.e. GO:0003700 or Transcription factor activity)

The database can be searched using a GO accession ID or term. Searching by GO term will return all database entries annotated by that GO term.

Found 472 search results for: GO:0003700

No.	Gene	Accession	Parent Accession	Locus Tag	Product Qualifier
1	Etv3,Mets,Pe1	EGW07031	JH000344	I79_009579	ETS translocation variant 3
2	Etv3l	EGW07032	JH000344	I79_009580	ETS translocation variant 3-like protein
3	Mef2d	EGW07051	JH000344	I79_009600	Myocyte-specific enhancer factor 2D

Searching for “caspase” will return 13 entries from the database displayed in a table format.

General information for the returned entries includes gene names, protein accession IDs, scaffolds (parent accession), locus tag, and protein name/description (product qualifier).

To find more details about a single entry, click on a gene symbol or accession number.

Found 13 search results for: caspase

No.	Gene	Accession	Parent Accession	Locus Tag	Product Qualifier
1	Casp8	EGV99437	JH000741	I79_014611	Caspase-8
2	Pacap	EGV97626	JH000776	I79_014961	Proapoptotic caspase adapter protein
3	Card11	EGV95787	JH000039	I79_001680	Caspase recruitment domain-containing protein 11
4	Casp7_Mch3	EGW14942	JH003907	I79_024625	Caspase-7
5	Card9	EGW12771	JH001398	I79_019269	Caspase recruitment domain-containing protein 9
6	Card14_Bimp2	EGW08250	JH000695	I79_014172	Caspase recruitment domain-containing protein 14
7	Casp3_Cpp32	EGW06745	JH002126	I79_021949	Caspase-3
8	Casp12	EGW01061	JH001131	I79_017574	Caspase-12
9	Casp9_Mch6	EGV96884	JH001574	I79_020089	Caspase-9
10	Casp14	EGV95473	JH000504	I79_011797	Caspase-14
11	Casp2_Ich1	EGV93989	JH000477	I79_011415	Caspase-2
12	Casp14l	EGV93647	JH000995	I79_016717	Putative caspase-14-like protein
13	Card6	EGV92856	JH000378	I79_010067	Caspase recruitment domain-containing protein 6

3) Clicking on either [Casp8](#) or [EGV99437](#) will open the Gene Details page for caspase 8.

The **General Information** section contains information on gene names, symbols, IDs, and descriptions.

The **Genomic Information** section provides the coordinates of the coding region of the gene and the WGS scaffold on which the gene is located.

The **Sequence Information** section provides links to download the nucleotide and protein sequence in FASTA format.

The **External Links** section provides links to CHO protein entries in the NCBI and EMBL databases.

The **Homologs** section provides links to homologous proteins in other species in the UniProt databases.

The **Annotation** section provides functional information and GO terms assigned to the gene product. The GO terms link to the AmiGO gene ontology browser.

Home	General Info	Genomes	Resources	Community	Partners
General Information					
Name	Caspase-8				
Symbol	Casp8				
Synonyms	(not available)				
Locus tag	I79_014611				
Description	(temporary unavailable)				
Genomic Information					
WGS Scaffold	JH000741				
CDS	complement(join(JH000741:197973..198108,JH000741:199527..200025,JH000741:203332..203473,JH000741:205312..205376,JH000741:206895..206945,JH000741:207026..207164,JH000741:209567..209675,JH000741:214348..214652))				
Graphics	JH000741				
Sequence Information					
Nucleotide sequence	Download sequence				
Protein sequence	Download sequence				
External Links					
NCBI (protein)	EGV99437				
NCBI (protein graphics)	EGV99437				
EMBL-EBI (protein)	EGV99437				
Homologs					
UniProtKB/Swiss-Prot	O89110				
UniProtKB/TrEMBL	B2CMK5				
Annotation					
GO Terms	GO_process: GO:0042981 - regulation of apoptosis [Evidence IEA] GO_process: GO:0006915 - apoptosis [Evidence IEA] GO_process: GO:0006508 - proteolysis [Evidence IEA] GO_function: GO:0008234 - cysteine-type peptidase activity [Evidence IEA] GO_function: GO:0005515 - protein binding [Evidence IEA] GO_function: GO:0004197 - cysteine-type endopeptidase activity [Evidence IEA]				
Please report any issue <a href="#">Provide Feedback</a>					

Questions or comments can be sent by clicking on the [Provide Feedback](#) button located at the bottom of every Gene Details page.

4) The **Genomic Information** section provides information about the WGS scaffold on which the selected gene is located.

Genomic Information	
WGS Scaffold	JH000741
CDS	complement(join(JH000741:197973..198108,000741:205312..205376,JH000741:206895..675,JH000741:214348..214652))
Graphics	JH000741

**Cricetus griseus cell line CHO-K1 unplaced genomic scaffold scaffold880, whole genome shotgun sequence**

GenBank: JH000741.1

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

[Details](#)

LOCUS JH000741 84704 bp DNA linear CON 25-AUG-2011

DEFINITION Cricetus griseus cell line CHO-K1 unplaced genomic scaffold scaffold880, whole genome shotgun sequence.

ACCESSION JH000741 REVISION: 153999..245652

VERSION JH000741.1 GI:344144324

ISLINES Project: 87793

KEYWORDS WGS.

SOURCE Cricetus griseus (Chinese hamster)

ORGANISM Cricetus griseus

In the **WGS Scaffold** field, clicking on the scaffold accession ID opens the NCBI database entry.

In the **CDS** field, the scaffold coordinates of the coding sequence are listed.

In the **Graphics** field, clicking on the scaffold accession ID opens the NCBI Map Viewer to view genome annotation in this region of the scaffold.

5) The **Sequence Information** section provides links to download the nucleotide and protein sequence in FASTA format.

Sequence Information	
Nucleotide sequence	<a href="#">Download sequence</a>
Protein sequence	<a href="#">Download sequence</a>

>EGV99437

atggatttcagaactgtctttacaagattgctgaaaaactgtagacagtgactggctgc

tttcaggtagctgcaggaaggactgttgaggagacacatctgtcctcctgaaggagct

gagctgacaggttcagacaaagccaggttctgctacacaggtcagctctttaaagcttc

ctgtatgatgataactcagagcctgcttgatattttgtggaatggagaagagtcacct

>EGV99437

MDFRNCLYKIAENLDSSELAALKFLCLDYIPQKKQEPKIDALMFQVLQKGLLEEDNLSFL

LYDNNSSLLDIFVEMEKRVILRENLDAKLSICAVNKSLLKIEAYEKSSRGMSLEGGA

GTNYDEEALTEFKELHFEVVKPYRDTANKIHEVLRFYQSMHDKDKDCIFCILLSHGDTGII

DEADFLGMATVKNCSYRDPQTGTWYIQLCQSLRKRCFGDDILSILTGVDVSNKDR

6) The **External Links** and **Homologs** sections provide links to CHO proteins and homologs in the NCBI, EMBL, and UniProt databases.

External Links	
NCBI (protein)	EGV99437
NCBI (protein graphics)	EGV99437
EMBL-EBI (protein)	EGV99437

**EMBL-EBI EGV99437.1: Cricetus griseus (Chinese hamster) Caspase-8**

View [FASTA](#) [XML](#)

Overview [Source/Features](#) [Other Features](#) [Genes](#) [Sequences](#)

Organism: Cricetus griseus | Molecule type: genomic DNA | Topology: linear | Data class: CON | Taxonomic Division: ROD

Sequence length: 1,446 | Sequence version: 1

Linkage: Eukarya Metazoa Chordata Cephalochordata Eumetazoa Mammalia Eutheria Euarchontaria Glires Rodentia Soricomorpha Soricidae Soricinae Soricini

Annotations: Overview

Overview: Forward strand: 84,704 bp

Features: Forward strand: 14,620 bp

Comp: AT001047111.1

Source: Cricetus griseus

Genes: C8\_014611

mRNA: C8\_014611

CDS: C8\_014611

**089110 (CASP8\_MOUSE)** Reviewed, UniProtKB/Swiss-Prot

Last modified November 16, 2011. Version 110 [History](#)

Clusters with 100%, 90%, 50% identity | Documents (3) | Third-party data

Names Attributes General annotation Ontologies Interactions Sequence annotation

**Names and origin**

Protein names: Recommended name: Caspase-8 | Short name: CASP-8 | EC: 3.4.22.61 | Cleaved into the following 2 chains: 1. Caspase-8 subunit p18 | 2. Caspase-8 subunit p10

Gene names: Name: Casp8

Organism: Mus musculus (Mouse)

Taxonomic identifier: 10090 [NCBI]

## Searching the Chinese hamster mitochondrial genome

- 1) Select the **Chinese hamster mitochondrial genome** from the **Genomes** menu.

Home General Info Genomes Resources

Gene Name:  search (use % as Gene Name to list all DB records)

CHO-K1  
Chinese Hamster Mitochondria

- 2) Type a keyword into the search box at the top of the page.  
Typing **%** in the search box will list all 13 entries currently in the database.  
The Chinese hamster mitochondrial genome database can be searched only by **Gene symbol** (i.e. COX).

Searching for “COX” will return 3 entries from the database displayed in a table format.

General information for the returned entries includes gene symbols, NCBI Entrez Gene IDs, gene names and genomic coordinates (start and end positions).

Found 3 search results for: COX

No.	Gene Symbol	Gene ID	Name	Start position	End position
1	COX1	3979185	cytochrome c oxidase subunit I [Cricetulus griseus]	5309	6853
2	COX2	3979186	cytochrome c oxidase subunit II [Cricetulus griseus]	6993	7676
3	COX3	3979189	cytochrome c oxidase subunit III [Cricetulus griseus]	8586	9369

To find more details about a single entry, click on a gene symbol.

- 3) Clicking on **COX1** will open the Gene Details page for cytochrome c oxidase subunit I.

The **General Information** section contains information on gene names, symbols, IDs, and descriptions.

The **Genomic Information** section provides the gene coordinates in the mitochondrial genome

The **Sequence Information** section provides links to download the nucleotide and protein sequence in FASTA format.

The **Homologs Information** section provides links to homologous proteins in other species in the NCBI and Ensembl databases.

The **External Links** section provides links to CHO gene and protein entries in the NCBI databases.

The **Annotation** section reports GO terms and provides links to the AmiGO gene ontology browser.

Home	General Info	Genomes	Resources	Community
General Information				
Gene Symbol	COX1			
Name	cytochrome c oxidase subunit I [Cricetulus griseus]			
Gene ID	3979185			
Description	(temporary blank)			
Genomic Location				
Genomic Accession (start position)	5309			
Genomic Accession (end position)	6853			
Map	YP_537121.1			
Sequence Information				
Nucleotide Sequence	Download sequence			
Protein Sequence	Download sequence			
Homologs Information				
Ensembl Gene	Human	Mouse	Rat	
	ENSG00000198804	ENSMUSG00000064351	ENSRNOG00000034234	
Entrez Gene	Human	Mouse	Rat	
	4512	17708	26195	
External Links				
Entrez Gene	3979185			
Entrez Protein	YP_537121.1			
Annotation				
GO Terms	GO:0016020 membrane			
	GO:0070469 respiratory chain			
	GO:0005743 mitochondrial inner membrane			
	GO:0016021 integral to membrane			



