

# Web Apollo/JBrowse

- JBrowse is a web based genome browser
  - Visualize features that are mapped to a genome
  - These features are displayed as tracks
  - Many different types of data may be
- Web Apollo adds editing functions to JBrowse
  - Manual gene curation
  - Changes automatically saved back to server
  - Edits are visible to other annotators in real-time

# Web Apollo

1

The screenshot displays the Apollo genome browser interface. The browser address bar shows the URL: [https://apollo.nal.usda.gov/lepdec\\_training/jbrowse/?loc=Scaffold10%3A1702617..1710066&tracks=DNA%2CAnnotations%2CGC](https://apollo.nal.usda.gov/lepdec_training/jbrowse/?loc=Scaffold10%3A1702617..1710066&tracks=DNA%2CAnnotations%2CGC). The interface includes a top navigation bar with 'File', 'View', 'Help', and 'Tools' menus. The main view shows a genomic track for Scaffold10:1702617..1710066 (7.45 Kb). The tracks displayed are:

- User-created Annotations**: A yellow bar at the top of the track.
- GC Content**: A bar chart showing GC content across the scaffold.
- Gaps in assembly**: A black bar indicating gaps in the assembly.
- LDEC\_v0.5.3-Models**: A track showing LDEC models, with a specific model labeled 'LdecTmpM002528-RA'.
- est\_gff:cufflinks\_i5k**: A track showing transcript links, with a specific link labeled '1:CUFF.221.1\_a'.
- augustus\_masked**: A track showing masked regions, with a specific region labeled 'augustus\_masked-Scaffold10-abinit-gene-17.0-mRNA-1'.
- snap\_masked**: A track showing masked regions, with a specific region labeled 'snap\_masked-Scaffold10-abinit-gene-17.10-mRNA-1'.
- Female RNA-Seq coverage**: A bar chart showing RNA-Seq coverage for females, with a y-axis scale from 0 to 70.
- Female RNAseq**: A track showing RNA-Seq reads for females.

On the left side, there is a sidebar with the following sections:

- 3. Mapped Proteins**: 41 items. Includes 'Other' (1) and 'Protein2genome' (20). The 'Protein2genome' section lists various taxonomic groups with checkboxes, such as 'protein2genome\_Annelida', 'protein2genome\_Arthropoda', etc.
- Supplementary: BlastX**: 20 items.
- 4. Transcriptome**: 1 item. Includes 'Assembly' (1) with a checked 'est\_gff:cufflinks\_i5k' option.
- transcriptome**: 4 items.
- Coverage plots (BigWig)**: 2 items. Includes 'Female RNA-Seq coverage' (checked) and 'Male RNA-Seq coverage' (unchecked).
- Mapped Reads**: 2 items. Includes 'Female RNAseq' (checked) and 'Male RNAseq' (unchecked).

On the right side, there are three large numbers: 2, 3, and 3, which appear to be annotations or markers.

# Viewing evidence

1. The “Available Tracks” panel lists all tracks
  - Tracks are grouped hierarchically
    - Study or analysis
    - Categories based on evidence type
2. The Navigation panel describes the viewed region
  - Coordinate ranges for the entire sequence on top
  - Navigation controls including pan, zoom and search in the middle
  - Coordinate ranges for the current view along the bottom
3. The Evidence panel shows the features in genomic context
  - Each type of data is a separate track
  - Tracks may be displayed in different ways

# Viewing evidence

- Turning tracks on or off
  - Click a tracks checkbox to turn it on or off
  - Click 'X' by the track label also turns the track off

The screenshot displays the Apollo genome browser interface. The top navigation bar includes the Apollo logo and menu options: File, View, Tools, and Help. The main view shows a genomic region on Scaffold10, with coordinates ranging from 0 to 2,000,000. A red box highlights the coordinate 800,000. The tracks shown include:

- GC Content:** A track showing GC content across the region, with a checkbox and an 'X' icon next to the label.
- User-created Annotations:** A track showing annotations, including a blue bar labeled "LdecTmpM002494-RA".
- est\_gff:cufflinks\_i5k:** A track showing transcript evidence, with a checkbox and an 'X' icon next to the label. Red arrows point to the 'X' icon and the checkbox.

The left sidebar shows the "Available Tracks" panel, which is organized into categories:

- 0. Reference Assembly:** 2 tracks (GC Content, Gaps in assembly).
- BCM\_v0.5.3:** 47 tracks (Gene Sets, Evidence, Mapped Proteins, Other, cegma, Protein2genome, Supplementary: BlastX).
- 4. Transcriptome:** 1 track (Assembly).

The "est\_gff:cufflinks\_i5k" track is checked, and the "X" icon next to its label is also checked, indicating it is turned off. Red arrows point to these elements in the screenshot.

# Login

- For scaffold selection view:

The screenshot shows the Apollo web application interface. The browser address bar displays `https://apollo.nal.usda.gov/lepdec_training/selectTrack.jsp`. The application header includes the 'pollo' logo and 'File Tools' menu. Below the header, there is a search bar and a 'Filter:' input field. The main content area shows a table with columns 'Organism', 'Name', and 'Length'. The table is currently empty, displaying 'Showing 0 to 0 of 0 entries'. A 'Login' dialog box is overlaid on the table, containing the following fields and buttons:

**Login**

User name

Password

The screenshot shows the Apollo web application interface after a successful login. The browser address bar displays `https://apollo.nal.usda.gov/lepdec_training/selectTrack.jsp`. The application header includes the 'pollo' logo and 'File Tools' menu. Below the header, there is a search bar and a 'Filter:' input field. The main content area shows a table with columns 'Organism', 'Name', and 'Length'. The table is populated with 10 scaffold entries. A user profile icon labeled 'demo' is visible in the top right corner of the application header.

Organism	Name	Length
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold1</a>	3793193
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10</a>	2193025
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold100</a>	1122226
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold1000</a>	265361
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10000</a>	15875
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10001</a>	7998
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10002</a>	10857
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10003</a>	15698
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10004</a>	10589
<input type="checkbox"/> Leptinotarsa decemlineata	<a href="#">Scaffold10005</a>	10524

Showing 1 to 10 of 24,393 entries

# Login

- Directly from JBrowse genome browser



# Login

- Directly from JBrowse genome browser

The screenshot displays the Apollo genome browser interface. The main window shows a genomic track for Scaffold10:664302..704616 (40.32 Kb). The track includes GC Content, Gaps in assembly, LDEC models (LdecTmpB002493-RA and LdecTmpM002494-RA), Augustus masked regions (including abinit-gene-B.1-mRNA-1 and Scaffold10-abinit-gene-7.0-mRNA-1), CEGMA BUSC1815, cufflinks (1:CUFF.188.1 and 1:CUFF.189.1), and RNA-Seq coverage for Female and Male. A 'Login' dialog box is overlaid in the center, containing the following fields and buttons:

- Login** (dialog title)
- User name: demo
- Password: \*\*\*\*
- Buttons: Login, Clear

The left sidebar shows a navigation menu with categories like 'Supplementary: BlastX' (containing 20 items) and 'Transcriptome' (containing 4 items). The 'Transcriptome' section is expanded to show 'Assembly' (1 item) and 'Coverage plots (BigWig)' (2 items), with 'Female RNA-Seq coverage' and 'Male RNA-Seq coverage' checked. The 'Mapped Reads' section (2 items) is also visible.

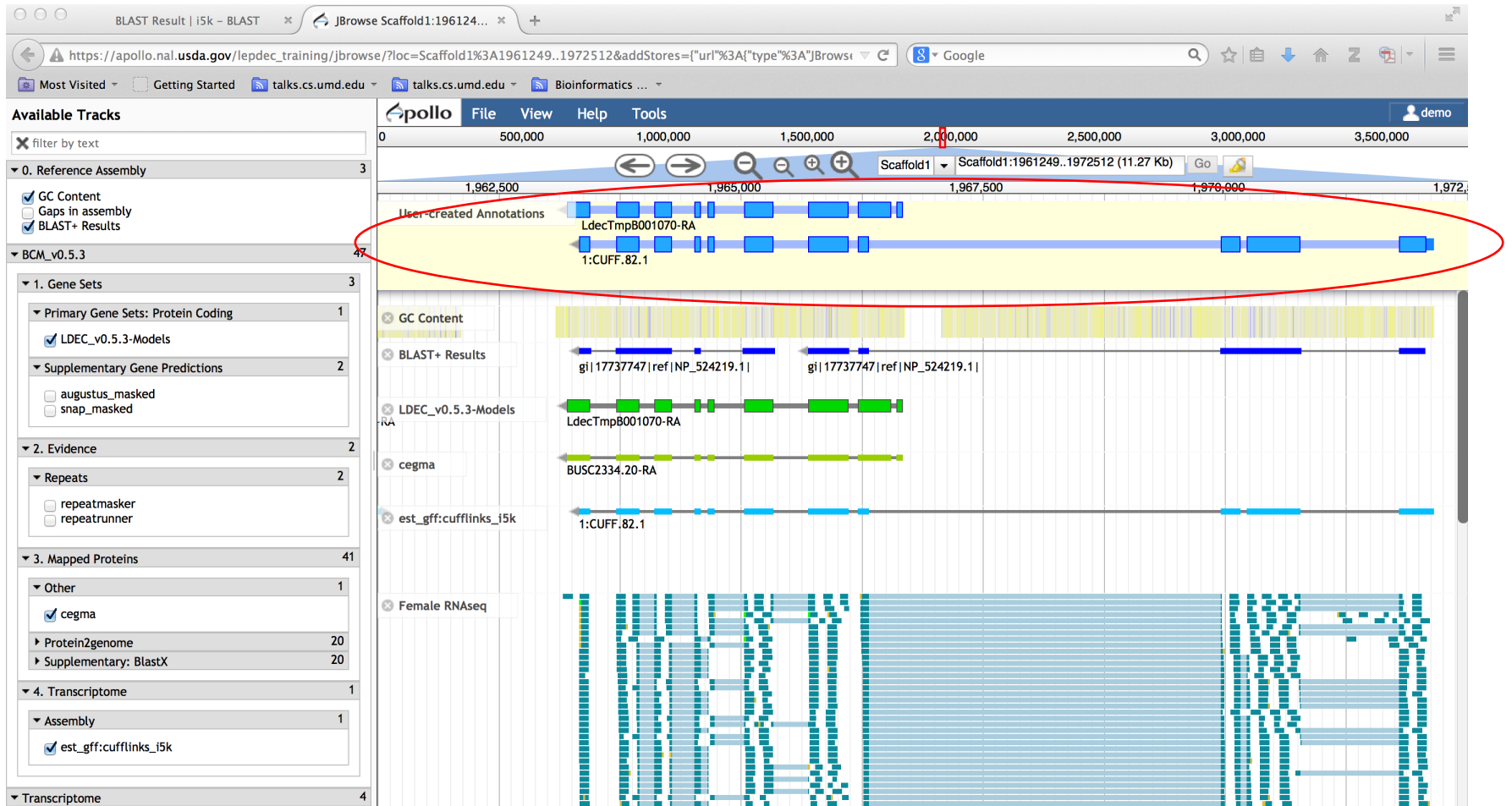
# Login

- Directly from JBrowse genome browser

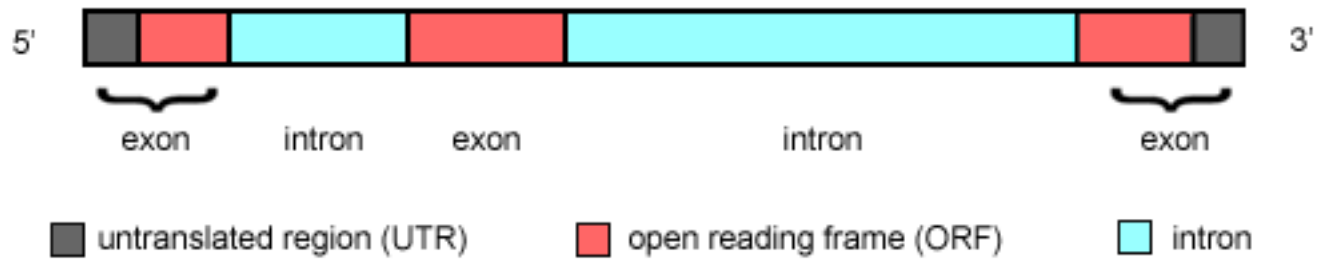
The screenshot displays the JBrowse genome browser interface. The top navigation bar includes the 'apollo' logo, 'File', 'View', 'Tools', and 'Help' menus, and a user profile dropdown menu labeled 'demo'. The main display area shows a genomic track for Scaffold10, with a scale from 0 to 2,000,000. The track displays various annotations, including 'User-created Annotations' (yellow bar), 'GC Content' (yellow bar chart), 'Gaps in assembly' (black bars), 'LDEC\_v0.5.3-Models' (gene models for 'LdecTmpB002493-RA' and 'LdecTmpM002494-RA'), 'augustus\_masked' (mRNA models for 'augustus\_masked-Scaffold10-abinit-gene-7.0-mRNA-1'), 'cegma' (model for 'BUSC1815.32-RA'), 'est\_gff:cufflinks\_15k' (models for '1:CUFF.188.2', '1:CUFF.186.1', '1:CUFF.188.1', and '1:CUFF.189.1'), 'Female RNA-Seq coverage', and 'Male RNA-Seq coverage'. The left sidebar contains 'Available Tracks' with categories like 'Reference Assembly', 'Gene Sets', 'Evidence', and 'Mapped Proteins'. The top navigation bar shows the URL 'https://apollo.nal.usda.gov/lepdec\_training/jbrowse/?loc=Scaffold10%3A664302..704203&tracks=DNA%2CAnnotations%2CGC%20Content%2CGaps%20in%20assembly%2Clepdec...' and a 'demo' user profile icon.



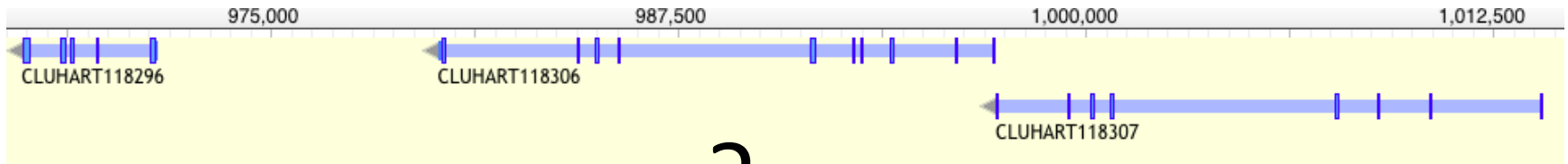
# Here is your annotation



# Functional annotation



Right, now we have our genes, but what do they do?



?

?

?

Insulin receptor?

Vesicle-trafficking protein?

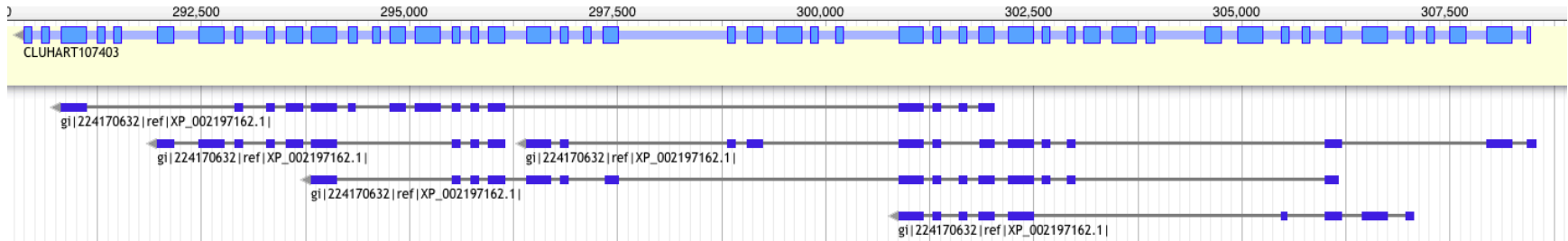
Alcohol dehydrogenase?

Aquaporin?

Transcription factor

MAP kinase kinase kinase?

# But we have used proteins in our annotation!



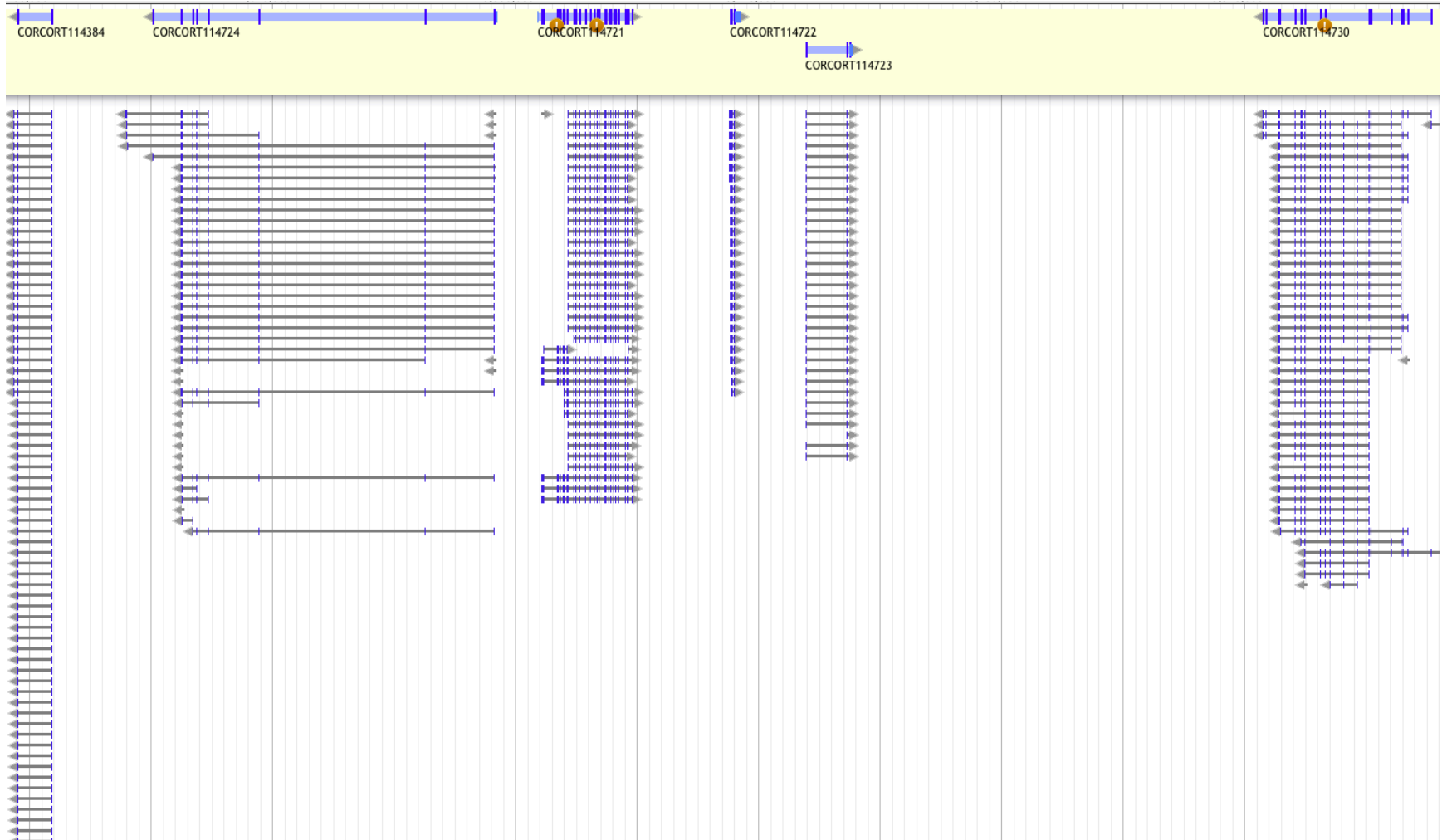
It is actually kind of complex...

... and Maker does not do this for you.

## Extract sequences -> functional annotation

- Extract sequences from Webapollo
- Annotate the sequences functionally using Blast and InterProScan

# Blast-based annotation - proteins



## Blast-based approach

- Fairly fast and easy
- Orthology not certain - best blast-hit does not equal orthologous!

# Interproscan

www.ebi.ac.uk/interpro/interproscan.html

Genome asse...b Wiki Sill BILS projects Wikipedia Google Maps

An Introduction to the Gene Ontology

About InterProScan 5 < InterPro < EMBL-EBI

EMBL-EBI

Services Research Training About us

InterPro  
Protein sequence analysis & classification

Search InterPro... Search

Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

Home Release notes Training & tutorials FAQs Download About InterPro Contact

## About InterProScan

### What is InterProScan?

InterProScan is the software package that allows sequences (protein and nucleic) to be scanned against InterPro's signatures. Signatures are predictive models, provided by several different databases (referred to as member databases), that make up the InterPro consortium.

The software is available:

- As a web-based tool, using the sequence search box on the [InterPro homepage](#), for the analysis of single protein sequences (also available in the [EBI tool section](#))
- Programmatically via Web services that allow up to 25 sequences to be analysed per request (both [SOAP](#) and [REST](#)-based services are available)
- As a downloadable package for local installation from the EBI's FTP server, for instructions see the [detailed documentation pages](#).

InterProScan is run regularly against UniProtKB and the results are made available via the InterPro website.

### More information

For more information, and for instructions on how to obtain, install and run InterProScan, please see the [detailed documentation pages](#).

### Publications



[InterProScan 5: genome-scale protein function classification](#)  
Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sebastien Pesseat, Antony F. Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter

*Bioinformatics*, Jan 2014  
(doi:10.1093/bioinformatics/btu031)  
[HTML](#) - [PDF \(324Kb\)](#)

EMBL-EBI

News  
Brochures  
Contact us  
Intranet

Services

- By topic
- By name (A-Z)
- Help & Support

Research

- Overview
- Publications
- Research groups
- Postdocs & PhDs

Training

- Overview
- Train at EBI
- Train outside EBI
- Train online
- Contact organisers

Industry

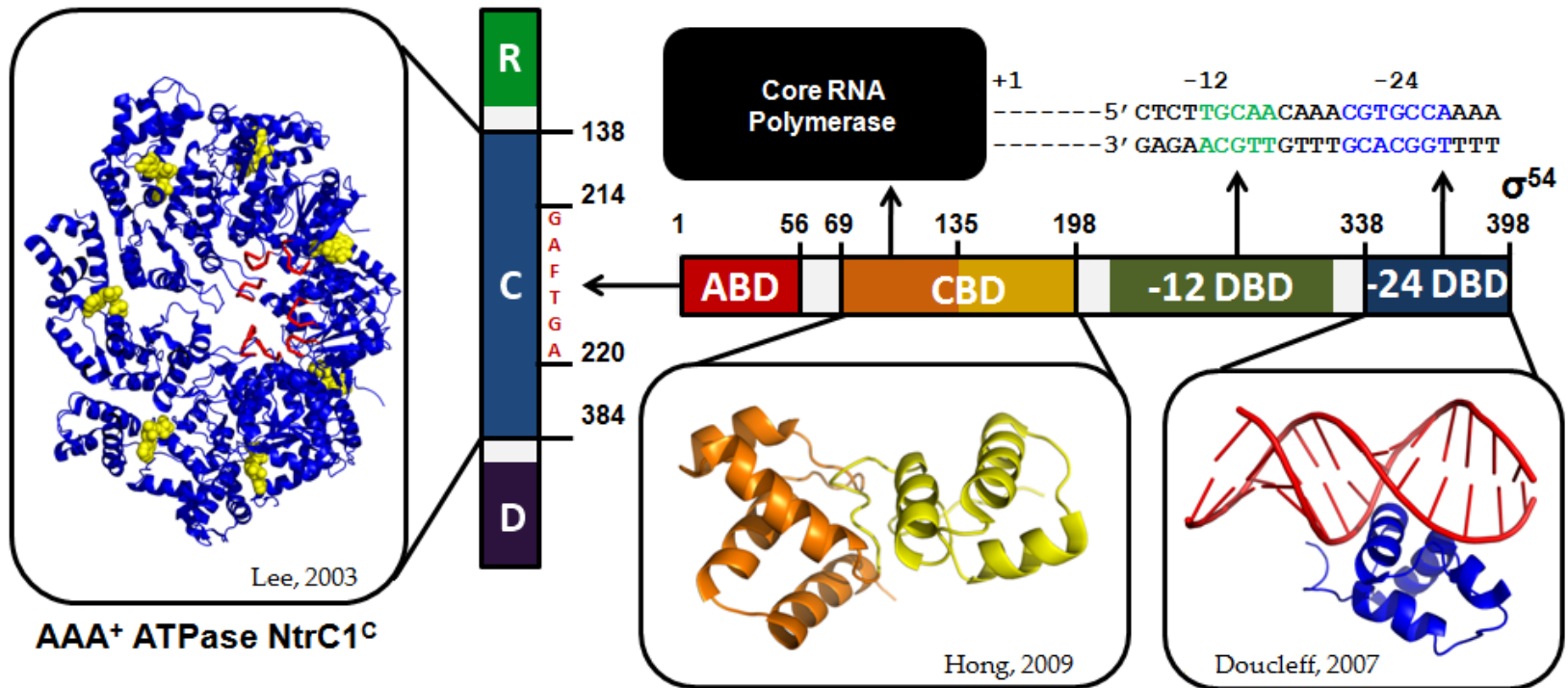
- Overview
- Members Area
- Workshops
- SME Forum
- Contact Industry programme

About us

- Overview
- Leadership
- Funding
- Background
- Collaboration
- Jobs
- People & groups
- News



# Sequence domains

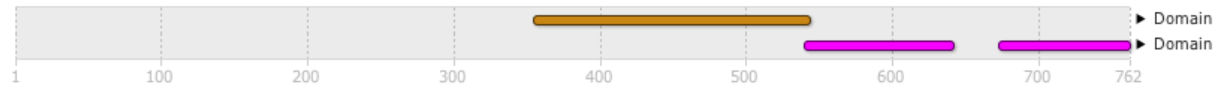


# Interproscan results

## Protein family membership

- F Crotonase superfamily (IPR001753)
- F Fatty acid oxidation complex, alpha subunit, mitochondrial (IPR012803)

## Domains and repeats



## Detailed signature matches



# Interproscan results - GO terms

## GO term prediction

### Biological Process

- [GO:0006631](#) fatty acid metabolic process
- [GO:0006635](#) fatty acid beta-oxidation
- [GO:0008152](#) metabolic process
- [GO:0055114](#) oxidation-reduction process

### Molecular Function

- [GO:0003824](#) catalytic activity
- [GO:0003857](#) 3-hydroxyacyl-CoA dehydrogenase activity
- [GO:0004300](#) enoyl-CoA hydratase activity
- [GO:0016491](#) oxidoreductase activity
- [GO:0016616](#) oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
- [GO:0050662](#) coenzyme binding

### Cellular Component

- [GO:0005739](#) mitochondrion
- [GO:0016507](#) mitochondrial fatty acid beta-oxidation multienzyme complex

# Gene Ontology

The screenshot shows a browser window titled "The Gene Ontology" with the address bar containing "www.geneontology.org". The website header features the "the Gene Ontology" logo and a search bar. A navigation menu includes links for Downloads, Tools, Documentation, Projects, About, and Contact. The main content area begins with a "Welcome to the Gene Ontology website!" message, followed by a paragraph describing the project's goal of standardizing gene and gene product attributes. A search box is provided for the Gene Ontology Database, with instructions to use AmiGO. Below this, there is a paragraph encouraging community input and another paragraph about funding sources. On the right side, there are two sidebar sections: "Quick Links" with various tools and download options, and "News" with social media links for Twitter, RSS, and Facebook. The footer contains copyright information and links to Helpdesk, Cite, Terms of use, News, and RSS. A status bar at the bottom indicates the current page is "http://www.geneontology.org/GO.downloads.annotations.shtml".

The Gene Ontology

Search  go!

[Downloads](#) [Tools](#) [Documentation](#) [Projects](#) [About](#) [Contact](#)

## Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides [a controlled vocabulary of terms](#) for describing gene product characteristics and [gene product annotation data](#) from GO Consortium members, as well as [tools to access and process this data](#). [Read more about the Gene Ontology...](#)

### Search the Gene Ontology Database

Search for genes, proteins or GO terms using [AmiGO](#):

GO!

[AmiGO](#) is the official GO browser and search engine.

The Gene Ontology project very much encourages input from the community into both the content of the GO and annotation using GO. We are very happy to work with others to ensure that the GO is both complete and accurate, and we also very much encourage communities to submit GO annotations for inclusion in the GO database. [Please contact us.](#)

The Gene Ontology Consortium is supported by a U41 grant from the National Human Genome Research Institute (NHGRI) [grant HG002273]. [See the full list of funding sources](#). The Gene Ontology Consortium would like to acknowledge the assistance of many more people than can be listed here. Please visit the [acknowledgements page](#) for the full list.

#### Quick Links

- Tools
- [AmiGO browser](#)
- [Submit GO Annotations](#)
- [OBO-Edit ontology editor](#)
- [Ontology downloads](#)
- [Annotation downloads](#)
- [Database downloads](#)
- [Documentation](#)
- [GO FAQ](#)
- [GO on SourceForge](#)
- [Contact GO](#)

#### News

- [GO on Twitter](#)
- Finding updates...
- [GO newsdesk](#)
- [GO news RSS feed](#)
- [GO on Facebook](#)

Copyright © 1999-2013 [the Gene Ontology](#)

[Helpdesk](#) • [Cite](#) • [Terms of use](#) • [News](#) • [RSS](#)

Member of the [Open Biological and Biomedical Ontologies](#)

Open "http://www.geneontology.org/GO.downloads.annotations.shtml" in a new tab

## Gene Ontology

- C: Cellular Component, e.g., endoplasmatic reticulum, integral to plasma membrane
- P: Biological Process, e.g., pyrimidine metabolic process
- F: Molecular Function, e.g., catalytic acticity, transporter activity

# Use the eBioKit

MRS

Home Blast Align Status Admin

Search All Databanks for Search

Show menu

## Welcome!

This is MRS, a search engine for biological and medical databanks. Use it to search well over a terabyte of indexed text.

## Usage

Usage of MRS should be straightforward. Just type some search terms in the search field at the top of the window and hit the Search button. More complex queries can be used as well.

## New version 6

MRS 6, or M6, is a complete rewrite of MRS.

## Feedback

If you have suggestions for improvement, please mail to M.L. Hekkelman. Or even better, use the mailinglist to post and discuss them. You can find the mailinglist at

<https://lists.berlios.de/mailman/listinfo/mrs-user>

## SOAP access

This MRS server can be accessed using SOAP, the wsdl's are located at:

<http://172.26.27.13:18090/mrsws/search/wsdl>  
<http://172.26.27.13:18090/mrsws/blast/wsdl>

## Open Source

MRS was designed and implemented by Maarten Hekkelman at the CMBI with the help and contributions from many others. Version 6 is distributed under a Boost license. You can find the source code for this version at

<ftp://ftp.cmbi.ru.nl/pub/software/mrs/>

## Abstract

The biological data explosion of the 'omics' era requires fast access to many data types in rapidly growing data banks. The MRS software provides the tools to rapidly and reliably download, store, index, and query flat-file databanks. Data stored and indexed by MRS takes considerably less space on disk than the raw data, despite that these raw data are included. The MRS index information is part of the stored data. Therefore, public and private data can be combined by simple concatenation and thus without computational overheads.

When using this server or the software, please refer to:

*MRS: A fast and compact retrieval system for biological data. Hekkelman M.L., Vriend G. Nucleic Acids Research 2005 33(Web Server issue):W766-W769; doi:10.1093/nar/gk442.*