

# Module 5

## NCBI's Zebrafish Resources



## Workshop Overview

- How to access information for a known gene
- Viewing genes in context of a Genome Browser
- Managing data tracks in the Genome Browser
- BLAST within the genome browser
- Comparative genomics
- NCBI resources for other fish species
- Additional NCBI resources

## How do I search NCBI for my gene of interest?

- Start from the NCBI homepage ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov))
- Choose the 'Gene' database from the dropdown menu
- Query by organism name combined with gene name or gene symbol
  - Example: enter 'Zebrafish pitx2'

Query by organism and gene name/symbol

Query by organism and gene name/symbol

Search results are displayed as a tabular list of Gene records that correspond to a single gene from a specific organism. The table can be further sorted by defined filters in the left column or by taxon on the right column.

Tabular list of results

Result counts sorted by taxonomy

Query by advanced searching

Filters allow you to narrow your search results by defined criteria

| Name/Gene ID                          | Description   | Location   | Aliases                                    | MIM    |
|---------------------------------------|---|--|--|--------|
| <a href="#">pitx2</a><br>ID: 30164    | paired-like homeodomain 2 [ <i>Danio rerio</i> (zebrafish)]               | Chromosome 14, NC_007125.7 (36220460..36233365)              | pitx2ac, zgc:110508, pitx2                 |        |
| <a href="#">tdgf1</a><br>ID: 30304    | teratocarcinoma-derived growth factor 1 [ <i>Danio rerio</i> (zebrafish)] | Chromosome 10, NC_007112.7 (1955474..1961930, complement)    | cb95, crypto, oep, wu.fe01604, zgc:109629  |        |
| <a href="#">dtx2b</a><br>ID: 30657    | distal-less homeobox 2b [ <i>Danio rerio</i> (zebrafish)]                 | Chromosome 1, NC_007112.7 (30977936..30979683, complement)   | dx5  |        |
| <a href="#">ikzf5</a><br>ID: 402893   | IKAROS family zinc finger 5 [ <i>Danio rerio</i> (zebrafish)]             | Chromosome 17, NC_007128.7 (21817383..21822198)              | wu.fb56f09, zgc:92405                      |        |
| <a href="#">PAX6</a><br>ID: 5080      | paired box 6 [ <i>Homo sapiens</i> (human)]                               | Chromosome 11, NC_000011.10 (31784792..31817961, complement) | AN, AN2, ASGD5, D11S812E, FVH1, MGDA, WAGR | 607108 |
| <a href="#">Hoxa1</a><br>ID: 15394    | homeobox A1 [ <i>Mus musculus</i> (house mouse)]                          | Chromosome 6, NC_000072.6 (52155367..52159286, complement)   | ERA1, Hox-1.6                              |        |
| <a href="#">PHACTR1</a><br>ID: 221692 | phosphatase and actin regulator 1 [ <i>Homo sapiens</i> (human)]          | Chromosome 6, NC_000006.12 (12716635..13290462)              | RPEL, RPEL1, dJ257A7.2                     | 608723 |

## What resources will I find on the Gene page?

The Gene record provides access to gene-specific information including nomenclature, reference sequences, genomic sequence viewer, publications and functional data.

### pitx2 paired-like homeodomain 2 [ *Danio rerio* (zebrafish) ]

Gene ID: 30164, updated on 3-Jun-2018

#### Summary

**Official Symbol** [pitx2](#) provided by [ZNC](#)  
**Official Full Name** [paired-like homeodomain 2](#) provided by [ZNC](#)  
**Primary source** [ZFIN:ZDB-GENE-990714-27](#)  
**See related** [Ensembl:ENSDARG00000036194](#) [Vega:OTTARG00000021245](#)  
**Gene type** protein coding  
**RefSeq status** PROVISIONAL  
**Organism** [Danio rerio](#)  
**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio  
**Also known as** [pitx2a](#); [pitx2c](#); [zgc:110508](#)  
**Orthologs** [human](#) [mouse](#) [all](#)

Nomenclature, taxonomy and links to related external resources

Links to orthologs

#### Genomic context

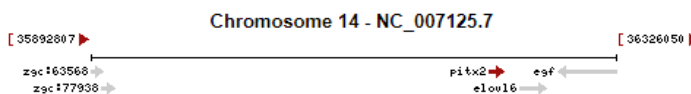
See [pitx2](#) in [Genome Data Viewer](#)

Location: chromosome: 14

Exon count: 5

| Annotation release  | Status            | Assembly                                      | Chr | Location                            |
|---------------------|-------------------|---|-----|-------------------------------------|
| <a href="#">106</a> | current           | GRCz11<br>( <a href="#">GCF_000002035.6</a> ) | 14  | NC_007125.7<br>(36220460..36233365) |
| <a href="#">105</a> | previous assembly | GRCz10<br>( <a href="#">GCF_000002035.5</a> ) | 14  | NC_007125.6<br>(35880168..35889514) |

Genomic location of NCBI's gene annotation on both current and past assemblies



Graphical display of genomic context

#### Genomic regions, transcripts, and products

#### Bibliography

#### Variation

#### Pathways from BioSystems

#### General gene information

#### General protein information

#### NCBI Reference Sequences (RefSeq)

Additional resources including publications, GeneRIFs, Gene Ontology (from GO Consortium) and RefSeq transcripts that are derived from a combination of automation and manual curation

## How do I access the Reference Sequences (RefSeqs) for my gene of interest?

- Expand the 'NCBI Reference Sequences (RefSeq)' section of the Gene record
- Expanded view shows both curated (NM\_, NR\_ and NP\_) and model (XM\_, XR\_ and XP\_) RefSeqs

NM\_/NR\_/NP\_ accessions represent curated RefSeqs

NC\_ accessions represent the RefSeq genomic assembly. Quick access to the FASTA for the genomic range of the gene

XM\_, XR\_, XP\_ accessions represent RefSeqs generated by NCBI's automated pipeline

**NCBI Reference Sequences (RefSeq)**

RefSeqs maintained independently of Annotated Genomes

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

**mRNA and Protein(s)**

- [NM\\_130975.2](#) → [NP\\_571050.1](#) pituitary homeobox 2  
[See identical proteins and their annotated locations for NP\\_571050.1](#)

Status: PROVISIONAL

Source sequence(s) [AF132446](#)

UniProtKB/Swiss-Prot [Q9W5Z2](#)

Related [ENSDDARP0000052568.6](#), [OTTDDARP00000022079](#), [ENSDART0000052569.7](#), [OTTDA](#), [T00000026935](#)

Conserved Domains (2) [summary](#)

|                           |                           |
|---------------------------|---------------------------|
| <a href="#">pfam00046</a> | Homeobox; Homeobox domain |
| Location:86 → 138         |                           |
| <a href="#">pfam03826</a> | OAR; OAR domain           |
| Location:272 → 289        |                           |

RefSeqs of Annotated Genomes: *Danio rerio* Annotation Release 106 [details...](#)

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

### Reference GRCz11 Primary Assembly

#### Genomic

- [NC\\_007125.7](#) Reference GRCz11 Primary Assembly  
Range 36220460..36233365  
Download [GenBank](#), [FASTA](#), [Sequence](#)

#### mRNA and Protein(s)

- [XM\\_005157307.4](#) → [XP\\_005157364.1](#) pituitary homeobox 2  
[See identical proteins and their annotated locations for XP\\_005157364.1](#)
- UniProtKB/Swiss-Prot [Q9W5Z2](#)
- Related [ENSDDARP00000120155.1](#)
- Conserved Domains (2) [summary](#)
- |                           |                           |
|---------------------------|---------------------------|
| <a href="#">pfam00046</a> | Homeobox; Homeobox domain |
| Location:41 → 93          |                           |
| <a href="#">pfam03826</a> | OAR; OAR domain           |
| Location:227 → 244        |                           |

Nucleotide

GenBank

Send to

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Articles about the *pitx2* gene

Functional characterization of zebrafish orthologs of the human *[PLoS One. 2017]*

Mutations of conserved non-coding elements of *PITX* *[Hum Mol Genet. 2017]*

Epigenetic regulation of left-right asymmetry by DNA meth *[EMBO J. 2011]*

See all

Pathways for the *pitx2* gene

Canonical wnt signaling

TGF-beta signaling pathway

NCBI Reference Sequence: [NM\\_130975.2](#)

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

[Go to](#)

LOCUS [NP\\_130975](#) 1963 bp mRNA linear VRT 28-MAR-2018

DEFINITION *Danio rerio* paired-like homeodomain 2 (*pitx2*), mRNA.

ACCESSION [NP\\_130975](#)

VERSION [NP\\_130975.2](#)

KEYWORDS RefSeq

SOURCE *Danio rerio* (zebrafish)

ORGANISM *Danio rerio*  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; *Danio*.

REFERENCE 1 (bases 1 to 1963)  
AUTHORS Wang L, Liu Z, Lin H, He D, Tao Q and Liu F.  
TITLE Epigenetic regulation of left-right asymmetry by DNA methylation  
JOURNAL *EMBO J.* 36 (20), 2987-2997 (2017)

PUBMED 28832887

REFERENCE 2 (bases 1 to 1963)  
AUTHORS Heh E, Takeuchi H, Hübelsen S, Halktunger SS and Semina RV.  
TITLE Functional characterization of zebrafish orthologs of the human Beta 3-Glucosyltransferase B3GLCT gene mutated in Peters Plus Syndrome  
JOURNAL *PLoS ONE* 12 (9), e0184903 (2017)

PUBMED 28225587

REMARK Publication Status: Online-Only

REFERENCE 3 (bases 1 to 1963)  
AUTHORS Protas HE, Heh E, Footz T, Kasberger J, Baraban SC, Levin AV, Katz LJ, Ritch R, Walter PH, Semina RV and Gould DB.  
TITLE Mutations of conserved non-coding elements of *PITX2* in patients with ocular dysgenesis and developmental glaucoma  
JOURNAL *Hum. Mol. Genet.* 26 (18), 3630-3638 (2017)

PUBMED 28211203

Nucleotide record provides access to FASTA sequence, feature annotation and sequence analysis tools – BLAST and Genome Data Viewer

### Exploring the Genomic context section of the Gene Page

The default view of the 'Genomic regions, transcripts and products' section of the Gene page which shows the gene annotation on the current reference assembly. This section provides a quick link to NCBI's Genome browser as well as access to download and BLAST the Refseqs.

Current and previous annotation

Link to view annotation in NCBI's Genome Data Viewer (GDV)

The screenshot displays the 'Genomic context' section of a gene page. At the top, it shows the location on chromosome 14 and the exon count (5). Below this is a table comparing the current annotation (Release 106, GRCz11) with the previous assembly (Release 105, GRCz10). A genomic map shows the gene structure on chromosome 14 with exons and introns. The 'Genomic regions, transcripts, and products' section is expanded, showing the genomic sequence and various tracks including RefSeq gene models (pitx2), RNA-seq exon coverage, and RNA-seq intron features. A tooltip window is open over the RefSeq track, providing detailed information about the gene and links to download and BLAST the RefSeqs.

| Annotation release  | Status            | Assembly                                   | Chr | Location                         |
|---------------------|-------------------|--|-----|----------------------------------|
| <a href="#">106</a> | current           | GRCz11 ( <a href="#">GCF_000002035.6</a> ) | 14  | NC_007125.7 (36220460..36233365) |
| <a href="#">105</a> | previous assembly | GRCz10 ( <a href="#">GCF_000002035.5</a> ) | 14  | NC_007125.6 (35880168..35889514) |

Chromosome 14 - NC\_007125.7

Genomic regions, transcripts, and products

Genomic Sequence: [NC\\_007125.7](#) Chromosome 14 Reference GRCz11 Primary Assembly

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

**Gene: pitx2**  
**Title:** paired-like homeodomain 2  
**Location:** 36,220,460..36,233,365  
**Length:** 12,906  
**Position:** 540,321,874  
**Merged features:** NP\_571050.1 and NM\_130975.2  
**Download:** [NP\\_571050.1](#), [NM\\_130975.2](#)

**Links & Tools**  
[View GeneID: 30164 \(pitx2\)](#)  
[View ZFIN: ZDB-GENE-990714-27](#)

**BLAST Genome-specific:** [NC\\_007125.7 \(36,220,460..36,233,365\)](#)  
**BLAST Genomic:** [NC\\_007125.7 \(36,220,460..36,233,365\)](#)  
**FASTA View:** [NC\\_007125.7 \(36,220,460..36,233,365\)](#)  
**GenBank View:** [NC\\_007125.7 \(36,220,460..36,233,365\)](#)

Hovering over the RefSeq opens a window that allows for quick access to download and BLAST

## How do I search for my gene of interest in NCBI's Genome Data Viewer (GDV)?

Gene searches can also be performed in NCBI's new genome browser which offers easy access to the expansive range and number of organisms and assemblies represented in the GDV browser.

- Starting from the landing page
  - Select an organism by entering an organism name (common or scientific) in the search box
    - Example: 'Zebrafish'
  - Or select an organism from the tree. More organism can be exposed by clicking on the nodes within the tree
  - The knowledge panel to the right will update to the selected organism

The screenshot displays the NCBI Genome Data Viewer (GDV) interface. At the top, the NIH logo and navigation links are visible. The main heading is "Genome Data Viewer". Below this, there is a "Select organism" search box containing "Danio rerio (zebrafish)". To the left is a taxonomic tree with various organisms, including yeast, nematode, zebrafish, chicken, rat, mouse, pig, sheep, cattle, dog, horse, maize, rice, Arabidopsis, and Plasmodium falciparum 3D7. A callout box points to the search box with the text "Search by common or scientific names at multiple taxonomic levels". Another callout box points to the right-hand panel with the text "Species-specific knowledge panel". A third callout box points to the zebrafish node in the tree with the text "Update knowledge panel to a new species by clicking on the organism". A fourth callout box points to the expanded tree view with the text "Clicking the tree nodes expands to show more organisms".

The right-hand panel, titled "Danio rerio (zebrafish) genome", contains a search box for the genome, a dropdown menu for the assembly (GRCz11), and buttons for "Browse genome" and "BLAST genome". Below this, there are sections for "Assembly details" and "Annotation details".

**Assembly details**

|                   |                             |
|-------------------|-----------------------------|
| Name              | GRCz11                      |
| RefSeq accession  | GCF_000002035.6             |
| GenBank accession | GCA_000002035.4             |
| Download via FTP  | RefSeq, GenBank             |
| Submitter         | Genome Reference Consortium |
| Level             | Chromosome                  |
| Category          | Reference genome            |

**Annotation details**

|                    |            |
|--------------------|------------|
| Annotation Release | 106        |
| Release date       | 2017-06-26 |

At the bottom of the knowledge panel, there is a chromosome ideogram showing 25 chromosomes.

## Using GDV's knowledge panel

The GDV knowledge panel provides multiple modes of access to the genome browser

- Search by query term. The search box accepts a broad range of search terms such as gene names and symbols, protein names, dbSNP ids (human only), phenotypes or sequence accessions.
  - Example query 'pitx2'
- Select Assembly version 'GRCz11' which defaults to the most current reference assembly
- Alternatively, use the 'Browse Genome' button which allows for searching directly in the genome browser view
- The 'BLAST Genome' button allows for a nucleotide or translated protein search against the selected genome

**Danio rerio (zebrafish) genome**

Search in genome  
pitx2

Examples: myod1, chr25:31420000-31425000, DNA repair

Assembly  
GRCz11

[Browse genome](#) [BLAST genome](#)

**Assembly details**

|                          |  |
|--------------------------|--|
| <b>Name</b>              | GRCz11   |
| <b>RefSeq accession</b>  | <a href="#">GCF_000002035.6</a>                  |
| <b>GenBank accession</b> | <a href="#">GCA_000002035.4</a>                  |
| <b>Download via FTP</b>  | <a href="#">RefSeq</a> , <a href="#">GenBank</a> |
| <b>Submitter</b>         | <a href="#">Genome Reference Consortium</a>      |
| <b>Level</b>             | Chromosome                                       |
| <b>Category</b>          | Reference genome                                 |

**Annotation details**

|                           |            |
|---------------------------|------------|
| <b>Annotation Release</b> | 106        |
| <b>Release date</b>       | 2017-06-26 |

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Callouts:

- Search by gene symbol, name genomic location
- Select assembly you want to search
- FTP download access to the curated RefSeq and archival GenBank assemblies
- Detailed information on the RefSeq annotation including gene counts and RNA-seq data used in the annotation
- Select browser view by chromosome



## What is displayed in the GDV default view?

A text search for 'pitx2' returns the default GDV browser view zoomed into the pitx2 gene on chr 14. The browser display consists of a series of widgets on the left that control the graphical display on the right. You can customize the look and feel of the control panel. Use the toggle icon found at the left side of each widget header to collapse or expand the various panels. Re-order the panels by clicking in the headers and dragging and dropping them into their desired location. The Sequence Viewer displays the chromosomal location and the annotated gene features.


Page Widgets can be used to update location shown in the Sequence Viewer

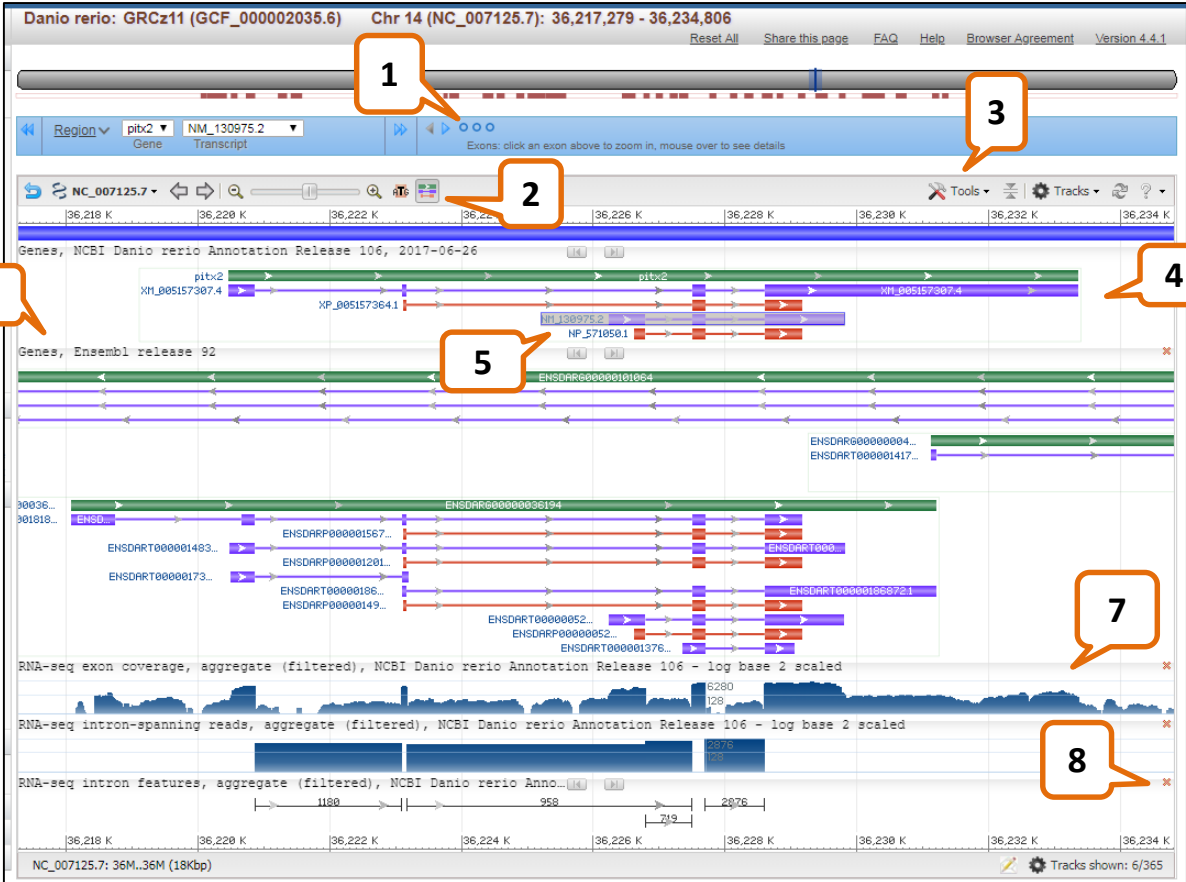
Chromosomal location displayed in the sequence viewer



Sequence viewer panel shows the current annotation of the selected gene along with default tracks for Ensembl's annotation and RNA-seq intron and exon aggregate reads

## How do I adjust my view within sequence viewer?

1. Exon navigator allows you to jump between gene features within a view or between exons
2. The 'Show All' icon  expands the green gene bar to show transcripts (purple) and proteins (red)
3. 'Tools' allows you to adjust the gene orientation, view sequence text, and create pdfs
4. RefSeq annotation – default view shows both model and curated RefSeqs
5. RefSeq transcripts shaded in grey indicate discrepancies with the reference genome
6. Ensembl annotation
7. RNA-seq exon and intron aggregate tracks
8. Clicking on the orange 'x' removes the data track



The screenshot displays the NCBI Zebrafish genome browser interface for the *pitx2* gene on Chromosome 14 (NC\_007125.7). The interface includes several tracks and tools:

- 1. Exon navigator:** A horizontal bar at the top showing the gene structure with exons and introns.
- 2. Show All icon:** A small icon in the navigation bar that expands the gene bar to show transcripts (purple) and proteins (red).
- 3. Tools menu:** A dropdown menu in the navigation bar for adjusting gene orientation, viewing sequence text, and creating PDFs.
- 4. RefSeq annotation:** A track showing the RefSeq model and curated RefSeqs.
- 5. RefSeq transcripts:** Transcripts are shown with grey shading to indicate discrepancies with the reference genome.
- 6. Ensembl annotation:** A track showing the Ensembl annotation for the gene.
- 7. RNA-seq exon coverage:** A track showing RNA-seq exon coverage, aggregate (filtered), NCBI Danio rerio Annotation Release 106 - log base 2 scaled.
- 8. RNA-seq intron features:** A track showing RNA-seq intron features, aggregate (filtered), NCBI Danio rerio Anno... - log base 2 scaled.

## How do I display additional data tracks?

In addition to the default display, you can view other data tracks by selecting the 'Tracks' button which opens a dialog box containing additional data tracks. In the example shown, the 'Expression' tab offers the choice of tissue-specific RNA-seq intron and exon reads.

The screenshot displays the NCBI Zebrafish genome browser interface for the *pitx2* gene on Chromosome 14 (NC\_007125.7). The 'Tracks' button is highlighted with an orange box. A 'Configure Page' dialog box is open, showing the 'Expression' tab selected. The 'Active Tracks' section lists several RNA-seq tracks, with 'RNA-seq exon coverage, aggregate (filtered), NCBI Danio rerio Annotation Release 106' and 'RNA-seq intron-spanning reads, aggregate (filtered), NCBI Danio rerio Annotation Release 106' selected. The 'Track Settings' section shows the selected track settings: 'RNA-seq exon coverage, aggregate (filtered), NCBI Danio rerio Annotation Release 106'. The main browser view shows the gene structure and the selected RNA-seq tracks, with a callout box pointing to the tracks and the text 'Tissue-specific RNA-seq intron and exon tracks'.

**Configure Page**

Tracks Custom Data

Active Tracks

| Active                              | Track name  |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | RNA-seq exon coverage, aggregate (filtered), NCBI Danio rerio Annotation Release 106                      |
| <input type="checkbox"/>            | RNA-seq exon coverage, aggregate (filtered, unique hits), NCBI Danio rerio Annotation Release 106         |
| <input checked="" type="checkbox"/> | RNA-seq intron-spanning reads, aggregate (filtered), NCBI Danio rerio Annotation Release 106              |
| <input type="checkbox"/>            | RNA-seq intron-spanning reads, aggregate (filtered, unique hits), NCBI Danio rerio Annotation Release 106 |
| <input checked="" type="checkbox"/> | RNA-seq intron features, aggregate (filtered), NCBI Danio rerio Annotation Release 106                    |
| <input type="checkbox"/>            | RNA-seq intron features, aggregate (filtered, unique hits), NCBI Danio rerio Annotation Release 106       |
| <input type="checkbox"/>            | RNA-seq intron features, aggregate (unfiltered), NCBI Danio rerio Annotation Release 106                  |

Track Settings: RNA-seq exon coverage, aggregate (filtered), NCBI Danio rerio Annotation Release 106

Exon coverage of RNA-seq alignments, filtered to remove low abundance alignments and some apparent retained-intron alignments

**Tissue-specific RNA-seq intron and exon tracks**

## How do I view protein and mRNA feature annotation?

Gene features can be projected onto the transcript and protein products by selecting the 'Genes' tab within the 'Tracks' menu then highlighting 'Genes, Known RefSeqs, NCBI Danio rerio Annotation release 106' and using the 'Track Settings' section to configure the display. Selecting 'Product Features' option displays feature annotation in the GDV Sequence Viewer.

The screenshot shows the NCBI Genome Data Viewer (GDV) interface. The main view displays genomic tracks for the *pitx2* gene region. A 'Configure Page' dialog box is open, showing the configuration for the 'NCBI Genes Track'. The 'Active Tracks' section lists 'Genes, Known RefSeqs, NCBI Danio rerio Annotation Release 106'. The 'Track Settings' section shows 'Rendering options' set to 'Show all' and 'Product Features' checked under 'Other settings'.

Annotations in the main view include:

- Genes, NCBI Danio rerio Annotation Release 106, 2017-06-26
- Genes, Ensembl release 92
- Transcripts: XM\_005157307.4, XP\_005157364.1, NP\_005157364.1
- Protein products: NP\_005157364.1

The screenshot shows a zoomed-in view of the NCBI Genome Data Viewer (GDV) interface. The main view displays genomic tracks for the *pitx2* gene region. A 'Feature annotation' dialog box is open, showing detailed information for a specific DNA base contact.

Annotations in the main view include:

- Genes, NCBI Danio rerio Annotation Release 106, 2017-06-26
- Genes, Ensembl release 92
- Transcripts: XM\_005157307.4, XP\_005157364.1, NP\_005157364.1
- Protein products: NP\_005157364.1
- Other features: DNA binding site L., Homeobox, specific DNA base ...

Feature annotation details:

- Feature:** specific DNA base contacts [nucleo...
- site:** other
- Title:** specific DNA base contacts [nucleotide binding]
- Location:** 85..135
- [Length]**
- Span:** 51
- Placed:** 6
- Position:** 491,482,387
- Links & Tools**
- View CDD:** [238038](#)
- BLAST Genome-specific:** [NP\\_571050.1 \(85..135\)](#)
- BLAST protein:** [NP\\_571050.1 \(85..135\)](#)
- FASTA View:** [NP\\_571050.1 \(85..135\)](#)
- GenBank View:** [NP\\_571050.1 \(85..135\)](#)

## How can I view RefSeq annotation on a previous version of the reference assembly?

RefSeq annotations on previous versions of the reference assembly can be viewed using the 'Comparative genomics' tab and selecting previous assemblies. The annotations can be projected onto the assembly alignment using the 'Project features' setting.

In some cases cross-species alignments can be provided if the assemblies are similar enough.

The screenshot displays the NCBI RefSeq browser interface for the gene *tunor* (NC\_007128.7) on chromosome 17 of *Danio rerio*. The interface is divided into several sections:

- Top Panel:** Shows the gene name, transcript, and a genomic map with coordinates from 3,050 K to 3,020 K.
- Configuration Page (Modal Window):**
  - Tracks Tab:** Lists active tracks. Three tracks are selected:
    - Assembly-assembly alignments, GRCz11 (GCF\_000002035.6) vs. GRCz10 (GCF\_000002035.5)
    - Assembly-assembly alignments, GRCz11 (GCF\_000002035.6) vs. Zv7 (GCF\_000002035.1)
    - Assembly-assembly alignments, GRCz11 (GCF\_000002035.6) vs. Zv8 (GCF\_000002035.3)
    - Assembly-assembly alignments, GRCz11 (GCF\_000002035.6) vs. Zv9 (GCF\_000002035.4)
  - Track Settings:** For the selected track, settings include:
    - Alignment Display: Adaptive
    - Pileup Display: Match/Mismatch graph (count)
    - Score method: Show Differences
    - Unaligned Tails: Show Tail Length
    - Sort alignments by: No sorting
    - Other settings:  Link Mate Pairs,  Show Labels,  Show pileup,  Show Identical Bases,  Project features
- Main View:** Shows the genomic tracks for:
  - GRCz11:** The current reference assembly.
  - GRCz10:** A previous reference assembly.
  - Zv9:** A previous assembly version.

## How can I use GDV to view BLAST results in a graphical display?

GDV recently added BLAST functionality to the genome browser. Clicking on the 'BLAST genome' button opens a BLAST page that offers blastn, tblastn and tblastx searches against the selected reference assembly. The BLAST results page provides a direct link to the GDV alignment.

For this example we will perform a tblastn search with the RefSeq nucleotide accession NM\_131311.2

**Danio rerio (zebrafish) genome**

Search in genome  
Location, gene or phenotype

Assembly: GRCz11

[Browse genome](#) **BLAST genome**

**Assembly details**

- Name: GRCz11
- RefSeq accession: GCF\_000002035.6
- GenBank accession: GCA\_000002035.4
- Download via FTP: RefSeq, GenBank
- Submitter: Genome Reference Consortium
- Level: Chromosome
- Category: Reference genome

**BLAST** >> blastn suite

Danio rerio (zebrafish) RefSeq assembly GCF\_000002035.6

blastn **tblastn** tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s): **NM\_131311.2**

Or, upload file: Choose File No file chosen

Job Title: Enter a descriptive title for your BLAST search

Database: GRCz11 RefSeq assembly [GCF\_000002035.6] (1923 sequences)

Program Selection

Optimize for:

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

**BLAST** Search database genomic/7955/GCF\_000002035.6 using Megablast (Optimize for highly similar sequences)  Show results in a new window

Adjust search parameters

Graphic Summary

Distribution of the top 3 Blast Hits on 1 subject sequences

Color key for alignment scores

- <40
- 40-50
- 50-80
- 80-200
- >=200

Descriptions

Alignments

Download GenBank Graphics sort by: E value

Danio rerio strain Tuebingen chromosome 9, GRCz11 Primary Assembly  
Sequence ID: NC\_007120.7 Length: 56459846 Number of Matches: 3

| Range                 | Score            | Expect | Identities       | Gaps        | Strand | Plus/Minus |
|-----------------------|------------------|--------|------------------|-------------|--------|------------|
| 1: 3397407 to 3398443 | 1916 bits (1037) | 0.0    | 1037/1037 (100%) | 0/1037 (0%) |        |            |

Features: homeobox, crstn, Dlx2a

Query 668 AGGTTAAAATCTGGTTCCAGAACTGCTCCTTCAAAAGTTCAGAAAGTTGTGGAAAAGTGGAG 727  
Sbjct 3398443 AGGTTAAAATCTGGTTCCAGAACTGCTCCTTCAAAAGTTCAGAAAGTTGTGGAAAAGTGGAG 3398384

Query 728 AGATCCACCCGAGCAGATGTGGCTCCAGTGGAGTCTCCACCTCACCCCTCACCGCTC 787  
Sbjct 3398383 AGATCCACCCGAGCAGATGTGGCTCCAGTGGAGTCTCCACCTCACCCCTCACCGCTC 3398324

Query 788 TGCCTCCGCGCTTGGACTTCGGCACAGCAGAGATGACACTGTAAACTGGGTTTGT 847

**Related Information**

[New Genome Data Viewer - aligned genomic context](#)

Links to GDV

### How do I visualize my BLAST results in GDV?

**Genome Data Viewer**

- Ideogram View
- Search
- Your Data
- BLAST

HXPYZA2P015 (NM\_131311.2) Tools

Details

Alignments

| Query                      | Hit location | Score | Identity |
|----------------------------|--------------|-------|----------|
| <i>Chr9 (NC_007120.7)</i>  |              |       |          |
| 1 >> 484                   | 3,400.5K (-) | 874   | 100%     |
| 484 >> 671                 | 3,399.6K (-) | 340   | 100%     |
| 523 >> 670                 | 3,388.9K (+) | 133   | 80%      |
| 668 >> 1,704               | 3,397.9K (-) | 1871  | 100%     |
| 1,173 >> 1,211             | 46,249M (-)  | 39    | 82%      |
| <i>Chr1 (NC_007112.7)</i>  |              |       |          |
| 169 >> 304                 | 30.979M (-)  | 84    | 73%      |
| 484 >> 675                 | 30.979M (-)  | 158   | 78%      |
| 666 >> 767                 | 30.979M (-)  | 90    | 79%      |
| 1,173 >> 1,206             | 21.697M (-)  | 45    | 88%      |
| 1,173 >> 1,209             | 22.100M (-)  | 41    | 83%      |
| 1,173 >> 1,206             | 3,409.6K (-) | 39    | 85%      |
| <i>NW_018394493.1</i>      |              |       |          |
| 169 >> 304                 | 135.57K (-)  | 84    | 73%      |
| 484 >> 675                 | 135.09K (-)  | 158   | 78%      |
| 666 >> 767                 | 134.83K (-)  | 90    | 79%      |
| <i>Chr19 (NC_007130.7)</i> |              |       |          |
| 484 >> 672                 | 41.471M (-)  | 147   | 77%      |
| 518 >> 671                 | 41.466M (-)  | 70    | 70%      |

Only alignments on the current assembly are shown

Add Tracks

Using the 'BLAST Genome' button automatically populates the BLAST widget with recent BLAST searches. In this example we selected RID: HXPYZA2P015 which is then displayed in the GDV Sequence Viewer

The 'Tools' dropdown menu takes you to the BLAST page to run a new search

New BLAST  
Show BLAST  
Copy RID

BLAST alignment results are grouped by chromosomal location and can be sorted by any column or filtered by identity

The 'accessions menu' allows you to quickly jump within the table to specific sequence on which alignments are found.

| Sequence             | Ident | Cover |
|----------------------|-------|-------|
| ✓ Chr9 (NC_007120.7) | 100%  | 100%  |
| NW_018394493.1       | 79%   | 25%   |
| Chr1 (NC_007112.7)   | 79%   | 25%   |
| Chr12 (NC_007123.7)  | 87%   | 12%   |
| Chr19 (NC_007130.7)  | 77%   | 11%   |
| Chr3 (NC_007114.7)   | 71%   | 8%    |
| Chr14 (NC_007125.7)  | 74%   | 4%    |
| Chr13 (NC_007124.7)  | 79%   | 3%    |
| Chr17 (NC_007128.7)  | 84%   | 2%    |

## Viewing the BLAST result in the Sequence Viewer?

Continuing with this example (RID: HXPYZA2P015) we will use the 'alignment inspector' to select a BLAST hit to view in the Sequence Viewer.

**BLAST**

HXPYZA2P015 (NM\_131311.2) Tools

**Details**

Danio rerio distal-less homeobox 2a (dlx2a), mRNA

Program blastn  
Database GPIPE/7955/106/ref\_top\_level  
Length 1704

**Alignments**

| Query              | Hit location | Score | Identity |
|--------------------|--------------|-------|----------|
| Chr9 (NC_007120.7) |              | 874   | 100%     |
|                    |              | 340   | 100%     |

Selecting the 'BLAST Alignment Inspector' opens a detailed display of alignments located on the currently displayed assembly sequence. It provides a graphical display of the relationship of a query's alignments to the assembly and NCBI's current RefSeq gene annotation.

Selecting the 'Gene Hit' opens the BLAST alignment against the gene feature in the Sequence Viewer

Hovering over any exon or aligned region creates a highlight in the Alignment Inspector column and at the corresponding region of the sequence viewer display in GDV.

**BLAST Alignment Inspector** NM\_131311.2: Danio rerio distal-less homeobox 2a (dlx2a), mRNA

Query Length Score Gaps Identity Hit overview

**Gene Hit: dlx2a** NM\_131311.2

|              |       |       |    |      |
|--------------|-------|-------|----|------|
| 1 >> 484     | 484   | 874   | 0% | 100% |
| 484 >> 671   | 188   | 340   | 0% | 100% |
| 668 >> 1,704 | 1,037 | 1,871 | 0% | 100% |

**Gene Hit: dlx1a** NM\_131305.1

|            |     |     |    |     |
|------------|-----|-----|----|-----|
| 523 >> 670 | 148 | 132 | 1% | 80% |
|------------|-----|-----|----|-----|

**Gene Hit: hdac4** XM\_005172529.4

|                |    |    |    |     |
|----------------|----|----|----|-----|
| 1,173 >> 1,211 | 39 | 39 | 0% | 87% |
|----------------|----|----|----|-----|

3,396,506 << 3,396,693  
NC\_007120.7: 3,400,314  
Exon  
CDS range: 3,400,246 - 3,400,570  
Exon range: 3,400,246 - 3,400,729

**BLAST alignment**

NC\_007120.7: 3.4M..3.4M (3.5Kbp) C



## How can I use the BLAST widget to search for paralogs?

The BLAST widget can also be used to search to location of lower quality hits that may represent paralogs or gene family members. In this example the *dlx2a* transcript, NM\_131311.2 (RID# HXPYZA2P015) was searched against the zebrafish reference assembly. Additional genomic alignments can be selected from the Ideogram view or from the BLAST widget.

**BLAST hits to the full assembly are shown in the Ideogram view**

**Gene feature aligning to the BLAST hit**

**Mismatches against the genome are displayed in red**

The screenshot displays the NCBI Genome Data Viewer interface for the zebrafish reference assembly. The main window shows the BLAST Alignment Inspector for the query NM\_131311.2 (Danio rerio distal-less homeobox 2a (dlx2a), mRNA). The BLAST Results section shows several alignments, with the top hit being NM\_131207.2 (dlx2b) with a score of 874 and 100% identity. Other hits include zcche7 (NM\_001301388.2) and dnah8 (XM\_021479325.1). The BLAST Results section also displays RNA-seq exon coverage and intron-spanning reads for the query transcript. The Ideogram view on the left shows the genome with BLAST hits highlighted in orange. The BLAST Alignment Inspector shows the alignment of the query transcript to the genome, with mismatches highlighted in red.

| Gene Hit      | Length         | Score | Gaps | Identity | Hit overview |
|---------------|----------------|-------|------|----------|--------------|
| <i>dlx2b</i>  | 169 >> 304     | 136   | 84   | 0%       | 73%          |
|               | 484 >> 675     | 192   | 158  | 0%       | 78%          |
|               | 666 >> 767     | 102   | 89   | 0%       | 79%          |
| <i>zcche7</i> | 1,173 >> 1,206 | 34    | 44   | 0%       | 88%          |
| <i>dnah8</i>  | 1,173 >> 1,206 | 39    | 46   | 0%       | 85%          |

| Query              | Hit location   | Score        | Identity |      |
|--------------------|----------------|--------------|----------|------|
| Chr9 (NC_007120.7) | 1 >> 484       | 3,400.6K (-) | 874      | 100% |
|                    | 484 >> 671     | 3,386.5K (-) | 340      | 100% |
|                    | 523 >> 670     | 3,388.4K (+) | 133      | 80%  |
|                    | 666 >> 1,704   | 3,397.4K (-) | 1871     | 100% |
|                    | 1,173 >> 1,211 | 46.2.4M (+)  | 38       | 87%  |
| Chr1 (NC_007112.7) | 169 >> 304     | 30,979M (-)  | 84       | 73%  |
|                    | 484 >> 675     | 30,979M (-)  | 158      | 78%  |
|                    | 666 >> 767     | 30,979M (-)  | 90       | 79%  |
|                    | 1,173 >> 1,206 | 21,697M (-)  | 45       | 88%  |
|                    | 1,173 >> 1,209 | 22,100M (-)  | 41       | 83%  |
|                    | 1,173 >> 1,206 | 3,409.6K (-) | 39       | 85%  |

## How do I access NCBI's resources for other fish species?

Access to RefSeq annotated genomes for other fish species can be found at [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/all/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/). Genomes are organized by taxonomic groups. Opening the dropdown menu for each group provides a species list and links to FTP, BLAST (**B**), a detailed annotation report (**AR**) and the Genome Data Viewer (**GDV**).

Genome

[Limits](#) [Advanced](#)

[Eukaryotic Annotation Home](#) | [Documentation](#) | [Annotated Genomes](#) | [Annotation Policy](#) | [Request Annotation](#)

### Eukaryotic genomes annotated at NCBI

Hundreds of eukaryotic genomes have been annotated by the NCBI Eukaryotic Genome Annotation Pipeline (see [graphs](#)). The latest annotation release available for each genome is shown in the tables below. The tables are organized by taxonomic group and provide links to the annotation report, FTP site, genome BLAST page, and Genome Data Viewer page.

Only completed annotations are shown here. Please browse the [annotation runs currently in progress](#) to see what will become available in a few days.


[Show/Hide All](#)


- [Featured \(6\)](#)
- [Primates \(25\)](#)
- [Rodents \(20\)](#)
- [Even-toed ungulates and whales \(Cetartiodactyla\) \(21\)](#)
- [Other Mammals \(46\)](#)
- [Birds \(62\)](#)
- [Fish \(53\)](#)
- [Other Vertebrates \(17\)](#)
- [Insects \(99\)](#)
- [Other Invertebrates \(29\)](#)

FTP - FTP Download   **B** - Organism-specific BLAST   **AR** - Annotation Report   **GDV** - Genome Data Viewer

| Species   | RefSeq assembly(ies)  | Annotation Release | Freeze Date | Release Date | Links  |
|---|---|--------------------|-------------|--------------|--|
| <a href="#">Acanthochromis polyacanthus (spiny chromis)</a>                   | <a href="#">ASM210954v1 (GCF_002109545.1)</a>                 | 100                | 2017-08-03  | 2017-08-07   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Amphiprion ocellaris (clown anemonefish)</a>                      | <a href="#">AmpOce1.0 (GCF_002776465.1)</a>                   | 100                | 2017-12-11  | 2017-12-13   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Astyanax mexicanus (Mexican tetra)</a>                            | <a href="#">Astyanax_mexicanus-2.0 (GCF_000372685.2)</a>      | 102                | 2017-09-27  | 2017-10-03   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Austrofundulus limnaeus (bony fishes)</a>                         | <a href="#">Austrofundulus_limnaeus-1.0 (GCF_001266775.1)</a> | 100                | 2015-09-02  | 2015-09-17   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Boleophthalmus pectinirostris (great blue-spotted mudskipper)</a> | <a href="#">BP_fa (GCF_000788275.1)</a>                       | 100                | 2017-04-26  | 2017-04-29   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Callorhynchus milii (elephant shark)</a>                          | <a href="#">Callorhynchus_milii-6.1.3 (GCF_000165045.1)</a>   | 100                | 2014-04-29  | 2014-05-09   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Clupea harengus (Atlantic herring)</a>                            | <a href="#">ASM96633v1 (GCF_000966335.1)</a>                  | 100                | 2015-06-10  | 2015-06-12   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |
| <a href="#">Cynoglossus semilaevis (tongue sole)</a>                          | <a href="#">Cse_v1.0 (GCF_000523025.1)</a>                    | 102                | 2018-05-02  | 2018-05-16   | <a href="#">FTP</a> <a href="#">B</a> <a href="#">AR</a> <a href="#">GDV</a> |

Species are listed alphabetically along with information on the assembly submission name, date of submission and annotation and links to NCBI resources for that species

 U.S. National Library of Medicine  
National Center for Biotechnology Information

18 

## Searching NCBI's Assembly database?

NCBI's Assembly database is an archival database of submitted genome assemblies across a broad range of organisms. The database also tracks changes to assemblies that are updated by submitting groups over time with a versioned Assembly accession number. The Assembly interface also provides easy access to download genomic datasets.

Example:

- Enter 'fishes' in the Assembly search field
- Use the filter on the left side to select 'Latest RefSeq'
- Select one or multiple assemblies
- Use the 'Download Assemblies' button to Select 'RefSeq' or 'GenBank' datasets
- Use the file type dropdown menu to select from a selection of FTP download choices

The screenshot displays the NCBI Assembly database search results for the query 'fishes'. The interface includes a search bar at the top with the text 'fishes' and a 'Download Assemblies' button. On the left, there is a sidebar with various filters such as 'Organism group', 'Status', 'Assembly level', 'RefSeq category', and 'Exclude'. The main search results area shows a list of items, with the first item being 'GRCz11' (Genome Reference Consortium Zebrafish Build). A dropdown menu is open next to the 'Download Assemblies' button, showing a list of file types including 'Genomic GFF', 'Protein FASTA', and 'RepeatMasker output'. Annotations with orange boxes and arrows point to the 'Assembly List' (the search results area), 'Filters' (the left sidebar), 'Download options' (the dropdown menu), and the 'Download Assemblies' button.

## Using the Assembly database to find more information on a submitted assembly?

The individual assembly pages provide information on the structure and quality of the assembled genome in addition to assembly names and other meta-data, statistical reports, and links to genomic sequence data.

In this example the Assembly page for coho salmon (*Oncorhynchus kisutch*) is shown.

The screenshot shows the NCBI Assembly database page for **Okis\_V1** (*Oncorhynchus kisutch*). The page is divided into several sections:

- Metadata:** Organism name, isolate, sex, bio sample, project, submitter, date, assembly level, genome representation, RefSeq category, GenBank accession, RefSeq accession, and assembly identification.
- Global statistics:** A table summarizing sequence length, assembly gaps, scaffolds, contigs, and chromosomes.
- Access to FTP download of the RefSeq and GenBank Assemblies:** A callout box pointing to the 'Access the data' section, which includes links for downloading the RefSeq assembly, GenBank assembly, full sequence report, and statistics report.
- Information and links to the Assembly submission:** A callout box pointing to the 'Assembly Information' section, which includes links for Assembly Help, Assembly Basics, and the NCBI Assembly Data Model.
- Detailed Assembly Statistics:** A callout box pointing to the 'Global statistics' table.
- The annotation report provides detailed information on the data used for the RefSeq annotation including gene counts:** A callout box pointing to the 'Gene and feature statistics' section, which includes a table of feature counts.

**Global statistics**

| Statistic                                    | Value         |
|--|---------------|
| Total sequence length                        | 2,369,932,239 |
| Total assembly gap length                    | 109,511,248   |
| Gaps between scaffolds                       | 1,566         |
| Number of scaffolds                          | 22,813        |
| Scaffold N50                                 | 1,266,128     |
| Scaffold L50                                 | 433           |
| Number of contigs                            | 97,074        |
| Contig N50                                   | 58,118        |
| Contig L50                                   | 8,653         |
| Total number of chromosomes and plasmids     | 31            |
| Number of component sequences (WGS or clone) | 22,813        |

**Gene and feature statistics**

Counts and length of annotated features are provided below for each assembly.

**Feature counts**

| Feature               | Okis_V1 |
|-----------------------|---------|
| Genes and pseudogenes | 46,096  |
| protein-coding        | 36,425  |
| non-coding            | 4,754   |
| pseudogenes           | 4,917   |
| genes with variants   | 11,142  |
| mRNAs                 | 57,579  |
| fully-supported       | 50,348  |
| with > 5% ab initio   | 3,600   |
| partial               | 3,189   |
| with filled gap(s)    | 1,373   |
| known RefSeq (NM_)    | 0       |
| model RefSeq (NM_)    | 57,579  |
| Other RNAs            | 6,737   |
| fully-supported       | 6,090   |
| with > 5% ab initio   | 0       |

## How do I use NCBI resources to find orthologs?

When available, ortholog sets can be found as a link on the Gene record. Clicking on 'all' opens a tabular list of Gene records for all orthologs within the set. The entire ortholog dataset can be downloaded as a nucleotide or protein FASTA file by following links to the Nucleotide and Protein databases using the 'related data' dropdown menu to the right of the Gene search page.

**myd88 myeloid differentiation primary response 88 [ *Danio rerio* (zebrafish) ]**  
 Gene ID: 403145, updated on 20-May-2018

**Summary**

**Official Symbol** myd88 provided by ZNC  
**Official Full Name** myeloid differentiation primary response 88 provided by ZNC  
**Primary source** ZFIN:ZDB-GENE-040219-3  
**See related** Ensembl:ENSDARG00000010169 Vega:OTTDARG00000021382  
**Gene type** protein coding  
**RefSeq status** PROVISIONAL  
**Organism** [Danio rerio](#)  
**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Cyprinodontiformes; Cyprinodontiformes; Danio  
**Also known as** zgc:103541  
**Orthologs** [human](#) [mouse](#) [all](#)

Select 'all' to open a list of all Gene records in the ortholog set

Select a particular species

ortholog\_gene\_4615[group]

Tabular 20 per page Sort by Relevance

**Search results**  
 Items: 1 to 20 of 234  
 Showing Current items.

| Name/Gene ID                        | Description  | Location   | Aliases    | MIM    |
|-------------------------------------|--|--|------------|--------|
| <a href="#">Myd88</a><br>ID: 17874  | myeloid differentiation primary response gene 88 [ <i>Mus musculus</i> (house mouse)]  | Chromosome 9, NC_000075.6 (119335988..119340040, complement) |            |        |
| <a href="#">MYD88</a><br>ID: 4615   | myeloid differentiation primary response 88 [ <i>Homo sapiens</i> (human)]             | Chromosome 3, NC_000003.12 (38137439..38143022)              | MYD88D     | 602170 |
| <a href="#">Myd88</a><br>ID: 301059 | myeloid differentiation primary response 88 [ <i>Rattus norvegicus</i> (Norway rat)]   | Chromosome 8, NC_005107.4 (128022512..128027462, complement) |            |        |
| <a href="#">myd88</a><br>ID: 403145 | myeloid differentiation primary response 88 [ <i>Danio rerio</i> (zebrafish)]          | Chromosome 24, NC_007135.7 (20188914..20192854, complement)  | zgc:103541 |        |
| <a href="#">MYD88</a><br>ID: 444881 | myeloid differentiation primary response 88 [ <i>Bos taurus</i> (cattle)]              | Chromosome 22, NC_037349.1 (11609334..11613790)              |            |        |
| <a href="#">MYD88</a><br>ID: 420420 | myeloid differentiation primary response 88 [ <i>Gallus gallus</i> (chicken)]          | Chromosome 2, NC_006089.5 (4892500..4905115)                 |            |        |
| <a href="#">MYD88</a><br>ID: 398826 | myeloid differentiation primary response 88 [ <i>Spizella socialis</i> (song sparrow)] | Chromosome 13, NC_010455.5 (2397654..2397658)                |            |        |

**Results by taxon**  
 Top Organisms [\[Tree\]](#)  
[Mus musculus \(1\)](#)  
[Mus caroli \(1\)](#)  
[Mus pahari \(1\)](#)  
[Rattus norvegicus \(1\)](#)  
[Meriones unguiculatus \(1\)](#)  
 All other taxa (229)

**Find related data**  
 Database: **Nucleotide**  
 Select  
 BioProject  
 BioSystems  
 Books  
 ClinVar  
 Conserved Domains  
 dbGaP  
 dbVar  
 EST  
 Gene  
 Genome  
 GEO Profiles  
 GSS  
 GTR  
 HomoloGene  
 MedGen  
**Nucleotide**  
 OMIM  
 PMC  
 Probe

Select a particular species

**Nucleotide**

Species: Animals (336)  
 Molecule types: mRNA (336)  
 Source databases: RefSeq (336)

Items: 1 to 20 of 336  
 Filters activated: mRNA, RefSeq. Clear all

1. [PREDICTED: Alligator sinensis myeloid differentiation primary response 88](#)  
 5,217 bp linear mRNA  
 Accession: XM\_006039500.3 GI: 1394705023  
[BioProject](#) [Protein](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

2. [PREDICTED: Python bivittatus myeloid differentiation primary response 88 \(MYD88\)\\_mRNA](#)  
 1,199 bp linear mRNA  
 Accession: XM\_025168299.1 GI: 1391806804  
[BioProject](#) [Protein](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Download features:  
 Format: FASTA Nucleotide  
 Create File

Select Nucleotide allows you to download FASTA

## How do I use BLAST in GDV to locate orthologs of my gene?

Potential orthologs can be searched using the BLAST widget which supports translated protein sequence to the genomic sequence (tblastn). In this example the zebrafish MyD88 protein, NP\_997979.2 (RID # HYBHRHS9014) was searched against the coho salmon (*Oncorhynchus kisutch*) reference assembly.

Job title: ref|NP\_997979.2| (284 letters)

RID: HYBHRHS9014 (Expires on 06-13 10:21 am)

Query ID: NP\_997979.2  
 Description: myeloid differentiation primary response protein MyD88 [Danio rerio]  
 Molecule type: amino acid  
 Query Length: 284

Database Name: genomic/8019/GCF\_002021735.1  
 Description: Okis\_V1 RefSeq assembly [GCF\_002021735.1]  
 Program: TBLASTN 2.8.0+ -> Citation

Other reports: Search Summary [Taxonomy reports]

Graphic Summary

Distribution of the top 8 Blast Hits on 4 subject sequences

Color key for alignment scores

|     |       |       |        |       |
|-----|-------|-------|--------|-------|
| <40 | 40-50 | 50-80 | 80-200 | >=200 |
|-----|-------|-------|--------|-------|

Descriptions

Alignments

Download GenBank Graphics Sort by: E value

Oncorhynchus kisutch isolate 150728-3 unplaced genomic scaffold, Okis\_V1 Un\_scaffold04938, whole genome shotgun sequence  
 Sequence ID: NW\_018090668.1 Length: 245405 Number of Matches: 4

| Range               | Score          | Expect | Method                       | Identities | Positives  | Gaps     | Frame |
|---------------------|----------------|--------|------------------------------|------------|------------|----------|-------|
| 1: 150947 to 151237 | 105 bits (263) | 1e-23  | Compositional matrix adjust. | 52/97(54%) | 70/97(72%) | 1/97(1%) | -2    |

Query 6: SIDHEATPITALNCSFRKLGFLFNDPTNVAADVIRTVAFVDFNYLEIKMFEKDCPFEE 65  
 S-D IP+ ALN + RK-LGLFLNP NTVA+DW +VAE H F+YLEEKNE P K  
 Sbjct 151237 SLDLWNLPLRALNINWRKRLGLFLNPNRYVADSMVSAEMGFSYLEEKNYEDCQDPTRK 151058  
 Query 66: VLTQMETR-PDATVANILSLLEKAERKQVISELKEIL 101

Related Information  
[New Genome Data Viewer - aligned genomic context](#)

Genome Data Viewer

Oncorhynchus kisutch: Okis\_V1 (GCF\_002021735.1) NW\_018090668.1 (unplaced): 138,810 - 165,445

myd88 XM\_020474013.1

BLAST Alignment Inspector NP\_997979.2: myeloid differentiation primary response protein MyD88 [Danio rerio]

| Gene Hit  | myd88 | XM_020474013.1 |
|-----------|-------|----------------|
| 6 > 101   | 97    | 105            |
| 102 > 124 | 23    | 37             |
| 211 > 268 | 58    | 35             |
| 233 > 270 | 38    | 73             |

tblastn alignment

BLAST Results for: ref|NP\_997979.2| (284 letters)

Cleaned Alignments - BLAST Results for: ref|NP\_997979.2| (284 letters)

Genes, NCBI Oncorhynchus kisutch Annotation Release 100, 2017-03-14

myd88 NP\_997979.2

NW\_018090668.1: 139K-155K (17Kbp)

## Coming Soon to the GDV browser

- Load your own data
  - For example: GFF3, BED, BAM formats
- Track Hubs
  - Search Track Hub Registry from GDV

**Genome Data Viewer** Danio rerio: GRCz10 (GCF\_000002035.5) Chr 1 (NC\_007112.6): 13,626,351 - 1

Region: LOC553528 Gene, NM\_001328067.1 Transcript

NC\_007112.6 13,640 K 13,660 K 13,680 K 13,700 K 13,720 K

Genes, NCBI Danio rerio Annotation Release 105, 2016-06-24

wfs1a XM\_690160.8 LOC567498 XM\_005159850.3 XP\_005159907.1 ch

crmp1 [+4] fip11tb [+6] NM\_201120.1

**User Data and Track Hubs**

Principal Splice Isoforms APPRIS APPRIS Tools

Hub Name PrincipalsplicingisoformsAPPRIS

Short Label Principal Splice Isoforms APPRIS

Long Label Selects a single CDS variant for each gene as the 'PRINCIPAL' isoform based on the range of protein ... [more](#)

Contact [appris.cnio@gmail.com](mailto:appris.cnio@gmail.com)

More [Description](#)

Remote data tracks 7

bigbed 7

Disconnect Open

Supported File/Data types

**APPRIS - Principal Isoforms**

ENSDART00000090... ENSDART00000103... ENSDART00000030...

ENSDART00000064... ENSDART00000046... ENSDART00000030...

NC\_007112.6: 14M..14M (230Kbp)

## Additional NCBI resources

- Bulk Data Download
  - FTP
    - <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/>
    - <ftp://ftp.ncbi.nlm.nih.gov/refseq/>
  - NCBI Genome Pages – provides direct access to datasets  
<https://www.ncbi.nlm.nih.gov/genome/?term=danio+rerio>
  - NCBI Assembly Database – allows for bulk download of data across multiple assemblies.  
<https://www.ncbi.nlm.nih.gov/assembly/?term=Neopterygii>
- Programmatic access - Eutils/Edirect
  - <https://www.ncbi.nlm.nih.gov/books/NBK25500/>
  - <https://www.ncbi.nlm.nih.gov/books/NBK179288/>
  - YouTube Videos <https://www.youtube.com/user/NCBINLM>
  - Tutorials  
[ftp://ftp.ncbi.nih.gov/pub/factsheets/Factsheet\\_bulk\\_download.pdf](ftp://ftp.ncbi.nih.gov/pub/factsheets/Factsheet_bulk_download.pdf)



## NCBI news and blogs

- <https://ncbiinsights.ncbi.nlm.nih.gov/>
- <https://www.facebook.com/ncbi.nlm>
- <https://www.ncbi.nlm.nih.gov/home/learn/>

## Contact Us!

NCBI Help Desk

<https://support.nlm.nih.gov>

RefSeq

<https://www.ncbi.nlm.nih.gov/projects/RefSeq/update.cgi>

Nuala O'Leary – RefSeq Curator, Zebrafish Champion

[olearyna@ncbi.nlm.nih.gov](mailto:olearyna@ncbi.nlm.nih.gov)

# Thank you.

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

## RefSeq/Gene

### Terence Murphy

Eric Cox  
Catherine Farrell  
Tamara Goldfarb  
Diana Haddad  
John Jackson  
Vinita Joardar  
Kelly McGarvey  
Michael Murphy  
Nuala O'Leary

Shashi Pujar  
Bhanu Rajput  
Sanjida Rangwala  
Lillian Riddick  
Barbara Robbertse  
Brian Smith-White  
Pooja Strobe  
Anjana Vatsan  
David Webb

## RefSeq Developers

Alex Astashyn  
Olga Ermolaeva  
Vamsi Kodali  
Craig Wallin

## Annotation Pipeline

### Francoise Thibaud-Nissen

Paul Kitts  
Mike Dicuccio  
Wratko Hlavina  
Avi Kimchi

Jinna Choi  
Boris Kiryutin  
Patrick Masterson  
Eyal Mozes  
Anton Perkov  
Dan Rausch  
Robert Smith  
Alexandre Souvorov

## GDV/Remap/GBench

### Valerie Schneider

Peter Meric  
Nathan Bouk  
Hsiu-Chuan Chen  
Cliff Clausen  
Anatoliy Kuznetsov  
Andrei Shkeda

## A cast of thousands

Ken Katz  
Michael Ovetsky  
Lukas Wagner  
Andrei Shkeda  
Donna Maglott  
**Kim Pruitt**  
Jim Ostell