

January 29, 2024

MedakaBase User Manual

Table of Contents:

1. Introduction
2. Gene Name Search
3. Genome Browser (JBrowse, WebApollo)
4. BLAST
5. GMAP - cDNA Alignment
6. Genome Slicer
7. Gene Fetcher
8. Download
9. Terms of Use

1. Introduction

In 2022, [NBRP Medaka](#) incorporated a new mission related to the dissemination of genomic information into its activities involving organisms familiar to Japanese, Medaka. To launch [MedakaBase](#) web server, [Dr. Masaru Yagura](#), a former researcher of the NBRP project at the National Institute of Genetics, Molecular Life History Laboratory, collaborated under the supervision of Professor Yasukazu Nakamura, Genome Informatics Laboratory at the National Institute of Genetics. Building on the experience of operating the [Cats-I web server](#), MedakaBase service was initiated at the end of 2022.

The primary mission of NBRP Medaka's activities is to facilitate the sharing of existing genomic resources rather than acquiring new ones. Consequently, a platform was established to promote the effective utilization of already available data in public databases, along with navigation tools to facilitate user engagement. This document aims to clarify the various tools provided by MedakaBase.

As of 2024, Dr. Kenji Morikami, a researcher at Molecular Life History Laboratory, is responsible for maintaining and making improvements to the MedakaBase web server as needed, to enhance user-friendliness.

Related Sites:

- **Medaka – annex:** <https://github.com/Squalomix/medaka-annex>
Compiles various Medaka-related sites, including MedakaBase.
- **Medaka Shigen:** <https://shigen.nig.ac.jp/medaka/>
Provides diverse and useful biological information about Medaka including sequence information.

Latest Information:

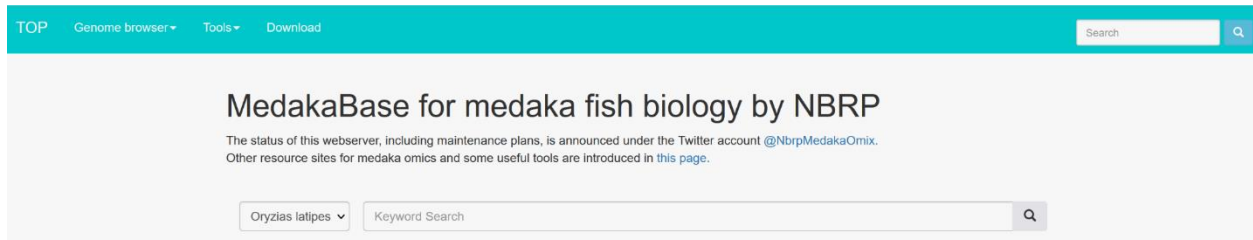
The latest information about MedakaBase is announced in [Medaka Omics News](#), mainly in Japanese, and systematic information about server operation, such as scheduled downtime for maintenance, is communicated through [MedakaBase Twitter account](#).

Contact: nbrp.medaka.info@gmail.com
(Kenji Morikami, Mana Sato, Shigehiro Kuraku)

2. Gene Name Search <https://medakabase.nbrp.jp/>

Perform the gene name search from the top page of this site or by using the search window on the right side of the menu bar.

The search is conducted against the gene names assigned to Medaka assembly in Ensembl entries. Information from Ensembl entries is based on Ensembl Release 107 (July 2022; gene build May 2018).



For tips on search techniques, please refer to the guide just below.

Depending on how gene names are specified, such as utilizing wildcards (*) it becomes possible to search for multiple members of a gene family simultaneously.

Caution

As information retrieval across species is not supported, if search involving multiple species simultaneously is required, please perform the search on the Ensembl website.

Search tips

Example	Search results
ENSORLG00000003517	The specific target is shown
"GMP synthase"	The full name including the space in between is taken as a search word
foxl*	**' is taken as a wildcard and all of foxl1, foxl2 and foxl3 are shown
"elongation factor" +transcription	Both words are taken as search words
"elongation factor" -transcription	Matches to the second word are excluded from those to the first word

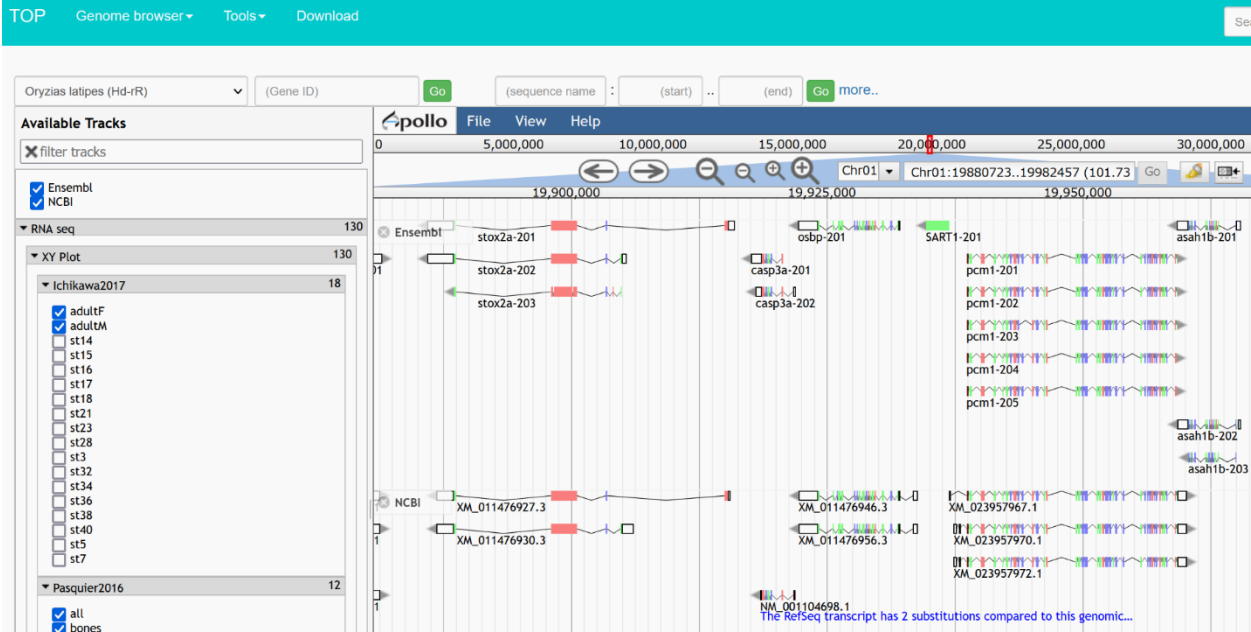
3. Genome Browser

3.1. JBrowse <https://medakabase.nbrp.jp/viewer/Hd-rR/>

Choose items you want to view from the checkboxes listed on the left, and they will be displayed on the browser. From left lists, you can select gene models from Ensembl and/or NCBI, while mapping RNA-seq reads available in public as BAM files. The displayed RNA-seq data is sourced from the following papers; [Ichikawa et al., Nat Comm 2017](#) , [Pasquier et al., BMC Genomics, 2016](#) , & [Uchida et al. BMC Biol. 2022](#).

Caution

The gene models starting with “liftoff_” (available for HNI and HSOK) were generated by using [Liftoff](#) by the NBRP Medaka Team. Please note that they are based on cross-species inference and are not supported by endogenous evidence for the individual species.



The genome browser capable of displaying both NCBI and Ensembl gene models simultaneously is exclusive to MedakaBase.

Furthermore, NBRP Medaka originally optimized GTF files (not GFF files) by extending UTRs for advanced molecular analysis including single-cell analyses, and the files are now publicly accessible [here](#). However, display of the new gene model on this JBrowse is currently not supported.

3.2. WebApollo

A tool that allows modification of the structure of gene models (e.g., start and end points of exons). However, it is not in general use now.

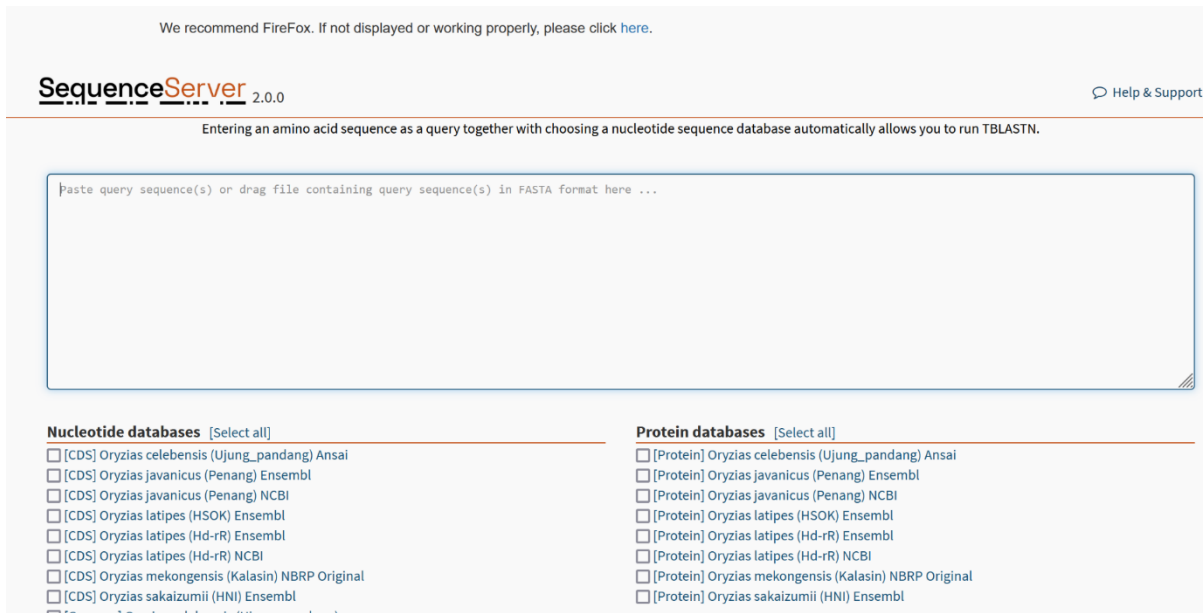
4. **BLAST** <https://medakabase.nbrp.jp/tools/blast/>

Enter your own nucleotide or amino acid sequence into the box as the query and the sequence type is automatically determined. Then, select either the nucleotide or protein database listed in the lower section, and a similarity search will be performed. The discrimination of sequence type is done automatically. Briefly, BLASTN is used for nucleotide vs. nucleotide, BLASTX for nucleotide vs. amino acid, TBLASTN for amino acid vs. nucleotide, and BLASTP for amino acid vs. amino acid.

Caution

- 1) Please avoid selecting unnecessary databases, as it may prolong the analysis time. Keep the selection to the minimum necessary.

- 2) Please note that in the nuclear genome sequence of *Oryzias latipes* (Hd-rR strain), there are around 900 sequences (unanchored contigs) that are not integrated into chromosomes and are not part of the publicly available assembly. These sequences are included in the search targets of [Genome] in the Nucleotide databases. When hits are found in BLAST searches, they are explicitly labeled as “unanchored contig”, distinct from chromosomes.



Interpretation

For detailed information about the BLAST program, NCBI provides [guides](#). The search results are displayed by ranking the similarity of sequences. However, it's important to note that BLAST is not a tool for determining that the sequences have the same ancestor (i.e., homology) or for identifying the orthology. Especially for orthology determination, additional steps such as drawing and adequately interpreting molecular phylogenetic trees are necessary.

Notes

As part of NBRP's activities in 2023, the genome sequence of *Oryzias mekongensis* was curated and added to the search target databases. Furthermore, the Molecular Life History Laboratory has acquired the genome information of Pacific saury serving as one of the outgroups closely related to the *Oryzias* genus. The assembly of Pacific saury is already accessible through the NCBI genome database and [custom genome browser](#).

When presenting research results based on results from this web site, please not only mention MedakaBase but also cite the origin of these sequences. For example, in the case of *Oryzias celebensis*, cite [Ansai et al., Nat Comm 2021](#). For more details, please carefully read the "9. Terms of Use" mentioned later.

5. GMAP - cDNA alignment <https://medakabase.nbrp.jp/tools/gmap/>

A tool for aligning your transcript sequences onto the genome, considering information about the introns.

Input your sequence, such as cDNA, in the provided box, and choose the target species from the [Genome] popup menu, and then select the desired display format from the [Format] popup menu. The mapping results will be displayed. When you download the file by clicking “Full document” at the bottom, note that the file does not have any extensions, but you can open it with a text editor.

Application

This tool allows mapping cDNA sequences to the genome while considering intron information. By using this, it becomes possible to grasp, for example, the exon ends necessary for experimental design. Specifically, by selecting “only show alignment” at the bottom of the [Format] popup menu, you can accurately determine, at a 1bp level, where introns are inserted into the cDNA sequence. Although the intron sequences are not displayed in this data, the insertion positions of introns are indicated by numbers. Using these numbers, you can extract the nucleotide sequence with the Genome Slicer (refer to “6. Genome Slicer” described below). This can be applied to easily design gRNA on introns, 5’UTRs, or 3’UTRs.

GMAP

cDNA alignment to genome.

Genome **Format**

Enter FASTA sequence(s) below:

Full documentation for GMAP is available [here](#).
Reference: Thomas D. Wu and Colin K. Watanabe. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 2005 21: 1859-1875.

6. Genome Slicer https://medakabase.nbrp.jp/tools/genome_slicer/

This tool allows you to cut and display a specific region of a genome sequence based on specified coordinates (start to end) for a given ID.

To use, refer to the example provided in the lower section. Input the genome sequence ID (chromosome number) and define the start and end points. The term “Complement” in the “Example” section indicates the method for acquiring the complementary strand sequence. By specifying numerical values in descending order, you can obtain the sequence of the complementary strand.

Genome Slicer

Extract a sequence of a particular genomic region of interest.

Genome

Oryzias latipes (Hd-rR) ▼

Output Format

HTML ▼

Enter genomic location below (comma separated or one per line):

Submit

Example:

Basic usage: Chr01:1-1000 or ChrM:1001..2000

Complement: Chr01:1000-1 (start > left)

Multiple sequences: Chr01:1001-2000,Chr01:2000-1001,Chr02:10001-12000,ChrM:101-200

7. Gene Fetcher https://medakabase.nbrp.jp/tools/gene_fetcher/

This tool enables the extraction of specific regions of a gene.

Select the target organism from the [Genome] popup menu and choose the desired type of sequence from “**Gene sequence**”. In the “**Gene**” section, select either FASTA or GenBank format for the output. You can also specify numerical values in the empty fields next to “5’-upstream/3’-downstream” to obtain the flanking sequence, including upstream and downstream sequences of the target gene, based on the specified number of bases.

Enter the Gene ID of the target gene, which must be an Ensembl Transcript ID starting with “ENSORLT”, into the box.

Gene Fetcher

Extract a coding (and UTR) sequence of a particular gene of interest.

Genome

Gene sequence

mRNA (CDS + UTR) CDS Protein

Gene region

FASTA GenBank 5'-upstream 3'-downstream

Enter gene IDs below (comma separated or one per line):

Example:

Ensembl gene: ENSORLT00000031622

The information for Ensembl entries is based on Ensembl Release 107 (July 2022; gene build May 2018).

8. Download <https://medakabase.nbrp.jp/download/>

You can obtain the sequences for the desired organism.

Select the target organism from the popup menu, and click on the file name. Determine the type of sequence (Genome, CDS, or Protein) based on the file extensions.

Download Genome Resources

Oryzias latipes (Hd-rR) ▾

FileName	Description	Size
Hd-rR_Ensembl.protein.faa.gz	-	10.3 Mbyte
Hd-rR_genome.fna.gz	-	229.4 Mbyte
Hd-rR_Ensembl.cds.fna.gz	-	15.9 Mbyte

9. Terms of use

When reporting the research results based on output from MedakaBase, it is crucial not only to mention MedakaBase but also to cite the source of the sequence data. For example, in the case of *Oryzias celebensis*, please refer to the paper by [Ansai et al., Nat Comm 2021](#).

The sources of information utilized in MedakaBase are shown on the TOP page as follows. Information from Ensembl entries is based on Ensembl Release 107 (July 2022).

Caution

Regarding the sequence information for *Oryzias mekongensis*, as it is unpublished, please contact NBRP Medaka Team (nbrp.medaka.info@gmail.com) for any inquiries related to its disclosure.

Assembly and annotation versions adopted in this site

The current version of genome assembly and annotation is as follows.

For use of the data indicated with '#' in publications and data presentations, please contact us (see below for the contact detail).

Species	Genome	Annotation
<i>Oryzias latipes</i> (Hd-rR)	ASM223467v1	Ensembl: Release 107, NCBI: Release 103
<i>Oryzias sakaizumii</i> (HNI)	ASM223471v1	Ensembl: Release 107
<i>Oryzias latipes</i> (HSOK)	ASM223469v1	Ensembl: Release 107
<i>Oryzias javanicus</i> (Penang)	OJAV_1.1	Ensembl: Release 107, NCBI: OJAV_1.1
<i>Oryzias celebensis</i> (Ujung pandang)	OryCel_1.0	Ansai: 1.0
<i>Oryzias mekongensis</i> (Kalasin)	fOryMek1.1 [#]	NBRP Original: Helixer 1.0 [#]