

2024年1月29日版

MedakaBase ユーザーマニュアル

目次

1. 導入
2. 遺伝子名検索
3. Genome Browser (JBrowse、WebApollo)
4. BLAST
5. GMAP - cDNA alignment
6. Genome Slicer
7. Gene Fetcher
8. Download
9. 利用条件

1. 導入

概要

[NBRP メダカ](#)では、日本発の生物ならではの活動として、2022年度より新たにゲノム情報に関する発信をミッションに追加した。そのプラットフォームとして、国立遺伝学研究所・分子生命史研究室 NBRP プロジェクト特任研究員（当時）の[矢倉勝博士](#)が、遺伝研・大量遺伝情報研究室（中村保一教授）の協力の下で、[Cats-I ウェブサーバー](#)の機能を踏襲したウェブサーバー [MedakaBase](#) を立ち上げ、2022年末より運用を開始した。

NBRP の活動におけるおもなミッションは、新規データ取得ではなく、リソース公開である。そのため、公共データベースで公開されている既存データの活用を促進を目指したプラットフォームの構築と、ユーザーによる効果的な利用のためのナビゲーションを行っている。

2024年現在、遺伝研・分子生命史研究室 NBRP 特任研究員の森上賢治博士が、ユーザービリティの改善を施すとともにウェブサーバーを管理している。本ドキュメントにて [MedakaBase](#) の各機能を説明する。

関連サイト

- Medaka-annex <https://github.com/Squalomix/medaka-annex>
MedakaBase を含むメダカ関連のサイトをまとめている。
- Medaka Shigen <https://shigen.nig.ac.jp/medaka/>
オミクス時代以前からの配列情報を含む多様なリソースを提供している。

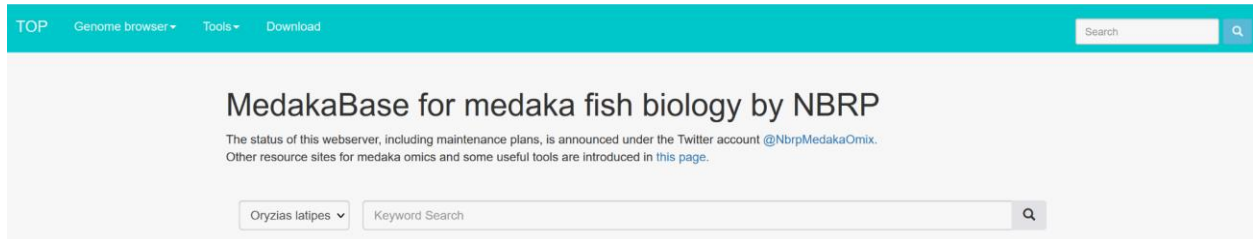
最新情報

Medaka Omics News（例、[第一号](#)）にてアナウンスするとともに、とくにウェブサーバーの稼働状況や休止予定などは [Twitter アカウント](#)にて発信している。

連絡先 nbrp.medaka.info@gmail.com（森上賢治、佐藤茉菜、工樂樹洋）

2. 遺伝子名検索 <https://medakabase.nbrp.jp/>

TOP ページあるいは各ツールのページにて、メニューバー右の検索窓から実行する。検索は、Ensembl エントリに振られたメダカの遺伝子名に基づいて行われる。Ensembl のエントリの情報は、Ensembl Release 107 (July 2022; genebuild May 2018)による（以降の機能でも同様）。



ワイルドカード（*）の利用などによる検索については、すぐ下のガイドを参照していただきたい。遺伝子名の指定の仕方によって、遺伝子ファミリーのメンバー複数を一度に検索することなどが可能となる。

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

しかし、種間を跨ぐ情報検索はできないため、種間比較が必要な場合は、Ensembl のサイトで。

3. Genome Browser

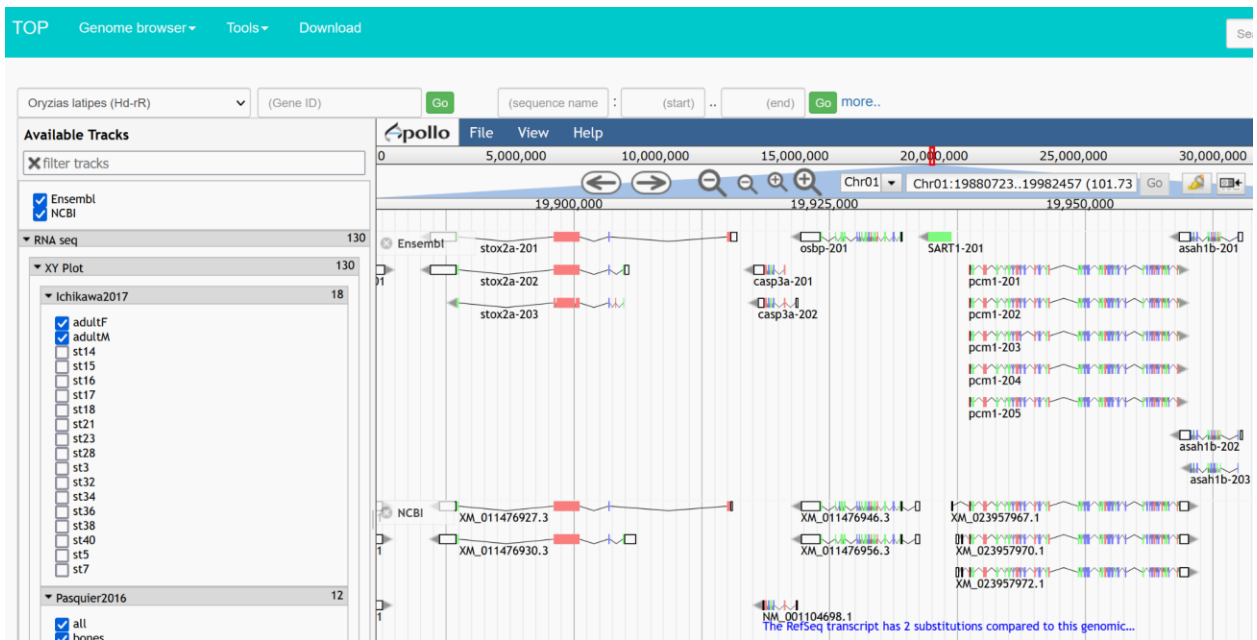
3.1 JBrowse <https://medakabase.nbrp.jp/viewer/Hd-rR/>

左に羅列されたチェックボックスから閲覧したいものを選択すると、ブラウザに表示される。上から、EnsemblとNCBIの遺伝子モデル、そしてRNA-seqリードのマッピング結果（.BAMファイルの可視化）を選択できる。選択可能なRNA-seqデータは下記に由来する。

[Ichikawa et al., Nat Comm 2017](#) , [Pasquier et al., BMC Genomics, 2016](#) , & [Uchida et al. BMC Biol. 2022](#)

* 注意 *

“liftoff_”からはじまる遺伝子モデルは（HNIとHSOKについて選択可能）、MedakaBaseのゲノムブラウザで表示するために、NCBIに登録されているHd-rRの遺伝子モデルを基に、[Liftoff](#) ツールを使って新たに作られたものである。すなわち、Hd-rRでの情報に基づき、生物種を跨いで推定された遺伝子情報であるため、利用には注意が必要となる。



NCBIとEnsemblの遺伝子モデルを並列に閲覧できるサイトは、MedakaBaseのみである。

また、NBRPでは、既存の遺伝子モデルについて、主に3'UTR領域の伸長を目的としたGTFファイルの改変を行い（[UTRの延伸](#)）、これにより得られたGTFファイルは現在、[オンラインで公開](#)している。ただし、現時点ではJBrowseでの表示には対応していない。

3.2 WebApollo

可視化されている遺伝子モデルの構造（例、Exonの始点や終点）を改変することが可能なツール。現在は一般利用はしていない。

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

類似した配列を検索するツール。

手持ちの塩基配列あるいはアミノ酸配列を、クエリー (Query) 配列 (見出し配列ともいう) として空欄に入力すると、その別が自動で判定される。つづいて、下段に羅列された検索対象のデータベースを選択すると、その組み合わせに応じて相同性検索が実行される。すなわち、塩基 vs 塩基では BLASTN、塩基 vs アミノ酸では BLASTX、アミノ酸 vs 塩基では TBLASTN、アミノ酸 vs アミノ酸では BLASTP が実行される。

不必要に多くのデータベースを選ぶと時間がかかるため、必要最小限の数に留めること。

BLAST プログラムの詳細については、NCBI がガイド (例) を提供している。検索結果は、配列の類似性に基づいてランク付けし表示される。しかし、この結果だけでは、相同性、すなわち配列が同一起源をもつのかどうかや、オーソロジー (種分岐に由来し、種間で対応する遺伝子) について判定できないことに注意。とくにオーソロジーの判定もは、別途、分子系統樹の推定と吟味を必要とする。

検索対象データベースのうち、*Oryzias mekongensis* は、2023 年に NBRP の活動の一環としてゲノム配列を整備し、追加したものである。また、分子生命史研究室では、*Oryzias* 属のアウトグループのひとつとして、サンマのゲノム情報を取得し別ページにて公開しており、配列の利用が可能となっている (概要記事)。

Oryzias latipes Hd-rR の核ゲノム配列では、染色体規模の 24 本以外に、染色体に組み入れられていない約 900 本の配列 (unanchored contigs) が存在することに注意。それらの配列は、Nucleotide databases 内の検索対象である [Genome] 内に含まれており、BLAST 検索でヒットした際には、染色体とは区別され「unanchored contig」と明示される。

検索に基づく研究結果を発表する際には、MedakaBase について触れるだけでなく、配列の由来について引用をすること。（例、*Oryzias celebensis* の場合は [Ansai et al., Nat Comm 2021](#) を引用）詳細については、後述の「9. 利用条件」を精読していただきたい。

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

イントロンを考慮して転写産物配列をゲノムにマップ（アライン）するツール。

空欄に手持ちの cDNA などの配列を入力して、**Genome** ポップアップメニューから対象メダカ種を選び、次に、**Format** ポップアップメニューから希望する表示フォーマットを選ぶと、マッピング結果が表示される。

たとえば、実験デザインに必要な Exon 端の把握には、**Format** ポップアップメニューの最下段「only show alignment」を選択するとよい。これによって cDNA 配列のどの部分にイントロンが挿入されるのかを 1bp レベルで正確に知ることができる。イントロン配列はこのデータには表示されないが、イントロン挿入位置について数字で表示されるため、その数字を用いることで **Genome slicer**（後述「6. Genome slicer」を参照）により、塩基配列を抽出することができる。これを応用して、イントロンや 5'UTR もしくは 3'UTR 上に gRNA を簡便に設計できる。

The screenshot shows the GMAP web interface. At the top, it says "GMAP" and "cDNA alignment to genome." Below this, there are two dropdown menus: "Genome" and "Format". The "Genome" dropdown is set to "Oryzias latipes (Hd-rR)" and the "Format" dropdown is set to "GFF3 gene format". Below these menus, there is a text input field labeled "Enter FASTA sequence(s) below:". At the bottom left of the input field, there is a "Search" button. At the bottom of the page, there is a link for "Full documentation for GMAP is available here." and a reference: "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."

下段の **Full document** をダウンロードすると得られるファイルには拡張子が付いていないが、「メモ帳」などのテキストビューワーで開くことができる。

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

指定した ID を持つゲノム配列の、指定した位置（始点—終点）の配列部分を切り取り表示するツール。

下段にある **Example** に示した例に従い、ゲノム配列 ID（染色体番号）と、始点および終点を指定する。**Example** 内にある **Complement** は、相補鎖の配列を取得する際の指定方法である。降順に数値を指定することで相補鎖の配列を取得できる。

Genome Slicer

Extract a sequence of a particular genomic region of interest.

Genome	Output Format
<input type="text" value="Oryzias latipes (Hd-rR)"/>	<input type="text" value="HTML"/>

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

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/

遺伝子の特定の領域を抜き出すことができるツール。

Genome ポップアップメニューから対象の生物種を選び、**Gene sequence** から取得したい配列の種類を選択する。

Gene region では、結果の出力形式として、FASTA 形式あるいは GenBank 形式のいずれかを選ぶことができる。また、5'-upstream/3'-downstream 横の空欄に数値を指定することで、指定した塩基数に応じて、目的遺伝子上流配列および下流配列をふくむ配列を取得することができる。

空欄には、事前に入手している目的の遺伝子 ID を入力する。ただし、「ENSORLT」から始まる Ensembl Transcript ID に限られる。

Gene Fetcher

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

Genome

Oryzias latipes (Hd-rR) ▼

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):

Submit

Example:

Ensembl gene: ENSORLT00000031622

Ensembl のエントリの情報は、Ensembl Release 107 (July 2022; genebuild May 2018)に基づく。

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

目的の生物種の配列が取得できる。

ポップアップメニューから目的の生物種を選び、ファイル名をクリックする。

Genome、CDS、Protein の別はファイル名から判断すること。

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. 利用条件

MedakaBase の利用に基づく研究結果を発表する際には、MedakaBase について触れるだけでなく、配列の由来について引用すること。（例、*Oryzias celebensis* の場合は [Ansai et al., Nat Comm 2021](#) を引用）

MedakaBase に採用している情報のソースは、下記のように TOP ページに掲載している。Ensembl のエントリの情報は、Ensembl Release 107 (July 2022) に基づくが、その実体は、それ以前から Ensembl が公開している遺伝子推定結果（Ensembl では「genebuild」と呼ばれる）の May 2018 版である。

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 [#]

Oryzias mekongensis の配列については、論文としては未発表であるため、公表については当 NBRP メダカ生命情報チーム (nbrp.medaka.info@gmail.com) へご相談ください。