

# ゲノム、遺伝子の比較

清水謙多郎

[shimizuk@fc.jwu.ac.jp](mailto:shimizuk@fc.jwu.ac.jp)

# ゲノム配列をどう決めるか？

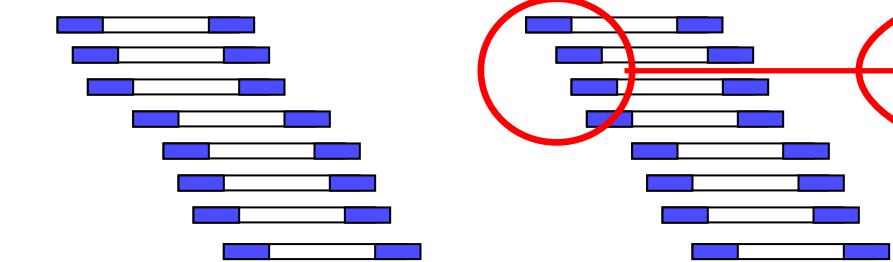
## DNAシークンサ

配列断片が生成される

数百bp(ショートリード)  
～数万bp(ロングリード)



現在のシークンサでは、1回の実験で $10^{12}$ 塩基を越える解析が可能なものが存在



CGGAGTCAACTTACCTATA-----  
TTACCTATATTCTAATCG----  
CTATATTCTAATCGTAG--  
TATTCTAATCGTAGTA

## アセンブリ

配列断片の重なりをもとにつなぐ  
読み取りエラーも考慮

## コンティグ配列



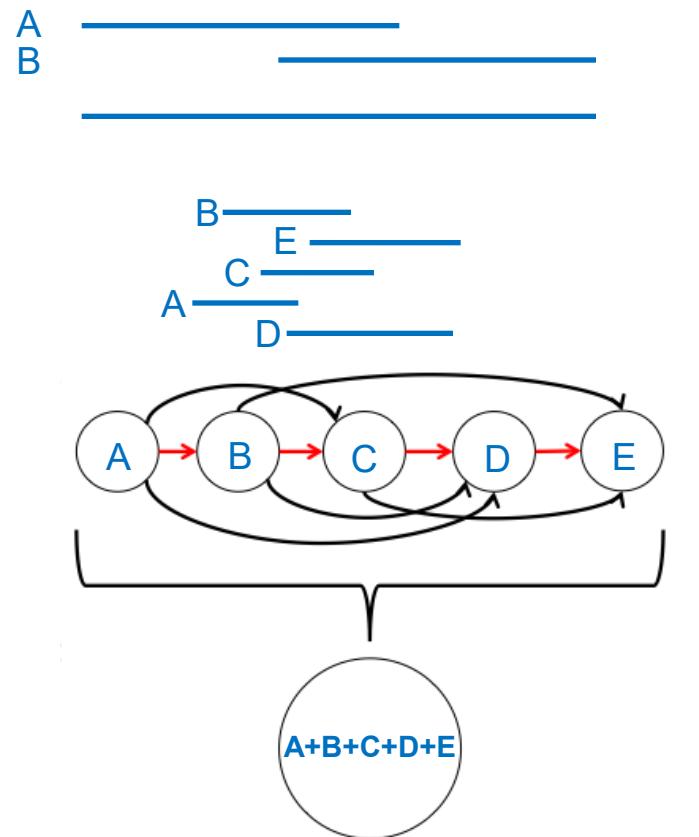
## スキヤフォールド

距離情報をもとに相対的な位置を決める

距離情報をもとに相対的な位置を決める

# ゲノム配列の決定手法

- Overlap-Layout-Consensus (OLC)
  - ロングリードに対して用いられる
- Overlap
  - リードの重なりを検出し、リード間のつながりの関係をグラフの有向辺で表現
- Layout
  - リードを矛盾なく順序づけ、コンティグを形成する
  - リード間の包含関係、つながり関係を整理
- Consensus
  - リードのつながり関係をもとに配列を決定する



TCCTCATTCTTGATCCGGCTC

# Consensusの例

ベースコールエラーに対する対処

	-ACCGT--	
	---CGTGC	
	<b>TACGGT--</b>	
コンセンサス	<hr/> <b>TACCGTGC</b>	<b>塩基置換エラー</b>
	-ACC-GT--	
	---CAGTGC	
	<b>TACC-GT--</b>	
コンセンサス	<hr/> <b>TACCGTGC</b>	<b>挿入エラー</b>
	-ACCGT--	
	---C-TGC	
	<b>TACCGT--</b>	
コンセンサス	<hr/> <b>TACCGTGC</b>	<b>欠失エラー</b>

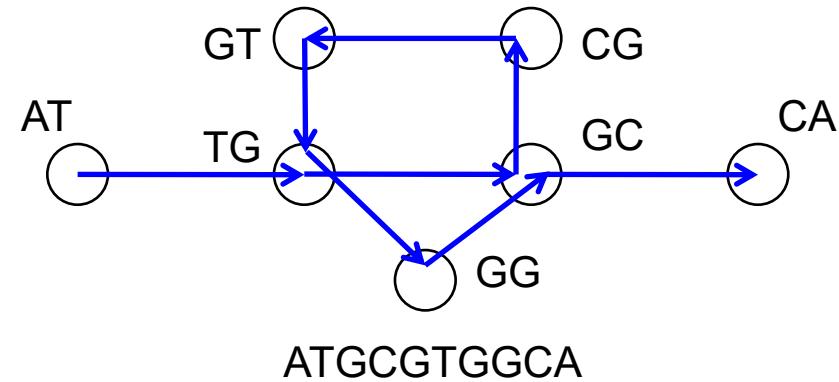
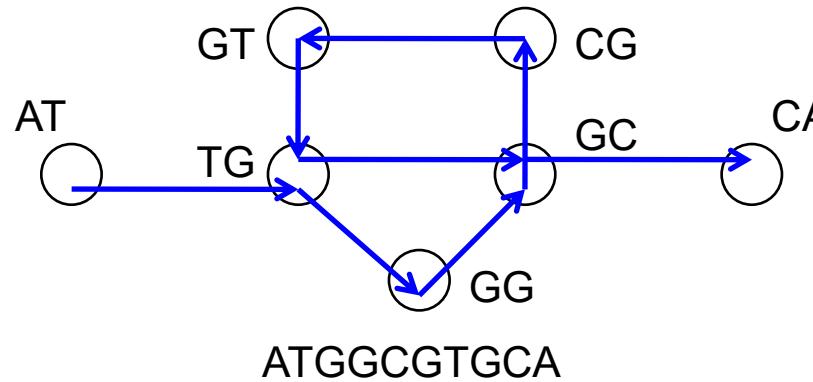
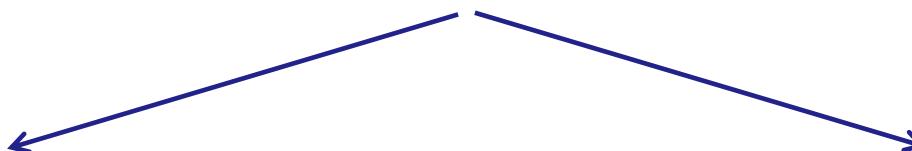
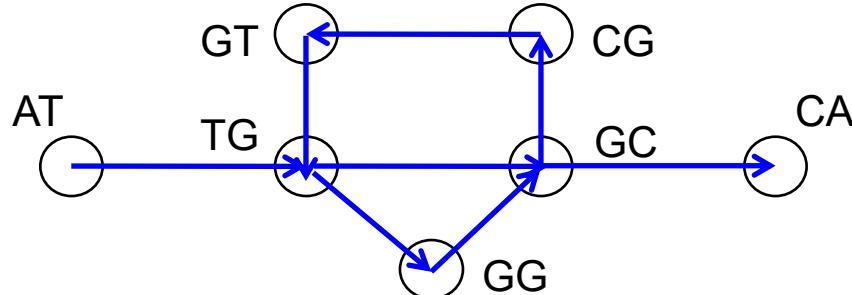
- 多数決
- 重み付き投票方式  
ベースコール（各塩基（A, T, G, C）の読み取り精度）のクオリティ情報（QV値）の利用が考えられる

# de Bruijn graph

- de Bruijn graph (デュブラングラフ)
  - ショートリードに対して用いられる
  - 配列の断片を  $k$ -mer (長さ $k$ の部分配列) に分割
  - $(k-1)$ 文字を頂点とする
  - 頂点をつないで  $k$ -merとなるよう有向辺を引く
- エッジをオイラー経路 (一筆書き) のように全部たどると、元の配列が得られる
  - オイラー経路: すべての辺を1回ずつ訪れるパス
  - オイラー閉路: ある頂点から出発し、同じ頂点に戻るオイラー経路
- ゲノムシーケンシングでの利用
  - 短いリードを膨大な数 ( $10^7 \sim 10^9$ ) 扱うとき、リードどうしを全部比較 (OLC法) するのは重すぎる
  - $k$ -merの共有で高速につなげることができ、現実に(基本的な手法として) よく用いられている

# オイラー経路の探索例

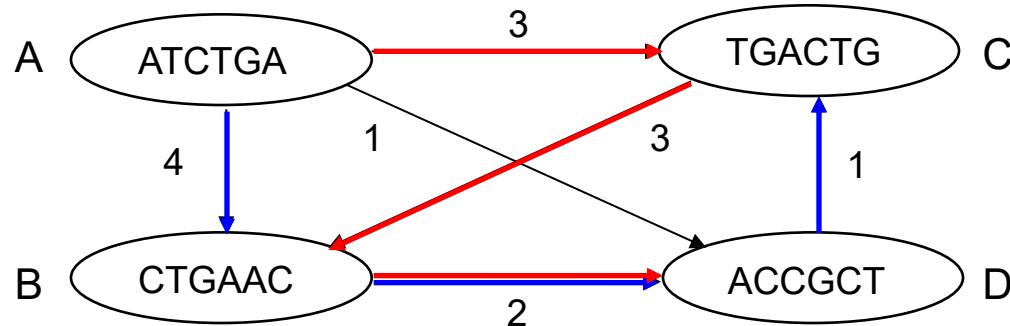
$S = \{ ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT \}$



# ハミルトン経路

- 最大重みハミルトン経路

- 配列断片（リード）を頂点とし、頂点 $p$ の末尾 $k$ 文字と頂点 $q$ の先頭 $k$ 文字が一致しているとき、有向辺 $p \rightarrow q$ をもたせ、重みを $k$ とする
- ハミルトン経路: グラフ上の全ての頂点を1回ずつ通る経路
- 重み付きハミルトン経路: 経路上の辺の重みを合計したもの
- 最大重みのハミルトン経路を求め、その頂点を重ね合わせたものがShortest Common Superstring



貪欲法によるハミルトン経路:  $A \rightarrow B \rightarrow D \rightarrow C$  重み: 7

コンティグ: ATCTGAACCGCTGACTG 長さ: 17

重みを最大にするハミルトン経路:  $A \rightarrow C \rightarrow B \rightarrow D$  重み: 8

コンティグ: ATCTGACTGAACCGCT 長さ: 16

# ハミルトン経路とオイラー経路

- ハミルトン閉路では、頂点の数を $n$ とすると、 $O(n!)$ の時間をする
  - 最大重みハミルトン経路は、ゲノムアセンブリの初期に用いられたモデル
  - 数百万～数千万リードを扱うゲノム解析には適用できない
- オイラー閉路では、辺の数を $n$ とすると、 $O(n)$ の時間
  - 現在の主要なアセンブリ手法は、オイラー閉路を適用

# Genome Data Viewer

- [NCBI Genome Data Viewer](#)のページを開いてください
  - Web検索で「Genome Data Viewer」、「NCBI GDV」などと入力すれば、恐らくトップでヒットします
  - この講義では、資料にリンクを埋め込んでいます

## Web検索



# ALDH2遺伝子の検索

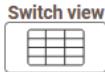
An official website of the United States government [Here's how you know](#) ▾



shimizu.jwu@gmail...

## Genome Data Viewer

生物種（「Home Sapiens」（ヒト）がデフォルト）



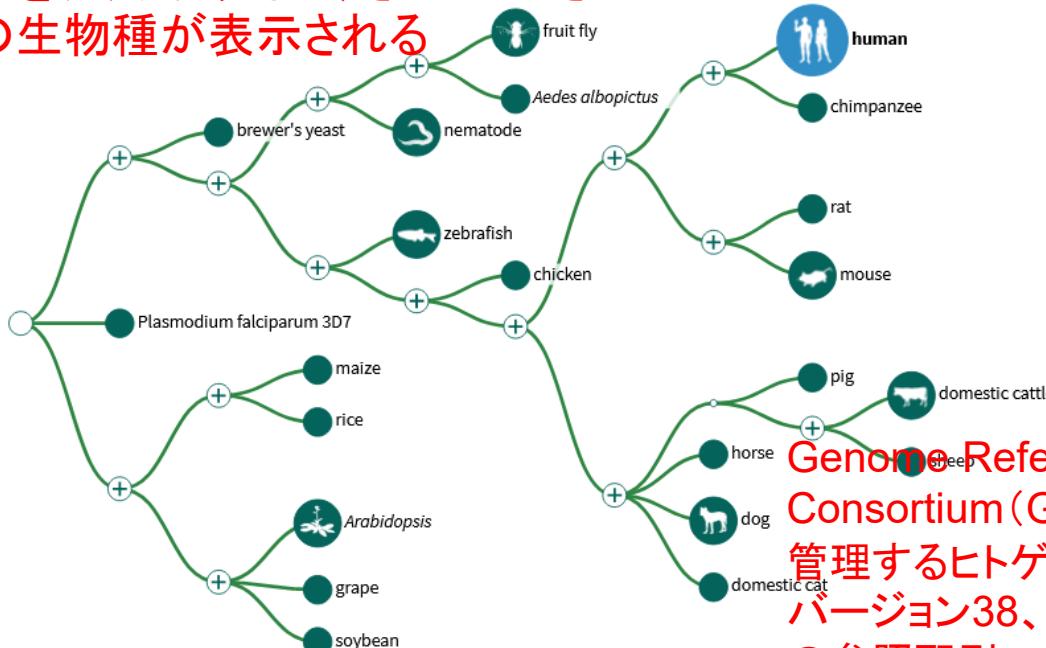
Search organisms

Homo sapiens (human)

To view more organisms in the tree, click on nodes that have '+' signs. Press and hold the '+' to expand and reveal all the subgroups.  
Or, search for an organism using the search box above.

New! Click on Switch view at the top to see another way of navigating genomes.

+をクリックすると、さらにたくさんの生物種が表示される



いろいろな情報をを使って  
検索できる

GDV supports the exploration and analysis of [NCBI-annotated](#) and selected non-NCBI annotated eukaryotic genome assemblies. Currently, assemblies from over 3320 organisms are available.

3320以上の生物を登録

The screenshot shows the detailed information for the Homo sapiens (human) genome assembly. The assembly is GRCh38.p14. The page includes a search bar, assembly details (Name: GRCh38.p14, RefSeq accession: GCF\_000001405.40, GenBank accession: GCA\_000001405.29, Submitter: Genome Reference Consortium, Level: Chromosome, Category: Reference genome), annotation details (Annotation Release: RS\_2024\_08, Release date: Aug 26, 2024), and a chromosome map. A red arrow points from the text "Genome Reference Consortium (GRC)が管理するヒトゲノムのバージョン38、パッチ14の参照配列" to the assembly name "GRCh38.p14". Another red arrow points from the text "いろいろな情報をを使って検索できる" to the search bar.

# ゲノムブラウザで調べてみよう

## Genome Data Viewer

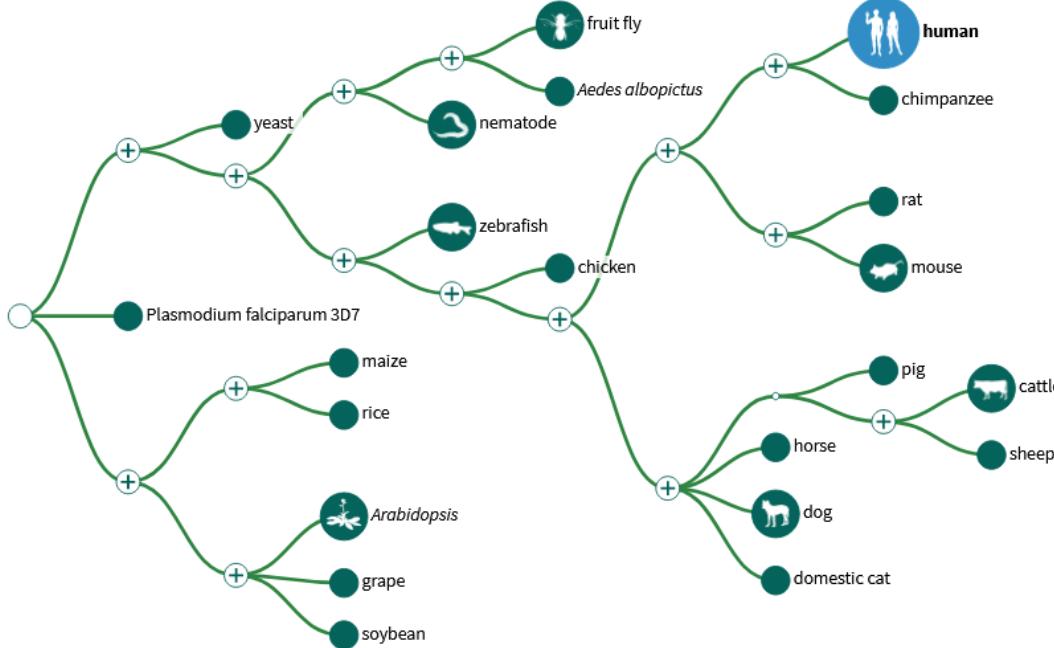


Search organisms

Homo sapiens (human)

To view more organisms in the tree, click on nodes that have '+' signs. Press and hold the '+' to expand and reveal all the subgroups. Or, search for an organism using the search box above.

New! Click on Switch view at the top to see another way of navigating genomes.



ALDH2と入力

**Homo sapiens (human)**

Search in genome: ALDH2

Assembly: GRCh38.p14

Browse genome Compare genomes ...

**Assembly details**

Name	GRCh38.p14
RefSeq accession	GCF_000001405.40
GenBank accession	GCA_000001405.29
Submitter	Genome Reference Consortium
Level	Chromosome
Category	Reference genome

**Annotation details**

Annotation Release	RS_2023_03
Release date	Mar 20, 2023

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

# 他の生物種とゲノムを比較する

An official website of the United States government [Here's how you know](#)



shimizu.jwu@gmail...

Compare genomes with NCBI's Comparative Genome Viewer (CGV) or the brand new Multiple Comparative Genome Viewer (MCGV).

Home Share this page Reset All More Tools More Info

More Tools Compare genomes Download genomes

Genome Data Viewer

Homo sapiens (human)

Assembly BLAST genome 0001405.40 Chr 12 (NC\_000012.12)

NC\_000012.12: 111,761,872 - 111,822,591

p13.33 p13.32 p13.31 p13.2 p12.3 p12.2 p12.1 p11.23 p11.21 p11.1 q12 q13.11 q13.12 q13.13 q13.2 q13.3 q14.1 q14.3 q15 q21.1 q21.2 q21.31 q21.32 q21.33 q22 q23.1 q23.2 q23.3 q24.11 q24.13 q24.21 q24.23 q24.31 q24.32 q24.33

Gene Transcript Exons: click an exon to zoom in, mouse over to see details

Region ALDH2 NM\_000690.4

NC\_000012.12 111,765 K 111,770 K 111,775 K 111,780 K 111,790 K 111,800 K 111,805 K 111,810 K 111,815 K 111,820 K

Tools Tracks Download ?

Genes, MANE Project (release v1.3)

NM\_000690.4 ALDH2 NP\_000681.2

NCBI RefSeq Annotation GCF\_000001405.40-RS\_2024\_08

NM\_000690.4 ALDH2 NP\_000681.2

NM\_001204889.2 MIR6761 (+31)

NM\_001191818.1

Biological regions, aggregate, NCBI RefSeq Annotation GCF\_000001405.40-RS\_2024\_08

enhancer enhancer enhancer enhancer enhancer

Genes, Ensembl release 112

ENSG00000111275 ENSG00000257767 ENSP0000450353.4

ENSP0000450353.4

Cited Variations, dbSNP b156 v2

Ideogram View

Unplaced/unlocalized scaffolds: 166

Alt loci/patches: 514

1 2 3 4 5 6 7 8 9 10 11 12 13 14

15 16 17 18 19 20 21 22 X Y MT

Tracks and User Data

BLAST

Tracks by Accession

Assembly Region Details

「More Tools」をクリック、さらに「Compare genomes」をクリック

# 他の生物種とゲノムを比較する

An official website of the United States government [Here's how you know](#)



Log in

CGV Home Help Release Notes

## Comparative Genome Viewer

This tool allows you to compare two genomes based on assembly-assembly alignments provided by NCBI.

### Set up your view

Make a selection in each of these four steps to view assembly comparison.

1. Select a species

Homo sapiens (human)

2. Select a second species

Pan troglodytes (chimpanzee)

「Pan troglodytes (chimpanzee)」を指定

3. Select an assembly

GRCh38.p14 (GCF\_000001405.40)

4. Select a second assembly

GRI\_mPanTro3-v1.1-hic.freeze\_pri (GCF\_028858775.1)

アセンブリ(ゲノム解析の結果得られる配列の断片をまとめたもの)

ヒト

Not finding your alignment of interest?  
[Fill out the form](#) to request more alignments.

**View Comparison**

GRCh38.p14 (GCF\_000001405.40)

チンパンジー

GRI\_mPanTro3-v1.1-hic.freeze\_pri (GCF\_028858775.1)



Connect with NLM



National Library of Medicine  
8600 Rockville Pike  
Bethesda, MD 20894

Web Policies  
FOIA  
HHS Vulnerability Disclosure

Help  
Accessibility  
Careers

NLM | NIH | HHS | USA.gov

Feedback

# 他の生物種とゲノムを比較する

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Pan troglodytes*  
NHGRI\_mPanTro3-v1.1-hic.freeze\_pri (GCF\_028858775.1) and *Homo sapiens*  
GRCh38.p14 (GCF\_000001405.40). ⓘ

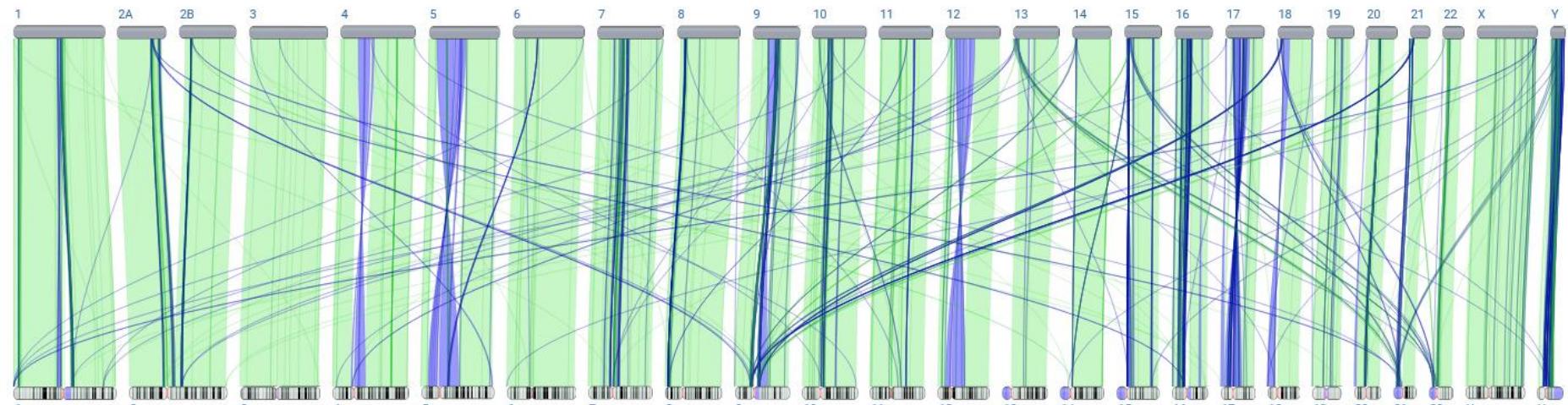
Find a gene in this alignment

Type gene symbol or name, for example Ace2, or ...

Search

緑は、順方向に対応している  
青は、逆方向に対応している  
領域としてねじれているような部分は逆位

*Pan troglodytes* NHGRI\_mPanTro3-v1.1-hic.freeze\_pri (GCF\_028858775.1)



*Homo sapiens* GRCh38.p14 (GCF\_000001405.40)

Reset to genome view A A



Download data

Download image

# 他の生物種とゲノムを比較する



Log in

CGV Home Help Release Notes

## Comparative Genome Viewer

This tool allows you to compare two genomes based on assembly-assembly alignments provided by NCBI.

### Set up your view

Make a selection in each of these four steps to view assembly comparison.

#### 1. Select a species

Homo sapiens (human)

#### 2. Select a second species

Homo sapiens (human)

「Mus musculus (house mouse)」を指定

Mus musculus (house mouse)

Balaenoptera musculus (Blue whale)

Bos taurus (cattle)

Callithrix jacchus (white-tufted-eared marmoset)

Camelus dromedarius (Arabian camel)

[Clear Form](#)

[View Comparison](#)

Not finding your alignment of interest?

[Fill out the form](#) to request more alignments.

FOLLOW NCBI



Connect with NLM



National Library of Medicine  
8600 Rockville Pike  
Bethesda, MD 20894

Web Policies  
FOIA  
HHS Vulnerability Disclosure

Help  
Accessibility  
Careers

NLM | NIH | HHS | USA.gov

Feedback

# 他の生物種とゲノムを比較する

[Log in](#)[CGV Home](#)[Help](#)[Release Notes](#)

## Comparative Genome Viewer

This tool allows you to compare two genomes based on assembly-assembly alignments provided by NCBI.

### Set up your view

Make a selection in each of these four steps to view assembly comparison.

1. Select a species

Homo sapiens (human)

2. Select a second species

Mus musculus (house mouse)

3. Select an assembly

GRCh38.p14 (GCF\_000001405.40)

4. Select a second assembly

GRCm39 (GCF\_000001635.27)

[Clear Form](#) [View Comparison](#)

「View Comparison」をクリック

Not finding your alignment of interest?

[Fill out the form](#) to request more alignments.

FOLLOW NCBI



Connect with NLM



National Library of Medicine  
8600 Rockville Pike  
Bethesda, MD 20894

Web Policies  
FOIA  
HHS Vulnerability Disclosure

Help  
Accessibility  
Careers

NLM | NIH | HHS | USA.gov

Feedback

# 他の生物種とゲノムを比較する

CGV Home Help Release Notes

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Mus musculus* GRCm39 (GCF\_000001635.27) and *Homo sapiens* GRCh38.p14 (GCF\_000001405.40). ⓘ

Find a gene in this alignment

Type gene symbol or name, for example Ace2, or ...

Search

[Go to dotplot view](#)

[Reset to genome view](#)

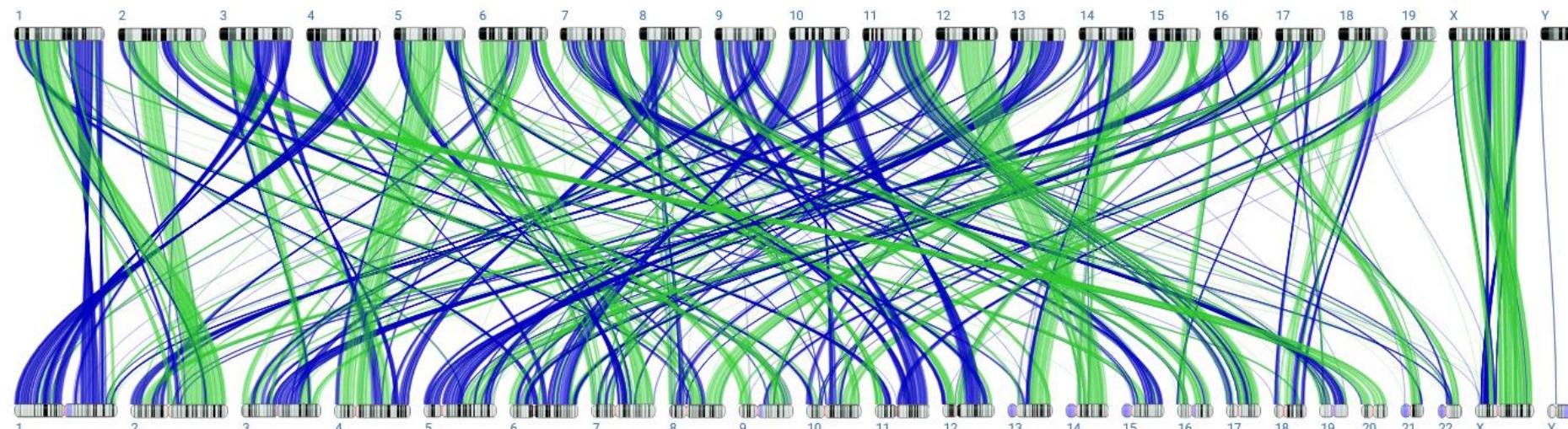


A

A



*Mus musculus* GRCm39 (GCF\_000001635.27)



*Homo sapiens* GRCh38.p14 (GCF\_000001405.40)

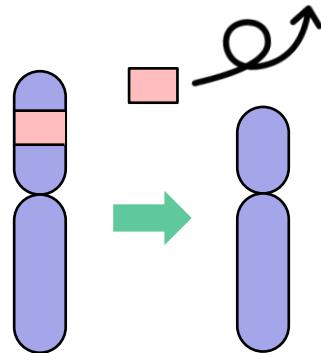


[Download data](#)

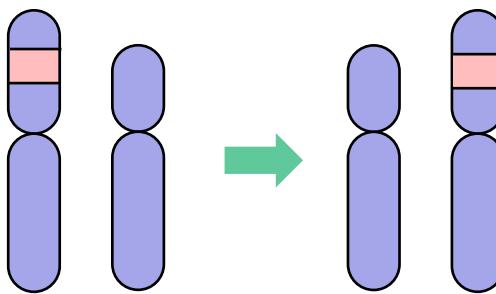
[Download image](#)

# 染色体の再編成

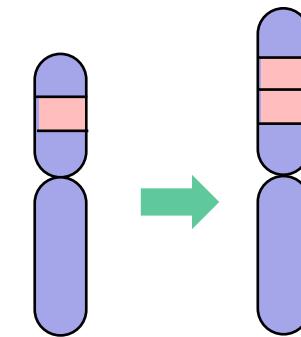
疾患の原因と進化的多様化の原動力の両面をもつ重要なゲノム変化



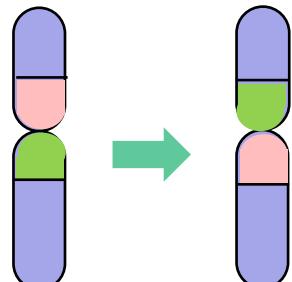
欠失



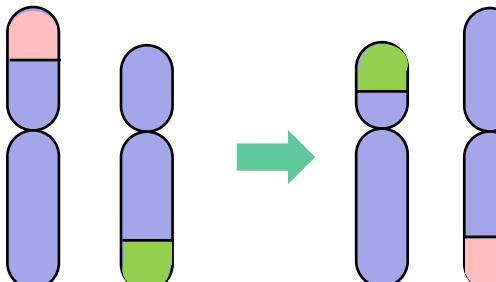
挿入



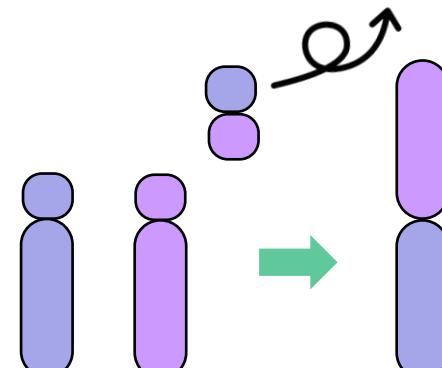
重複



逆位



転座



Robertson型転座

# 遺伝子の位置の比較

CGV Home Help Release Notes

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Mus musculus* GRCm39 (GCF\_000001635.27) and *Homo sapiens* GRCh38.p14 (GCF\_000001405.40). ⓘ

Find a gene in this alignment

HOX

Search

[Go to dotplot view](#)

Search results 「HOX」と入力

Assembly GRCm39 - 100 genes shown

Gene	Description	Location
Hoxb1	homeobox B1	Chr11: 96256547..96259082
Hoxd13	homeobox D13	Chr2: 74498569..74501947
Phox2b	paired-like homeobox 2b	Chr5: 67251751..67256399

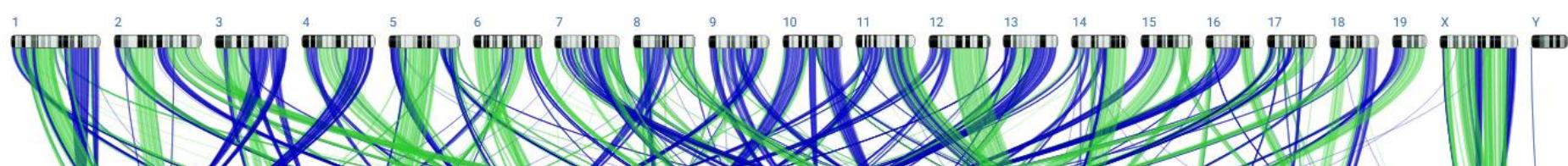
Assembly GRCh38.p14 - 73 genes shown

Gene	Description	Location
HOTAIR	HOX transcript antisense RNA	Chr12: 53962312..53974954
HOXA10	homeobox A10	Chr7: 27170605..27179861
HOXA9	homeobox A9	Chr7: 27162438..27165537

染色体上の位置でソート

枠を広げて見て下さい

*Mus musculus* GRCm39 (GCF\_000001635.27)



# HOX遺伝子の比較

- ホメオボックス遺伝子群 (Hox gene) : 動物の体節形成や前後軸の形成、手足などのパターン形成を司る遺伝子群
- ヒトでは、染色体7番にHOXAクラスタが存在、マウスでは、染色体6番に存在
- エンハンサーHOTTIP (HOXA distal transcript antisense RNA) は染色体7番、ヒトではHOXAクラスターの近くに存在 → 手指の発達においてHOXAクラスターの遺伝子が一貫したパターンで発現し、手指の細かい運動や構造に関係している可能性がある
- エンハンサー: 特定の遺伝子の転写を活性化するDNA配列
- 遺伝子クラスター: よく似た遺伝子が、染色体上で互いに近い位置にまとまって並んでいる状態

# 他の生物種とゲノムを比較する

CGV Home Help Release Notes

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Mus musculus* GRCm39

(GCF\_000001635.27) and *Homo sapiens* GRCh38.p14 (GCF\_000001405.40). ⓘ

Find a gene in this alignment

Type gene symbol or name, for example Ace2, or ...

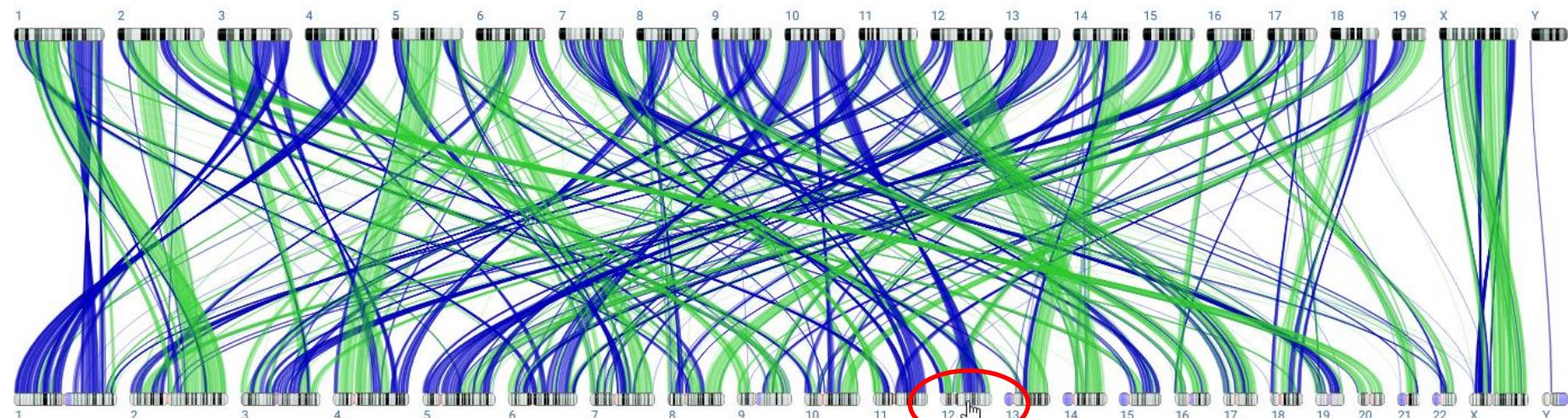
Search

[Go to dotplot view](#)

[Reset to genome view](#)

A A

*Mus musculus* GRCm39 (GCF\_000001635.27)



*Homo sapiens* GRCh38.p14 (GCF\_000001405.40)

12番染色体(Ch12)を選ぶ

[Download data](#)

[Download image](#)

Adjust your view

Reciprocal best placed alignments (forward and reverse) are shown by default.

# 他の生物種とゲノムを比較する

CGV Home Help Release Notes

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Mus musculus* GRCm39

(GCF\_000001635.27) and *Homo sapiens* GRCh38.p14 (GCF\_000001405.40). ⓘ

Find a gene in this alignment

ALDH2

Search

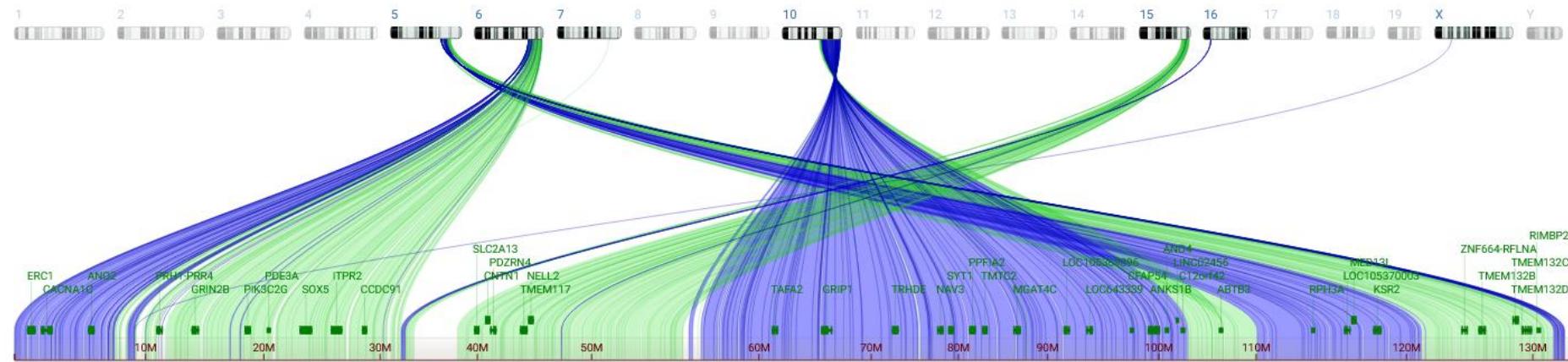
[Go to dotplot view](#)

[Reset to genome view](#)

A A

「ALDH2」と入力

*Mus musculus* GRCm39 (GCF\_000001635.27)



Chr 12

*Homo sapiens* GRCh38.p14 (GCF\_000001405.40)

[Download data](#)

[Download image](#)

# 他の生物種とゲノムを比較する

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Mus musculus* GRCm39

(GCF\_000001635.27) and *Homo sapiens* GRCh38.p14 (GCF\_000001405.40). ⓘ

Find a gene in this alignment

[Go to dotplot view](#)

### Search results

Assembly GRCm39 - 1 gene shown

Gene	Description	Location
Aldh2	aldehyde dehydrogenase 2, mitochondrial	Chr5: 121704090..121731887

Assembly GRCh38.p14 - 1 gene shown

Gene	Description	Location
ALDH2	aldehyde dehydrogenase 2 family member	Chr12: 111766933..111817532

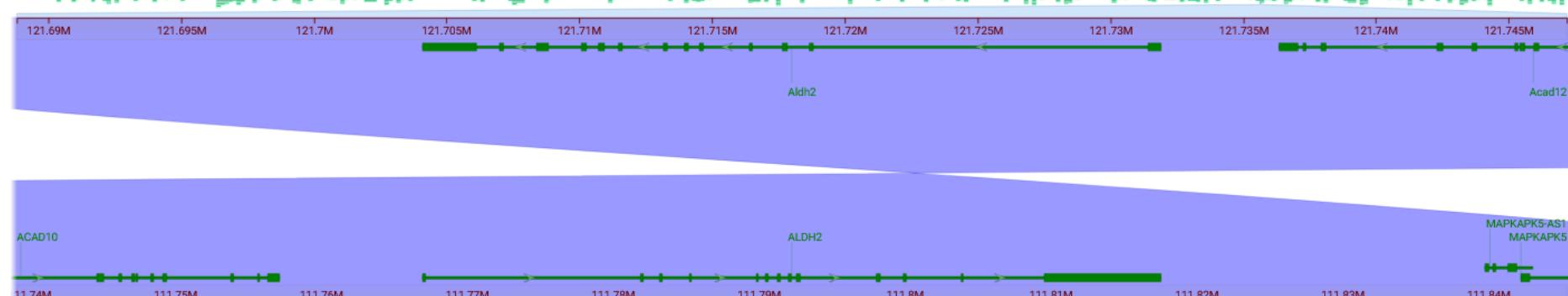
ヒト(下)とマウス(上)のそれぞれの染色体上の位置が表示される

*Mus musculus* GRCm39 ([GCF\\_000001635.27](#))



[Reset to genome view](#)

A A



*Homo sapiens* GRCh38.p14 ([GCF\\_000001405.40](#))

# 他の生物種とゲノムを比較する

Find a gene in this alignment

ALDH2

Search

[Go to dotplot view](#)

## Search results

Assembly GRCm39 - 1 gene shown

Gene	Description	Location
Aldh2	aldehyde dehydrogenase 2, mitochondrial	Chr5: 121704090..121731887

Assembly GRCh38.p14 - 1 gene shown

Gene	Description	Location
ALDH2	aldehyde dehydrogenase 2 family member	Chr12: 111766933..111817532

[Reset to genome view](#)



Mus musculus GRCm39 ([GCF\\_000001635.27](#))

Chr 5

121.69M 121.695M 121.7M 121.705M 121.71M 121.715M 121.72M 121.725M 121.73M 121.735M 121.74M 121.745M

Aldh2

Aced12

Alignment details ...  
Zoom to alignment  
**View alignment sequence**  
Download alignment FASTA  
View alignment on GRCm39 in GDV  
View alignment on GRCh38.p14 in GDV

ACAD10

11.74M 111.75M 111.76M 111.77M 111.78M 111.79M 111.8M 111.81M 111.82M 111.83M 111.84M

ALDH2

MAPKAPK5-AS1  
MAPKAPK5

Chr 12

ブラウザの表示画面で右クリック  
→「View alignment sequence」を選択

## 他の生物種とゲノムを比較する

## Comparative Genome Viewer

You are ready to explore whole genome alignment between *Mus musculus* GRCm39

(GCF\_000001635.27) and *Homo sapiens* GRCh38.p14 (GCF\_000001405.40).

### Find a gene in this alignment

## ALDH2

Search

## Go to dotplot view

*Mus musculus* GRCm39 (GCF\_000001635.27)

Chr 5

121.69M

Aldh

Acad11

Chr 5: 121,672,363..122,145,451

CCAGGGAGGTAAAGG

表示部分(橙色)を、エキソン部分(紅白)に移動する  
両方の配列が存在するエキソン部分を任意に選んで、ヒトとマウスの配列を並べて表示してみよう

*Homo sapiens* GRCh38.p14 (GCF\_000001405.40)

Chr 5: 121,672,363..122,145,45

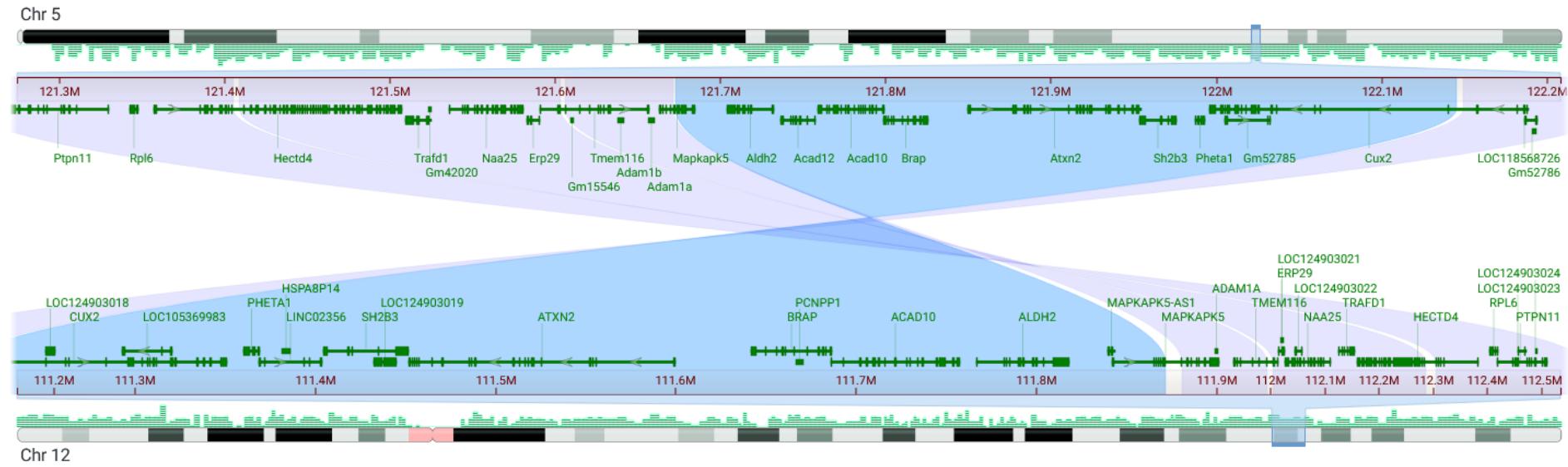
121211410

Chr 12

# 他の生物種とゲノムを比較する

ズームアウトして、周辺の遺伝子を見てみよう

Mus musculus GRCm39 (GCF\_000001635.27)



Homo sapiens GRCh38.p14 (GCF\_000001405.40)



Download data Download image

# 遺伝子配列の取得

## Comparative Genome Viewer

You are ready to explore whole genome alignment between ***Mus musculus* GRCm39**

(GCF\_000001635.27) and ***Homo sapiens* GRCh38.p14** (GCF\_000001405.40). ⓘ

Find a gene in this alignment

ALDH2 Search

[Go to dotplot view](#)

### Search results

Assembly GRCm39 - 1 gene shown

Gene	Description	Location
Aldh2	aldehyde dehydrogenase 2, mitochondrial	Chr5: 121704090..121731887

Assembly GRCh38.p14 - 1 gene shown

Gene	Description	Location
ALDH2	aldehyde dehydrogenase 2 family member	Chr12: 111766933..111817532

他の画面でもよいが、遺伝子名の表示された状態にあることが必要

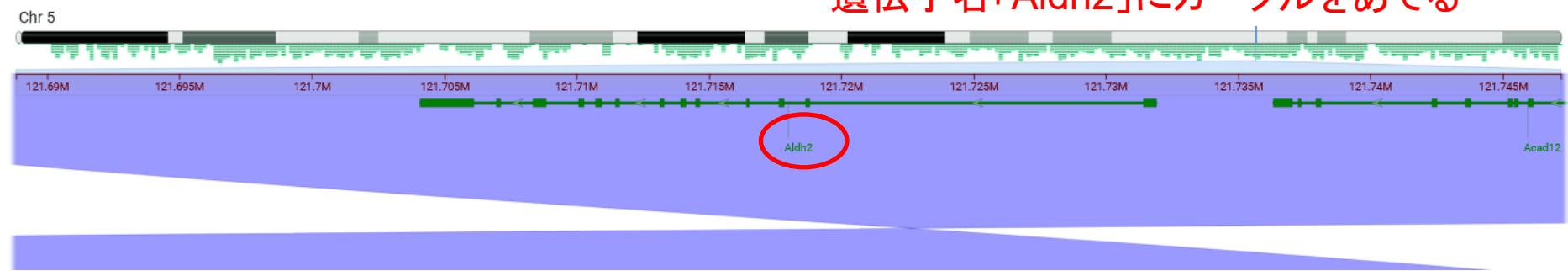
***Mus musculus* GRCm39 (GCF\_000001635.27)**

[Reset to genome view](#)

A A



遺伝子名「Aldh2」にカーソルをあてる



# 遺伝子配列の取得

## Comparative Genome Viewer

You are ready to explore whole genome alignment between ***Mus musculus* GRCm39**

(GCF\_000001635.27) and ***Homo sapiens* GRCh38.p14** (GCF\_000001405.40). ⓘ

Find a gene in this alignment

[Go to dotplot view](#)

### Search results

Assembly GRCm39 - 1 gene shown

Gene	Description	Location
Aldh2	aldehyde dehydrogenase 2, mitochondrial	Chr5: 121704090..121731887

Assembly GRCh38.p14 - 1 gene shown

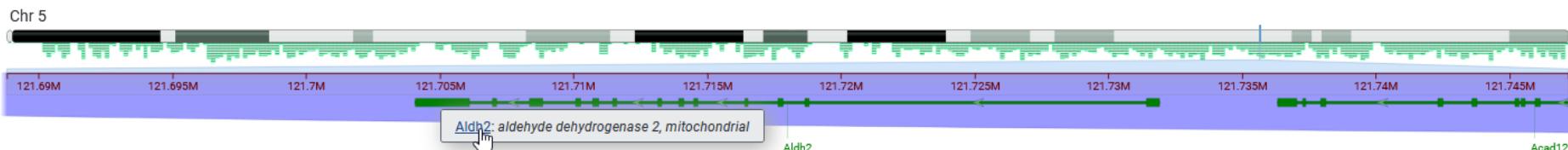
Gene	Description	Location
ALDH2	aldehyde dehydrogenase 2 family member	Chr12: 111766933..111817532

[Reset to genome view](#)

A A



***Mus musculus* GRCm39 (GCF\_000001635.27)**



「Aldh2」をクリック

# 遺伝子配列の取得

Gene

Gene

Advanced

Search

Hide sidebar >>

Full Report

## 遺伝子Aldh2(マウス)のページ

Send to:

Download Datasets

配列をダウンロードするには  
「Download Datasets」をクリック

Aldh2 aldehyde dehydrogenase 2, mitochondrial [ *Mus musculus* (house mouse) ]  
Gene ID: 11669, updated on 6-Oct-2023

Summary

Official Symbol	Aldh2 provided by MGI
Official Full Name	aldehyde dehydrogenase 2, mitochondrial provided by MGI
Primary source	MGI:MGI:99600
See related	<a href="#">Ensembl</a> :ENSMUSG00000029455 <a href="#">AllianceGenome</a> :MGI:99600
Gene type	protein coding
RefSeq status	VALIDATED
Organism	<i>Mus musculus</i>
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mus; Mus
Also known as	Ahd5; ALDH1; Ahd-5; AHD-M1; ALDH-E2
Summary	Predicted to enable NADH binding activity; aldehyde dehydrogenase (NAD+) activity; and identical protein binding activity. Predicted to be involved in several processes, including behavioral response to ethanol; cellular detoxification of aldehyde; and intrinsic apoptotic signaling pathway in response to oxidative stress. Located in mitochondrion. Is expressed in several structures, including alimentary system; genitourinary system; nervous system; respiratory system; and sensory organ. Human ortholog(s) of this gene implicated in several diseases, including alcohol use disorder; artery disease (multiple); diabetes mellitus (multiple); diabetic neuropathy; and liver disease (multiple). Orthologous to human ALDH2 (aldehyde dehydrogenase 2 family member). [provided by Alliance of Genome Resources, Apr 2022]
Expression	Broad expression in adrenal adult (RPKM 574.9), liver adult (RPKM 432.0) and 18 other tissues <a href="#">See more</a>
Orthologs	<a href="#">human</a> <a href="#">all</a>

NEW

[Try the new Gene table](#)

[Try the new Transcript table](#)

Table of contents  
Summary

- Genomic context
- Genomic regions, transcripts, and products
- Expression
- Bibliography
- Variation
- Pathways from PubChem
- Interactions
- General gene information
  - Markers, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

Genome Browsers

Genome Data Viewer

Ensembl

UCSC

Related information

- Order cDNA clone
- BioAssay by Target (List)
- BioAssay, by Gene target
- BioAssays, RNAi Target, Tested
- BioProjects
- BioSystems

Chromosome 5 - NC\_000071.7

# 遺伝子配列の取得



Log in

Gene Gene Search Advanced Help

Full Report Send to: Hide sidebar >>

**Aldh2 aldehyde dehydrogenase 2, mitochondrial [ *Mus musculus* (house mouse) ]**

Gene ID: 11669, updated on 6-Oct-2023

**Summary**

Official Symbol Aldh2 provided by MGI  
Official Full Name aldehyde dehydrogenase 2, mitochondrial provided by MGI  
Primary source MGI/MGI:99600  
See related Ensembl:ENSMUSG00000029455 Alliance Genome HG:99600  
Gene type protein coding  
RefSeq status VALIDATED  
Organism *Mus musculus*  
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Murinae; Mus; Mus  
Also known as Ahd5; ALDH1; Ahd-5; AHD-M1; ALDH-E2  
Summary Predicted to enable NADH binding activity; aldehyde dehydrogenase (NAD<sup>+</sup>) activity; and identical protein binding activity. Predicted to be involved in several processes, including behavioral response to ethanol; cellular detoxification of aldehyde; and intrinsic apoptotic signaling pathway in response to oxidative stress. Located in mitochondrion. Is expressed in several structures, including alimentary system; genitourinary system; nervous system; respiratory system; and sensory organ. Human ortholog(s) of this gene implicated in several diseases, including alcohol use disorder; artery disease (multiple); diabetes mellitus (multiple); diabetic neuropathy; and liver disease (multiple). Orthologous to human ALDH2 (aldehyde dehydrogenase 2 family member). [provided by Alliance of Genome Resources, Apr 2022]  
Expression Broad expression in adrenal adult (RPKM 574.9), liver adult (RPKM 432.0) and 18 other tissues See more  
Orthologs human all Try the new Gene table Try the new Transcript table

**Genomic context**

Location: 5 F; 5 61.86 cm Exon count: 13

Annotation release	Status	Assembly	Chr	Location
RS_2023_04	current	GRCm39 (GCF_000001635.27)	5	NC_000071.7 (121704090..121731887, complement)
108.20200622	previous assembly	GRCm38.p6 (GCF_000001635.26)	5	NC_000071.6 (121566027..121593824, complement)

Chromosome 5 - NC\_000071.7

**Download Datasets**

Gene Sequences (FASTA)  
Transcript sequences (FASTA)  
Protein sequences (FASTA)

In addition, your package will include a detailed data report in both TSV and JSONL formats.

File name: Aldh2\_datasets.zip

Download

すべてにチェックを入れて、「Download」ボタンを押す

しかし、ここでは、別な方法でダウンロードを行う

Table of contents  
Summary  
Context  
Regions, transcripts, and products  
From PubChem  
Gene information  
Homology, Gene Ontology  
Protein information  
NCBI Reference Sequences (RefSeq)  
Related sequences  
Additional links  
Genome Browsers  
Genome Data Viewer  
Ensembl  
UCSC  
Related information  
Order cDNA clone  
BioAssay by Target (List)  
BioAssay by Gene target  
BioAssays, RNAi Target, Tested  
BioProjects  
BioSystems

# 遺伝子配列の取得

Gene

Gene

Advanced

Search

Help

Full Report

## 遺伝子Aldh2(マウス)のページ

Send to:

Hide sidebar >>

### Aldh2 aldehyde dehydrogenase 2, mitochondrial [ *Mus musculus* (house mouse) ]

Gene ID: 11669, updated on 6-Oct-2023

Download Datasets

#### Summary

#### Table of contents

Summary

Genomic context

Genomic regions, transcripts, and products

Expression

Bibliography

Variation

Pathways from PubChem

Interactions

General gene information

Markers, Homology, Gene Ontology

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Additional links

#### Genome Browsers

Genome Data Viewer

Ensembl

UCSC

Official Symbol Aldh2 provided by MGI

Official Full Name aldehyde dehydrogenase 2, mitochondrial provided by MGI

Primary source MGI:MGI:99600

See related Ensembl:ENSMUSG00000029455 AllianceGenome:MGI:99600

Gene type protein coding

RefSeq status VALIDATED

Organism *Mus musculus*

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mus; Mus

Also known as Ahd5; ALDH1; Ahd-5; AHD-M1; ALDH-E2

Summary Predicted to enable NADH binding activity; aldehyde dehydrogenase (NAD<sup>+</sup>) activity; and identical protein binding activity. Predicted to be involved in several processes, including behavioral response to ethanol; cellular detoxification of aldehyde; and intrinsic apoptotic signaling pathway in response to oxidative stress. Located in mitochondrion. Is expressed in several structures, including alimentary system; genitourinary system; nervous system; respiratory system; and sensory organ. Human ortholog(s) of this gene implicated in several diseases, including alcohol use disorder; artery disease (multiple); diabetes mellitus (multiple); diabetic neuropathy; and liver disease (multiple). Orthologous to human ALDH2 (aldehyde dehydrogenase 2 family member). [provided by Alliance of Genome Resources, Apr 2022]

Expression Broad expression in adrenal adult (RPKM 574.9), liver adult (RPKM 432.0) and 18 other tissues [See more](#)

Orthologs [human](#) [all](#)

NEW

Try the new [Gene table](#)

Try the new [Transcript table](#)

「Transcript table」をクリック

#### Genomic context

Location: 5 F, 5 61.86 cM

See Aldh2 in [Genome Data Viewer](#)

Exon count: 13

Annotation release	Status	Assembly	Chr	Location
RS_2023_04	current	GRCm39 ( <a href="#">GCF_000001635.27</a> )	5	NC_000071.7 (121704090..121731887, complement)
108.20200622	previous assembly	GRCm38.p6 ( <a href="#">GCF_000001635.26</a> )	5	NC_000071.6 (121566027..121593824, complement)

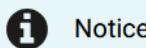
Chromosome 5 - NC\_000071.7

# 遺伝子配列の取得



Search NCBI ...

Log in



## Notice

Because of a lapse in government funding, the information on this website may not be up to date, transactions submitted via the website may not be processed, and the agency may not be able to respond to inquiries until appropriations are enacted. The NIH Clinical Center (the research hospital of NIH) is open. For more details about its operating status, please visit [cc.nih.gov](http://cc.nih.gov). Updates regarding government operating status and resumption of normal operations can be found at [opm.gov](http://opm.gov).

NCBI Datasets Taxonomy Genome Gene Command-line tools Documentation

## Transcripts and Proteins

Aldh2 – aldehyde dehydrogenase 2, mitochondrial

Mus musculus (house mouse)

The screenshot shows a table of gene and transcript information. The columns are: Gene ID, Symbol, Transcript, Length (nt), Protein, Length (aa), Protein name, and Action. There are two transcripts listed: NM\_001308450.1 and NM\_009656.4. The NM\_009656.4 row has a checked checkbox in the first column and a 'RefSeq Select' button in the fifth column. A red circle highlights the checked checkbox in the first column of the second row.

	Gene ID	Symbol	Transcript	Length (nt)	Protein	Length (aa)	Protein name	Action
<input type="checkbox"/>	11669	Aldh2	NM_001308450.1	2,060	NP_001295379.1	471	aldehyde dehydrogenase 2, mitochondrial	...
<input checked="" type="checkbox"/>	11669	Aldh2	NM_009656.4 <span>RefSeq Select</span>	3,883	NP_033786.1	519	aldehyde dehydrogenase 2, mitochondrial	...

「RefSeq Select」を選択  
アイソフォーム1が選択される

# 遺伝子配列の取得

An official website of the United States government [Here's how you know](#)



Search NCBI ...

Log in

## Notice

Because of a lapse in government funding, the information on this website may not be up to date, transactions submitted via the website may not be processed, and the agency may not be able to respond to inquiries until appropriations are enacted. The NIH Clinical Center (the research hospital of NIH) is open. For more details about its operating status, please visit [cc.nih.gov](http://cc.nih.gov). Updates regarding government operating status and resumption of normal operations can be found at [opm.gov](http://opm.gov).

NCBI Datasets Taxonomy Genome Gene Command-line tools Documentation

## Transcripts and Proteins

Aldh2 – aldehyde dehydrogenase 2, mitochondrial

Mus musculus (house mouse)

2 Transcripts 1 selected							
<input type="checkbox"/>	Gene ID	Symbol	Transcript	Length (nt)	Protein	Length (aa)	Protein name
<input type="checkbox"/>	11669	Aldh2	NM_001308450.1	2,060	NP_001295379.1	471	aldehyde dehydrogenase 2, mitochondrial
<input checked="" type="checkbox"/>	11669	Aldh2	NM_009656.4 <small>RefSeq Select</small>	3,883	NP_033786.1	519	aldehyde dehydrogenase 2, mitochondrial

「Download」を選択

Download  
See in Viewer

# 遺伝子配列の取得

An official website of the United States government [Here's how you know](#)



Search NCBI ...

Log in

## Notice

Because of a lapse in government funding, the agency may not be able to respond to requests for information about its operating status, please visit [www.nih.gov](#).

NCBI Datasets

Taxonomy

Gene

ALDH2 - aldehyde dehydrogenase 2

Homo sapiens

Summary

Transcripts and proteins

Download

datasets

API

Gene name

Symbol

Gene ID

Organism

Gene type

Download Package

1 gene selected for download

「Gene sequence」も選択しておいて下さい

Select sequence files

- Gene sequences (FASTA)
- Transcript sequences (FASTA)
- Protein sequences (FASTA)
- CDS (FASTA)
- 5' UTR (FASTA)
- 3' UTR (FASTA)

Select data report files

- Gene report
- Product report

Select data report file format(s)

- JSONL
- JSONL and TSV

Your selected data will be downloaded as a ZIP archive

Estimated file size is 1 MB

Name your file

ncbi\_dataset.zip

Cancel

Download

Other identifiers

Ensembl ENSG00000111275

HGNC HGNC:404

OMIM 100650

UniProt P05091

Feedback

# 遺伝子配列の取得

- ncbi\_dataset.zipという圧縮ファイルとして、ダウンロードされる
  - 内容によらず、同じ名前のファイルになるので注意
- これを解凍する

# 遺伝子配列の取得

- 圧縮ファイルの中の`ncbi_dataset/data/`の下に配列が存在
  - `gene.fna`
  - `rna.fna`
  - `protein.faa`

拡張子	意味
<code>.fasta, .fas, .fa, .fsa, .seq</code>	とくに内容を指定しないfasta ファイル
<code>.fna</code>	遺伝子/RNAの塩基配列のfastaファイル
<code>.ffn</code>	ゲノムの翻訳領域(coding region)を含む fastaファイル
<code>.faa</code>	アミノ酸配列のfastaファイル
<code>.mpna</code>	複数のアミノ酸配列を含むfastaファイル
<code>.frn</code>	Non-coding RNAのfastaファイル

# 遺伝子配列の取得

- Gene Sequences (FASTA): 遺伝子の塩基配列
  - .fnaファイル
- Transcript Sequences (FASTA): mRNAの塩基配列
  - .rnaファイル
  - cDNA (complementary DNA) の配列
  - cDNA: mRNAから逆転写酵素を用いた逆転写反応によって合成された二本鎖DNA  $\leftrightarrow$  データベース上はエキソン領域 (coding + UTR) を正しい順序でつなげた「成熟mRNA」に対する塩基 (ACGT) の配列
- Protein Sequences (FASTA): タンパク質のアミノ酸配列
  - .proteinファイル

# 遺伝子配列の取得

- マウスのgene.fnaの内容

```
>NC_000071.7:c121731887-121704090 Aldh2 [organism=Mus musculus]
[GeneID=11669] [chromosome=5]
ATTCTCTCGCCGCCATATCTGCACAGATGTGAGCCTAGGCGCCAGCCACCCCTGCTAGGAGCGCACACC
ACTCTGGCTAGGCTTCAGGGTTCTGCAAACCTCCATCTGACTTGGCTTGGAGCCAGGGTCGCG
CCCCTAGGCCGTGAGGGCTGGGACTCCCTGACCACGCCCGTGTCTCCGCCTCCATTGGCGGCTGC
AGGGGGCGGAGGCAGGGACTTGTCTCAACGCTGCAGTCGCCCTCCGATCGGAAGGCTCTCGGCT
CCGTTGGCTGGCTGCCATTTCAGTTCAAGCTGGCTCAGTTCAGCATGCTGC
GCGCCGCACTCACCACGTCCGCCGGACCGCGCCTGAGCCGCTGTTGCCGCCACCAGCGC
GGTGCCAGCCCCAACCATCAGCCTGAGGTCTTGCAACCAGGTGAGACTCACTTCGCCTCCTGGGCC
CTAGGGCGGGACGGAGCGGGCGACGCCAGCTCAGTTCCCGCCTCCGCGCCAGCGATCTTGCGCGTCT
TCCCTGCCCATGTGGCCGCAGTGTGCGCTCTGGCCTGGACAGTCCTGGCCTCTTACCCGCAT
CGATGCAGCCTGCTCCTCCCCGCCGGTCCGGGTTACCCAGGAAGGCCAGCCAGCTCCTCTC
TGCAGACCTCCCATACACCCCTACCCGATTACGCAGACCTGAGCCAGAGTTCAGAAGCAAAAGCTGGG
TCAGGACCCCTCTGTGGCCGCTTACCCAACCGGGACTTGGAAATGTTGCTTAGCTTTGAGTTTG
GTAAAGTCTGGAATCACATGAAAGAACGTGGTTGTGAAGTATTCTTTGTTGTCAACAAGTGGGACT
TGGCTGCCCTTCTGTAAAATGGGAAAGCAGTCAGGGCTGATATCTTGATGTGTTATTTGAAAGAAGA
TCTCATTGTAGCCAAGGCTGGCCGGATTCTGTGTTCCGTCAGCCACTCCCACCCACTTGG
TTTGTGAGGCCAGACACGCAGCTACCACCTCCGGTCTCAGCTTCTACTTTCTTGGAAACAG
.....
ACTCACCTAGTGAGTGAACAAGGCGTGGAGAGCAAGCTGCCATCACAGGCACAAGAACGGACGGTGAGC
TTAGCTTAGAAACTAGCCAGTCAGAGGCAGAGCTGAGGGTAGAAGGCTGATGAAGCCCTGAAGTTGTCCT
TCGACCTCCATATACACATCCCTGTATGTGATGCGCACTCAATGAAATAAGTAAATACAATTG
AAAGATCA
```

# 遺伝子配列の取得

- マウスのrna.fnaの内容

```
>NM_009656.4 Aldh2 [organism=Mus musculus] [GeneID=11669]
[transcript=1]
ATTCTTCGCCGCCATATCTGCACAGATGTGAGCCTAGGCCAGCCACCCCTGCTAGGGAGCGCACACC
ACTCTGGCTAGGCTTCTCAGGGTTCTGCAAACCTCCATCTGACTTGGCTTGAGGCCAGGGGTCGCG
CCCCTAGGCCGTGAGGGGCTGGGACTCCCTGACCACGCCCGTGTCTCCGCCTCCATTGGCGGCTGC
AGGGGGCGGAGGCGAGGAAGTGTCTTCAACGCTGCAGTCGCCCTCCGATCGGAAGGCTCTCGGCT
CCGTCGGCTCGGCTCGCCATTTCAGTTCAAGCTCCGCTCAGTTCAAGCATGCTGC
GCGCCGCACTCACCCTGTCGCCGGACCGCGCCTGAGCCGCCTGTTGTCCGCCGCCACCAGCGC
GGTGCAGCCCCAACCATCAGCCTGAGGTCTTCTGCAACCAGATCTCATTAAACAATGAGTGGCACGAC
.....
ACGGTGAGCTTAGCTTAACTAGCCAGTCAGAGGCAGAGCTGAGGGTAGAAGGCTGATGAAGCCCTGA
AGTTGTCCTTCGACCTCCATATACACATCCCTGTATGTGCATGCGCACTCAATGAAATAATAAGTAAAT
ACAATTTAAAGATCAAAAAAAAAAAAAAAA
```

```
>NM_001308450.1 Aldh2 [organism=Mus musculus] [GeneID=11669]
[transcript=2]
ATTCTTCGCCGCCATATCTGCACAGATGTGAGCCTAGGCCAGCCACACCAGAGGCCAGGGCTGCTAGGGAGCGCACACC
ACTCTGGCTAGGCTTCTCAGGGTTCTGCAAACCTCCATCTGACTTGGCTTGAGGCCAGGGGTCGCG
CCCCTAGGCCGTGAGGGGCTGGGACTCCCTGACCACGCCCGTGTCTCCGCCTCCATTGGCGGCTGC
AGGGGGCGGAGGCGAGGAAGTGTCTTCAACGCTGCAGTCGCCCTCCGATCGGAAGGCTCTCGGCT
CCGTCGGCTCGGCTCGCCATTTCAGTTCAAGCTCCGCTCAGTTCAAGCATGCTGC
GCGCCGCACTCACCCTGTCGCCGGACCGCGCCTGAGCCGCCTGTTGTCCGCCGCCACCAGCGC
GGTGCAGCCCCAACCATCAGCCTGAGGTCTTCTGCAACCAGATCTCATTAAACAATGAGTGGCACGAC
.....
TGTGACGGTGAGAGCCCAGGCTGAGCATAAACAACTGCTCAGCGCCTGCTTCACTCACCAGTTAGGATTTAA
GAAAGAGCAAACGTGGATGGGCTGCATAGAGATCAGTACCCCTGCTGACTGGTTGGAATAAAATAGT
CAGCTACTGTGGAAAAAAAAAAAAAAA
```

アイソフォームを指定しないと、2つの  
アイソフォームの配列が取得される

# 遺伝子配列の取得

- マウスのprotein.faaの内容

>NP\_001295379.1 Aldh2 [organism=Mus musculus] [GeneID=11669] [isoform=2 precursor]

MLRAALTTVRRGPRLSRLLSAAATSAVPAPNHQPEVFCNQIFINNEHDAVSRKTFPTVPNSTGEVICQV  
AEGNKEDVDKAVKAARAALQLGSPWRRMDASDRGRLLYRLADLIERDRTYLAALETLDNGKPYVISYLVD  
LDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCQIIPWNFPLLMQAWKLGPALATGNVVVM  
KVAEQTPLTALYVANLIKEAGFPPGVVNIVPGFGPTAGAAIASHEGVDKVAFTGSTEVGHLIQVAAGSSN  
LKRVTLELGGKSPNIIMSDADMDWAVEQAHFALFFNQGQCCCAGSRTFVQENVYDEFVERSVARAKSRVV  
GNPFDSRTEQGPQVDETQFKKILGYIKSGQQEGAKLLCGGGAAADRGYFIQPTVFGDVKDGMTIAKEEIF  
GPVMQILKFKTIEEVVGRANDSKYGLAAAVFTKDLKANYLSQALQAGTVW

>NP\_033786.1 Aldh2 [organism=Mus musculus] [GeneID=11669] [isoform=1 precursor]

MLRAALTTVRRGPRLSRLLSAAATSAVPAPNHQPEVFCNQIFINNEHDAVSRKTFPTVPNSTGEVICQV  
AEGNKEDVDKAVKAARAALQLGSPWRRMDASDRGRLLYRLADLIERDRTYLAALETLDNGKPYVISYLVD  
LDMVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCQIIPWNFPLLMQAWKLGPALATGNVVVM  
KVAEQTPLTALYVANLIKEAGFPPGVVNIVPGFGPTAGAAIASHEGVDKVAFTGSTEVGHLIQVAAGSSN  
LKRVTLELGGKSPNIIMSDADMDWAVEQAHFALFFNQGQCCCAGSRTFVQENVYDEFVERSVARAKSRVV  
GNPFDSRTEQGPQVDETQFKKILGYIKSGQQEGAKLLCGGGAAADRGYFIQPTVFGDVKDGMTIAKEEIF  
GPVMQILKFKTIEEVVGRANDSKYGLAAAVFTKDLKANYLSQALQAGTVW  
アイソフォームを指定しないと、2つの  
アイソフォームの配列が取得される

# FASTA形式

- 1行の説明とその後に続く配列データから構成される
  - ヘッダ行: 先頭の1行。'>'で始まる。配列の説明。
  - 配列データ: 配列を表す文字列。複数行にわたってよい。
    - IUB/IUPACで規定されているアミノ酸または核酸コード
    - ただし、
      - 小文字は大文字に変換される
      - '-'でギャップを表す
      - 数字は受け付けない

A	adenosine	M	A C (amino)
C	cytidine	S	G C (strong)
G	guanine	W	A T (weak)
T	thymidine	B	G T C
U	uridine	D	G A T
R	G A (purine)	H	A C T
Y	T C (pyrimidine)	V	G C A
K	G T (keto)	N	A G C T (any)
		-	gap of indeterminate length

A	alanine	P	proline
B	aspartate or asparagine	Q	glutamine
C	cystine	R	arginine
D	aspartate	S	serine
E	glutamate	T	threonine
F	phenylalanine	U	selenocysteine
G	glycine	V	valine
H	histidine	W	tryptophan
I	isoleucine	Y	tyrosine
K	lysine	Z	glutamate or glutamine
L	leucine	X	any
M	methionine	*	translation stop
N	asparagine	-	gap of indeterminate length

>sp|P61626|LYSC\_HUMAN Lysozyme C OS=Homo sapiens GN=LYZ PE=1 SV=1  
MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRST  
DYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRDVR  
QYVQGCGV

>ref|NC\_000012.11|:69742134-69748013 Homo sapiens chromosome 12, GRCh37  
AAATACTGGGCCAGCTCACCTGGTCAGCCTAGCACTCTGACCTAGCAGTCAACATGAAGGCTCTCATT  
GTTCTGGGCTTGTCCCTTCTGTTACGGCCAGGGCAAGGTCTTGAAAGGTGTGAGTTGGCAGAA  
CTCTGAAAAGATTGGGAATGGATGGCTACAGGGGAATCAGCCTAGCAAACGTAAAGTCACTCTCCATAA  
TTCCAGAGAATTAGCTACGTATGGAACAGACACTAGGAGAGAAGGAAGAAGAAGGGCTTGAGTGA

(途中まで)

IUPAC: International Union of Pure and Applied Chemistry(国際純正応用化学連合)

IUB: International Union of Biochemistry(国際生化学連合)

# 配列の準備

- 取得したファイルは以下のように名前を変更する
  - rna.fna → mouse-rna.fna
  - protein.faa → mouse-protein.faa
- 同様に、ヒトの配列をダウンロードして下さい
- NCBIの「gene」から検索する一般的な方法で取得する
- 同様に、取得したファイルの名前を変更する
  - rna.fna → human-rna.fna
  - protein.faa → human-protein.faa

# 遺伝子配列の取得

以下、作業を飛ばして  
結構です

An official website of the United States government [Here's how you know](#) ▾



Log in

All Databases ▾

NCBIの「gene」から検索する一般的な方法で取得する

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

### Submit

Deposit data or manuscripts into NCBI databases



### Download

Transfer NCBI data to your computer



### Learn

Find help documents, attend a class or watch a tutorial



### Develop

Use NCBI APIs and code libraries to build applications



### Analyze

Identify an NCBI tool for your data analysis task



### Research

Explore NCBI research and collaborative projects



## COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [FAQs](#)

## Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

## NCBI News & Blog

[Comparing Yeast Species Used in Beer Brewing and Bread Making](#)  
29 Sep 2023

[Using the NIH Comparative Genomics Resource \(CGR\) to gain knowledge](#)  
27 Sep 2023

[Join NCBI at ASHG 2023](#)  
November 1-5 in Washington, D.C. We look forward to seeing you in person at the American Society for Human

[BLAST ClusteredNR Database is Now Available for Download!](#)  
26 Sep 2023

[Now available! You can download the ClusteredNR protein database](#)

「Gene」をクリックする

# 遺伝子配列の取得

An official website of the United States government [Here's how you know](#)

**NIH** National Library of Medicine  
National Center for Biotechnology Information

Gene  Advanced   Help

「ALDH2」と入力し、  
「Search」をクリック

**Using Gene**

- [Gene Quick Start](#)
- [FAQ](#)
- [Download/FTP](#)
- [RefSeq Mailing List](#)
- [Gene News](#)
- [Factsheet](#)

**Gene Tools**

- [Submit GeneRIFs](#)
- [Submit Correction](#)
- [Statistics](#)
- [BLAST](#)
- [Genome Workbench](#)
- [Splign](#)

**Other Resources**

- [OMIM](#)
- [RefSeq](#)
- [RefSeqGene](#)
- [Protein Clusters](#)

**Representative queries**

Find genes by...	Search text
free text	<a href="#">human muscular dystrophy</a>
chromosome and symbol	<a href="#">(1[chr] OR 2[chr]) AND adh*[sym]</a>
partial name and multiple species	<a href="#">alive[prop] AND transporter[title] AND ("Drosophila melanogaster"[organ] OR "Mus musculus"[organ])</a>
associated sequence accession	<a href="#">M11313[accn]</a>
gene name (symbol)	<a href="#">BRCA1[sym]</a>
publication (PubMed ID)	<a href="#">11331580[PMID]</a>

# 遺伝子配列の取得

An official website of the United States government [Here's how you know](#).



Log in

Gene Gene ALDH2  
Create RSS Save search Advanced

Help

Gene sources

Tabular ▾ 20 per page ▾ Sort by Relevance ▾

Send to: ▾

Filters: [Manage Filters](#) Hide sidebar >>

Genomic

Categories

Alternatively spliced

Annotated genes

Non-coding

Protein-coding

Sequence content

CCDS

Ensembl

RefSeq

RefSeqGene

Status

clear

Current

[Clear all](#)

[Show additional filters](#)

GENE  
[ALDH2 – aldehyde dehydrogenase 2 family member](#)

*Homo sapiens (human)*

Also known as: ALDH-E2, ALDHI, ALDM

Gene ID: 217

[RefSeq transcripts \(2\)](#) [RefSeq proteins \(2\)](#) [RefSeqGene \(1\)](#) [PubMed \(794\)](#)

[Orthologs](#) [Genome Data Viewer](#) [BLAST](#) [Download](#)

Was this helpful?

まず、提示された候補  
をクリック(ヒト)

RefSeq Sequences

Search results

Items: 1 to 20 of 734

<< First < Prev Page 1 of 37 Next > Last >>

[See also 23 discontinued or replaced items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">ALDH2</a> ID: 217	aldehyde dehydrogenase 2 family member [ <i>Homo sapiens (human)</i> ]	Chromosome 12, NC_000012.12 (111766933..11817532)	ALDH-E2, ALDHI, ALDM	100650
<input type="checkbox"/> <a href="#">Aldh2</a> ID: 11689	aldehyde dehydrogenase 2, mitochondrial [ <i>Mus musculus (house mouse)</i> ]	Chromosome 5, NC_000071.7 (121704090..121731887, complement)	AHD-M1, ALDH-E2, ALDHI, Ahd-5, Ahd5	
<input type="checkbox"/> <a href="#">Aldh2</a> ID: 29539	aldehyde dehydrogenase 2 family member [ <i>Rattus norvegicus (Norway rat)</i> ]	Chromosome 12, NC_051347.1 (34949549..34982527)		
<input type="checkbox"/> <a href="#">ALDH2</a> ID: 508629	aldehyde dehydrogenase 2 family member [ <i>Bos taurus (cattle)</i> ]	Chromosome 17, NC_037344.1 (62300434..62326158, complement)		
<input type="checkbox"/> <a href="#">ALDH2</a> ID: 100171598	aldehyde dehydrogenase 2 family member [ <i>Pongo abelii (Sumatran orangutan)</i> ]	Chromosome 12, NC_071997.1 (110510360..110553361)	CR201_G0001094	
<input type="checkbox"/> <a href="#">ALDH2</a>	aldehyde dehydrogenase 2 family	Chromosome 26, NC_051830.1		

Top Organisms [Tree](#)

*Homo sapiens* (194)

*Mus musculus* (51)

*Rattus norvegicus* (10)

*Bos taurus* (4)

*Zeugodacus cucurbitae* (3)

All other taxa (472)

[More...](#)

Find related data

Database: [Select](#)

[Find items](#)

Search details

ALDH2[All Fields] AND alive[prop]

[Search](#)

See more...

Recent activity

[Turn Off](#) [Clear](#)

ALDH2 AND (alive[prop]) (734)

Gene

ALDH2 (1634)

Protein

ALDH2 aldehyde dehydrogenase 2 family member [*Homo sapiens*]

Gene

Aldh2 aldehyde dehydrogenase 2, mitochondrial [*Mus musculus*]

Gene

# 遺伝子配列の取得



Search NCBI ...

Log in



Because of a lapse in government funding, the information on this website may not be up to date, transactions submitted via the website may not be processed, and the agency may not be able to respond to inquiries until appropriations are enacted. The NIH Clinical Center (the research hospital of NIH) is open. For more details about its operating status, please visit [cc.nih.gov](http://cc.nih.gov). Updates regarding government operating status and resumption of normal operations can be found at [opm.gov](http://opm.gov).

NCBI Datasets Taxonomy Genome Gene Command-line tools Documentation

## ALDH2 - aldehyde dehydrogenase 2 family member BETA

Homo sapiens

Summary Transcripts and proteins Orthologs Gene ontology

Download

datasets

API

Gene name aldehyde dehydrogenase 2 family member

Symbol ALDH2

Gene ID 217

Organism Homo sapiens (human)

Gene type protein-coding

### Other identifiers

Ensembl ENSG00000111275

HGNC HGNC:404

OMIM 100650

UniProt P05091



# 遺伝子配列の取得

An official website of the United States government. [Here's how you know.](#)

 National Library of Medicine  
National Center for Biotechnology Information

Search NCBI ... Log in

NCBI Datasets Taxonomy Genome Gene Command-line tools Documentation

ALDH2 - aldehyde dehydrogenase 2 family member BETA

Homo sapiens

Summary Transcripts and proteins Orthologs Gene ontology

Transcripts and proteins

	Gene ID	Symbol	Transcript	Length (nt)	Protein	Length (aa)	Protein na	Action
<input type="checkbox"/>	217	ALDH2	NM_000690.4 MANE Select	9,561	NP_000681.2	517	aldehyde c	⋮
<input type="checkbox"/>	217	ALDH2	NM_001204889.2	9,420	NP_001191818.1	470	aldehyde c	⋮

「MANE Select」を選択  
アイソフォーム1が選択される

FOLLOW NCBI

# 遺伝子配列の取得

NCBI Datasets Taxonomy Genome Gene **Gene** Command-line tools Documentation

ALDH2 - aldehyde dehydrogenase 2 family member BETA

Homo sapiens

Summary **Transcripts and proteins** Orthologs Gene ontology

Transcripts and proteins

<input type="checkbox"/>	Gene ID	Symbol	Transcript	Length (nt)	Protein	Length (aa)	Protein na	Action
<input checked="" type="checkbox"/>	217	ALDH2	NM_000690.4 <small>MANE Select</small>	9,561	NP_000681.2	517	aldehyde c	⋮
<input type="checkbox"/>	217	ALDH2	NM_001204889.2	9,420	NP_001191818.1	470	aldehyde c	⋮

2 Transcripts 1 selected

Download CSV Select columns

「Download」を選択

Download CSV

See in Viewer

FOLLOW NCBI

# 遺伝子配列の取得

- ヒトのgene.fnaの内容

>NC\_000012.12:111766933-111817532 ALDH2 [organism=Homo sapiens]  
[GeneID=217] [chromosome=12]

GAGACCCTAGCTCTGCTCTGGTCCGCTCGCTGTCCGCTAGCCCGCTGCGATGTTGCGCGCTGCCGCCG  
CTTCGGGCCCGCCTGGGCCGCCCTTGTCAAGCCGCCACCCAGGCCGTGCCTGCCCAACCAG  
CAGCCCAGGGTCTTCTGCAACCAGGTGAGCCCACCCGGCCGGCTCGCGTTGTTCCGGCCCGAGTCC  
CCCGCAGGCCCTAGGAAGGCCCGCGCCGTGGCCTTAGTGTACTCATCTGGGGCTCGAGGGGTTT  
GCAGAGGCTGACCTGGAAGCACATCTGCCCTCCTCGCAAGCATTAGCCCCCTCGCCGCCTCT  
GACAGTCCCCGTCCCATTCCCCGTGACTTGGGCCCTCTCGTCTCGCCAGCCTCGGTCTGTTTC  
CAGGGCCAACCTCGGGGTTCCGTTCTCCCATGGTCCTTGCTTCGGGTCTCCGCAGGGTCCCCACCC  
.....

GGGACAGATTAACAGGAATCCATAGCCCAGGGATGCGACCCAAAATTATCAAAGTAGAGACATTATGTGT  
TTGCAATGTGCTATGATTAATGCAGTGACATAACTGGCAAGATTACAGGTCAATTGGATATTGTTACC  
TGGAGCTGGCCTTCTTAGTTGCCAAAAAAGACATAAGGATTAAAAAGACAAACCATAAA

>NC\_060936.1:111743846-111794443 ALDH2 [organism=Homo sapiens]  
[GeneID=217] [chromosome=12]

GAGACCCTAGCTCTGCTCTGGTCCGCTCGCTGTCCGCTAGCCCGCTGCGATGTTGCGCGCTGCCGCCG  
CTTCGGGCCCGCCTGGGCCGCCCTTGTCAAGCCGCCACCCAGGCCGTGCCTGCCCAACCAG  
CAGCCCAGGGTCTTCTGCAACCAGGTGAGCCCACCCGGCCGGCTCGCGTTGTTCCGGCCCGAGTCC  
CCCGCAGGCCCTAGGAAGGCCCGCGCCGTGGCCTTAGTGTACTCATCTGGGGCTCGAGGGGTTT  
GCAGAGGCTGACCTGGAAGCACATCTGCCCTCCTCGCAAGCATTAGCCCCCTCGCCGCCTCT  
GACAGTCCCCGTCCCATTCCCCGTGACTTGGGCCCTCTCGTCTCGCCAGCCTCGGTCTGTTTC  
CAGGGCCAACCTCGGGGTTCCGTTCTCCCATGGTCCTTGCTTCGGGTCTCCGCAGGGTCCCCACCC  
.....

GACAGATTAACAGGAATCCATAGCCCAGGGATGCGACCCAAAATTATCAAAGTAGAGACATTATGTGTTT  
GCAATGTGCTATGATTAATGCAGTGACATAACTGGCAAGATTACAGGTCAATTGGATATTGTTACCTG  
GAGCTGGCCTTCTTAGTTGCCAAAAAAGACATAAGGATTAAAAAGACAAACCATAAA

# 遺伝子配列の取得

- ヒトのrna.fnaの内容

```
>NM_000690.4 ALDH2 [organism=Homo sapiens] [GeneID=217] [transcript=1]
GAGACCCTAGCTCTGCTCTGGTCCGCTCGCTGTCCGCTAGCCCGCTGCGATGTTGCGCGCTGCCGCCG
CTTCGGGCCCGCCTGGGCCGCCCTTGTCAAGCCGCCACCCAGGCCGTGCCTGCCCAACCAT
CAGCCCGAGGTCTTCTGCAACCAGATTTCATAAAACAATGAATGGCACGATGCCGTAGCAGGAAAACAT
TCCCCACCGTCAATCCGTCCACTGGAGAGGTCATCTGTCAGGTAGCTGAAGGGGACAAGGAAGATGTGGA
CAAGGCAGTGAAGGCCGCCGGCCCTCCAGCTGGCTCACCTGGCGCCGATGGACGCATCACAC
AGGGGCCGGCTGCTGAACCGCCTGGCGATCTGATCGAGCGGGACCACCTACCTGGCGGCCTGGAGA
CCCTGGACAATGGCAAGCCCTATGTCATCTCCTACCTGGTGGATTGGACATGGTCCTCAAATGTCTCCG···
...
CCATAGCCCAGGGATGCGACCCAAAATTATCAAAGTAGAGACATTATGTGTTGCAATGTGCTATGATTA
ATGCAGTGACATAACTGGCAAGATTACAGGTCAATTGGATATTGTTACCTGGAGCTGGTCCTTCTTAG
TTGCCAAAAAAGACATAAGGATTAAGGACAAACCATAAA
>NM_001204889.2 ALDH2 [organism=Homo sapiens] [GeneID=217]
```

[transcript=2]

```
GAGACCCTAGCTCTGCTCTGGTCCGCTCGCTGTCCGCTAGCCCGCTGCGATGTTGCGCGCTGCCGCCG
CTTCGGGCCCGCCTGGGCCGCCCTTGTCAAGCCGCCACCCAGGCCGTGCCTGCCCAACCAT
CAGCCCGAGGTCTTCTGCAACCAGATTTCATAAAACAATGAATGGCACGATGCCGTAGCAGGAAAACAT
TCCCCACCGTCAATCCGTCCACTGGAGAGGTCATCTGTCAGGTAGCTGAAGGGGACAAGGCCTGGAGAC
CCTGGACAATGGCAAGCCCTATGTCATCTCCTACCTGGTGGATTGGACATGGTCCTCAAATGTCTCCGG
TATTATGCCGGCTGGCTGATAAGTACCAACGGGAAACCATCCCCATTGACGGAGACTTCTCAGCTACAC
CACGCCATGAACCTGTGGGGGTGTGCGGGCAGATCATTCCGTGGAATTCCGCTCCTGATGCAAGCATG
.....
CATAGCCCAGGGATGCGACCCAAAATTATCAAAGTAGAGACATTATGTGTTGCAATGTGCTATGATTA
TGCAGTGACATAACTGGCAAGATTACAGGTCAATTGGATATTGTTACCTGGAGCTGGTCCTTCTTAGT
TGCCAAAAAAGACATAAGGATTAAGGACAAACCATAAA
```

アイソフォームを指定しないと、2つの  
アイソフォームの配列が取得される

# 遺伝子配列の取得

- ヒトのprotein.faaの内容

```
>NP_000681.2 ALDH2 [organism=Homo sapiens] [GeneID=217] [isoform=1  
precursor]
```

```
MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEHDASRKTPTVNPSTGEVICQVAE  
GDKEDVDKAVKAARAAFQLGSPWRRMDASHRGRLLNRLADLIERDRTYLAALETLDNGKPYVISYLVLDL  
MVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCGQIIPWNFPLLMQAWKLGPALATGNVVVMKV  
AEQTPLTALYVANLIKEAGFPPGVNIVPGFGPTAGAAIASHEDVDKVAFTGSTEIGRVIQVAAGSSNLK  
RVTELEGGKSPNIIMSDADMDWAVEQAHFALFFNQGQCCCAGSRTFVQEDIYDEFVERSVARAKSRVVG  
PFDSKTEQGPQVDETQFKKILGYINTGKQEGAKLLCGGGIAADRGYFIQPTVFGDVQDGMTIAKEEIFGP  
VMQILKFKTIEEVGRANNSTYGLAAAFTKDLDKANYLSQALQAGTVWVNCYDVFQGAQSPFGGYKMSGS  
GRELGEGYGLQAYTEVKTVTVKVPQKNS
```

```
>NP_001191818.1 ALDH2 [organism=Homo sapiens] [GeneID=217] [isoform=2  
precursor]
```

アイソフォームを指定しないと、2つの  
アイソフォームの配列が取得される

```
MLRAAARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEHDASRKTPTVNPSTGEVICQVAE  
GDKALETLNGKPYVISYLVLDLMDVLKCLRYYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCGQIIPWNF  
PLLMQAWKLGPALATGNVVVMKVAEQTPLTALYVANLIKEAGFPPGVNIVPGFGPTAGAAIASHEDVDK  
VAFTGSTEIGRVIQVAAGSSNLKRVTELEGGKSPNIIMSDADMDWAVEQAHFALFFNQGQCCCAGSRTFV  
QEDIYDEFVERSVARAKSRVVGNPFDSTEQGPQVDETQFKKILGYINTGKQEGAKLLCGGGIAADRGYF  
IQPTVFGDVQDGMTIAKEEIFGPVMQILKFKTIEEVGRANNSTYGLAAAFTKDLDKANYLSQALQAGT  
VWVNCYDVFQGAQSPFGGYKMSGSRELGEYGLQAYTEVKTVTVKVPQKNS
```

# EMBL-EBI

- European Bioinformatics Institute (欧洲バイオインフォマティクス研究所)
- 欧州分子生物学研究所（EMBL）の下部組織、生命科学の研究所で、バイオインフォマティクスのデータおよびツールのリソースを提供
- <https://www.ebi.ac.uk/>

The screenshot shows the EMBL-EBI homepage with a blue header featuring the EMBL-EBI logo and a search bar. Below the header is a large banner with a microscopic image of cells. The main content area includes a search bar, a navigation menu with links to home, services, research, training, and about us, and four main buttons: 'Find data resources', 'Submit data', 'Explore our research', and 'Train with us'. A red box highlights the 'Find data resources' button. To the right of the buttons, there is Japanese text: '遺伝子、タンパク質などの情報を総合的に検索'. Below the buttons are four news cards: 'Global participation in biodata: Building equitable partnerships' (8 Oct 2025), 'EMBL-EBI and Google DeepMind renew partnership and release update to AlphaFold Database' (7 Oct 2025), 'We are EMBL: Mila Coniff on teamwork and tech' (2 Oct 2025), and 'How to rapidly search the world's microbial DNA' (30 Sep 2025).

EMBL's European Bioinformatics Institute

**EMBL-EBI**

Unleashing the potential of big data in biology

Find a gene, protein or chemical

All

Search

Example searches: blast keratin bfl1 | About EBI Search

Find data resources →

Submit data →

Explore our research →

Train with us →

遺伝子、タンパク質などの  
情報を総合的に検索

Latest news →

Global participation in biodata: Building equitable partnerships  
08 Oct 2025

EMBL-EBI and Google DeepMind renew partnership and release update to AlphaFold Database  
07 Oct 2025

We are EMBL: Mila Coniff on teamwork and tech  
02 Oct 2025

How to rapidly search the world's microbial DNA  
30 Sep 2025

## Services

Data resources and analysis tools to support life science research

EMBL's European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Find a data resource or a tool

Search

Explore all our data resources and tools →

Our full range of data resources and data analysis tools are essential for supporting life science research.

### Featured data resources



#### GWAS Catalog

The NHGRI-EBI GWAS Catalog is a quality-controlled, manually curated, literature-derived collection of all published genome-wide association studies.

全ゲノム関連解析のカタログ  
Web API | EBI Terms of use  
GWAS (全ゲノム関連解析)：ヒトゲノム全体の遺伝的変異を解析し、特定の疾患や形質との関連を調べる方法

Featured tools



BiNChE



UniProt ID Mapping



Zooma

生命科学データに一貫したアクセス方法を提供  
Web API | CC-BY  
URLs



#### Identifiers.org

Identifiers.org provides a manually maintained Registry of biomolecular data collections and a service to map URLs to URLs.

#### International Genome Sample Resource

A deep catalog of shared human genetic variation in population groups worldwide that follows the International Genome Sample

ヒトの遺伝的変異の詳細なカタログ  
Web API | CC-BY  
URLs

Terms of use

この下を見ると…

Data submission →

Sequence analysis →

Support and feedback →

Service teams →

EMBL-EBI's Data Stewardship Principles

## 配列解析

### Open

EMBL's Open Science policy makes databases, code and software freely available to scientists whenever possible. Our Terms of Use impose no additional restrictions on data use, beyond those imposed by the data owner. Our Licensing policy aims to implement machine-readable cc0 licenses for data resources wherever possible. Our teams work open-source and many of our software systems can be downloaded and installed locally.

### Long-Term Resilience

Delivery of our mission, to help scientists realise the potential of big data in biology, depends on the long-term resilience of our data resources. This is ensured through resource life cycle management, staffing and infrastructure continuity measures, data backup and delivery through international partners. Our approach to data preservation and service continuity.

### Expert Added Value

Our communities of expert biologists enhance the data we steward via curation and annotation. This includes integration with data from other sources such as the scientific literature. Increasingly, curators use emerging technologies such as text mining, machine learning, large-language models and other AI tools to automate and scale, subject to rigorous evaluation and quality control. Our work integrating the latest advances in AI

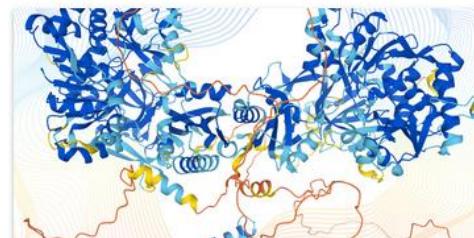
## Covid-19のデータ ポータル Highlight projects



COVID-19 Data Portal →

The European COVID-19 Data Platform facilitates data sharing and analysis in order to accelerate coronavirus research.

## タンパク質モデル構造 データベース



AlphaFold DB →

Enjoy open access to over 990,000 protein structure predictions for the human proteome and other key proteins.

## ヒトの細胞の遺伝学に 基づく参照マップ



Human Cell Atlas →

Aiming to map every cell type in the human body, the Human Cell Atlas will transform our understanding of biology and disease.

# 配列の比較

Explore Sequence Analysis Tools with  
**Job Dispatcher**  
EMBL's European Bioinformatics Institute

Job Dispatcher Help & Privacy Your Jobs Feedback

Welcome to the new **Job Dispatcher** website. We'd love to hear your [feedback](#) about the new webpages! X

The Job Dispatcher at EMBL-EBI offers free access to a range of bioinformatics tools and biological datasets through its web and programmatic interfaces. It also powers various popular sequence analysis services hosted at the EMBL-EBI, including InterProScan, UniProt, and Ensembl Genomes.

**Retrieve Your Job**

Retrieve the job results  **Search**

Example searches: fasta-R20230421-121005-0548-43453433-p1m

---

**Your Recent Jobs**

Job ID	Status	Last update	Delete
There is no saved job			

---

**Tool Categories**

 **Pairwise Sequence Alignment**  
Identify regions of similarity between two biological sequences

 **Multiple Sequence Alignment**  
Identify conserved regions between multiple selected

この下を見ると…

# 配列の比較

## Tool Categories



### Pairwise Sequence Alignment

Identify regions of similarity between two biological sequences.

[Needle](#) | [Stretcher](#) | [GGSEARCH2SEQ](#) | [Water](#) | [Matcher](#) | More...

「Pairwise Sequence Alignment」を選ぶ



### Multiple Sequence Alignment

Identify conserved sequence patterns from multiple related sequences.

[Clustal Omega](#) | [Cons](#) | [Kalign](#) | [MAFFT](#) | [MUSCLE](#) | [T-Coffee](#) | More...



### Sequence Similarity Search

Find sequences in databases based on similarity.

[NCBI BLAST](#) | [PSI-BLAST](#) | [FASTA](#) | [SSSEARCH](#) | [PSI-Search](#) | More...



### Sequence Translation

Emboss sequence translation and back translation tools.

[Transeq](#) | [Sixpack](#) | [Backtranseq](#) | [Backtranambig](#)



### Sequence Statistics

Set of tools for performing sequence statistics.

[Pepinfo](#) | [Pepstats](#) | [Pepwindow](#) | [SAPS](#) | [Cpgplot](#) | More...



### Phylogeny

Phylogenetic tree generation using the ClustalW2 program.

[Simple Phylogeny](#)



### Protein Functional Analysis

Assign biological or biochemical roles to proteins.

[InterProScan](#) | [PfamScan](#) | [HMMER3 hmmscan](#) | More...



### Sequence Operations

Generate checksums for protein and nucleotide sequences.

[SeqChecksum](#)



### RNA Analysis

Set of tools for performing various RNA analysis.

[Infernal cmscan](#) | [R2DT](#)



### Sequence Format Conversion

Read different biological sequence formats and convert them to other formats.

[Seqret](#) | [MView](#)



### Dbfetch

Single point of access to retrieve entries from various EMBL-EBI databases.



### EMBOSS Tools

Set of selected EMBOSS tools for sequence analysis.

[Cons](#) | [Needle](#) | [Stretcher](#) | More...

# 配列の比較

手順としては3つ先の  
スライドに飛びます

EMBL-EBI home Services Research Training About us EMBL-EBI

BETA This beta version is fully functional but is still in active development. If you use these webpages, we'd love to hear your [feedback!](#)

Explore Sequence Analysis Tools with

## Job Dispatcher

EMBL's European Bioinformatics Institute

Job Dispatcher Help & Privacy Pairwise Sequence Alignment (PSA)

Feedback

### Pairwise Sequence Alignment

Pairwise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid). By contrast, Multiple Sequence Alignment (MSA) is the alignment of three or more biological sequences of similar length. From the output of MSA applications, homology can be inferred and the evolutionary relationship between the sequences studied.

### Global Alignment

Global alignment tools create an end-to-end alignment of the sequences to be aligned.

#### EMBOSS Needle

EMBOSS Needle creates an optimal global alignment of two sequences using the Needleman-Wunsch algorithm.

[Launch EMBOSS Needle](#)

#### EMBOSS Stretcher

EMBOSS Stretcher uses a modification of the Needleman-Wunsch algorithm that allows larger sequences to be globally aligned.

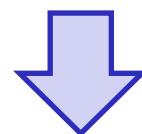
[Launch EMBOSS Stretcher](#)

# 配列を比較する方法

遺伝子やタンパク質の配列を並べて比較する  
アラインメントという方法が使われる

**ACGAAGCTCTA**

**ACCAGAGTCA**



ギャップ

**ACGA-AGCTCTA**

|| | || | | |

**ACCAGAG-TC-A**

配列を、類似した部分が並ぶよう、配列要素（文字）を対応づける操作

長さをそろえて、  
文字と文字の「最適な」  
対応関係を調べる  
必要に応じてギャップ文字「-」  
を入れる

- 対応する文字がないということ
- 進化の過程で文字が挿入されたり失われたりすることがあるため

# グローバルアラインメントとローカルアラインメント

- **グローバルアラインメント**: 配列全体にわたり類似性を考慮してアラインメント
  - 全体的に類似した配列の異なる部位を調べる
    - 同じタンパク質の生物種間の違いを調べる
    - 進化の解析を行う
  - 最適なペアワイスアラインメントのアルゴリズムとして **Needleman-Wunsch**のアルゴリズムがある
- **ローカルアラインメント**: 局所的な類似部分をアラインメント
  - 機能に関わる配列パターンを考慮したアラインメント
  - 長さが大きく異なる配列、配列類似性が低い配列の比較に用いられる
  - 最適なペアワイスアラインメントのアルゴリズムとして **Smith-Waterman**のアルゴリズムがある

# アラインメントの例

## グローバルアラインメント(ヒトのヘモグロビンとLupinレグヘモグロビン)

HBA_HUMAN	1	-MVLSPADKTNVKAAGKVGGAHAGEYGAELERMFLSF---PTTKTYFP	45
		.. :.....  :: .:.. :...: .. . :..  .. .. .	
LGB2_LUPLU	1	MGALTESQAALVKSSWEEFNANIPKH---THRFFILVLEIAPAAKDLFS	46
		.. :.....  :: .:.. :...: .. . :..  .. .. .	
HBA_HUMAN	46	HF----DLSHGSAQVKGHGKKVADALTNA----VAHVDDMPNALSALSD	86
		.. :.....  :: .:.. :...: .. . :..  .. .. .. ..	
LGB2_LUPLU	47	FLKGTSEVPQNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGS	96
		.. :.....  :: .:.. :...: .. . :..  .. .. .. ..	
HBA_HUMAN	87	LHAHKLRVPVNFKLLSHCLLVTLAAHLPAEFTPASLDKFLASVSTV	136
		: ...  . .... .:.. .:.. . .... :.. :.. :.. :..	
LGB2_LUPLU	97	VHVSX-GVADAHFPVVKEAILKTIKEVVGAKWSEELNSAWTIAYDELAIV	145
		: ...  . .... .:.. .:.. . .... :.. :.. :.. :..	
HBA_HUMAN	137	LTSKYR---	142
		...:..	
LGB2_LUPLU	146	IKKEMNDAA	154
			# Matrix: EBLOSUM62 # Gap_penalty: 10.0 # Extend_penalty: 0.5 # Length: 159 # Identity: 26/159 (16.4%) # Similarity: 60/159 (37.7%) # Gaps: 22/159 (13.8%) # Score: 40.5

## ローカルアラインメント(ヒトのヘモグロビンとLupinレグヘモグロビン)

HBA_HUMAN	3	LSPADKTNVKAAGKVGGAHAGEYGAELERMFLSF---PTTKTYFPFH-	47
		:.....  :: .:.. :...: .. . :..  .. .. ..	
LGB2_LUPLU	4	LTESQAALVKSSWEEFNANIPKH---THRFFILVLEIAPAAKDLFSFLK	49
		.. :.....  :: .:.. :...: .. . :..  .. .. .. .	
HBA_HUMAN	48	--DLSHGSAQVKGHGKKVADALTNA----VAHVDDMPNALSALSDLHA	89
		.. :.....  :: .:.. :...: .. . :..  .. .. .. .. .	
LGB2_LUPLU	50	GTSEVPQNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHV	99
		.. :.....  :: .:.. :...: .. . :..  .. .. .. .. .	
HBA_HUMAN	90	HKLRVPVNFKLLSHCLLVTLAAHLPAEFTPAS	125
		.  . .... .:.. .:.. . .... :.. :.. :..	
LGB2_LUPLU	100	SK-GVADAHFPVVKEAILKTIKEVVGAKWSEELNSA	134
		# Matrix: EBLOSUM62 # Gap_penalty: 10.0 # Extend_penalty: 0.5 # Length: 136 # Identity: 25/136 (18.4%) # Similarity: 55/136 (40.4%) # Gaps: 18/136 (13.2%) # Score: 48.5	

EMBOSS(The European Molecular Biology Open Software Suite)  
Pairwise Sequence Alignment

# 配列の比較

BETA This beta version is fully functional but is still in active development. If you use these webpages, we'd love to hear your [feedback!](#)

EMBOSS Needle  
Pairwise Sequence Alignment (PSA)

Job Dispatcher Help & Privacy Input form Feedback

## ヒトとマウスのmRNAの配列の比較

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

**Input Sequence**

PROTEIN  DNA

Paste in your sequence or use the example

参考... ファイルが選択されていません。

aldh2-human-rna.fna

Paste in your sequence or use the example

参考... ファイルが選択されていません。

aldh2-mouse-rna.fna

Use the example Clear Sequence Find more example inputs

**Parameters**

OUTPUT FORMAT①

pair

More Options

# 配列の比較

Input sequence ⓘ

Sequence type

Protein  DNA

Paste your first sequence here - or use the example sequence

```
>NM_001204889.2 ALDH2 [organism=Homo sapiens] [GeneID=217] [transcript=2]
GAGACCTAGCTCTGCTCTGGTCCGCTCGCTGTCCGCTAGCCGCTGCGATGTTGCGCG
CTGCCGCCGCTTCGGCCCCGCTGGCGCCCTTGTCAGCCGCCACCCAGG
CCGTGCCTGCCCAACCAGCAGCCGAGGTCTTGTCAACCAGATTTCATAAACATG
AATGGCACGATGCCGTAGCAGGAAAACATTCCCCACCGTCAATCGTCCACTGGAGAGG
TCATCTGTCAGGTAGCTGAAGGGACAAGGCCTTGGAGACCCCTGGACAATGGCAAGCCCT
ATGTCATCTCCTACCTGGGATTTGGACATGGCCTCAAATGTCCTGGTATTATGCCG
```

ファイルが選択...ていません。

Paste your second sequence here - or use the example sequence

```
>NM_009656.4 Aldh2 [organism=Mus musculus] [GeneID=11669] [transcript=1]
ATTCTCTCGCCGCCATATCTGCACAGATGTGAGCCTAGGCAGCCACCCCTGCTAGG
AGCGCACACCACTCTGGCTAGGCTTCAGGGTCTGCAAACCTCCATCTGACTTGGC
TTGGGAGCCAGGGGTGCGCCCCCTTAGGCCGTGAGGGGCTGGACTCCCTGACCACGCC
CCCGTGTCTCCGCCCTCCATTGGCGCTGCAGGGGGCGGAGGACTTGTCTCAA
CGCTGCAGTCGCCCTCGATCGCAAGGCTTCTCGGCTCCGCTGGCTCGGCC
ATTCAGTTCAAGTCGGGTCAAGTTAAGCTCGCTCAGTTCAAGCATGCTCGCGCCACT
```

ファイルが選択...ていません。

More example inputs

Parameters

OUTPUT FORMAT ⓘ

pair

[More options](#) ▾

Submit

Title

EMBOSS Needle's job

# 配列の比較

Results for job: emboss\_needle-I20231009-150124-0957-45762922-p1m

# 配列の比較

BETA This beta version is fully functional but is still in active development. If you use these webpages, we'd love to hear your [feedback!](#)

## EMBOSS Needle

Pairwise Sequence Alignment (PSA)

Job Dispatcher Help & Privacy [Input form](#)

[Feedback](#)

### ヒトとマウスのタンパク質の配列の比較

Input Sequence

PROTEIN  DNA

Paste in your sequence or use the example

参照... ファイルが選択されていません。

aldh2-human-protein.faa

Paste in your sequence or use the example

参照... ファイルが選択されていません。

aldh2-mouse-protein.faa

[Use the example](#)

[Clear Sequence](#)

[Find more example inputs](#)

Parameters

OUTPUT FORMAT①

pair

[More Options](#)

# 配列の比較

Results for job: emboss\_needle-I20231009-150412-0014-63955803-p1m

Tool Output	Result Files	Submission Details
<pre>##### # Program: needle # Rundate: Mon 9 Oct 2023 15:04:17 # Commandline: needle #   -auto #   -stdout #   -asequence emboss_needle-I20231009-150412-0014-63955803-p1m.asequence #   -bsequence emboss_needle-I20231009-150412-0014-63955803-p1m.bsequence #   -datafile EBLOSUM62 #   -gapopen 10.0 #   -gapextend 0.5 #   -endopen 10.0 #   -endextend 0.5 #   -format pair #   -sprotein1 #   -sprotein2 # Align_format: pair # Report_file: stdout #####  ##### # Aligned_sequences: 2 # 1: NP_000681.2 # 2: NP_033786.1 # Matrix: EBLOSUM62 # Gap_penalty: 10.0 # Extend_penalty: 0.5 # # Length: 520 # Identity: 493/520 (94.8%) # Similarity: 504/520 (96.3%) # Gaps: 4/520 ( 0.8%) # Score: 2564.0 # #####  NP_000681.2    1 MLRAA---ARFGPRLGRRLLSAAATQAVPAPNQQPEVFCNQIFINNEWHD 47 NP_033786.1    1 MLRAALTTVRGRGPRLS-RLLSAAAATSAVPAPNHQPEVFCNQIFINNEWHD 49  NP_000681.2    48 AVSRKTFPTVNPSTGEVICQ/AEGDKEVDVKAVKAARAFAQLGSPIRRIID 97 NP_033786.1    50 AVSRKTFPTVNPSTGEVICQ/AEGDKEVDVKAVKAARAFAQLGSPIRRIID 99  NP_000681.2    98 ASHGRGRLNRLADLIERDRTYLAALETLDNGKPVVVISYLVLDMDVVKCLR 147 NP_033786.1    100 ASDRGRGLLYRLADLIERDRTYLAALETLDNGKPVVVISYLVLDMDVVKCLR 149  NP_000681.2    148 YYAGIADKYHGKTIPIDGQFFSYTRHEPVGVCGIIIPUNFPPLMQAKWLG 197 NP_033786.1    150 YYAGIADKYHGKTIPIDGQFFSYTRHEPVGVCGIIIPUNFPPLMQAKWLG 199  NP_000681.2    198 PALATGIVVVVKVAEQTPLTALYVANLIKEAGFPQPGVNZNVPGFQPTAGA 247 NP_033786.1    200 PALATGIVVVVKVAEQTPLTALYVANLIKEAGFPQPGVNZNVPGFQPTAGA 249  NP_000681.2    248 AIASHEDDKVAFGSTEIGRVIQVAAGSSNLKRVTLLELGKGSPINIIMSD 297 NP_033786.1    250 AIASHEGDVKAFGSTEVGHLLQVAAGSSNLKRVTLLELGKGSPINIIMSD 299  NP_000681.2    298 ADMDWAVEQAHFALFFNQGQCCCCAGSRTFVQEDIYDEFVERSVARAKSRV 347 NP_033786.1    300 ADMDWAVEQAHFALFFNQGQCCCCAGSRTFVQENVYDEFVERSVARAKSRV 349</pre>		

# 課題 1

---

ヒトとクジラ、ヒトとマグロのミオグロビンの配列について、アラインメントを比較してみよう。

ヒト myoglobin-human.fasta

クジラ myoglobin-whale.fasta

マグロ myoglobin-tuna.fasta

余裕があれば、ヒトのアザラシのミオグロビンについても比較してみよう

アザラシ myoglobin-seal.fasta

# 課題 1の解説

## ヒトとクジラ

```
#####
# Program: needle
# Rundate: Fri 10 Oct 2025 09:16:10
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20251010-091604-0708-43449506-p1m.asequence
#   -bsequence emboss_needle-I20251010-091604-0708-43449506-p1m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: MYG_HUMAN
# 2: A0A8C0D084_BALMU
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 154
# Identity: 132/154 (85.7%)
# Similarity: 143/154 (92.9%)
# Gaps: 0/154 ( 0.0%)
# Score: 710.0
#
#
=====
```

# 課題 1の解説

## ヒトとクジラのミオグロビン

MYG_HUMAN	1 MGLSDGEWQLVNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFHKL	50
AOA8C0D084_BA	1 MVLSDAEWQLVNLNIWAKVEADVAGHGQDILIRLFKGHPETLEKFDKFHKL	50
MYG_HUMAN	51 KSEDEMKAEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI	100
AOA8C0D084_BA	51 KTEAEMKAEDLKKHGNTVLTALGGILKKKGHHEELKPLAQSHATKHKI	100
MYG_HUMAN	101 PVKYLEFISECIIQVIQS <span style="border: 1px solid red;">K</span> HPGDFGADAQGAMNKALELFRKDMASNYKEL	150
AOA8C0D084_BA	101 PIKYLEFISDAII <span style="border: 1px solid red;">I</span> H <span style="border: 1px solid red;">V</span> IHSRHPGDFGADAQAMNKALELFRKDIAAKYKEL	150
MYG_HUMAN	151 GFQG	154
AOA8C0D084_BA	151 GFQG	154

タンパク質の表面

分子表面の正味電荷が+に寄る  
ミオグロビンを超高濃度に貯蔵しても凝集しにく  
いよう、表面の同電荷反発を強める？

# 課題 1の解説

## ヒトとマグロのミオグロビン

MYG_HUMAN	1 MGLSDGEWQLVNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFHKL	50
	...:...  ..  .    .... ..  .    .    ..:  .. .. ..:	
MYG_THUTH	1 ----MADFDAVLKCWGPVEADYTTIGGLVLTRLFKEHPETQKLFPKFAGI	46
MYG_HUMAN	51 KSEDEMKAEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI	100
	...  :...:...:...     .  .:  .  .  .  .  .  .  .  .  .  .  .	
MYG_THUTH	47 AQAD-IAGNAAVSAHGATVLKKLGELLKAKGSHAAILKPLANSHATKHKI	95
MYG_HUMAN	101 PVKYLEFISECIIQVLQSKHPGDFGADAQG--AMNKALELFRKDMASNYK	148
	:....     .:::  :...  .  .   .   :....:....  :.:	
MYG_THUTH	96 PINNFKLISEVLVKVMHEK----AGLDAGGQTALRNVMGIIIADEANYK	141
MYG_HUMAN	149 ELGFQG 154	
	.	
MYG_THUTH	142 ELGFSG 147	

## 課題 1の解説

## ヒトとアザラシのミオグロビン

MYG_HUMAN	1	MGLSDGEWQLVNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHL        .       . .:       . .	50
MYG_PHOVI	1	MGLSDGEWHLVNVWGKVETDLAGHGQEVLIRLFKSHPETLEKFDKFKHL	50
MYG_HUMAN	51	KSEDEMKAEDLKKGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI     : :.    :   .             :	100
MYG_PHOVI	51	KSEDDMRRSEDLRKGNTVLTALGGILKKKGHHEAELKPLAQSHATKHKI	100
MYG_HUMAN	101	PVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKEL  :       .   .   .    .:.       .   .       .  : .	150
MYG_PHOVI	101	PIKYLEFISEAIIHVLHSKHPAEFGADAQAAAMNKALELFRNDIAAKYKEL	150
MYG_HUMAN	151	GFQG 154   .	
MYG_PHOVI	151	GFHG 154	