

+

ウェブで実践する 配列比較解析

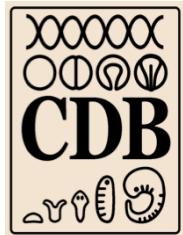
～相同性検索から
分子系統解析まで

統合データ
ベース講習会

AJACS
蝦夷 3

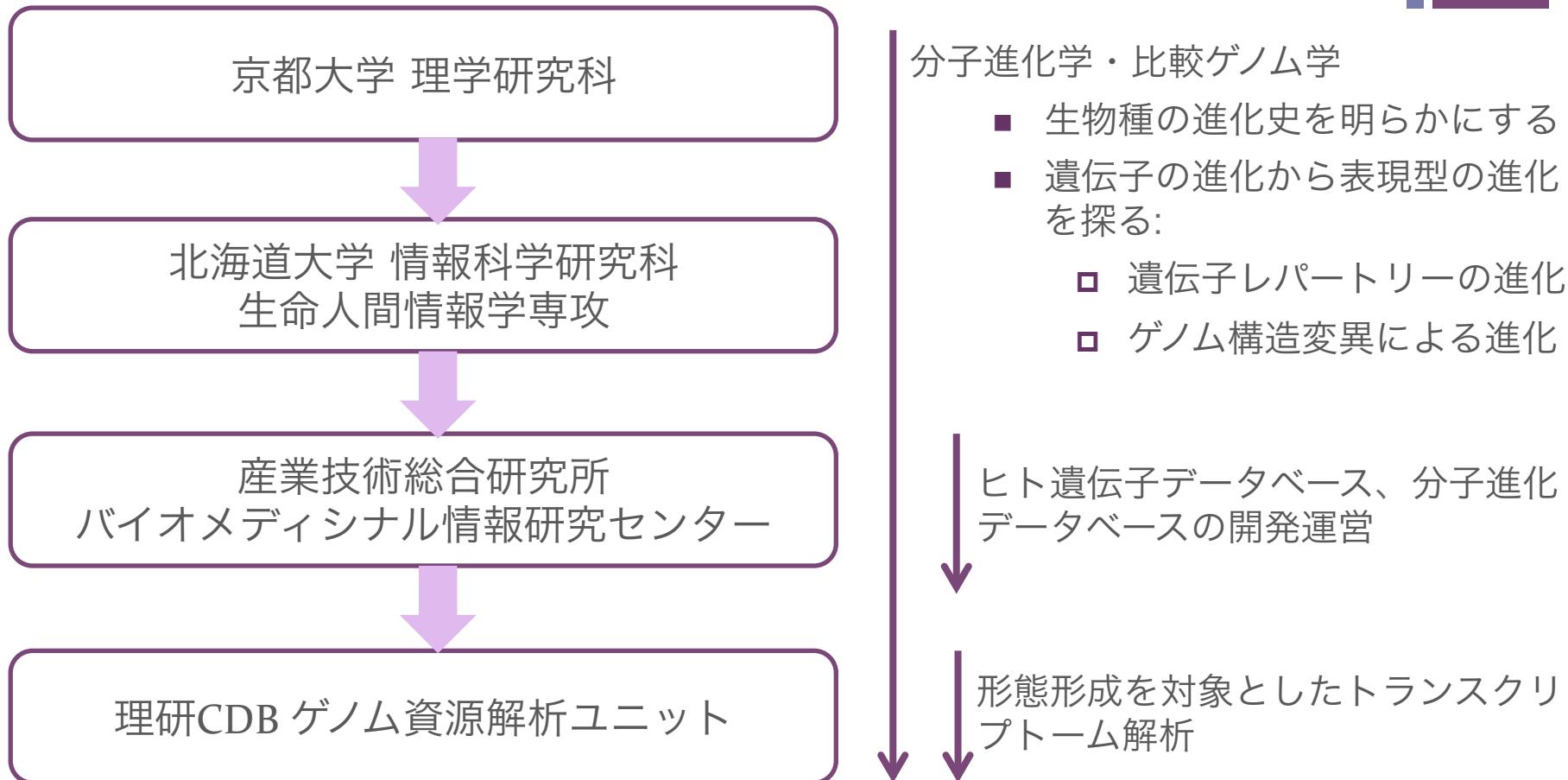
2013/11/06

北海道大学

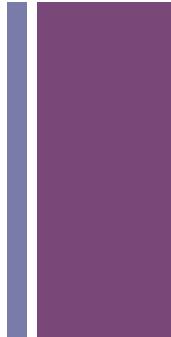


理研CDB 原雄一郎

+ 自己紹介



+ 今日お話しすること



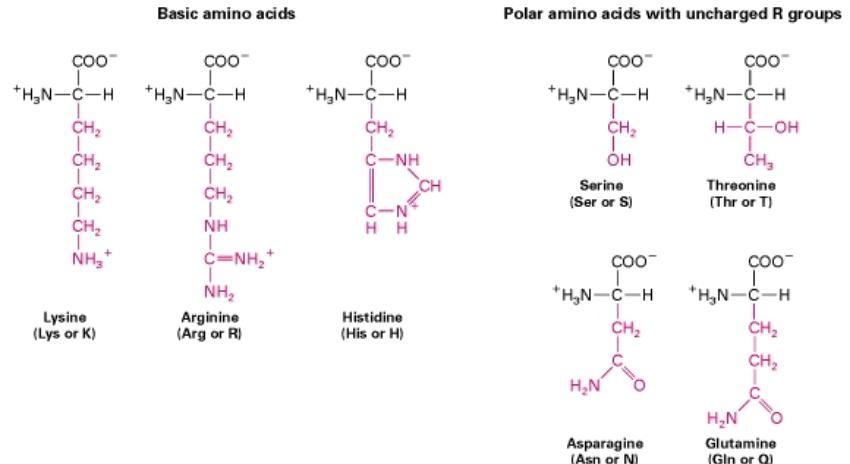
- ゲノムを構成する塩基配列、ゲノムから産出されるタンパク質の一次構造であるアミノ酸配列をコンピュータ上で比較する
 - ある塩基配列/アミノ酸配列に一致する/類似する既知の配列を探す
相同性検索
 - 似ている複数の配列を比較する
マルチプルアラインメント
 - 塩基配列やタンパクが経てきた進化の道筋を推定する
分子系統解析
- 説明+解析と実習を行います。



生物の情報を配列で表す

- DNA: 4種類の塩基をそれぞれ1文字のアルファベットで表す。
アデニン(Adenine)、チミン(Thymine)、グアニン(Guanine)、シトシン(Cytosine)
- RNA: チミンの代わりにウラシル(Uracil)を用いる
- タンパク: 一次構造のアミノ酸配列。各アミノ酸を1文字のアルファベットで表す

HYDROPHILIC AMINO ACIDS



+

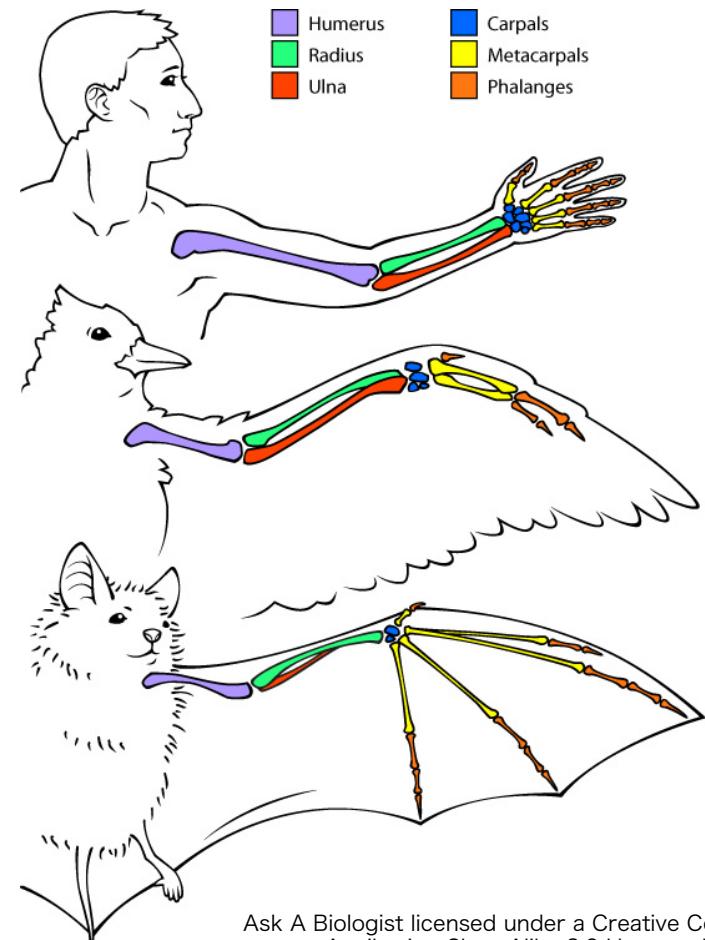
相同性検索とは

- ある(未知の)DNA/タンパク配列を既存の配列データベースに照らし合わせて配列や構造の類似性をもつ既存配列を探索し抽出すること
- 行っていることは、相同性検索(homology search)というよりは「類似性検索(similarity search)」が正しい



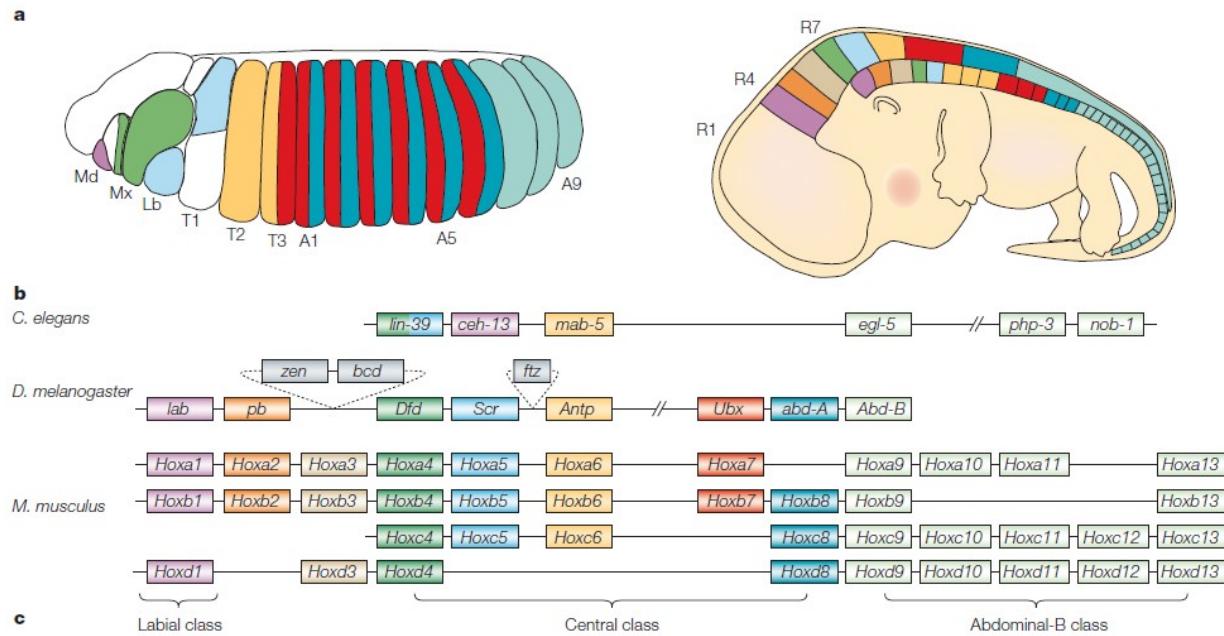
相同性(homology)とは

- 進化的起源が同一であるもの。
類似性を伴う
- トリやコウモリの翼は前足と相同
- 翼の構成物はトリとコウモリで独立に獲得された（羽毛と皮膜→アナロジー）



配列の相同性

- 類似した配列は同一の祖先から進化してきたという前提
相同な遺伝子は相同的な機能を持つことが多い



+

相同性と類似性

- 類似は定量的表現 高い \Leftrightarrow 低い
- 相同は定性的表現 ある \Leftrightarrow 無い
- 「相同性が高い」という表現は間違い
- 2者の類似度がある閾値より高いとき「相同」である

+

相同性検索で何ができるか

- 手元にある配列(問い合わせ配列、クエリ)の素性を調べられる
問い合わせ配列の生物種において、ゲノムや遺伝子配列が網羅的に
解読されていない→1 解読されている→2
 1. 既知の配列に類似: 既知配列の構造や機能が調べられていれば、
類似性とともに未知配列の構造や機能を予測できる
 2. 既知の配列に一致: その配列自体を特定できる
- 配列上の類似領域を特定できる
類似した領域: 保存的→共通した機能をもっている可能性がある
- 類似配列を収集する
似ている配列を集めて比較解析を行う

+ 相同性検索プログラム

■ Ssearch

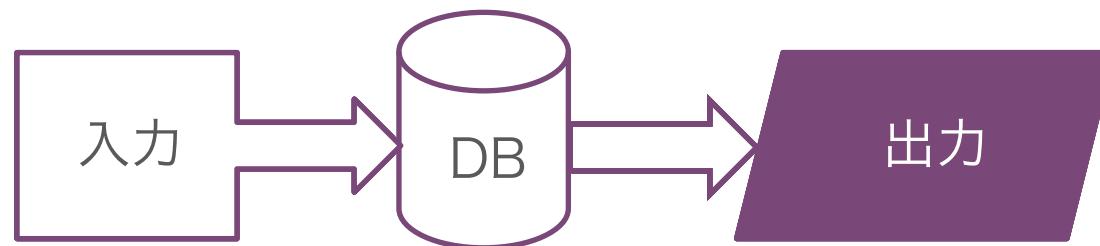
■ FASTA

■ BLAST

正確

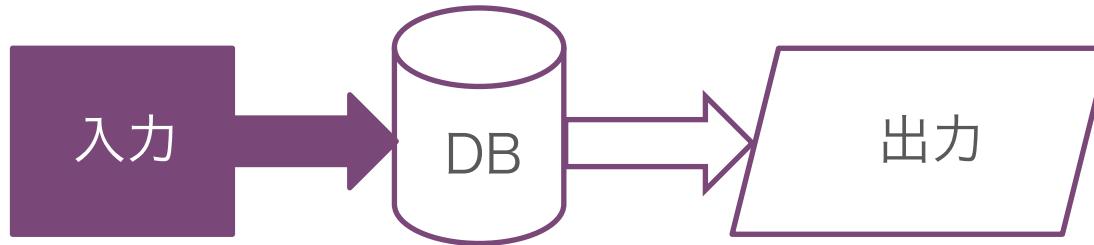
高速

Basic Local Alignment Search Tool
現在最も使われている相同性検索
プログラム





問い合わせ配列



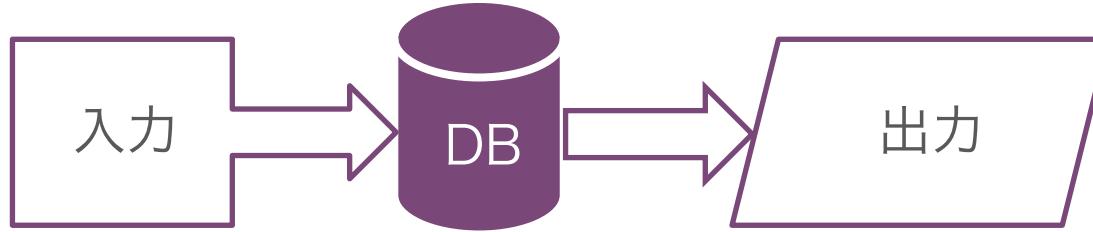
- 塩基・アミノ酸一文字表記のテキストデータ or 配列のID
- テキストベタ打ちあるいはFASTAフォーマットがよく用いられる

```
>gi|213536444|gb|FJ436046.1| Oryzias latipes Nanog (Nanog) mRNA, complete cds  
ATGGCGGAGTGGAAAACTCAGGTCAACTACAACCCCACATTCCATGCGTACACCTATGGCTTCGTGTATC  
AAACTGGGCCCGAACAGAACGACACGTAAACGGGAACGACTGGAGCCAAACTGTGAGCAGAACGGCTACAA  
CGGAGGACCCACGCAGTCTCATTTCCCCGCTAGGAGCCGGGAGGAGTCCCCACCACGCAGCCGGAGCAG  
CAGCCTGAGAGCGGCCACTATTACCAAGGACTCCGGGTGGTATAACATCAGAGAGGCCAGACGGCCGCT
```

```
>gi|213536445|gb|ACJ51123.1| Nanog [Oryzias latipes]  
MAEWKTQVNPNPTFHAYTYGFVYQTGPEQNHTGNDWSQNCEQNGYNGGPTQSHFPARSREESPPRSPEQ  
QPESGHYYQDSGVYIREAQTRGLVMAGQHRVGLDSGENCTRRTGSDSASDSEAHTSPDSWSSCSNYDRS  
VPQTDPVVWVKNEEQTGARSPDHSEDVSSSLMVESQSFAVQDTGDASSSTHAPFTTKQASSTPNAPKA
```

+

配列データベース



- 塩基/アミノ酸配列の集合体。目的をもって集められ、機能やゲノム構造などの注釈付け(アノテーション)をされる
 - 網羅的
 - DDBJ/EMBL/Genbank (塩基配列)、UniProt(タンパク)
 - 配列の特徴
 - miRBase (microRNA), IMGT/HLA (HLA領域)
 - 生物種
 - Ensembl (動物)、各種ゲノムプロジェクト

ほとんどのデータベースは、ウェブ上で実行できるBLASTをもっている

+ DDBJ/EMBL/Genbank

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Limits Advanced Help

Nucleotide

BLAST® Basic Local Alignment Search Tool My NCBI [Sign In] [Register]

NCBI BLAST Home BLAST finds regions of similarity between biological sequences. more... New DELTA-BLAST, a more sensitive protein-protein search Go

Using Nucleotide

Quick Start Guide
FAQ
Help
GenBank FTP
RefSeq FTP

You are here: NCBI > DNA & RNA > Nucleotide

GETTING STARTED

NCBI Education
NCBI Help Manual
NCBI Handbook
Training & Tutorials

BLAST Assembled RefSeq Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

<input type="checkbox"/> Human	<input type="checkbox"/> Oryza sativa	<input type="checkbox"/> Gallus gallus
<input type="checkbox"/> Mouse	<input type="checkbox"/> Bos taurus	<input type="checkbox"/> Pan troglodytes
<input type="checkbox"/> Rat	<input type="checkbox"/> Danio rerio	<input type="checkbox"/> Microbes
<input type="checkbox"/> Arabidopsis thaliana	<input type="checkbox"/> Drosophila melanogaster	<input type="checkbox"/> Apis mellifera

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a nucleotide database using a nucleotide query <i>Algorithms: blastn, megablast, discontiguous megablast</i>
protein blast	Search protein database using a protein query <i>Algorithms: blastp, psi-blast, phi-blast, delta-blast</i>
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins and T cell receptor sequences](#) (IgBLAST)
- Screen sequence for [vector contamination](#) (vecscreen)

News

Update to organism BLAST databases
The organism BLAST pages are being updated to use top-level (chromosome + unplaced and unlocalized scaffolds) RefSeq genomic records instead of scaffold records.
Thu, 17 Oct 2013 14:00:00 EST
[More BLAST news...](#)

Tip of the Day

Use Genomic BLAST to see the genomic context
If you are interested in the evolution of a particular gene or gene family it is often interesting to examine the intron-exon structure even across species.
[More tips...](#)

UniProt

The UniProt homepage features a large header with a purple background and a white logo. Below the header, there are two main sections of the website.

Top Navigation Bar: Includes links for Search, Blast, Align, Retrieve, ID Mapping, Downloads, Contact, and Documentation/Help.

Welcome Section: Contains a brief introduction to the mission of UniProt, followed by a "What we provide" section. This section lists four main databases: UniProtKB, UniRef, UniParc, and Supporting data, each with a brief description and a "complete and reference proteome sets" link.

Blast Search Form: A search interface for protein sequence similarity. It includes fields for "Sequence or UniProt identifier", "Database" (set to UniProtKB), "Threshold" (set to 10), "Matrix" (set to Auto), "Filtering" (set to None), and "Gapped" (set to yes) and "Hits" (set to 250). There are also "Blast" and "Clear" buttons, and a "« Options" link.

Help Section: Provides instructions for performing a sequence similarity search, mentioning Swiss-Prot and TrEMBL, and includes a "More..." link.

News Section: Headlines include "UniProt release 2013_10 - Oct 16, 2013" and "When the cat's away... I Cross-references to PRO". It also features a Twitter feed for @uniprot with 697 followers.

Site Tour Section: Shows a screenshot of the UniProt interface with a callout box highlighting a feature.

Protein Spotlight Section: Features an article about water, stating: "There is no life without water. However, living beings go through life with given amounts of water inside them. Which is why there has to be a system that sustains this balance, and prevents too much water from flowing in, or indeed pouring out...".

Footer: Displays the UniProt logo with its signature blue and green circular dots.

+ Ensembl

e!Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

Search all species...

e!Ensembl

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Search all species...

Browse a Genome

The Ensembl project provides genomic information for vertebrates and other eukaryotes. All information is freely available.

Popular genomes

Human
GRCh37

Zebrafish
Zv9

[★ Log in to customize](#)

All genomes

-- Select a species --

[View full list of all EnsemblGenomes](#)

Important Notice

We now use Blat as our default DNA search. This will make your query faster.

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:

Or Upload a file containing one or more FASTA sequences
 ファイル未選択

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
 Retrieve

Or Enter an existing ticket ID:
 Retrieve

dna queries
 peptide queries

Select the databases to search against

Select species:
Use 'ctrl' key to select multiple species

Gallus_gallus
 Gasterosteus_aculeatus
 Gorilla_gorilla
 Homo_sapiens

dna database
 protein database

Genomic sequence
Proteins

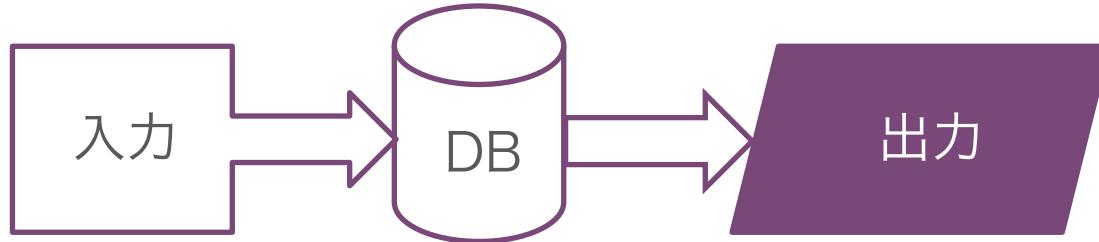
Select the Search Tool

configure

BLASTN
BLAT
TBLASTX

+

出力



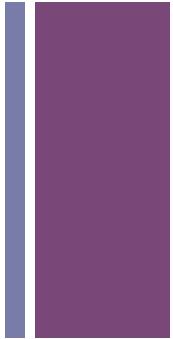
問い合わせ配列と類似配列の間の類似性を出力する

- 類似度
 - ビットスコア: 正規化されたアラインメントスコア。大きいほど類似度が高い
 - E-value: 偶然に類似してしまうと期待されるデータベース中の配列数。データベースの大きさによって変わる
 - 一致度 (%)
- 類似する領域
- アラインメント

Query 1	MAEWKTQV--NYNPTFHAYTYGFVYQTGPEQNHVTGNDWSQN--CEQNGYNGGPTQSHFP	56
	MA+WK+Q+ NYNP++HAY YG VYQ GPEQNH + W + + + YN G TQ+++	
Sbjct 1	MADWKSQISYNYNPSYHAYAYGLVYQPGPEQNHGNLSSWGETGVTDLNSYNSGVTQAYYA	60

+

実際に使ってみましょう！



+ NCBI BLASTにアクセス

The screenshot shows a Google search results page for the query "NCBI BLAST". The search bar at the top contains "NCBI BLAST". Below the search bar, there are several interface elements: "Google 検索", "I'm Feeling Lucky", the Google logo, and a search button. The main search results area has a heading "NCBI BLAST". Below this, there are several search filters: "ウェブ", "画像", "地図", "ショッピング", "動画", "もっと見る", and "検索ツール". A search count "約 17,800,000 件 (0.15 秒)" is displayed. A purple arrow points from the text "このページを訳す" down to the "BLAST: Basic Local Alignment Search Tool" link.

NCBI BLAST

Google 検索 I'm Feeling Lucky

Google NCBI BLAST

ウェブ 画像 地図 ショッピング 動画 もっと見る 検索ツール

約 17,800,000 件 (0.15 秒)

他のキーワード: [ncbi blast 使い方](#)

[**BLAST: Basic Local Alignment Search Tool**](#)

blast.ncbi.nlm.nih.gov/ このページを訳す

The Basic Local Alignment Search Tool (**BLAST**) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

[Nucleotide Blast \(BLASTn\)](#)

BLASTN programs search nucleotide databases using a ...

[Protein BLAST](#)

Search using Blastp (protein-protein BLAST) Show results in ...

[Blastx](#)

BLASTX search protein databases using a translated nucleotide ...

[Primer-BLAST](#)

NCBI/ Primer-BLAST: Finding primers specific to your PCR ...

+ BLASTの種類の選択

Basic BLAST

Choose a BLAST program to run.

[nucleotide blast](#)

Search a nucleotide database using a nucleotide query
Algorithms: blastn, megablast, discontiguous megablast

[protein blast](#)

Search protein database using a protein query
Algorithms: blastp, psi-blast, phi-blast, delta-blast

[blastx](#)

Search protein database using a translated nucleotide query

[tblastn](#)

Search translated nucleotide database using a protein query

[tblastx](#)

Search translated nucleotide database using a translated nucleotide query



+ クエリを入れる

Enter Query Sequence

BLASTP programs search protein databases using a protein query. [more...](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

[Clear](#)

Query subrange [?](#)

Or, upload file

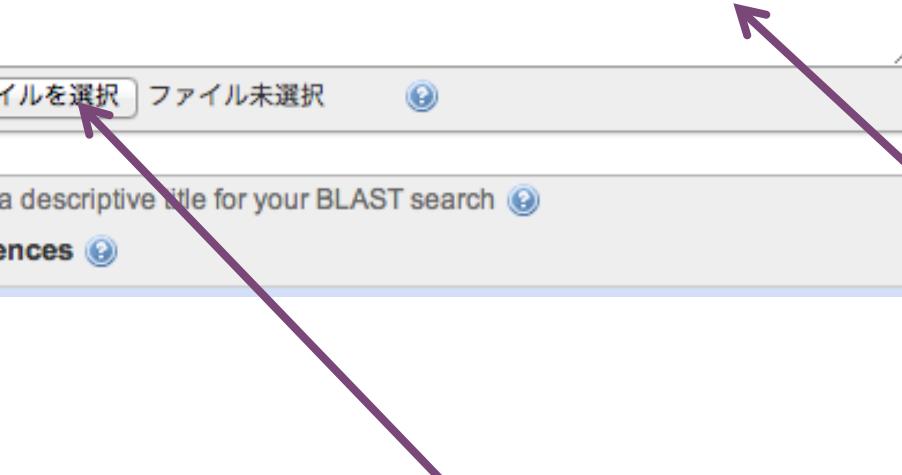
[ファイルを選択](#)

ファイル未選択 [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)



配列あるいはIDを入れる。
複数でもよい

あるいはファイルとしてアップロードする

+ データベースの選択

Choose Search Set

Database: Non-redundant protein sequences (nr) 

Organism
Optional: Enter organism name or id--completions will be suggested Exclude 

Enter organism common name, binomial, or tax id. Only 10 top taxa will be shown. 

Exclude
Optional: Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query
Optional: Enter an Entrez query to limit search 

- ✓ Non-redundant protein sequences (nr)
- Reference proteins (refseq_protein)
- UniProtKB/Swiss-Prot(swissprot)
- Patented protein sequences(pat)
- Protein Data Bank proteins(pdb)
- Metagenomic proteins(env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

全アミノ酸配列から冗長性を取り除いたデータ

モデル生物から選択した基準となる配列データ

+ よく使うオプション

The screenshot shows the 'Algorithm parameters' section of the NCBI BLAST search interface. It includes three tabs: 'General Parameters', 'Scoring Parameters', and 'Filters and Masking'. The 'General Parameters' tab is active.

- Max target sequences:** Set to 100. A purple arrow points from the Japanese text '示す類似配列の最大数' to this field.
- Short queries:** Contains a checked checkbox labeled 'Automatically adjust parameters for short input sequences'.
- Expect threshold:** Set to 10. A purple arrow points from the Japanese text 'E-valueの閾値' to this field.
- Word size:** Set to 3.
- Max matches in a query range:** Set to 0.

The 'Scoring Parameters' tab contains settings for the scoring matrix, gap costs, and compositional adjustments.

The 'Filters and Masking' tab contains settings for filtering low complexity regions and masking lowercase letters.

示す類似配列の
最大数

E-valueの閾値

リピート配列などの
フィルタリング

+ 実行！

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

Show results in a new window

NCBI/BLAST/blastp suite/Formatting Results - 6CHBC0YT01R [Formatting options]

Job Title: gb|ACJ51123.1| (420 letters)

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. 1 75 150 225 300 375 420

specific DNA base contacts
DNA binding site

Specific hits homeodomain

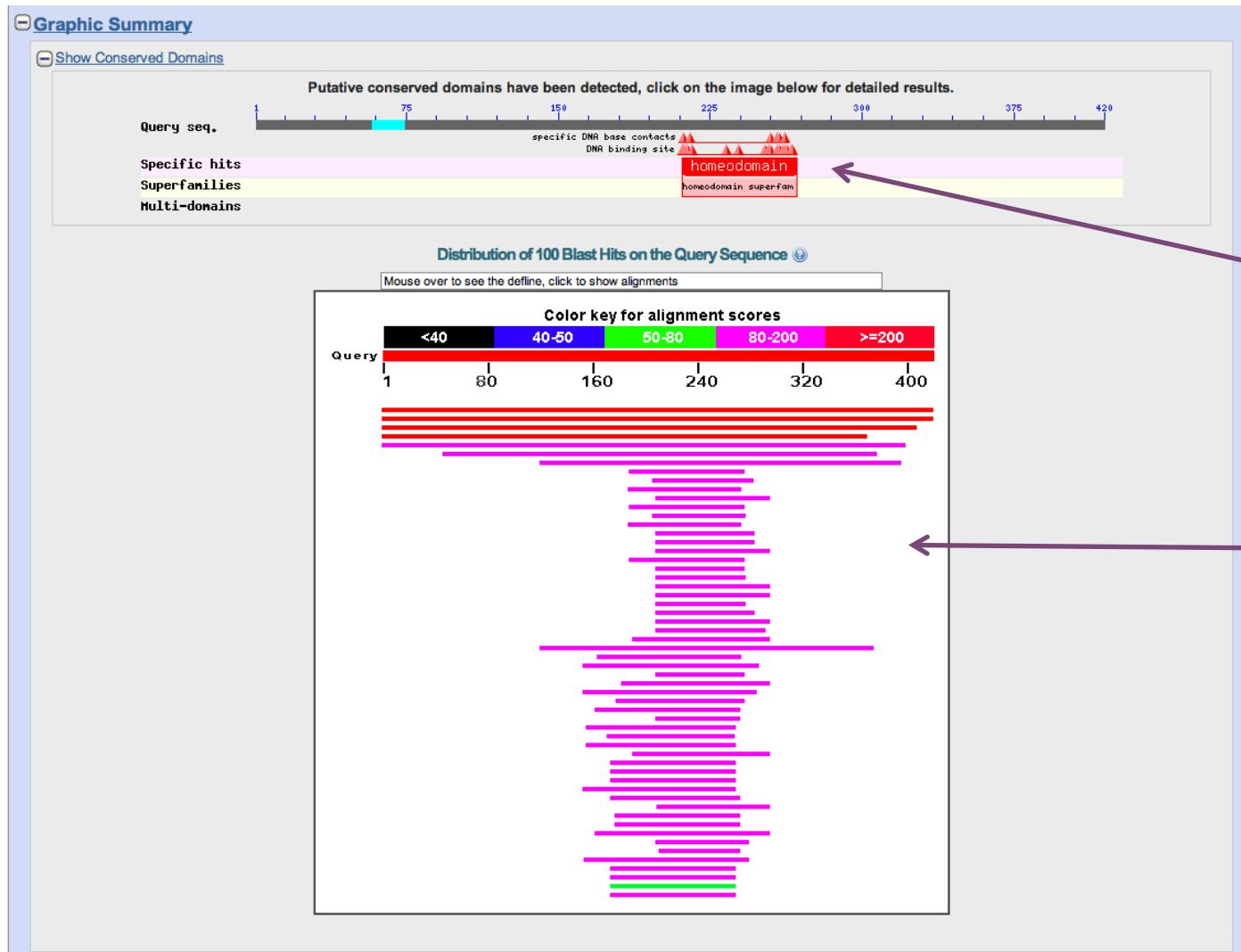
Superfamilies homeodomain superfam

Multi-domains

Request ID	6CHBC0YT01R
Status	Searching
Submitted at	Tue Oct 22 05:29:16 2013
Current time	Tue Oct 22 05:29:37 2013
Time since submission	00:00:20

+

結果 1 グラフィカル出力

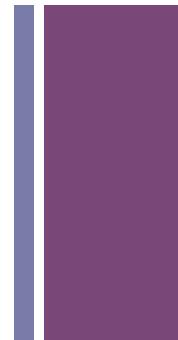


保存ドメインと
入力配列上の
位置

類似配列の類似度
(アライメントスコア)とアライン
される領域



結果2 サマリー



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Nanog protein [Oryzias latipes] >gb ACJ51123.1 Nanog [Oryzias latipes]	877	877	100%	0.0	100%	NP_001153902.1
<input type="checkbox"/>	Homeodomain transcription factor Nanog [Dicentrarchus labrax]	445	445	100%	2e-150	57%	CBN81816.1
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC101076503 [Takifugu rubripes]	291	291	96%	7e-91	46%	XP_003968034.1
<input type="checkbox"/>	unnamed protein product [Tetraodon nigroviridis]	236	236	88%	7e-70	39%	CAF93536.1
<input type="checkbox"/>	nanog homeobox [Danio rerio] >gb ABQ42198.1 ovary-expressed homeobox protein [Danio rerio] >gb AAI62315.1 Zgc:19 	196	196	95%	1e-54	38%	NP_001091862.1
<input type="checkbox"/>	nanog, partial [Carassius auratus]	171	171	78%	1e-45	37%	AEG74407.1
<input type="checkbox"/>	PREDICTED: homeobox protein pv.1-like isoform X1 [Oreochromis niloticus] >ref XP_005465993.1 PREDICTED: homeobox protein pv.1-like isoform X1 [Oreochromis niloticus] >ref XP_005465992.1 	140	140	65%	2e-35	37%	XP_005465992.1
<input type="checkbox"/>	PREDICTED: homeobox protein NANOG-like [Meleagris gallopavo]	94.4	94.4	20%	1e-18	52%	XP_003202657.1
<input type="checkbox"/>	PREDICTED: homeobox protein NANOG [Anas platyrhynchos]	93.2	93.2	18%	2e-18	53%	XP_005012489.1
<input type="checkbox"/>	Homeobox protein NANOG, partial [Columba livia]	92.0	92.0	20%	3e-18	52%	EMC82546.1
<input type="checkbox"/>	PREDICTED: homeobox protein vent1B-like [Falco peregrinus] >ref XP_005446101.1 PREDICTED: homeobox protein vent1B-like [Falco peregrinus] >ref XP_005446101.1 	92.8	92.8	20%	4e-18	51%	XP_005234495.1
<input type="checkbox"/>	PREDICTED: homeobox protein NANOG [Falco cherrug]	92.0	92.0	20%	4e-18	49%	XP_005446141.1



結果3 アラインメント

[Download](#) ▾ [GenPept](#) [Graphics](#)

Homeodomain transcription factor Nanog [Dicentrarchus labrax]

Sequence ID: [emb|CBN81816.1](#) Length: 436 Number of Matches: 1

Range 1: 1 to 436 GenPept Graphics					▼ Next Match	▲ Previous Match
Score	Expect	Method	Identities	Positives	Gaps	
445 bits(1145)	2e-150	Compositional matrix adjust.	250/436(57%)	308/436(70%)	16/436(3%)	
Query 1	MAEWKTQV--NYNPTFHAYTYGFVYQTGPEQNHVTGNDWSQN--CEQNGYNGGPTQSHFP		56			
Sbjct 1	MA+WK+Q+ NYNP++HAY YG VYQ GPEQNH + W + + + YN G TQ+++					
Query 57	MADWKSQISYNYNPSYHAYAYGLVYQPGPEQNHGNLSSWGETGVTDLNSNYNSGVTQAYYA		60			
Sbjct 61	--ARSREESPPRSPEQQPESGH-YYQDGSVVYIREAQTGRVLVMAGQHRVGLDSGENCTRR		113			
	AR+REESPP SPEQ +GH +YQ SGVVY+ + Q GRL++++GQHR D+ R					
Query 114	TTARAREESPPGSPEQHAANGHCHYQGSGVVYLGDTQAGRLLLSGQHRAAYDARTQEAGR		120			
Sbjct 121	TGSDSASDSEAHTSPDSWSSCSNYDRSPVQTDPVVVKNE--EQTGARSPDHSEDVSSL		171			
	GSDS SDSEAHTSPDSWSS S+ + S+PQ DP W K + ++ +RSPD EDV SSL					
Query 172	AGSDSTSDSEAHTSPDSWSSGSSREGSLPQADPATWAKQDLDEASSRSPDAGEDVPSSL		180			
Sbjct 181	MVESQSFAV---QDTGDASSSTHAPFTTCKQASSTPNAPKAKVRAAFSESQMSTLVQRF		228			
	E F V + T D ++S H P T KK +++++ N PKAK RAAFSESQM+ LVQRF					
Query 229	KEEPMPTVTGNEGTDDTATSVHVPLTAKKPSTTSGNNPKAKARAASFSESQMNALVQRF		240			
Sbjct 241	SVQRYLAPAEMKNLADVTGLTYKQVKTWFQNRRMKLRRHQKDTSWVSERYTINKDNTAAD		288			
	SVQRYL PAEMKNLA++TGLTYKQ+KTWFQNRRMKLRRHQKDTSWVSERYT NKD+					
Query 289	SVQRYLTPAEMKNLAEMTGLTYKQIKTWFQNRRMKLRRHQKDTSWVSERYTTNKDSPVRG		300			
Sbjct 301	TVF\$NVAPHVPPYQGDGMSHLRHHYNQHMMGAFKNT-PHNLAFYLAAMGNPPGTAGYPP		347			
	TVF+NV H+PPYQG+ L+ HYNQHMM +AFK T P NLAFYLAAMG+ G+AGYP					
Query 348	TVFTNVPSHIPPYQGEARPQLKEHYNQHMMESAFKKTAPQNLAFYLAAMGSAAGSAGYPS		360			
Sbjct 361	WSSSSPPQAAVPSRPQVPGWPPLPPGRSQFGFCPIPYDPSDAASLNFERNAIPDSKDGEWA		407			
	WSS PPQ AVP+RPQ GW +PPG S + + P ++ ++ AS N + +SKD ES					
Query 408	WSSGGPPQTAVPTRPQAAGWSMPPGISHYEYNPGAFNSANVASAYNTGHDMFSKDVESV		420			
Sbjct 421	NSRGSLNTAIVHNASQ		436			

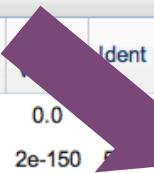
- ギャップはハイフンで表される
- 2配列間文字列は一致度を示す
 - 一致していればそのアミノ酸残基/塩基
 - アミノ酸配列の物理化学的性質が類似していれば”+”

+ 結果2 サマリー

Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	P-value	Ident	Accession
<input type="checkbox"/>	Nanog protein [Oryzias latipes] >gb ACJ51123.1 Nanog [Oryzias latipes]	877	877	100%	0.0	NP_001153902.1	
<input type="checkbox"/>	Homeodomain transcription factor Nanog [Dicentrarchus labrax]	445	445	100%	2e-150	NP_001153902.1	NP_81816.1
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC101076503 [Takifugu rubripes]	291	291	96%	7e-91	46%	XP_003968034.1
<input type="checkbox"/>	unnamed protein product [Tetraodon nigroviridis]	236	236	88%	7e-70	39%	CAF93536.1
<input type="checkbox"/>	nanog homeobox [Danio rerio] >gb ABQ42198.1 ovary-expressed homeobox protein [Danio rerio] >gb AAI62315.1 Zgc:19 	196	196	95%	1e-54	38%	NP_001091862.1
<input type="checkbox"/>	nanog, partial [Carassius auratus]	171	171	78%	1e-45	37%	AEG74407.1
<input type="checkbox"/>	PREDICTED: homeobox protein pv.1-like isoform X1 [Oreochromis niloticus] >ref XP_005465993.1 PREDICTED: homeobox protein pv.1-like isoform X1 [Oreochromis niloticus] >ref XP_005465992.1 	140	140	65%	2e-35	37%	XP_005465992.1
<input type="checkbox"/>	PREDICTED: homeobox protein NANOG-like [Meleagris gallopavo]	94.4	94.4	20%	1e-18	52%	XP_003202657.1
<input type="checkbox"/>	PREDICTED: homeobox protein NANOG [Anas platyrhynchos]	93.2	93.2	18%	2e-18	53%	XP_005012489.1
<input type="checkbox"/>	Homeobox protein NANOG, partial [Columba livia]	92.0	92.0	20%	3e-18	52%	EMC82546.1
<input type="checkbox"/>	PREDICTED: homeobox protein vent1B-like [Falco peregrinus] >ref XP_005446101.1 PREDICTED: homeobox protein vent1B-like [Falco peregrinus] >ref XP_005446101.1 	92.8	92.8	20%	4e-18	51%	XP_005234495.1
<input type="checkbox"/>	PREDICTED: homeobox protein NANOG [Falco cherrug]	92.0	92.0	20%	4e-18	49%	XP_005446141.1



類似配列のエントリー

Homeodomain transcription factor Nanog [Dicentrarchus labrax]

GenBank: CBN81816.1

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

Go to:

LOCUS CBN81816 436 aa linear VRT 01-FEB-2012
DEFINITION Homeodomain transcription factor Nanog [Dicentrarchus labrax].
ACCESSION CBN81816
VERSION CBN81816.1 GI:317419780
DBLINK BioProject: [PRJEA39865](#)
DBSOURCE embl accession [FQ310508.3](#)
KEYWORDS .
SOURCE Dicentrarchus labrax (European seabass)
ORGANISM [Dicentrarchus labrax](#)

FASTA形式で配列を
取得できる

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
Percoidei; Moronidae; Dicentrarchus.

REFERENCE 1
AUTHORS Kuhl,H., Tine,M., Hecht,J., Knaust,F. and Reinhardt,R.
TITLE Analysis of single nucleotide polymorphisms in three chromosomes of
European sea bass *Dicentrarchus labrax*
JOURNAL Comp. Biochem. Physiol. Part D Genomics Proteomics 6 (1), 70-75
(2011)
PUBMED [20452842](#)

REFERENCE 2 (residues 1 to 436)
AUTHORS Tine,M., Kuhl,H., Beck,A., Bargelloni,L. and Reinhardt,R.
TITLE Comparative analysis of intronless genes in teleost fish genomes:
insights into their evolution and molecular function
JOURNAL Mar Genomics 4 (2), 109-119 (2011)
PUBMED [21620332](#)

REFERENCE 3
AUTHORS Kuhl,H., Tine,M., Beck,A., Timmermann,B., Kodira,C. and
Reinhardt,R.

+ 種を限定してBLASTをかけたいときには

データベースを選ぶときに、種や分類群を入力して指定できる

Choose Search Set

Database: Non-redundant protein sequences (nr)

Organism (Optional): Vertebrates, Homo sapiens, Drosophila |

Exclude (Optional): Exclude +
 Exclude
 Exclude

Drosophila |
Drosophila (taxid:7215)
Drosophila Fallen, 1823 (taxid:7215)
Drosophila melanogaster (taxid:7227)
Drosophila (taxid:32281)
Drosophila simulans (taxid:7240)

20 top taxa will be shown. (?)

Program Selection

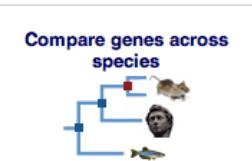
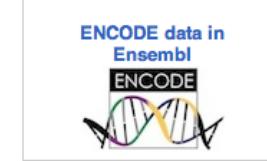
入力補完される

+ EnsemblのBLAST

- とくに特定の動物種を対象にし、類似配列のアノテーションを詳細に見たい場合には、EnsemblでBLASTをかけると便利

Ensembl Species

	Aardvark (preview - assembly only) <i>Orycteropus afer afer</i> OryAfe1
	Alpaca <i>Vicugna pacos</i> vicPac1
	Anole lizard <i>Anolis carolinensis</i> AnoCar2.0
	Armadillo (preview new assembly) <i>Dasnov3.0</i> Dasypus novemcinctus dasNov2
	Baboon (preview - assembly only) <i>Papio hamadryas</i> Pham
	Budgerigar (preview - assembly only) <i>Melopsittacus undulatus</i> MelUnd6.3
	Bushbaby <i>Otolemur garnettii</i> OtoGar3
	Ciona intestinalis <i>Ciona intestinalis</i> KH
	Ciona savignyi <i>Ciona savignyi</i> CSAV2.0
	Gibbon <i>Nomascus leucogenys</i> Nleu1.0
	Guinea Pig <i>Cavia porcellus</i> cavPor3
	Hedgehog <i>Erinaceus europaeus</i> HEDGEHOG
	Horse <i>Equus caballus</i> EquCab2
	Human <i>Homo sapiens</i> GRCh37
	Hyrax <i>Procavia capensis</i> proCap1
	Kangaroo rat <i>Dipodomys ordii</i> dipOrd1
	Lamprey <i>Petromyzon marinus</i> Pmarinus_7.0
	Platypus <i>Ornithorhynchus anatinus</i> OANA5
	Prairie vole (preview - assembly only) <i>Microtus ochrogaster</i> MicOch1.0
	Rabbit <i>Oryctolagus cuniculus</i> OryCun2.0
	Rat <i>Rattus norvegicus</i> Rnor_5.0
	Rhinoceros (preview - assembly only) <i>Ceratotherium simum simum</i> CerSimSim1
	Saccharomyces cerevisiae <i>Saccharomyces cerevisiae</i> EF4
	Sheep (preview - assembly only) <i>Ovis aries</i> Oar_v3.1
	Shrew (preview new assembly) <i>Sorex araneus</i> COMMON_SHREW1
	Sloth <i>Choloepus hoffmanni</i> choHof1



+ 入力・実行

配列を入力

DNA or
タンパク

種を選択
Ctrlを押しながら
複数選択できる

データベース
を選択

BLASTを選択

new **SETUP** **CONFIG** **RESULTS** **DISPLAY**

Important Notice
We now use Blat as our default DNA search. This will make your query faster.

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:

```
>gi|317419780|emb|CBN81816.1| Homeodomain transcription factor Nanog
MADWKSQIISNNPSYHAYAYGLVYQPGEQNHNGLSSWGETGVTDLSNYNSGVTQAY
PGSPEQHAANGHCYHQSGVVYLGDTQAGRLLLSQHRAAYDARTQEAGRAGSDSTS
GSREGSLPQADPATWAKQDLDEASSRSPDAGEDVPSSLKEEPMPFTVTGNEGTTDA
KPTTSGNNPKAKARAFAFSESQMNALVQRFSVQRYLTPAEMKNLAEMTGLTYKQIKTWF
```

Or Upload a file containing one or more FASTA sequences

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)

dna queries
 peptide queries

Select the databases to search against

Select species:
 Use 'ctrl' key to select multiple species
 Gallus_gallus
 Gasterosteus_aculeatus
 Gorilla_gorilla
 Homo_sapiens

dna database
 protein database

Select the Search Tool

BLASTP **RUN▶**

Search sensitivity:
 Optimise search parameters to find the following alignments
 Near-exact matches

About BlastView
 BlastView provides an integrated platform for sequence similarity searches against Ensembl databases, offering access to both BLAST and BLAT programs.

new **SETUP** **CONFIG** **RESULTS** **DISPLAY**

Run Search

Configuration for BLASTP

-E Maximum E-value for reported alignments

-B Maximum number of database hits to report

-filter Program used to filter query sequence

-sort_by Sort option for database hits

-statistics Statistics option for calculation of alignment score

-W Word size for seeding alignments

-wink Step-size for sliding-window used to seed alignments

-hitdist Max distance between words for two-hit seeding. (One-hit seeding by default)

-matrix Scoring matrix file

-Q Cost of first gap character

-R Cost of second and remaining gap characters

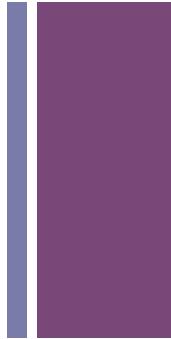
-nogap Turns off gapped alignments

-T Neighborhood word threshold score

-X Alignment extension cutoff

Additional Other options (not validated)

オプションを指定したい
ときにはconfigure



少し時間を
置いて押す

new SETUP CONFIG RESULTS DISPLAY

Retrieve result for ID:
BLA_KjSCqAcRe Retrieve

Alignment Display Options:
 Locations vs. Karyotype Locations vs. Query
 Summary Table

1: gil317419780|emb|CBN81816.1| (436 letters) Vs. PEP_ALL
Homo_sapiens 231 alignments, 100 hits [RawResult](#) **view ►**

refresh Online Help

Summary

- setup
 - Homo_sapiens
 - Proteins
 - BLASTP
 - Custom sensitivity
- configure
 - -E: 10
 - -B: 100
 - -filter: seg
 - -W: 4
 - -hitdist: 40
 - -matrix: BLOSUM80
 - -Q: 3
 - -R: 3
 - -T: 16
- results

display

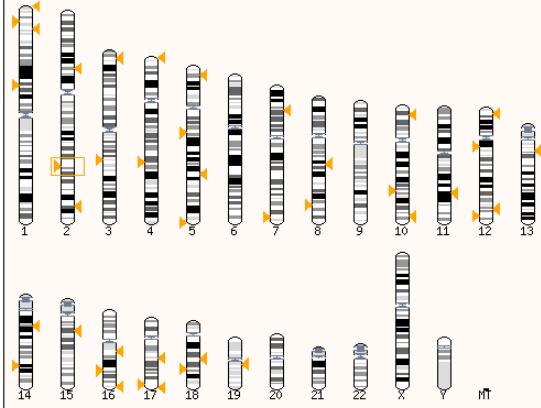
- ① Not yet initialised

viewボタンが
現れればOK

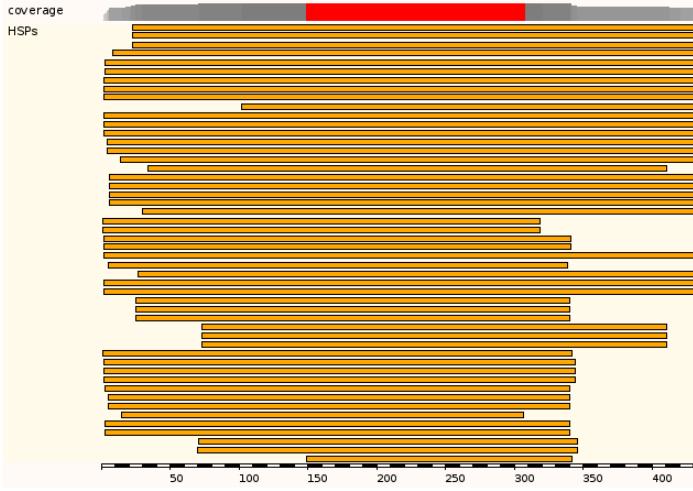


結果

Alignment Locations vs. Karyotype (click arrow to hide)



Alignment Locations vs. Query (click arrow to hide)



Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

refresh

Query	Subject	Chromosome	Supercontig	Clone	Contig	Lrg	Stats	Sort By
_off_Name	_off_Name	_off_Name	_off_Name	_off_Name	_off_Name	_off_Name	_off_Score	<Lrg
Start	Start	Start	Start	Start	Start	Start	E-val	>Lrg
End	End	End	End	End	End	End	P-val	<Score
[A] [S] [G] [C]	23 436 + ENSP00000427434	1	420 +	Chr:2 177033843	177036963 +	525	3.3e-41	25.60 496
[A] [S] [G] [C]	23 436 + ENSP00000386084	1	420 +	Chr:2 177033843	177036963 +	525	3.3e-41	25.60 496
[A] [S] [G] [C]	23 436 + ENSP00000245914	1	420 +	Chr:2 177033843	177036963 +	525	3.3e-41	25.60 496
[A] [S] [G] [C]	2 436 + ENSP00000311624	1	510 +	Chr:10 126714799	126716088 -	523	1.7e-40	26.31 517
[A] [S] [G] [C]	30 436 + ENSP00000312389	1	396 +	Chr:2 176945078	176948468 -	522	7.6e-41	26.60 470
[A] [S] [G] [C]	4 429 + ENSP00000050961	1	3104 +	Chr:8 77767252	77775409 +	510	1.3e-38	25.09 530
[A] [S] [G] [C]	4 435 + ENSP00000438926	1	2260 +	Chr:16 72822653	72828378 -	507	1.9e-38	26.35 520
[A] [S] [G] [C]	4 435 + ENSP00000268489	273	174 +	Chr:16 72822653	72828378 -	507	2.6e-38	26.35 520
[A] [S] [G] [C]	4 435 + ENSP00000399604	2654	14 +	Chr:8 77767252	77775397 +	497	2.2e-37	25.00 532
[A] [S] [G] [C]	4 435 + ENSP00000430848	2673	123 +	Chr:8 77767252	77775397 +	497	2.3e-37	25.00 532
[A] [S] [G] [C]	4 435 + ENSP00000430497	2699	3149 +	Chr:8 77767252	77775397 +	497	2.3e-37	25.00 532
[A] [S] [G] [C]	70 346 + ENSP00000407978	116	396 +	Chr:4 1396768	1399774 -	488	1.1e-37	30.00 340
[A] [S] [G] [C]	2 431 + ENSP00000387214	168	596 +	Chr:7 31617656	31683048 +	488	3.7e-37	25.85 530
[A] [S] [G] [C]	2 431 + ENSP00000384416	260	688 +	Chr:7 31617656	31683048 +	488	4.1e-37	25.85 530
[A] [S] [G] [C]	2 431 + ENSP00000395835	286	714 +	Chr:7 31617656	31683048 +	488	4.2e-37	25.85 530
[A] [S] [G] [C]	2 435 + ENSP00000331302	102	545 +	Chr:18 44560001	44561332 -	486	4.2e-37	24.62 520
[A] [S] [G] [C]	1 436 + ENSP00000420587	111	538 +	Chr:19 31769085	31770368 -	486	6.7e-37	24.57 525
[A] [S] [G] [C]	2 436 + ENSP00000402343	1961	2391 +	Chr:16 88499843	88501135 +	486	2.8e-36	27.27 528
[A] [S] [G] [C]	2 436 + ENSP00000456500	1989	2419 +	Chr:16 88499843	88501135 +	486	2.8e-36	27.27 528
[A] [S] [G] [C]	2 344 + ENSP00000382300	260	599 +	Chr:13 33692332	33703682 -	485	7.5e-37	27.80 410

+ Ensemblの画面へリンク

Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh37) ▾ Location: 2:177,015,122-177,037,825 Gene: HOXD3 Transcript: HOXD3-005

Search all species...

Transcript-based displays

- Transcript summary
- Supporting evidence (12)
- Sequence
 - Exons (4)
 - cDNA
 - Protein
- External References
 - General identifiers (17)
 - Oligo probes (43)
- Ontology
 - Ontology graph (16)
 - Ontology table (16)
- Genetic Variation
 - Variation table
 - Variation image
 - Population comparison
 - Comparison image
- Protein Information
 - **Protein summary**
 - Domains & features (9)
 - Variations (94)
- External data
 - Personal annotation
- ID History
 - Transcript history
 - Protein history

Protein summary ⓘ

Protein domains for ENSP00000424734.2

Ave. residue weight: 105.857 g/mol
Charge: 16.5
Isoelectric point: 8.9689
Molecular weight: 45,730.10 g/mol
Number of residues: 432 aa

Statistics

Configure this page

Add your data

Export data

Bookmark this page

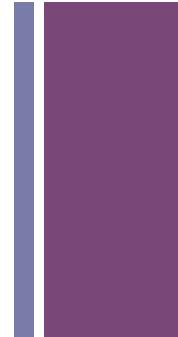
Share this page

+

課題

- 別ページの課題1を行います。

+ マルチプルアラインメント



- 3本以上の配列を相同な塩基/アミノ酸が対応するように整列させること
- 先ほど紹介した相同性検索プログラムは2配列をアラインする（ペアワイズアラインメント）プログラムでもある



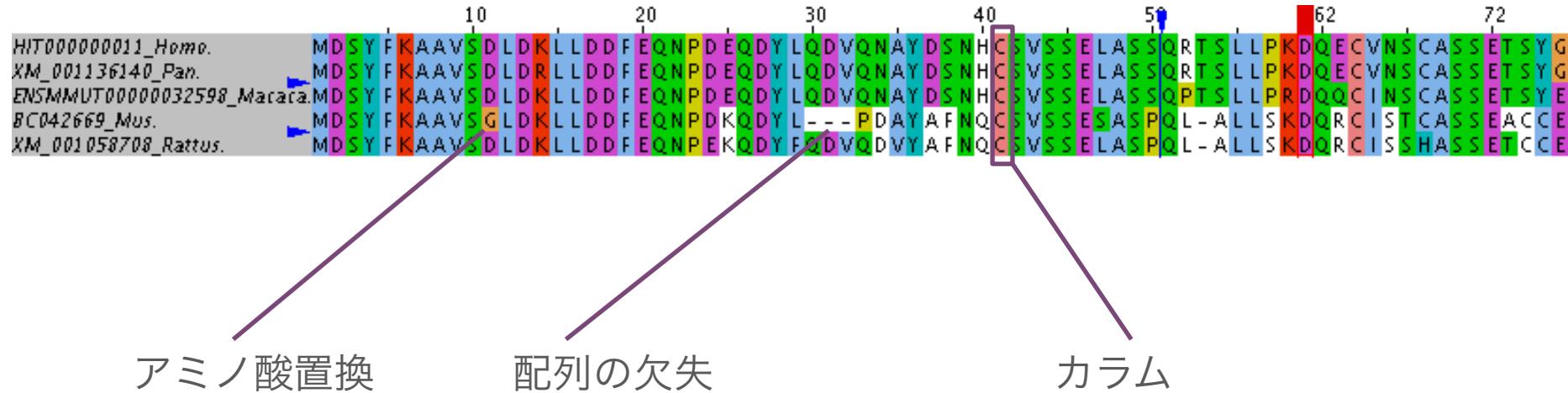
マルチプルアラインメント

- 3本以上の配列を相同な塩基/アミノ酸が対応するように整列させること

```
>HIT000251116_Homo. Nanog
MSVDPACPQSLPCFEASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMDL
LIQDSPDSSTSPKGKQPTSAEKSVAKKEDKVPVKQKTRTVFSSTQLCVLNDRFQRQKYL
SLQQMQELSNILNLNSYKQVKTWFQNQRMKSKRWQKNNWPKNNSNGVTQKASAAPTYPSLYSS
YHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNNSPF
YNCGEESLQSCMFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNM
QPEDV
>ENSPTRT00000008541_Pan. Nanog
MSVDPACPQSLPCFEASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMDL
LIQDSPDSSTSPKGKQPTSAENSVTKKEDKVPVKQKTRTVFSSTQLCVLNDRFQRQKYL
SLQQMQELSNILNLNSYKQVKTWFQNQRMKSKRWQKNNWPKNNSNGVTQKASAAPTYPSLYSS
YHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNNSPF
YNCGEESLQSCMFQPNSPASDLEAALEAAGEGLNVIQQTARYFSTPQTMDLFLNYSTNM
QPEDM
>ENSMUST00000112582_Mus. Nanog
MSVGLPGPHSLPSSEEASNNSGNASSMPAVFHPENYSCLQGSATEMLCTEAASPRPSSEDL
PLQGSPDSSTSPKQKLSSPEADKGPEEEENKVLARKQKMRTVFSQAQLCALKDRFQKQKY
LSLQQMQELSSILNLNSYKQVKTWFQNQRMKCKRWQKNQWLKTSNGLIQQKGSAPVEYPSIH
CSYPQGYLVNASGSLSMWGSQTWTNPTWSSQTWTNPTWSSQAWTAQSWNGQPWNAAPLHN
FGEDFLQPYVQLQQNFASDLEVNLREATRESHAHFSTPQALESFLNYSVTPPGEI
...
...
```

+ マルチプルアラインメント

- 3本以上の配列を相同な塩基/アミノ酸が対応するように整列させること



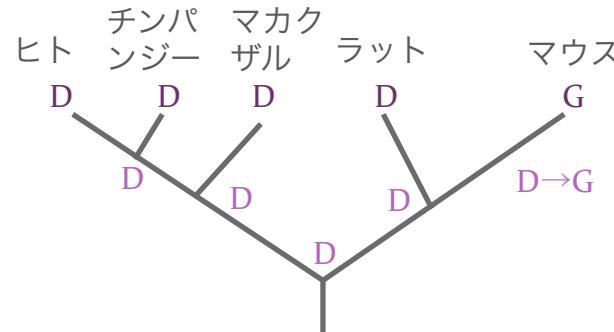
+ マルチプルアラインメントから何がわかる？

■ 類似性に基づく発見

- 多数の生物間で保存される領域
モチーフやドメイン
- ある配列特異的に多様化した領域
偽遺伝子、正の淘汰

■ 分子系統解析のソース

各カラムは遺伝子における相同な塩基/アミノ酸残基を示す



+ マルチプルアラインメントプログラム

下記いずれも塩基配列、アミノ酸配列に対応している

- これまでよく使われてきたプログラム
 - ClastalW
- 最近開発された高速かつ高精度なプログラム
 - Clastal Omega
 - MUSCLE
 - MAFFT
 - T-Coffee

MAFFT Server

- MAFFTはコンピュータにインストールとするソフトウェアとして開発されてきたが、ウェブでの使用も可能

MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences



Download version

[Mac OS X](#)

[Windows](#)

[Linux](#)

[Source](#)

Online version

[Alignment](#)

[mafft --add](#)

[Phylogeny](#)

[Rough tree](#)

Merits / limitations

Algorithms

Tips

Benchmarks

Feedback

About

MAFFT is a multiple sequence alignment program for unix-like operating systems. It offers a range of multiple alignment methods, L-INS-i (accurate; for alignment of <~200 sequences), FFT-NS-2 (fast; for alignment of <~10,000 sequences), etc.

Download and Installation

- [Mac OS X](#)
- [Linux](#)
- [Windows](#)
- [Source](#)
- [Changelog](#)

The latest version is 7.122 (2013/Oct/13).

The gap scoring scheme has been changed in version 7.1xx. The new version tends to insert more gaps into gap-rich regions. To disable this change, add the `--legacygappenalty` option.

Input Format

Fasta format. [example1 \(LSU rRNA\)](#), [example2 \(protein\)](#)

The type of input sequences (amino acid or nucleotide) is automatically recognized.





実行

Input:
Paste protein or DNA sequences in fasta format. [Example](#)

or upload a plain text file: ファイルを選択 ファイル未選択

Use structural alignment(s)

Allow unusual symbols (Selenocysteine "U", Inosine "i", non-alphabetical characters)

UPPERCASE / lowercase:

Same as input

Amino acid → UPPERCASE / Nucleotide → lowercase

Direction of nucleotide sequences:

Same as input

Adjust direction according to the first sequence (accurate enough for most cases) Beta

Adjust direction according to the first sequence (only for highly divergent data; extremely slow) Beta

Output order:

Same as input

Aligned

Notify when finished (optional; recommended when submitting large data):

Email address:

Submit

マルチFASTA形式の配列を入力

```
>HIT000251116_Homo. Nanog
MSVDPACPQSLPCFEASDCCKESSPMPVICGPEENPSLQMSSAEMPHETTVSPLPSSMDL
LIQDSDPSSTSPKGKQPTSAEKSVAKKEDKVPVKQKTRTVFSSTQLCVLNDRFQRQKYL
SLQQMQUELSNILNLNSYKQVKTWFQNQRMKSKRWQKNWPKNSNGVTQKASAPTPSLYSS
YHQGCLVNPTGNLPMWSNQTWNNTWSNQTQNIQSWSNHSWNTQTWCCTQSWNNQAWNNSPF
YNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSNMNMQPEDV

>ENSPRT00000008541_Pan. Nanog
MSVDPACPQSLPCFEASDCCKESSPMPVICGPEENPSLQMSSAEMPHETTVSPLPSSMDL
LIQDSDPSSTSPKGKQPTSAENSVTKKEDKVPVKQKTRTVFSSTQLCVLNDRFQRQKYL
SLQQMQUELSNILNLNSYKQVKTWFQNQRMKSKRWQKNWPKNSNGVTQKASAPTPSLYSS
YHQGCLVNPTGNLPMWSNQTWNNTWSNQTQNIQSWSNHSWNTQTWCCTQSWNNQAWNNSPF
YNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQTARYFSTPQTMDLFLNYSTNMQPEDM

>ENSMUST00000112582_Mus. Nanog
MSVGLPGPHSLPSSEEASNNSGNASSMPAVFHPENYSCLQGSATEMLCTEAASPRPSSEDL
PLQGSPDSSTSPKQKLSSPEADKGPEEEENKVLARKQKMRTVFSQAOQLCALKDRFQKQY
LSLQQMQUELSSILNLNSYKQVKTWFQNQRMKCKRWQKNQWLKTSNGLIQKGSAPVEYPSIH
CSYPQGYLVNASGSLSMWGSQTWTNPWTSSQTWTNPWTSSQAWTAQSWNGQPWNAAPLHN
FGEDFLQPYVOLQNFASDLEVNLATESHAHFSTPQALELFLNYSVTPPGEI
...
```

出力される配列の順序を
入力と同一にしたいときに

Submit

結果

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset](#) **New!**

Jalview

Reformat to GCG, PHYLIP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

GUIDANCE computes the residue-wise confidence scores and extracts well-aligned residues. **Beta**

Refine dataset

Phylogenetic tree

MAFFT-L-INS-i Result

CLUSTAL format alignment by MAFFT (v7.122b)

```
HIT000251116_Ho MSVDPACPQSLPCF-EASDCKESSPMPVICGPEENYPSLQMSSAEMPHTEVSPLPSSMD  
ENSPTRT00000008 MSVDPACPQSLPCF-EASDCKESSPMPVICGPEENYPSLQMSSAEMPHTEVSPLPSSMD  
ENSPPYT00000005 MSVDPACPQSLPCF-EASDCKESSPMPMICGPEENHPSLQMSSAEMPHTEVSPLPSSMD  
AB126938_Macaca MSVDPACPQSLPCL-EASDSKESSPMPVICGPEENYPSLQMSSAEMPHTEVSPLPSSMD  
ENSECAT00000013 MSVDPALPQSLPCP-EASNRSRGSSPVPEIYGPEENYASLQMSSAETPHMETVSPLPSSME  
DQ069776_Bos. MSVGPACPQPSLLGP-EASNRSSESSPM-----PEESYVSLQTSSADTLDTTVSPLPSSMD  
ENSCAFT00000022 LSAMPAGPQ-----APNSRDPSPMPEVYGRGNPASLPMSSAETPHAEETVSPLPSSMD  
ENSMUST00000112 MSVGLPGPHSLPSSEEASNSGNASSMPAVFHHP-ENYSCLQGSATEMLCTEAASPRPSSED  
AB162852_Rattus ALILLAIPS-----  
ENSMODT00000022 -----HHSAISPAPSSMD
```

```
HIT000251116_Ho LLIQDSPDSTSSTSPKGKQ-PTSAEKSVAKKEDKVPVKQKTRTVFSSTQLCVLNDRFQRQK  
ENSPTRT00000008 LLIQDSPDSTSSTSPKGKQ-PTSAENSVTKKEDKVPVKQKTRTVFSSTQLCVLNDRFQRQK  
ENSPPYT00000005 LLIQDSPDSTSSTSPKGKQ-PTAAENSATKKEDKVPVKQKARTVFSQAQLCVLNDRFQRQK  
AB126938_Macaca LLIQDSPDSTSSTSPRKVKEPLPASAETSEAKKEEVHGKKQKIRTVFSQTQLCVLNDRFQRQK  
ENSECAT00000013 LLIQDSPDSTSSTSPRKVKEPLPASAETSEAKKEEVHGKKQKIRTVFSQTQLCVLNDRFQRQK  
DQ069776_Bos. LLIQDSPDSTSSTSPRKVLP-SPSVEESTEKEETTVPVKKQKIRTVFSQTQLCVLNDRFQRQK  
ENSCAFT00000022 LLTDQDSPDSTSSTSPRKVLP-PPTSGEERTARKEDATQGKKQKMRTRVFSQTQLYVLNDRFQRQK  
ENSMUST00000112 LPLQGSPDSTSSTSPKQKLSSPPEADKGPEEEENKVLARKQKMRTRVFSQAQLCALKDRFQRQK  
AB162852_Rattus -----LSCLVPRLTRLARRKKTRSSPRKPKMRTRVFSQAQLCALKDRFQRQK  
ENSMODT00000022 KC1QDTPDSATSPTSNLSLSSONKPHTHOGKDOSPIKKPKMRTRVFSOAOLNVLNSRFVEOK
```

- * カラムの全ての配列が一致
- : カラムのアミノ酸残基が保存的
- . カラムのアミノ酸残基がまあまあ保存的

.. : * * *****: ** .*: ** .*: .*

+ FASTAフォーマット

```
>HIT000251116_Homo. hinv
MSVDPACPQSLPCF-EASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMD
LLIQDSPDSSTSPKGKQ-PTSAEKSVAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQK
YLSLQQMQELSNILNLSYKQVKTWFQNQRMKSKRWQKNN-WPKNSNGVTQ-KASAPTYPS
LYSSYHQGCLVNPTGNLPMWSN-----QTWNNSTWSNQTQNIQSWSNHSWNTQTW
CTQSWNNQAWNS----PFYNCGEESLQSCMQFQPNSPASDLEAAL-EAAGEGLNVIQQT
TRYFSTPQTM-DLFLNYSMNMQPEDV

>ENSPTRT0000008541_Pan. ensembl
MSVDPACPQSLPCF-EASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMD
LLIQDSPDSSTSPKGKQ-PTSAENSVTKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQK
YLSLQQMQELSNILNLSYKQVKTWFQNQRMKSKRWQKNN-WPKNSNGVTQ-KASAPTYPS
LYSSYHQGCLVNPTGNLPMWSN-----QTWNNSTWSNQTQNIQSWSNHSWNTQTW
CTQSWNNQAWNS----PFYNCGEESLQSCMQFQPNSPASDLEAAL-EAAGEGLNVIQQT
ARYFSTPQTM-DLFLNYSTNMQPEDM

>ENSPPYT0000005016_Pongo. ensembl
MSVDPACPQSLPCF-EASDCKESSPMPMICGPEENHPSLQMSSAEMPHTETVSPLPSSMD
LLIQDSPDSSTSPKGKQ-PTAAENSATKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQK
YLSLQQMQELSNILNLSYKQVKTWFQNQRMKSKRWQKNN-WPKNSNGVTQ-KASAPTYPS
LYSSYHQGCLVNPTGNLPMWSN-----QTWSNSSWSNQTQNIQSWSNHSWNTQTW
CTQSWNNQAWNS----PFYNCGEESLQSCMQFQPNSPASDLEAAL-EAAGEGLNVIQQT
ARYFSTPQTM-DLFLNYSTNMQPEDM
```

視認性は悪いが、コンピュータ上の解析に使いやすい



結果の可視化

- アラインメントをよりインタラクティブに見たい場合には、アラインメントビュワーを使用するとよい

Jalview

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset](#) **New!**

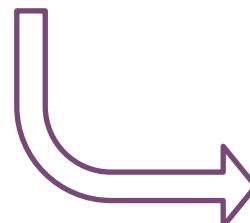
[Jalview](#)

[Reformat](#) to ETE, PHYLIP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

[GUIDANCE](#) computes the residue-wise confidence scores and extracts well-aligned residues. **Beta**

[Refine dataset](#)

[Phylogenetic tree](#)



注意のポップアップが出たら受け入れる

[Jalview](#)

[Start Jalview](#)

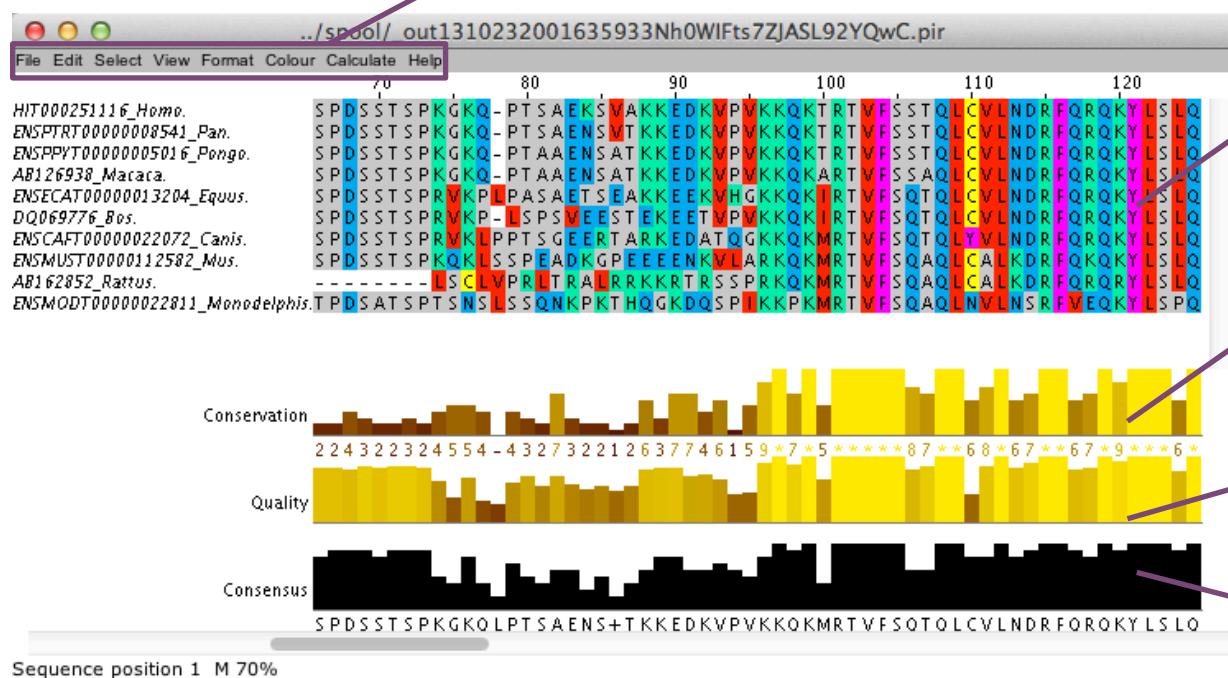
Reference:
[Waterhouse et al. \(2009\)](#)

Jalview home:
<http://www.jalview.org/>

+ Jalview

- OSを選ばず用いられる多機能なアラインメントエディタ
- スタンドアロンのみならず、ウェブブラウザに埋め込める

アラインメントの編集や表示方法の変更が可能



アミノ酸残基の物理化学的性質やカラムの一致度により色分けされる

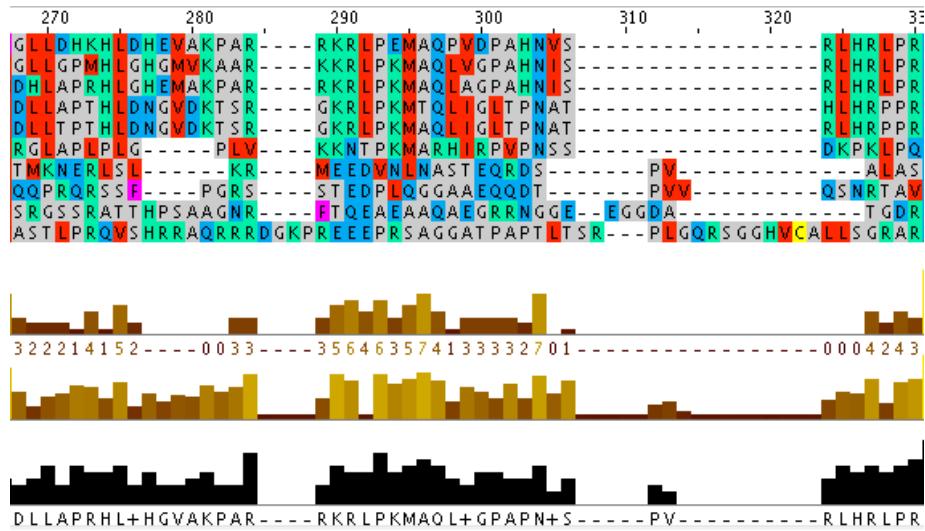
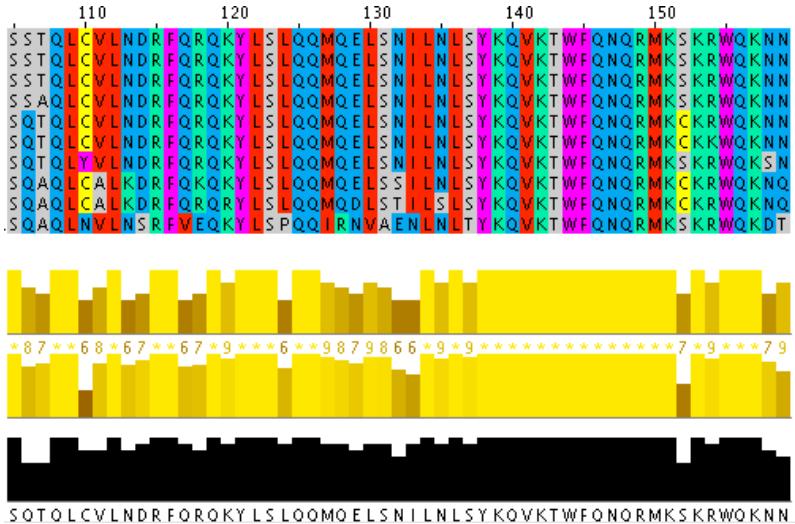
配列の一致度

アラインメントのクオリティー

コンセンサス配列

+ アラインメントの解釈

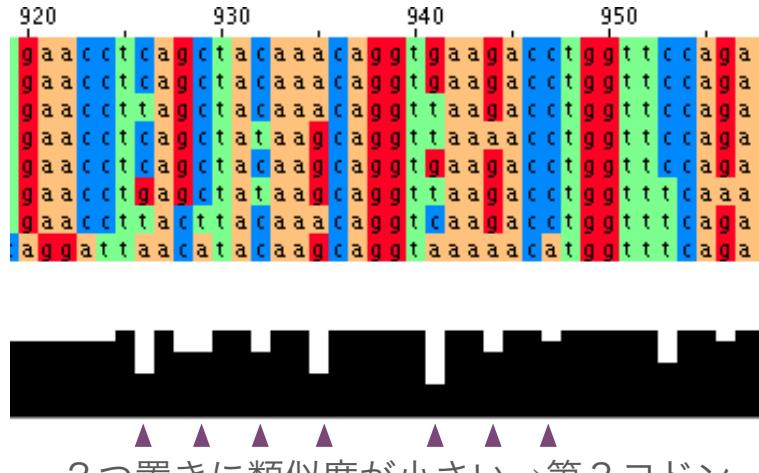
- 保存的かどうか
 - 保存的ならドメインやモチーフを持つ可能性がある





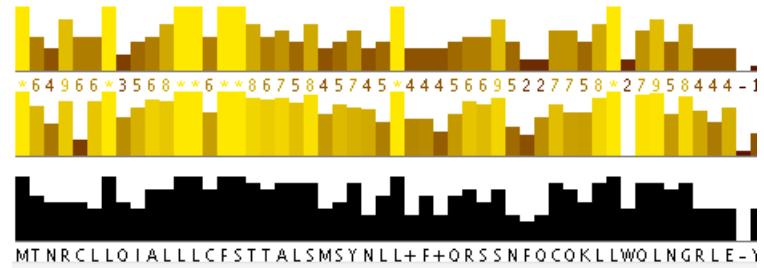
結果の解釈

- 塩基配列でも同様に配列間の保存度を調べられる
- アミノ酸残基の物理化学的性質の保存度
 - 親水性・疎水性



1U 2U 3U 4U 5U

MT NKCCLLQIAALLLCFSTTALSMSYNLLGFLQRSSNFOCQKLLWQLNGRLE-
MT NKCCLLQIAALLLCFSTTALSMSYSNLLGFLQRSSNFOCQKLLWQLNGRLE-
MT NKCCLLQIAALLLCFSTTALSMSYSNLLGFLQRSSSFOCQKLLWQLNGRLE-
MT HRCCLLQMVLLLCFSTTALSRSYSLLRFQQRRSLSAVCQKLLWQLPSTPQ-
MT SRCILQTTLLLYFSTMALAMSNDLLRSQLSSSSSLFCQELLQLQLNGTTE-
MTYRWILPMALLLCFSTTALSVNYDLRLRSQRSSNSACLMLLRQLQLNQAPQ-
MNNRWILHAAFLLCFSTTALSINYKQLQLQERTNIRKCQELLEQLNLNGKIN--
MANRWTLHIAFLLLCFSTTALSIDYKQLQLFQRQSTSIRTTCQKLLRQLQNLGRN--
MVYRGILYLALLLLFSPSISSKGYDSLRFHQRTTNQRSLMFLNKMTGKLHP-



10 20 30 40 50

MT NKCCLLQIAALLLCFSTTALSMSYNLLGFLQRSSNFOCQKLLWQLNGRLE-
MT NKCCLLQIAALLLCFSTTALSMSYSNLLGFLQRSSNFOCQKLLWQLNGRLE-
MT NKCCLLQIAALLLCFSTTALSMSYSNLLGFLQRSSSFOCQKLLWQLNGRLE-
MT HRCCLLQMVLLLCFSTTALSRSYSLLRFQQRRSLSAVCQKLLWQLPSTPQ-
MT SRCILQTTLLLYFSTMALAMSNDLLRSQLSSSSSLFCQELLQLQLNGTTE-
MTYRWILPMALLLCFSTTALSVNYDLRLRSQRSSNSACLMLLRQLQLNQAPQ-
MNNRWILHAAFLLCFSTTALSINYKQLQLQERTNIRKCQELLEQLNLNGKIN--
MANRWTLHIAFLLLCFSTTALSIDYKQLQLFQRQSTSIRTTCQKLLRQLQNLGRN--
MVYRGILYLALLLLFSPSISSKGYDSLRFHQRTTNQRSLMFLNKMTGKLHP-

N端における豊富な疎水性残基→シグナルペプチド

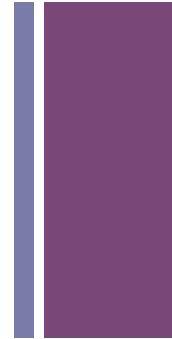
+

相同遺伝子群を取得するには

- マルチプルアランメントには3本以上の配列（相同遺伝子群）が必要
- 配列の類似性を基準に取得
 - 相同性検索で引っかかってきた遺伝子が候補となる
- 候補の範囲
 - 類似度: Bit Score, E-value, オーバーラップ率
 - 配列の長さ
 - 生物学的根拠: 生物の分類群、パラログ
- BLASTの結果から配列を1つ1つ取得するのは大変！



aLeaves



- ゲノムワイドに遺伝子が決定されている動物のアミノ酸配列が格納されており、相同性検索により相同配列を取得できる
- MAFFTと連動しており、MAFFT Serverの機能を用いてBLASTから取得した相同配列を絞りこめる

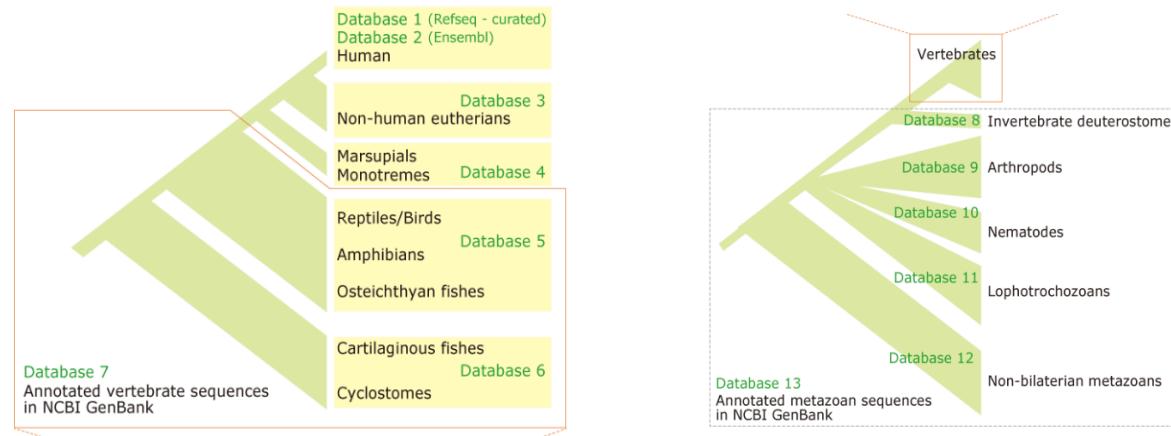
aLeaves - [əlɪfɪvz]

first step to build zoologically informative phylogenetic trees



Available databases

Phylogenetic ranges of available individual databases secondarily compiled are shown in the tree below.



+ BLASTを実行する

Start the search (powered by NCBI Blast)

Enter your "query" peptide sequence in the fasta format (example):

アミノ酸配列を入力

Or upload a file: ファイルを選択 ファイル未選択

Select database (one or more) (number of sequences in parentheses):

- 1. Human - Refseq (36,413) [detail]
- 2. Human - Ensembl (103,918) [detail]
- 3*. Non-human eutherians - Ensembl (743,883) [detail]
- 4. Non-eutherian mammals - Ensembl (83,640) [detail]
- 5. Non-mammalian vertebrates - Ensembl (372,299) [detail]
- 6. Cartilaginous fishes and cyclostomes (57,792) [detail]
- 7*. All vertebrate entries except mammals in NCBI Protein (1,155,700) [detail]
- 8. Invertebrate deuterostomes (135,321) [detail]
- 9. Arthropods (551,253) [detail]
- 10. Nematodes (252,265) [detail]
- 11. Other protostomes (262,588) [detail]
- 12. Non-bilaterian metazoans (cnidarians, placozoan & poriferan) (408,540) [detail]
- 13*. All metazoan entries except vertebrates in NCBI Protein (2,426,697) [detail]

用いる生物群をチェック

(caution: selecting the database marked with * will largely slow down the search)

Number of homologs to collect:

Threshold E-value:

Set low complexity filter or not [help of NCBI Blast]: no yes

取得配列数、E-value、
フィルタリングのオプション





BLAST結果

aLeaves - [əlī:vz]

first step to build zoologically informative phylogenetic trees

Top Database Species History Help About Links

Search finished!

Results

Collected sequences in multifasta format: [Open or download](#)
(File size = 0.008 Mbyte)

List of the collected sequences with database links: [Open or download](#)

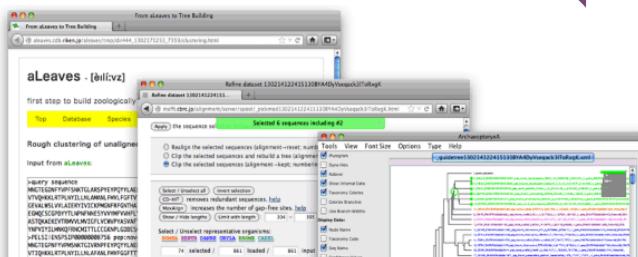
Raw result of the BLAST search: [Open or download](#)

(Caution: The results will be stored under the above links for only 7 days)

[Redo the search](#)

Proceed to tree building (powered by MAFFT server)

[Click here!](#) (open in a new window)



BLASTP 2.2.25 [Feb-01-2011]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,

Species ID	Latin species name	English common name	Category
HOMSA	<i>Homo sapiens</i>	human	Mammals
AILME	<i>Ailuropoda melanoleuca</i>	giant panda	Mammals
BOSTA	<i>Bos taurus</i>	cow	Mammals
CALJA	<i>Callithrix jacchus</i>	marmoset	Mammals
CANFA	<i>Canis familiaris</i>	dog	Mammals
CAVPO	<i>Cavia porcellus</i>	Guinea pig	Mammals
CHOHO	<i>Choloepus hoffmanni</i>	two-toed sloth	Mammals
DASNO	<i>Dasypus novemcinctus</i>	nine-banded armadillo	Mammals
DIPOR	<i>Dipodomys ordii</i>	kangaroo rat	Mammals
ECHTE	<i>Echinops telfairi</i>	small Madagascan hedgehog tenrec	Mammals
EQUCA	<i>Equus caballus</i>	horse	Mammals
ERIEU	<i>Erinaceus europaeus</i>	western European hedgehog	Mammals
FELCA	<i>Felis catus</i>	cat	Mammals
GORGO	<i>Gorilla gorilla</i>	Western lowland gorilla	Mammals
ICTTR	<i>Ictidomys tridecemlineatus</i>	thirteen-lined ground squirrel	Mammals
LOXAF	<i>Loxodonta africana</i>	African elephant	Mammals
MACMU	<i>Macaca mulatta</i>	rhesus monkey	Mammals
MICMU	<i>Microcebus murinus</i>	mouse lemur	Mammals

000000198362649	ECHTE	ENSETEP00000003796	pep:novel scaffold:TENR...	80	le-13
000000162388928	CHOHO	ENSCHOP00000000858	pep:known_by_projection...	79	2e-13
000000516339191	PROCA	ENSPCAP00000014417	pep:known_by_projection...	79	3e-13
000000302218043	LOXAF	ENSLAfp00000010815	pep:known_by_projection...	79	4e-13
000000562668242	SUSSC	ENSSSCP00000005352	pep:known_chromosome:Ss...	78	5e-13
000000543410200	RATNO	ENSRNPO00000063016	pep:known_chromosome:Rn...	77	7e-13

+ Rough treeから配列を選択

aLeaves - [èlívz] → MAFFT version 7

Rough clustering of unaligned sequences (powered by the MAFFT server)

Input from aLeaves:

```
>gi|317419780|emb|CBN81816.1| Homeodomain transcription factor Nanog [Dicentrarchus la  
MADWKSQISYNYNPSPHYAYAVGLVYOPGPEQNHGNLSWGETGVTLNSNNSGVTQAYAA  
TTARAREEPPGSPEQHANGCHYQGSGVVYLGDTQAGRLLSGQHRAAYDARTQEAGR  
AGSDSTSDSEAHSPDSWSGSSREGSLPQADPATWAKQDLDEASSRSPDAGEDVPSSL  
KEEPMPFTVTNEGTDDTATSVHVPLTAPKPKSTTSGNPKAKARAFAFSESQMNALVQRF  
SVQRYLTPAEMKNLAEMTGLTYKQIKTFQNRRMKLRRHQKDTSWVSERYYTNKDPVRG  
TVFTNVPSHIPYQGEARPOLKEHYNQHMMESAFKKTAPQNLAFYLAAMGSAAGSAGYPS  
WSSGPPQTAVPTPQAAWGSMPGISHYEYNPGAFNSANVASAYNTGHDMFSKDVESV  
NSRGSLNTAVHNASQ  
>XIPMA|ENXMAP0000006100 pep:novel scaffold:Xipmac4.4.2|JH556760.1|336167:338568:  
MAEWKTKVTVNYPSPYHTYASLVYQTVPQEQQHGNLPAWDDSGVPDYNCEPENLCAVA
```

Settings (help)

Distance measure:

- Rough (based on frequencies of shared 6-mers; suitable for very large dataset)
- Accurate (based on scores of pairwise global alignment)
- Accurate (based on scores of pairwise local alignment)

Clustering method:

- Average linkage (UPGMA)
- Minimum linkage

入力済のデータは
BLASTの相同配列

- ラフなクラスタリングに基づいた系統樹を表示
- 注意を受け入れて実行ボタンを押す

系統樹上で使用する配列を選択

ArchaeopteryxA

guidetree13102516498465KRzm2TcXsNCP08fwF9vrD.xml

Tools View Font Size Options Type Help

Phylogram
 Dyna Hide
 Rollover
 Show Internal Data
 Taxonomy Colorize
 Colorize Branches
 Use Branch-Widths

Display Data:
 Node Name
 Taxonomy Code
 Seq Name
 Confidence Values

Click on Node to:
Select/Unselect
Zoom:
Y+
F
+
Y-
Back to Super Tree
Order Subtrees
Uncollapse All
Search:

Select / Unselect all Invert selection
CD-HIT removes redundant sequences. [help](#)
Show / Hide lengths Limit with length : 161 - 436

Select / Unselect representative organisms:
DANRE HOMSA MUSMU ORYLA

65 selected / 100 loaded / 100 input
0 / 100 Jump to bottom ↕

1 gi|317419780|emb|CBN81816.1| Homeodomain_transcription_factor_Nanog_Dicentriarchus_lal
 5_GASAC|ENSGACP0000002950 pep:novel_group:BROADS1:groupXXI_3775351_3776625-1
 2_XIPMA|ENSXMAP0000006100 pep:novel_group:Xipmac4.4.2:JH5
 3_TAKRU|ENSTRUP0000032471 pep:novel_scaffold_FUGU4_scaffold_95_962854_965761_1_g
 4_TETNI|ENSTNIP0000007868 pep:novel_chromosome_TETRAODON_174356_176008_-1
 6_MELGA|ENSMGAP0000014747 pep:novel_chromosome_Galgal4_17366745_17571263_1_g
 9_MELGA|ENSGALP0000043037 pep:novel_chromosome_Galgal4_17366745_17571263_1_g
 10_ALIME|ENSMAP0000010793 pep:novel_group:AlM1_18271295_18271296_1_g
 11_PELSI|ENSPSIP0000013731 pep:novel_scaffold_PelSin_1.0_JH207982_1_231450_240970
 12_TAEGU|ENSTGUP0000013513 pep:novel_chromosome_taeGut_3_4_1_18232789_3623501
 13_ANOCA|ENSACAP0000015042 pep:novel_group:AnoCa_18232789_3623501_231450_240970
 14_MONDO|ENSMODP0000022421 pep:known_by_projection_chromosome_BROAD0_8_104628273_1
 15_MACEU|ENSMEUP0000009064 pep:known_by_projection_scaffold_Meug_1.0_Scaffold11284_1233
 16_SARHA|ENSSHAP0000004981 pep:known_by_projection_chromosome_Sarha_1.0_18232789_3623501_231450_240970
 17_TAEGU|ENSTGUP0000007868 pep:novel_scaffold_TENREC_scaffold_196299_4924_8605_1_gene_En
 18_PROCA|ENSPCAP0000144147 pep:known_by_projection_scaffold_proCap_1 scaffold_21276_8680
 19_ECHTE|ENSETEP0000003796 pep:known_by_projection_chromosome_EchTe_1.0_18232789_3623501_231450_240970
 20_CAVPO|ENSPCAP000007893 pep:novel_scaffold_cavpo3_scaffold_2167007_1272849_1_gene_E
 21_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 22_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 23_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 24_CAVPO|ENSPCAP000007893 pep:novel_scaffold_cavpo3_scaffold_2167007_1272849_1_gene_E
 25_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 26_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 27_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 28_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 29_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 30_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 31_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 32_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 33_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 34_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 35_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 36_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 37_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 38_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 39_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 40_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 41_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 42_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 43_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 44_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 45_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 46_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 47_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 48_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 49_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 50_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 51_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 52_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 53_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 54_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 55_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 56_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 57_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 58_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 59_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 60_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 61_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 62_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 63_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 64_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 65_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 66_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 67_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 68_MACEU|ENSMODP0000042937 pep:novel_chromosome_MH01_1_1088932_890502_2_1_gene_E
 69_BOSTA|ENSBTAP0000018597 pep:novel_chromosome_UDMS_1_15_34136078_34139761
 70_TURTR|ENSTTTP0000010595 pep:novel_by_projection_genescuff_turTrl_GensTrf_1
 71_DASHO|ENSDHNP0000005609 pep:novel_by_projection_scaffold_dashN02_scaffold_51726
 72_MACEU|ENSMEEUP0000005057 pep:novel_by_projection_genescuff_Meug_1.0_Scaffold11284_1233
 73_BRAFL|gi_260830242 ref_XP_002610070_1 brain_specific_homeobox_protein_Branchiosoma_flo
 74_BRAFL|gi_260826726 ref_XP_002603316_1 Barth-like_homeobox_protein_Branchiosoma_flo
 75_BRAFL|gi_260826726 ref_XP_002603317_1 H2_0-like_homeobox_protein_Branchiosoma_flo
 76_BRAFL|gi_260826726 ref_XP_002603318_1 hypothetical_protein_Branchiosoma_flo
 77_LIOAF|ENSLAP0000001086 pep:novel_supercoiled_lowAr3_scaffold_74_8924322_892921
 78_MACEU|ENSMODP0000003574 pep:novel_group:MaceU_18232789_3623501_231450_240970
 79_TAKRU|ENSTRUP0000016589 pep:novel_scaffold_FUGU4_scaffold_297_318071_321180_-1
 80_GASAC|ENSGACP0000013281 pep:novel_by_projection_group_BROADS1_groupVIII_9773
 81_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 82_GASAC|ENSGACP0000013281 pep:novel_by_projection_group_BROADS1_groupVIII_9773
 83_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 84_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 85_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 86_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 87_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 88_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 89_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 90_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 91_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 92_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 93_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 94_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 95_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 96_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 97_GASAC|ENSGACP0000013281 pep:novel_chromosome_onyCun_16_32211198_32215778
 98_CALJA|ENSCJAP0000011252 pep:novel_chromosome_C_jacchus_1_23752902_37881
 99_GORG|ENSGGP0000002659 pep:novel_chromosome_gorg_1_201197727_2012030
 100_GORG|ENSGGP0000002659 pep:novel_chromosome_gorg_1_201197727_2012030
 101_YIOLU|ENSLUP0000011342 pep:novel_scaffold_Yioli2_0_GL430022_1526685_152972
 102_PELSI|ENSPSIP0000014571 pep:novel_by_projection_scaffold_PelSin_1.0_JH207982_1_14
 103_PELSI|ENSPSIP0000014571 pep:novel_group_PelSin_1.0_JH207982_1_14
 104_BOSTA|ENSBTAP0000018597 pep:novel_chromosome_UDMS_1_15_34136078_34139761
 105_TURTR|ENSTTTP0000010595 pep:novel_by_projection_genescuff_turTrl_GensTrf_1
 106_DASHO|ENSDHNP0000005609 pep:novel_by_projection_scaffold_dashN02_scaffold_51726
 107_MACEU|ENSMEEUP0000005057 pep:novel_by_projection_genescuff_Meug_1.0_Scaffold11284_1233
 108_MACEU|ENSMEEUP0000005057 pep:novel_group_MaceU_18232789_3623501_231450_240970
 109_BRAFL|gi_260830242 ref_XP_002610070_1 brain_specific_homeobox_protein_Branchiosoma_flo
 110_BRAFL|gi_260826726 ref_XP_002603316_1 Barth-like_homeobox_protein_Branchiosoma_flo
 111_BRAFL|gi_260826726 ref_XP_002603317_1 H2_0-like_homeobox_protein_Branchiosoma_flo
 112_BRAFL|gi_260826726 ref_XP_002603318_1 hypothetical_protein_Branchiosoma_flo
 113_YIOLU|ENSLUP0000011342 pep:novel_scaffold_Yioli2_0_GL430022_1526685_152972
 114_PELSI|ENSPSIP0000014571 pep:novel_by_projection_scaffold_PelSin_1.0_JH207982_1_14
 115_PELSI|ENSPSIP0000014571 pep:novel_group_PelSin_1.0_JH207982_1_14
 116_GALGA|ENSGALP0000043037 pep:novel_chromosome_Galgal4_1:753
 117_TAEGU|ENSTGUP0000013513 pep:novel_chromosome_taeGut3_2.4:
 118_PELSI|ENSPSIP0000013731 pep:novel_scaffold_PelSin_1.0_JH207982_1_14
 119_TAEGU|ENSTGUP0000013515 pep:novel_chromosome_taeGut3_2.4:
 120_PELSI|ENSPSIP0000013422 pep:novel_scaffold_PelSin_1.0_JH207982_1_14
 121_ANOCA|ENSACAP00000015042 pep:known_by_projection_chromosome:
 122_MONDO|ENSMODP0000022421 pep:known_by_projection_chromosome:
 123_MACEU|ENSMEUP0000009064 pep:known_by_projection_scaffold:
 124_SARHA|ENSSHAP00000004981 pep:known_by_projection_scaffold:
 125_ECHTE|ENSETEP00000003796 pep:novel_scaffold:TENREC:scaffold:
 126_TETNI|ENSPCAP00000014417 pep:known_by_projection_scaffold:
 127_LIOAF|ENSLAP0000001086 pep:novel_group_lowAr3_scaffold_74_8924322_892921
 128_ALIME|ENSMAP0000010793 pep:novel_group_AlM1_18271295_18271296_1_g
 129_PELSI|ENSPSIP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 130_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 131_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 132_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 133_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 134_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 135_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 136_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 137_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 138_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 139_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 140_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 141_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 142_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 143_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 144_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 145_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 146_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 147_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 148_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 149_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 150_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 151_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 152_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 153_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 154_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 155_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 156_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 157_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 158_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 159_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 160_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 161_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 162_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 163_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 164_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 165_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 166_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 167_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 168_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 169_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 170_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 171_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 172_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 173_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 174_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 175_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 176_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 177_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 178_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 179_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 180_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 181_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 182_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 183_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 184_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 185_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 186_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 187_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 188_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 189_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 190_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 191_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 192_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 193_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 194_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 195_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 196_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 197_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 198_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 199_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 200_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 201_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 202_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 203_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 204_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 205_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 206_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 207_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 208_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 209_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 210_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 211_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 212_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 213_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 214_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 215_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 216_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 217_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 218_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 219_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 220_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 221_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 222_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 223_GASAC|ENSGACP0000013281 pep:novel_group_BROADS1_groupVIII_9773
 224_GASAC|ENSG

+

選択した配列セットをつくる

選択したデータセットでツリーを描き、再び配列選択画面へ

the sequence selection below and

- Build an alignment (reverse complementary sequences can be generated in the case of DNA)
- Rebuild a tree from the selected sequences
- Align the selected sequences

冗長な配列を取り除く

Select / Unselect all Invert selection
CD-HIT removes redundant sequences. [help](#)
Show / Hide lengths Limit with length : -
Select / Unselect representative organisms:
[DANRE](#) [HOMSA](#) [MUSMU](#)
 selected / loaded / input

+

選択した配列を用いてマルチプル アラインメント

Multiple sequence alignment and NJ / UPGMA phylogeny

New!! (2013/Sep/27)

We made a change in the scoring scheme in version 7.110.

For problems that require many gaps, alignment quality is (expected to be) improved.

For conserved dataset, the difference is small.

Scoring scheme:

- New
- Conventional

Input:

Data 131025171023422CQfEwbemJsVUF20IuXODd (selected from 131025170217459DkMdu0QpzIYlJNwvhB9j;

[View the input data](#)

- Use structural alignment(s)
- Allow unusual symbols (Selenocysteine "U", Inosine "i", non-alphabetical characters, etc.) [Help](#)

UPPERCASE / lowercase:

- Same as input
- Amino acid → UPPERCASE / Nucleotide → lowercase

Direction of nucleotide sequences:

- Same as input
- Adjust direction according to the first sequence (accurate enough for most cases) **Beta**
- Adjust direction according to the first sequence (only for highly divergent data; extremely slow) **Beta**

Output order:

- Same as input
- Aligned

Notify when finished (optional; recommended when submitting large data):

Email address:

+ さらに配列セットを編集

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset](#) [New](#)

Jalview

Reformat to GCG, PHYLIP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

GUIDANCE computes the residue-wise confidence scores and extracts well-aligned residues. **Beta**

Refine dataset

Phylogenetic tree

MAFFT-L-INS-i Result

CLUSTAL format alignment by MAFFT (v7.122b)

```

gi|317419780|em -- MADWKMQI -- SYYNPNPS--HAYAYGLVYQP--GPFQE
GASAC| ENSCAPC00 -- MADWKAQI -- SYYNPNPS--HAYAYGLVQ--GPEPQE
XIPMA| ENSMAPC00 -- MAEWKTQV -- TYYNPNPS--HTAYAQLVTP--VPEFQE
GADMO| ENSCMPO00 -- MAEWKTQL -- SYYNPNPSHHAAAYGLVIVQA--GSEQTQ
TAKRU| ENSTRUP00 -- MADWKTHI -- SCYGNPAY--LAAYAGHMFQ--PQCNA
TETNI| ENSTNP000 -- MADWKTHI -- NYNNYSAY--LAY--VFPQ--PQCNA
DANRE| ENSDAR000 -- MADWKMPV -- SYNFPNPS--HAYAYGLMYQP--SHEVG
GALGA| ENSGALP00 -- MSAHLAMPSS -- MSAHLAMPSS--GS
MELGA| ENSMGP00 -- MSAHLAMPSS -- MSAHLAMPSS--GA
TAEGU| ENSTGUP00 -- MSAHLAMPSS -- MSAHLAMPSS--YPCG
PELSI| ENNSPIS00 -- MSAHLAMPSS -- MSAHLAMPSS--YPAVG
TAEGU| ENSTGUP00 -- MSAHLAMPSS -- MSAHLAMPSS--YPCG
PELSI| ENNSPIS00 -- MSAHLAMPSS -- MSAHLAMPSS--YPAVG
ANCOA| ENSACAP00 -- MSAHLAMPSS -- MSAHLAMPSS--YPCG
MONDO| ENSMODP00 -- MSSCSHSQCQCCSC -- RPKG
MACEU| ENSMEUP00 -- MSSCSHSQCQCCSC -- RPKG
SARHA| ENSHSHA00 -- MTSCLSLQCQCCSS -- OPEG
ECHTE| ENSETEP00 -- MTSCLSLQCQCCSS -- OPEG
PROC4| ENSCAP00 -- MSVDPVASPNL -- LIA
LOXAF| ENSLAFP00 -- MSVDLASPOSL -- PEA
MACMU| ENSMMUP00 -- MSVDPACPOSLP -- CLEA
GORGO| ENSGCP00 -- MSVDPACPOSLP -- CFEA
PANTH| ENSPTRP00 -- MSVDPACPOSLP -- CFEA
HOMSA| ENSP00000 -- MSVDPACPOSLP -- CFEA
PONAB| ENSPPY000 -- MSVDPACPOSLP -- CFEA
NOMLE| ENSNLEP00 -- MSVDPACPOSLP -- CFEA
GORGO| ENSGCP00 -- MSVDPACPOSLP -- CFEA
PONAB| ENSPPY000 -- MSVDPACPOSLP -- CFEA
MACMU| ENSMMUP00 -- RSVDPACPOSLP -- CLEA
CALJA| ENSCJAP00 -- MSMDPVCSQSLP -- CSEA
MACMU| ENSMAP000 -- MSMDPVCSQSLP -- CSEA
OTOGA| ENSCJAP00 -- MSADPACPOCRS -- GPER
CHOMO| ENSMCP00 -- MNTDPSQPCPLP -- CPEA
BUSTA| ENSSGC00 -- MNTDPQAPOCPBP -- CPEA
HUSCC| ENSSGC00 -- MNTDPVQPCPLP -- CPEA
BOSTA| ENSBTPA00 -- MSADPACPOLP -- CSEA
EQUCA| ENSCAP00 -- MSVDPACPOSL -- GPER
AILME| ENSAMEP00 -- MSVDPALPOSLP -- CPEA
MUSPU| ENSMPU00 -- MSVDPAPCOPCB -- GPER
FELCA| ENSFCAP00 -- MNTDPQPCPLP -- CPEA
FELCA| ENSFCAP00 -- MNTDPVQPCPLP -- CPEA
FELCA| ENSFCAP00 -- MSVDPVQPCPLP -- CPEA
CANFA| ENSCAF00 -- LSAMPA -- GPOA
PTEVA| ENSVPA000 -- MSVDPACQSPP -- CPEA
MYOLU| ENSMLUP00 -- MSVDSAYQPSLP -- CSES
MYOLU| ENSMLUP00 -- MSVDSAYQPSLP -- CSES
MYOLU| ENSMLUP00 -- MSVDSAYQPSLP -- CSES
CHOHO| ENSCHOP00 -- MSVDSLASTQSLP -- CPET
AILME| ENSAMEP00 -- MSVDPARPCOPC -- GPER
ORYCU| ENSOCUP00 -- MSVDPACPALS -- ASE
TUBB| ENSTBEP00 -- LSAMPA -- GPOA
CAVPO| ENSCP000 -- MSVAPDCPSSL -- CSEA
OCHPI| ENSOPAP00 -- MSMAAACQCRP -- AMGC
RATNO| ENSPAP000 -- MSVDSLGHPLSPS -- CEEA
MUSPU| ENSMPU000 -- MSVGLPGHPLSPS -- SEEA
BRAFI| GI| 1260826 -- MTAACKLPP--ALSVENLIGARTTDVLDPIRASSPECLP -- PKHNR
BRAFI| GI| 1260826 -- MA -- GN -- HIFSIERLAQSSRRVPEK
BRAFI| GI| 1260826 -- MLEFDQNRCQTNFPAKLREIQLVEQ -- NAMTLQRSLDFGLFA -- PTPTK
OKID0| GSQID00000 -- MDE -- LYGNSSA1SLQIMAKTGTFPEFLTPVQ

```

51317419700 em QRFVSQYRLTAAEPMKLNLAEMTGLTYYQIKTWFQNRMRLRHHQKDTSWSVSERTYNTKNDSP
GASAC ENSGCAP00 QFRFSIQ—
XIPMA ENSMXA00 QFRFSAQ—
GADMO ENSMGMP00 QFRFA—
TAKRU ENSTRUP00 HKFSFQIYRLPESEMKAELAATGTLTYQKVRIDFEMIVPLRRHQKDTSWSVSERYIAHKDTS
TEITNI ENSTNP100 HKFSFQIYRLPESEMKAELAATGTLTYQKVRIDFEMIVPLRRHQKDTSWSVSERTYNTGTS
DANRE ENSDAR00 NFNRFQYRLPESEMKAELAGATGTLTYQKVTWFQNRMRLRHHQKDTSWSMTERYVVNN
GALGA ENSGALP00 RFRQSQYKLSPHQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MELGA ENSMGAP00 RFRQSQYKLSPHQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
TAEGU ENSTGUP00 RFRQSQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
FELSI ENSFSTP00 RFRQSQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
TAEGU ENSTGUP00 RFRQSQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
FELSI ENSFSTP00 RFRQSQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
ANGUA ENSACAP00 QRFQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MONDO ENSMOPD00 GFRFQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MACEU ENSMUPD00 GFRFQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
SARHA ENSHSHA00 SRFEQYKLSPQIRELAAALGLTYQKVTWFQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
ECHTE ENSETEP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
PROC4 ENSCPA00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
LOXAF ENSLAFP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MACMU ENSMUPD00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
GORGO ENSGGOP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
PANTR ENSPTRP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
HOMSA ENSPO0000 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
PONAB ENSPPY00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
NOMLE ENSNLEP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
GORGO ENSGGOP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
PONAB ENSPPY00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MACMU ENSMUPD00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
CALJA ENSCJAP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MACMU ENSMUPD00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
OTOGA ENSOGA00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MICMU ENSMCP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
SUSSC ENSSSCP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
BOSTA ENSBEP00 DRIQRQKSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
DRIQRQKSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
DRIQRQKSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
A1IME ENSAMEP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MUSHU ENSMUPD00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
FELCA ENSFCAP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
FELCA ENSFCAP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
FELCA ENSFCAP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
CANFA ENSCAPP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
PTEVA ENSVPA00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MYOLU ENSLNU00 EFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MYOLU ENSLNU00 EFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MYOLU ENSLNU00 EFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
CHOHO ENSCHOP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
A1IME ENSAMEP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
ORYCU ENSCUP00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
TUBEI ENSTEBP00 ARFQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
CAVFO ENSCP000 RFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
OCHPR ENSORP00 RFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
RATNN ENSRNPO00 RFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
MUSMU ENSMUPD00 DFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
BRAFL g_1260826 EFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
BRAFL g_1260826 EFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
BRAFL g_1260826 TRIFQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
BRAFL g_1260821 KFTOKYLSLAAERQQLRIGLTQDQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—
PKHOU ENSGTR00 RFRQOKYLSLSPQIMESELVNLYSKYQVKTFWQNRMRLRHHQKDTSWSVSERTYNTGKYLPP—

- 保存ドメインにも関わらず欠けてしまつた配列が2本
→重要ではない配列ならば取り除く

• * *

+ さらに配列セットを編集

Refine dataset

Click any node on the tree in the ArchaeopteryxA window.
Then the descendants of the node are selected / unselected in this page.
If no ArchaeopteryxA window appears, enable Java and [reload this page](#).

sequence selection below and

Realign the selected sequences. Reverse complementary sequences can be generated in the case of DNA. (alignment→reset; numbering→reset)
 Clip the selected sequences and rebuild a tree (alignment→kept; numbering→reset)
 Clip the selected sequences (alignment→kept; numbering→kept)

全選択に → Select / Unselect all
 removes redundant sequences. [help](#)
 increases the number of gap-free sites. [help](#)
 Limit with length : -

Select / Unselect representative organisms:
[DANRE](#) [HOMSA](#) [MUSMU](#)

57 selected / 59 loaded / 59 input
0 / 59

要らない配列の
チェックボックスを外す →

<input checked="" type="checkbox"/> 1	gi 317419780 emb CBN81816.1 Homeodomain transcription factor Nanog [Dicentrarchus labrax]
<input type="checkbox"/> 2	GASAC ENSGACP0000002950 pep:novel group:BROADS1:groupXXI:3775351:3776625:-1 gene:ENSGACG00000002275 transcript:ENSGACT00
<input checked="" type="checkbox"/> 3	XIPMA ENSMXMAP0000006100 pep:novel scaffold:Xipmac4.4.2:JH556760.1:336167:338568:1 gene:ENSMXMA00000006084 transcript:ENSM
<input type="checkbox"/> 4	GADMO ENSGMOP0000001893 pep:novel genescaffold:gadMor1:GeneScaffold_1975:25333:26467:-1 gene:ENSGM0G0000001786 transcript:ENSGM0G0000001786
<input checked="" type="checkbox"/> 5	TAKRU ENSTRUP00000032471 pep:novel scaffold:FUGU4:scaffold_95:962854:965761:-1 gene:ENSTRUG00000012825 transcript:ENSTRUG00000012825
<input checked="" type="checkbox"/> 6	TETNI ENSTNIP0000007868 pep:novel chromosome:TETRAODON8:6:574356:576008:-1 gene:ENSTNIG00000005192 transcript:ENSTNIT000
<input checked="" type="checkbox"/> 7	DANRE ENSDARP00000102376 pep:known chromosome:Zv9:24:12738148:12743225:1 gene:ENSDARG00000075113 transcript:ENSDART00001
<input checked="" type="checkbox"/> 8	GALGA ENSGALP00000043037 pep:known chromosome:Galga14:1:75366745:75371263:-1 gene:ENSGALG00000027772 transcript:ENSGALT00
<input checked="" type="checkbox"/> 9	MELGA ENSMGAP00000014747 pep:novel chromosome:UMD2:1:79355894:79358232:-1 gene:ENSMGAG00000013941 transcript:ENSMGAT00000
<input checked="" type="checkbox"/> 10	TAEGU ENSTGUP00000013513 pep:novel chromosome:taeGut3.2.4:1:88224132:88226482:1 gene:ENSTGUG00000013121 transcript:ENSTGUG00000013121
<input checked="" type="checkbox"/> 13	PELSI ENSPSIP00000013731 pep:novel scaffold:Pelsin_1.0:JH207983.1:235450:240970:-1 gene:ENSPSIG00000012334 transcript:ENSPSIG00000012334
<input checked="" type="checkbox"/> 11	TAEGU ENSTGUP00000013515 pep:novel chromosome:taeGut3.2.4:1:88232783:88235011:-1 gene:ENSTGUG00000013123 transcript:ENSTGUG00000013123
<input checked="" type="checkbox"/> 12	PELSI ENSPSIP00000013422 pep:novel scaffold:Pelsin_1.0:JH207983.1:219310:230273:1 gene:ENSPSIG00000012067 transcript:ENSPSIG00000012067
<input checked="" type="checkbox"/> 14	ANOCA ENSCACP00000015042 pep:known_by_projection chromosome:AnoCar2.0:2:80930150:80939167:-1 gene:ENSCACG00000015317 transcript:ENSCACG00000015317
<input checked="" type="checkbox"/> 15	MONDO ENSMODP00000022421 pep:known_by_projection chromosome:BROAD05:8:104628273:104648421:-1 gene:ENSMODG000000017979 transcript:ENSMODG000000017979
<input checked="" type="checkbox"/> 16	MACEU ENSMUEP00000009064 pep:known_by_projection scaffold:Meug_1.0:Scaffold11284:12337:13800:-1 gene:ENSMUEG00000009949 transcript:ENSMUEG00000009949
<input checked="" type="checkbox"/> 17	SARHA ENSSHAP0000004981 pep:known_by_projection scaffold:DEVIL7:0:GL861820.1:100942:104884:-1 gene:ENSSHAG00000004362 transcript:ENSSHAG00000004362

アラインし直すか
元のアラインメントを用いるか



編集の結果

gi|317419780|em QRFSVQRYLTPAEMKNLAEMTGLTYQIKTFQNRRMKLRRHQKDTSWSERYYTTNKDSP
GASAC|ENSGACP00 QRFSI-----
XIPMA|ENSXMAP00 QRFSVQRYLTPAEMKNLADMTGLTYKQVRHDFEMIVIPLRRHQKDTSWSERYAIHKDTS
GADMO|ENSGMOP00 QRFSAQ-----
TAKRU|ENSTRUP00 HKFSIQRYLPPSEMKNLAEVTGLTYK-----QNRRMKLRRHQKDSSWISERYAINNGAP
TETNI|ENSTNIP00 HKFSVQRYLPPAEMKNLAEVTGLTYK-----QNRRMKLRRHQKDTSWVSERYTINNGTS
DANRE|ENSDARP00 NRFNVQRYLTPAEMKTLAGATGLTYKQVKTFQNRRMKLRRHQDSSWMTERYVVN-----
GALGA|ENSGALP00 QRFSQSKYLSHQIRELAAALGLTYKQVKTFQNRRMKFRCQKEQSOWVDKG1YLP-----
MELGA|ENSMGP00 QRFSQSKYLSHQIRELAAALGLTYKQVKTFQNRRMKFRCQKEQSOWVDKG1YLP-----
TAEGU|ENSTGUP00 QRFSQSKYLSHQIRELAAALELTYKQVKTFQNRRMKFRCQKEQSOWMDKGMYLP-----
PELSI|ENSPSIP00 QRFSQSKYLSHQIRELGSALGLTYKQVKTFQNRRMKFRCQKEQTWMEEKGTCLS-----
TAEGU|ENSTGUP00 QRFSQSKYLSHQIRELAVALGLTYKQVKTFQNRRMKLRCRQSLWSERAQCLT-----
PELSI|ENSPSIP00 QRFSQSKYLSHQIRELALALGLTYKQVKTFQNRRMKLKTYYEKLHNLSERAQYLM-----
ANOCA|ENSACAP00 HRFNQNKYLSQQIRELAAALDTYKQIKTFQNRRMKFRTQKESLWLRKGMCP-----
MONDO|ENSMODP00 SRFVEQKYLSHQIRVNAENLNLTYKQVKTFQNRRMKSKRWQDTMTKNGNRNV-----
MACEU|ENSMUPO0 SRFVEQKYLSHQIRVNAENLNLTYKQVKTFQNRRMKSKRWQDTMTKNGNRNV-----
SARHA|ENSSHAP00 SRFLEQKYLSHQIRVNAENLNLTYKQVKTFQNRRMKSKRWQDSMSWSKNSNNMV-----
ECHTE|ENSETEP00 DHFQKQKYLSQLQMOELSEVLNLNSYKQVKTFQNRRMKCKRWKQTTNWNSNNHVT-----
PROCA|ENSPCAP00 DRFQKQKYLSQLQMOELSEVLNLNSYKQVKTFQNRRMKCKRWKQNTTWSKNGSGFT-----
LOXAF|ENSLAFTP00 DRFQKQKYLSQLQMOELSEALNLNTYKQVKTFQNRRMKCKRWKQNTTWSKNSNSVT-----
MACMU|ENSMMP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
GORGO|ENSGGPO00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
PANTR|ENSPTRP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
HOMSA|ENSP00000 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
PONAB|ENSPFPY00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
NOMLE|ENSNLEP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
GORGO|ENSGGPO00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
PONAB|ENSPFPY00 DRFQKQKYLSQLQTOESENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
MACMU|ENSMMP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGVT-----
CALJA|ENSCJAP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPRTSDVT-----
MACMU|ENSMMP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NPWPKNNSNGSG-----
OTOGA|ENSOGAPO0 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NWLNKNTNSMT-----
MICMU|ENSMICP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKSKRWQKN-----NWLNKNSNSLT-----
SUSSC|ENSSSCP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQKN-----HWPRNSNSV-----
BOSTA|ENSBTAP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKKWQKN-----NPWPNSNSGMP-----
EQUCA|ENSECAPO0 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKKWQKN-----NPWPRTNSVT-----
AILME|ENSEAMEP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQKN-----HPWKEKSVT-----
MUSPU|ENSPMP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQKN-----HPWKEKSVT-----
FELCA|ENSFCAPO0 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQKN-----NPWPNNNTVS-----
FELCA|ENSFCAPO0 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQKN-----NPWPNNNTVS-----
FELCA|ENSFCAPO0 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQNS-----NCPIH--NSVI-----
CANFA|ENSCAFC00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQNS-----NPWPESNSVT-----
PTEVA|ENSPVAP00 ERFQKQKYLSQLQMOELSENLNLNSYKQIKTFQNRRMKCKRWKQKN-----NPWPNSNCAM-----
MYOLU|ENSLMP00 ERFQKQKYLSQLQMOELSENLNLNSYKQIKTFQNRRMKCKRWKQKN-----NPWPNSHRVT-----
MYOLU|ENSLMP00 ERFQKQKYLSQLQMOELSSILNLNSYKQIKTFQNRRMKCKRWKQY-----NPWPNSHRVT-----
MYOLU|ENSLMP00 ERFQKQKYLSQLQMOELSSILNLNSYKQIKTFQNRRMKCKRWKQY-----NPWPNSHRVT-----
CHOHO|ENSCHOP00 DRFQKQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCCKWQKN-----NPWPNSNSV-----
AILME|ENSEAMEP00 DRFQKQKYLSQLQMOELSHNLNSYKQVKTFQNRRMKCCKRQNS-----NPWPESNSVT-----
ORYCU|ENSOCUP00 DRFQKQKYLSQLQMOELSHNLNSYKQVKTFQNRRMKCCKRQNS-----NPWPNICNSVT-----
TUPBE|ENSTBEP00 ARFQKQKYLSQLQMOELSDILNLNSYKQVKTFQNRRMKCKRWKQKN-----KPWNCSAVT-----
CAVPO|ENSCPOP00 DRFQKQKYLSQLQMOELSSILNLNSYKQVKTFQNRRMKCCKRWKQNS-----TGLONGAGAP-----
OCHPR|ENSPRPO0 DRFQKQKYLSQLQMOELSSILNLNTYKQVKTFQNRRMKCCKWQRN-----NPNNN---MI-----
RATNO|ENSRNPO0 DRFQKQYRLSQLQMOQDLSTILNLNSYKQVKTFQNRRMKCCKRWKQRN-----QWLTSNSGLT-----
MUSMU|ENSMUSP00 DRFQKQYKLSQLQMOELSSILNLNSYKQVKTFQNRRMKCCKRWKQRN-----QWLTSNSGLI-----
BRAFL|gi|260826 ERFLRQKYLSSAADDRETAKATGLTDEQVKTFQNRRMKLKRQQQDFATFPLHASVT-----
BRAFL|gi|260826 ERFLRQKYLSSAADDRETAKATGLTDEQVKTFQNRRMKLKRQQQDFATFPLHASVT-----
BRAFL|gi|260826 TRFSTQKYLSSAADDRELAHALDLTDQAQVKTFQNRRMKWKRQVDAESEARRFCMF-----
BRAFL|gi|260801 KRFQKQYLLAAEERQELAQRIGLTDTQVKTFQNRRMKWKRQQDEAQTLQALSMSLH-----
OIKDI|GSOIDT000 NRFRQYRLTPSDRDNVADNLGLTSTQVITWFQNRRAKLKRDKDEASPCKQKGRKS-----
*: * *

gi|317419780|em VQRYLTPAEMKNLAEVTGLTYQIKTFQNRRMKLRRHQKDTSWSERYYTTNKDSPVRGT
XIPMA|ENSXMAP00 VQRYLTPAEMKNLADMTGLTYQVKRHDDEMIVIPLRRHQKDTSWSERYAIHKDTSIHRP
TAKRU|ENSTRUP00 IQRYLPPSEMKNLAEVTGLTYK-----QNRRMKLRRHQKDSSWISERYAINNGAPVNGT
TETNI|ENSTNIP00 VQRYLPPAEMKNLAEVTGLTYK-----QNRRMKLRRHQKDTSWVSERYTINNGTSVQGT
DANRE|ENSDARP00 VQRYLTPAEMKTLAGATGLTYQVKTFQNRRMKLRRHQDSSWMTERYVVN-----
GALGA|ENSGALP00 SQKYLSPHQIRELAAALGLTYKQVKTFQNRRMKFRCQKEQSOWVDKG1YLP-----
MELGA|ENSMGP00 SQKYLSPHQIRELAAALGLTYKQVKTFQNRRMKFRCQKEQSOWVDKG1YLP-----
TAEGU|ENSTGUP00 SQKYLSPHQIRELAAALELTYKQVKTFQNRRMKFRCQKEQSOWMDKGMYLP-----
PELSI|ENSPSIP00 SQKYLSPQQIRELGLTYQVKTFQNRRMKFRCQKEQTSWMEKGTCLS-----
TAEGU|ENSTGUP00 SQKYLSPQQIRELAVALGLTYQVKTFQNRRMKLKRQCKQSLWSERAOCLT-----
PELSI|ENSPSIP00 SQKYLSTQEMLEALATGLTYQVKTFQNRRMKLKTYYEKHNLWSERAQYLM-----
ANOCA|ENSACAP00 NQKYLSPQQIRELAAALDTYQIKTFQNRRMKFRTQKESLWLRKGMCP-----
MONDO|ENSMODP00 EQKYLSPQQIRVNAENLNLTYQVKTFQNRRMKSKRWQDKDTMWTKNGNRNV-----
MACEU|ENSMUPO0 EQKYLSPQQIRVNAENLNLTYQVKTFQNRRMKSKRWQDKDTMWTKNSKGSSM-----
SARHA|ENSSHAP00 EQKYLSPQQIRVNAENLNLTYQVKTFQNRRMKSKRWQDKDSWSKNSNNMV-----
ECHTE|ENSETEP00 EQKYLSPQQQMELSEVNLNSYKQVKTFQNRRMKSKRWQDKTSWNSKNNHVT-----
PROCA|ENSPCAP00 KQKYLSPQQQMELSEVNLNSYKQVKTFQNRRMKCKRWKQTTNWNSNNHVT-----
LOXAF|ENSLAFTP00 KQKYLSQLQMQELSEALNLNTYKQVKTFQNRRMKCKRWKQNTTWSKNSNSVT-----
MACMU|ENSMMP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
GORGO|ENSGGPO00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
PANTR|ENSPTRP00 RQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
HOMSA|ENSP00000 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
PONAB|ENSPFPY00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
NOMLE|ENSNLEP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
GORGO|ENSGGPO00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
PONAB|ENSPFPY00 RQKYLSQLQTOESENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
MACMU|ENSMMP00 RQKYLSQLQMOELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
CALJA|ENSCJAP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSVT-----
MACMU|ENSMMP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSSG-----
OTOGA|ENSOGAPO0 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKTSNSMT-----
MICMU|ENSMICP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
SUSSC|ENSSSCP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
BOSTA|ENSBTAP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
EQUCA|ENSECAPO0 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
AILME|ENSEAMEP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
MUSPU|ENSPMP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
FELCA|ENSFCAPO0 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
FELCA|ENSFCAPO0 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
FELCA|ENSFCAPO0 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
CANFA|ENSCAFC00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
PTEVA|ENSPVAP00 RQKYLSQLQMQELSENLNLNSYKQVKTFQNRRMKCKRWKQNTNWSKNSNLTO-----
MYOLU|ENSLMP00 RQKYLSQLQMQELSSILNLNSYKQIKTFQNRRMKCKRWKQY-----NPWPNSHRVT-----
MYOLU|ENSLMP00 RQKYLSQLQMQELSSILNLNSYKQIKTFQNRRMKCKRWKQY-----NPWPNSHRVT-----
MYOLU|ENSLMP00 RQKYLSQLQMQELSSILNLNSYKQIKTFQNRRMKCKRWKQY-----NPWPNSHRVT-----
CHOHO|ENSCHOP00 RQKYLSQLQMQELSSILNLNSYKQVKTFQNRRMKCCKWQKN-----NPWPNSNSV-----
AILME|ENSEAMEP00 RQKYLSQLQMQELSSILNLNSYKQVKTFQNRRMKCCKWQKN-----NPWPESNSVT-----
ORYCU|ENSOCUP00 RQKYLSQLQMQELSSILNLNSYKQVKTFQNRRMKCCKRQNS-----NPWPNICNSVT-----
TUPBE|ENSTBEP00 RQKYLSQLQMQELSSILNLNSYKQVKTFQNRRMKCCKRQNS-----KPWNCSAVT-----
CAVPO|ENSCPOP00 RQKYLSQLQMQELSSILNLNSYKQVKTFQNRRMKCCKRQNS-----TGLONGAGAP-----
OCHPR|ENSPRPO0 RQKYLSQLQMQELSSILNLNTYKQVKTFQNRRMKCCKWQRN-----NLPNN---MI-----
RATNO|ENSRNPO0 RQKYLSQLQMQELSSILNLNTYKQVKTFQNRRMKCCKWQRN-----QWLTSNSGLT-----
MUSMU|ENSMUSP00 RQKYLSQLQMQELSSILNLNSYKQVKTFQNRRMKCCKRQNS-----QWLTSNSGLI-----
BRAFL|gi|260826 LQKYLSSAADDRETAKATGLTDEQVKTFQNRRMKLKRQQQDFATFPLHASVT-----
BRAFL|gi|260826 LQKYLSSAADDRETAKATGLTDEQVKTFQNRRMKLKRQQQDFATFPLHASVT-----
BRAFL|gi|260826 TQKYLSSAEEERQELAQRIGLTDQVKTFQNRRMKWKRQQDEAQTLQALSMSLH-----
BRAFL|gi|260801 TQKYLSSAEEERQELAQRIGLTDQVKTFQNRRMKWKRQQDEAQTLQALSMSLH-----
OIKDI|GSOIDT000 RQYLTSPDSDRDNVADNLGLTSTQVITWFQNRRAKLKRDKDEASPCKQKGRKS-----
*: ** : * : * : :

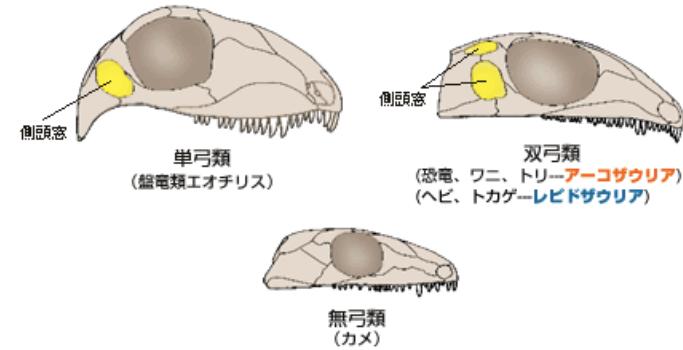
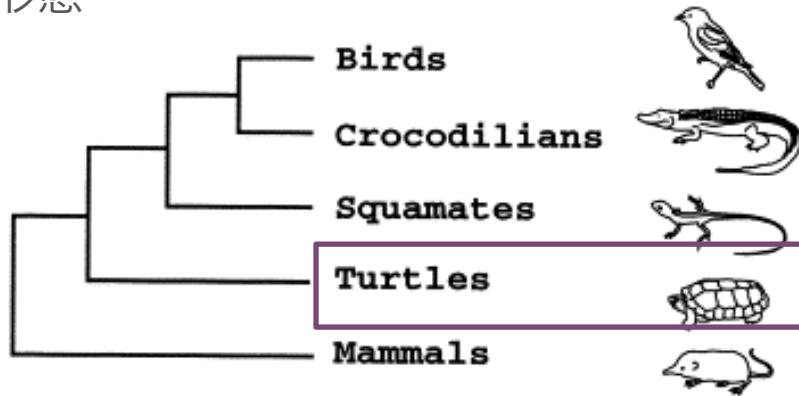
+

分子系統解析

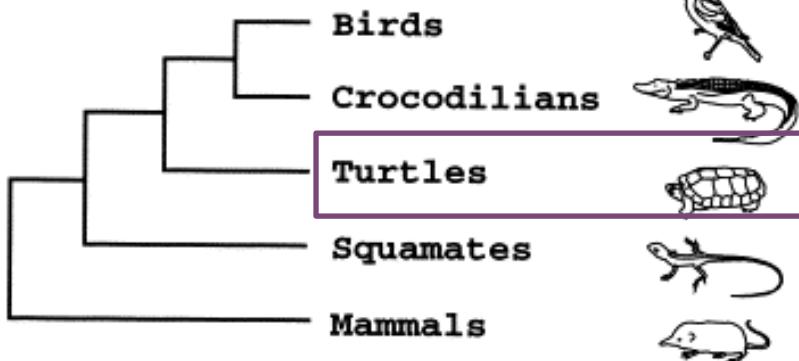
- 塩基配列やアミノ酸配列を用いて種の進化や遺伝子の進化を推定する
- 配列情報を用いて種や分類群の系統関係を精査する
- 遺伝子の系統樹と種の系統樹を照らし合わせることにより、遺伝子進化をより詳しく理解できる
 - 遺伝子重複
 - 自然選択の検定
 - 祖先配列の復元
 - 偽遺伝子化

+ 分子系統解析から種の進化を知る

形態

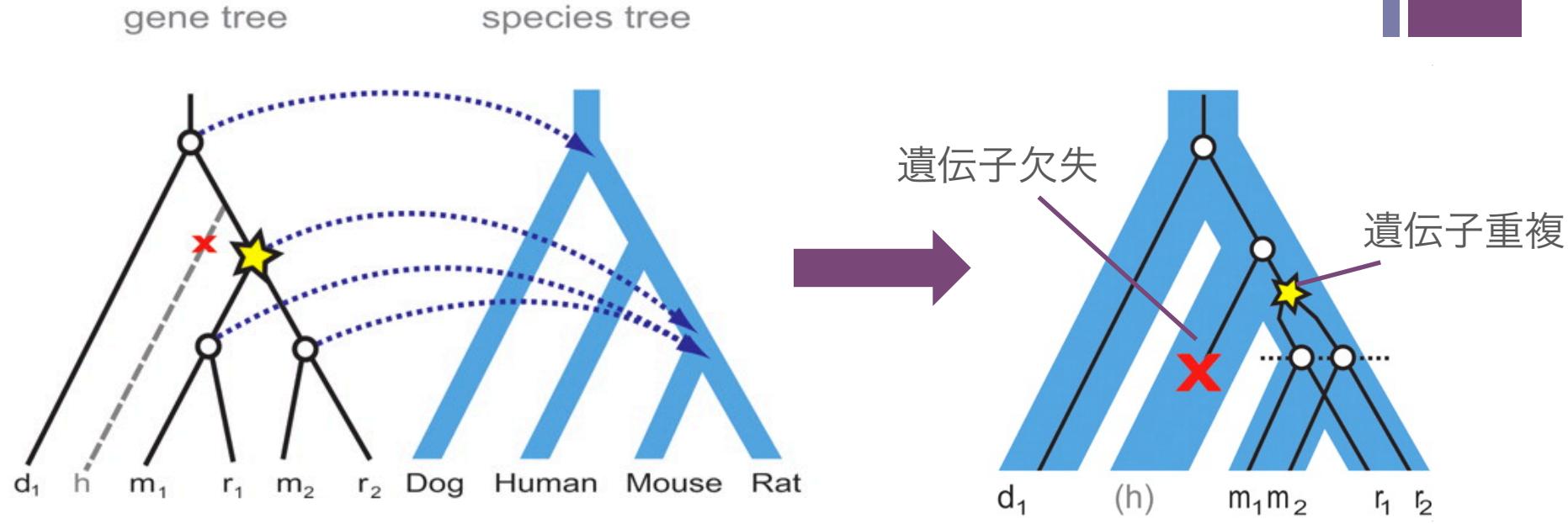


分子



遺伝子を用いた網羅的系統解析

+ 分子系統解析から遺伝子の進化を知る



Rasmussen and Kellis, 2007

- オーソログ 種分岐によって分かれた遺伝子
m₁はr₁のオーソログ
- パラログ 遺伝子重複によって分かれた遺伝子
m₁はr₂のパラログ

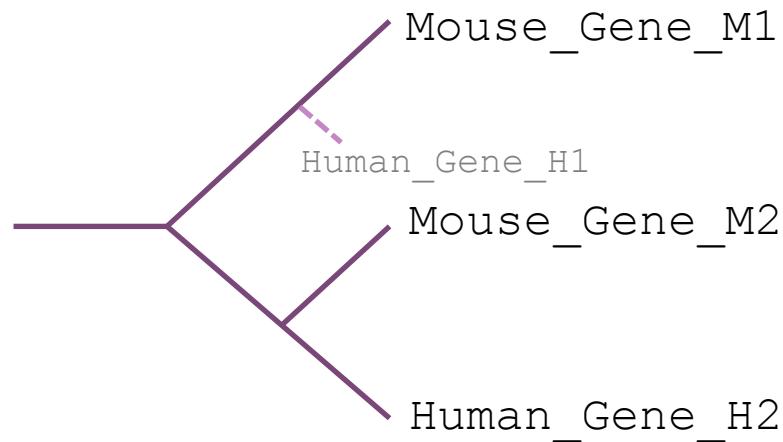
+ 相同性検索だけで十分？？

Query = Mouse_Gene_M1

Subject	bit score	E-value
Mouse_Gene_M1	360	2e-111
Mouse_Gene_M2	265	3e-70
Human_Gene_H2	260	8e-64

...

ヒト遺伝子のTop Hit
→H2がM1のオーソログ？



H2はM2のオーソログ
M1のオーソログH1は
欠失した

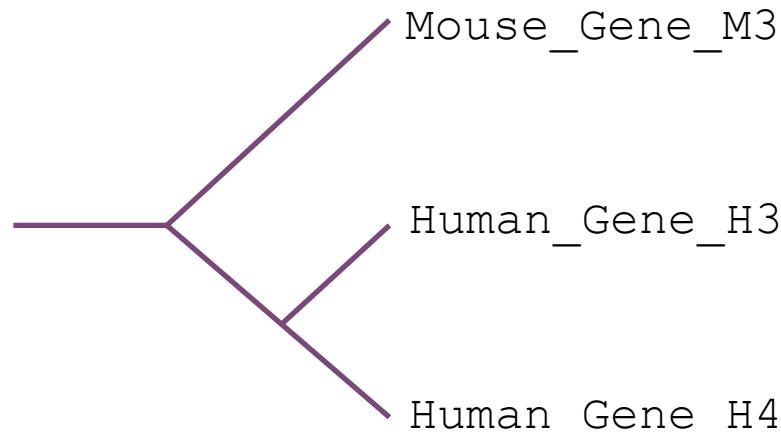
+ 相同性検索だけで十分？？

Query = Mouse_Gene_M3

Subject	bit score	E-value
Mouse_Gene_M3	360	2e-111
Human_Gene_H3	265	3e-70
Human_Gene_H4	260	8e-64

...

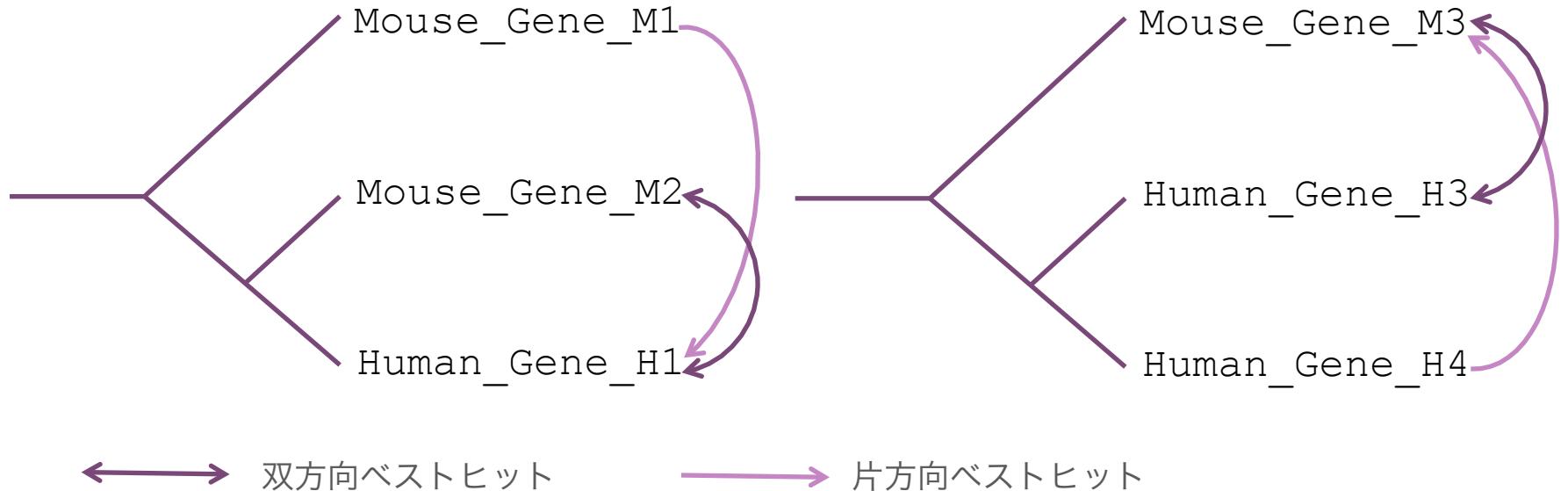
ヒト遺伝子のTop Hit
→H3がM3のオーソログ？



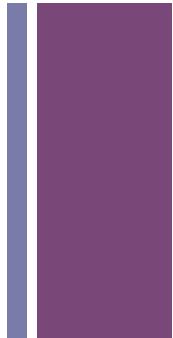
H3には系統特異的重複
遺伝子H4があり、H4も
H3のオーソログである

+ 双方向ベストヒット

- BLASTにおいて互いにベストヒットになる遺伝子の組み合わせ
- 簡易的なオーソログ同定に用いられる
- 複雑な系統関係を再現できない（複雑なのが当たり前）



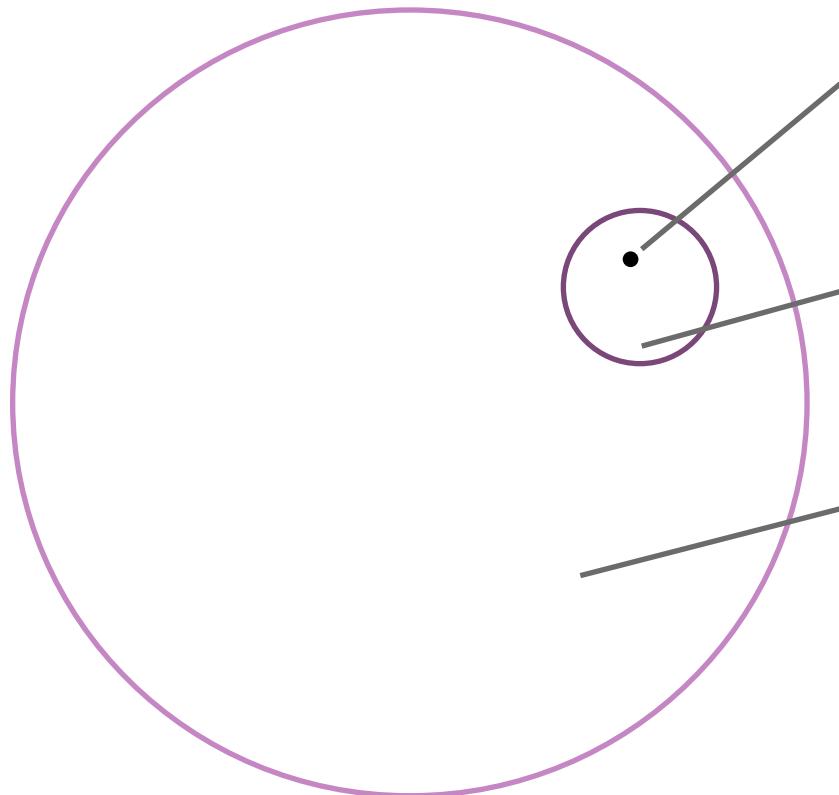
+ 分子系統解析の手順



- マルチプルアラインメント
- アラインメントが確からしい領域の選択
 - ギャップが無い、ミスマッチが頻発しすぎていない
- 系統樹推定
 - 近隣結合法: 最小進化する系統樹を準最適に求める。計算が速くマシンパワーを必要としない
 - 最尤法: 仮定する進化モデルに最適な系統樹を求める。多大な計算時間をするが、高速なアルゴリズムも開発されている。
RAxML: 現在広く用いられている高速かつ色々な進化モデルが選べる最尤法系統樹推定ソフトウェア

+

系統樹がとりうるトポロジーの数



近隣結合法: 発見的に1つのトポロジーを求める

最尤法（発見的）: 真の樹形に近い限られた範囲の樹形を比較して尤もらしい樹形を選択する

最尤法（全探索）: 全ての樹形を比較して尤もらしい樹形を選択する

配列数 $3 \rightarrow$ 樹形数 1

$$4 \rightarrow 3$$

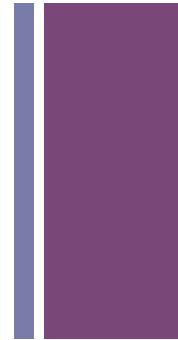
$$5 \rightarrow 15$$

$$10 \rightarrow 2,027,025$$

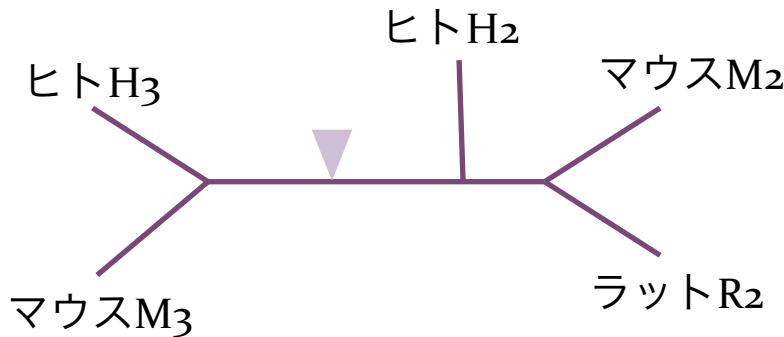
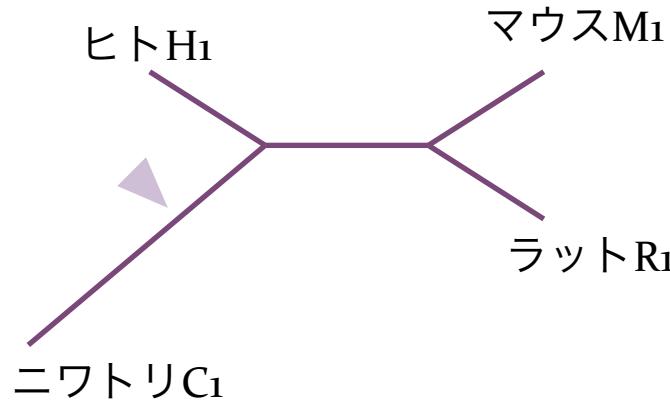
$$n \rightarrow (2n-5)!!$$

+

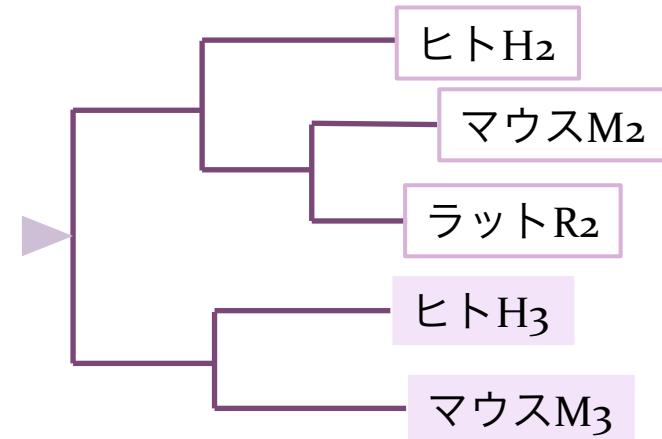
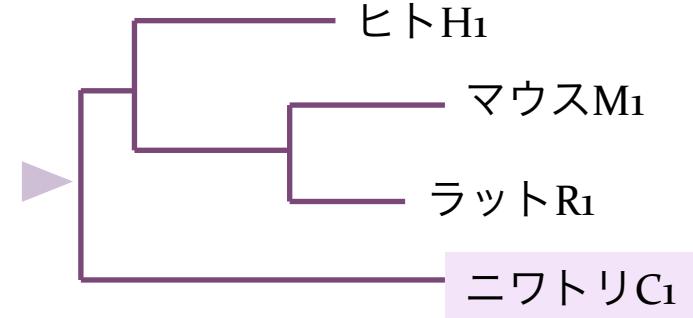
有根系統樹と無根系統樹



推定されるのは無根系統樹



外群を置き進化を反映した「根(root)」を与えると有根系統樹になる





近隣結合法の系統樹推定

■ MAFFTサーバにて実行

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset *New!*](#)

[Jalview](#)

[Reformat](#) to GCG, PHYLIP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

[GUIDANCE](#) computes the residue-wise confidence scores and extracts well-aligned residues. *Beta*

[Refine data](#) 

[Phylogenetic tree](#)

Given alignment

CLUSTAL format alignment by MAFFT (v7.122b)

gi|317419780|em MADWKSQISYNYNPSYHAYAGLVYQPGPEQNHNGLSSWGETGVTDLNSNYNSGVTOQAYYA
TAKRU|ENSTRUP00 MADWKTHISYGYNPQAYLAYAGHMFQPG---QNAGN---QGEPEVTDLSSNNVGVTONYYT
DANRE|ENSDARP00 MADWKMPVSYNFNPSPYHAYAGLMLYFQVSEHGPVLNS-WPDAAYT----HSGGVTAGYFT
BRAFL|gi|260826 -----
GALGA|ENSGALP00 MS-----
MONDO|ENSMODP00 MS-----
HOMSA|ENSP00000 MS-----
MUSMU|ENSMUSP00 MS-----

gi|317419780|em TTARARESPPGSPEQHAANGHCHYQGSGVVLGDTQAGRLLLSSQHRAAYDARTQEAGR
TAKRU|ENSTRUP00 TTAETQEGSPPRTEPHQVSNNGHYHQSGAVLYIDATQTSRFLLAGSRQPVYDERALETKR
DANRE|ENSDARP00 AQTA---QSPPWSPENGGASSTYSQ----YPGHSQNQRLFLS-YNKTEPDQKAKDAEQ
BRAFL|gi|260826 -----APLPNL-----
GALGA|ENSGALP00 AHLA----MPSY---GSVRCHGYYWP-----SFGSMDSSAEEAPA
MONDO|ENSMODP00 SCSSHQQCQSCR-PKGDAASRPYYTWI-----SASQPESLTSP-SRY
HOMSA|ENSP00000 VDPACPQSLPCF-EASDKC-----ESSPPMVICGPEEENY
MUSMU|ENSMUSP00 VGLPGPHSLPSSEEASNSG-----NASSMPAVFHFP-ENY

gi|317419780|em AGSDSTSDESAHTSPDSWSSGSSREGSLPQADPATWAKQDLDDDEASSRSPDAGE-DV---
TAKRU|ENSTRUP00 AGSDSTSDESETYSP-----DLKETSNKSPVASE-DV---
DANRE|ENSDARP00 TSSDTPSDSEAH-T-PDSWSSASSREG-VPLTNLNLPSPWRDRDYETDSGSPDSGERNL---
BRAFL|gi|260826 SAVTLRSDEGETS-----EDEEIDDVSESSSCD-DL-Q
GALGA|ENSGALP00 ADLSLTTEQKTPC-----HPDASPASSSSGTLIQ
MONDO|ENSMODP00 SNCHIQSPGKVQP-----NTAISPAPSSMDKCIQ
HOMSA|ENSP00000 -PSLQMSSAEMPH-----TETVSPLPSSMDLLIQ
MUSMU|ENSMUSP00 -SCLOGSATEMLC-----TEAASPRPSSEDLPLQ
*: :

ヨーロピアンシーバス
フグ
ゼブラフィッシュ
ナメクジウオ
ニワトリ
オボッサム
ヒト
マウス



近隣結合法の系統樹推定

■ MAFFTサーバにて実行

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset](#) **New!**

NJ or UPGMA tree (β)

8 sequences, 536 total sites, 162 gap-free sites, 114 conserved sites

Settings

Method:

- NJ -- Conserved sites (114 aa)
- NJ -- All gap-free sites (162 aa)
- Average linkage (UPGMA) -- All sites (536 aa and gaps)
- Minimum linkage -- All sites (536 aa and gaps)

NJのいずれかを選ぶ（基本的にはConserved sites）

Substitution model (valid when NJ is selected):

- JTT
- WAG **Alpha**
- Poisson

WAGの方が新しいアミノ酸置換モデル

Heterogeneity among sites (valid when JTT or WAG is selected):

- Ignore ($\alpha = \infty$)
- Estimate
- Specify; $\alpha =$ (0.10 - 5.0)

サイトごとに置換率がばらつくモデル。 α を推定することが望ましい

Bootstrap (valid for NJ):

- On
- Number of resampling: (5 - 1000)
(The number of sequences must be <1000 for Poisson model, or <100 for other models.)

急ぎで無いならブートストラップ試行回数は1000回に



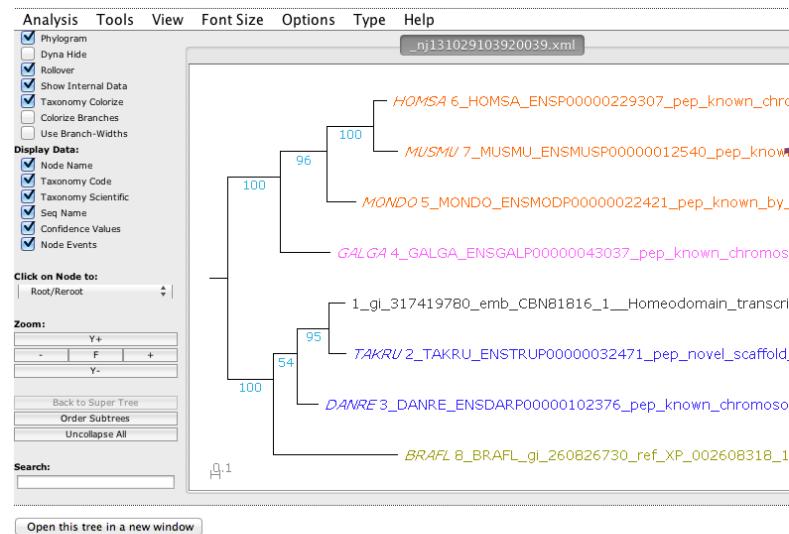
近隣結合法の系統樹推定

■ MAFFTサーバにて実行

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset](#) **New!**

[Refine dataset on this tree](#) **New!! (2012/Dec/17)**

Result



Archaeopteryx home:
<https://sites.google.com/site/cmzmasek/home/software/archaeopteryx>

References for Archaeopteryx:

- Zmasek and Eddy (2001)
- Han and Zmasek (2009)

Reference for CSDI (generalized speciation duplication inference):

- Zmasek and Eddy (2001)

Size = 8 sequences × 114 sites

Method = Neighbor-joining

Model = JTT

Alpha = 0.82 (estimated)

Bootstrap resampling = 100

Alignment id = 1310291016155426eXPdqORhMk62oejbKAcW

[Tree file \(Newick format\)](#)

系統樹ビュワーArchaeopteryx
による表示

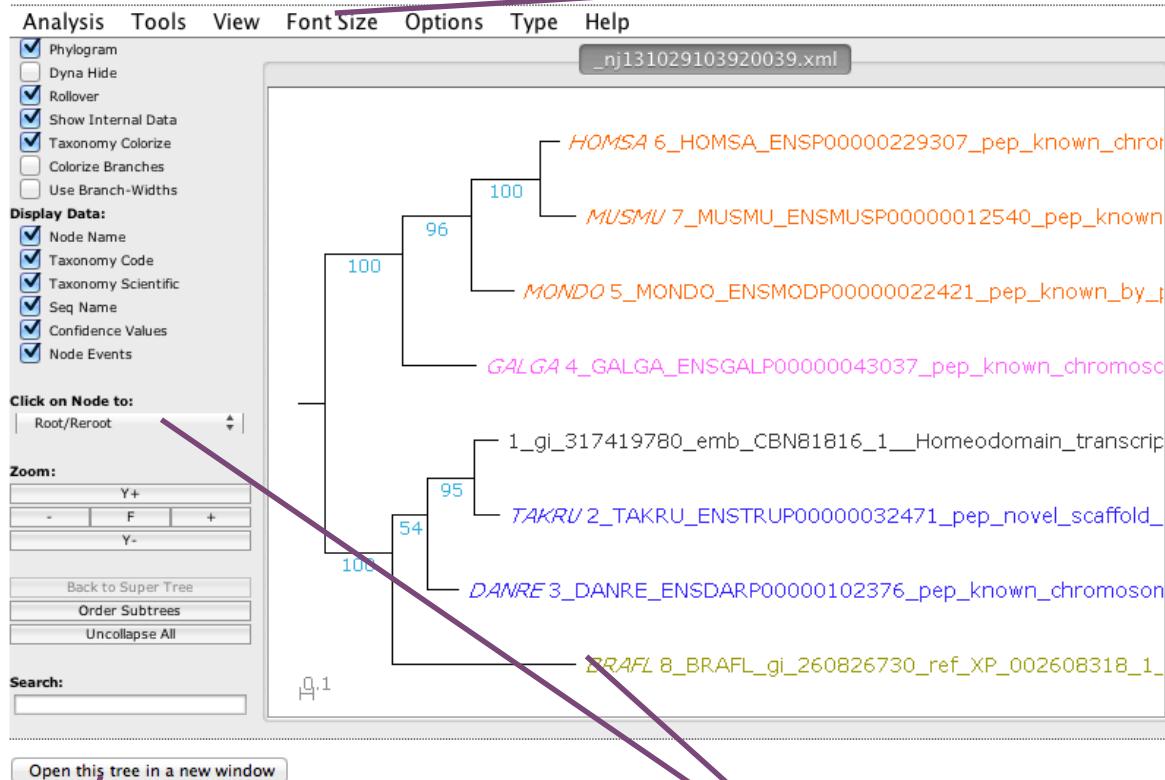
Newickフォーマットによる
系統樹のテキスト表示

```
(( 1_gi_317419780_emb_CBN81816_1__Homeodomain_transcripti
on_factor_Nanog__Dicentrarchus_labrax :0.1661,
2_TAKRU_ENSTRUP00000032471_pep_novel_scaffold_FUGU4_scaff
old_95_962854_965761_-1_gene_ENSTRUG00000012825_transcrip
t_ENSTRU00000032595_gene_biotype_protein_coding_transcrip
t_b :0.1757) :0.3413[&&NHX:B=95],
3_DANRE_ENSDARP00000102376_pep_known_chromosome_Zv9_24_12
738148_12743225_1_gene_ENSDARG00000075113_transcript_ENSD
ART00000114762_gene_biotype_protein_coding_transcript_bio
type :0.2242,
...)
```



Archaeopteryx

Result



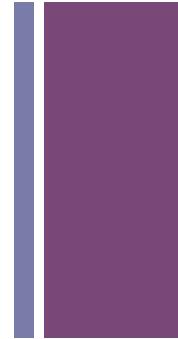
Large fontに

ポップアップとしてウイン
ドウを広げられる

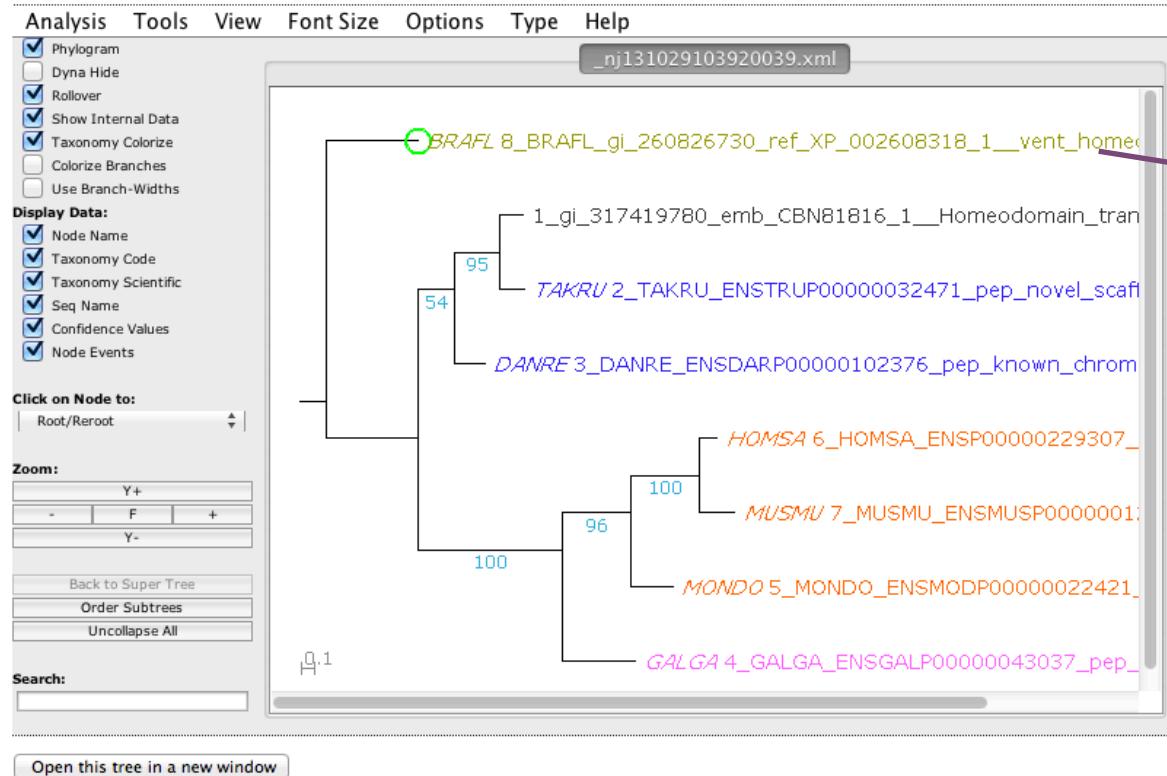
Root/Rerootを選び
枝の先を選択



Archaeopteryx



Result



ナメクジウオを外群に置くことにより、根をつけられる

+ RAxML

- Swiss Instituteが提供しているRAxML BlackBoxを使う
- アライメントフォーマットにPhylipを使用しているので、コンバーターを用いてFASTAから変換する。

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [Jalview](#) | [Tree](#) | [Refine dataset](#) **New!**

[Jalview](#)

[Reformat](#) to GCC, PHYLIP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

[GUIDANCE](#) computes the residue-wise confidence scores and extracts well-aligned residues. **Beta**

[Refine dataset](#)

[Phylogenetic tree](#)

Given alignment

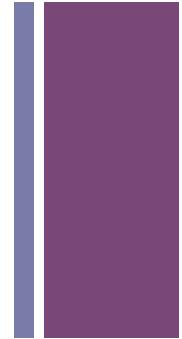
CLUSTAL format alignment by MAFFT (v7.122b)

```
gi|317419780|em MADWKSQISYNYNPSYHAYAYGLVYQPGPEQNHGNLSSWGETGVTDLSNYNSGVTQAYYA  
TAKRU|ENSTRUP00 MADWKTHISYGYNPAYLAYAYGHMFQPG--QNAGN---QGEPEVTDLSSNNVGVTQNYYT  
DANRE|ENSDAR00 MADWKMPVSYNFNPSYHAYAYGLMYPQVSEHGVPNLS-WPDAAYT---HSGGVTAGYFT  
BRAFL|gi|260826 -----  
GALGA|ENSGALP00 MS-----  
MONDO|ENSMODP00 MS-----  
HOMSA|ENSP00000 MS-----  
MUSMU|ENSMUSP00 MS-----
```

Submit Reset

Options	
Output sequence format:	<input type="button" value="Phylip Phylip4"/>
Return biosequence data:	<input type="radio"/> Download to file <input checked="" type="radio"/> View in browser
Change sequence case to:	<input checked="" type="radio"/> No change <input type="radio"/> lower <input type="radio"/> UPPER
Select	<input checked="" type="radio"/> all, or <input type="radio"/> sequences by number: <input type="text"/>
<input type="checkbox"/> Remove gap symbols: - <input type="checkbox"/> Calculate checksum of sequences	
Select <input checked="" type="radio"/> all, or <input type="radio"/> sequences by number: <input type="text"/>	
<input type="checkbox"/> Translate bases (list as from-base:to-base pairs) <input type="text"/>	

+ Phylip format



8 536

gi 3174197	MADWKSQISY	YNPNSYHAYA	YGLVYQPGPE	QNHGNLSSWG	ETGVTDLSNY
TAKRU ENST	MADWKTHISY	GYNPAYLAYA	YGHMFQPG--	QNAGN---QG	EPEVTDLSSN
DANRE ENSD	MADWKMPVSY	NFNPSYHAYA	YGLMYPQVSE	HGVPNLS-WP	DAAYT---H
BRAFL gi 2	-----	-----	-----	-----	-----
GALGA ENSG	MS-----	-----	-----	-----	-----
MONDO ENSM	MS-----	-----	-----	-----	-----
HOMSA ENSP	MS-----	-----	-----	-----	-----
MUSMU ENSM	MS-----	-----	-----	-----	-----
	NSGVTQAYYA	TTARAREESP	PGSPEQHAAN	GHCHYQGSGV	VYLGDTQAGR
	NVGVTQNYYT	TTAETQEGSP	PRTPEHQVSN	GHYHYQSAGV	LYIDATQTSR
	SGGVTAGYFT	AQTA---QSP	PWSPENGGAS	STYSQ-----	-YPGHSQNGR
	-----	-----	-----	-----	-----
	-----	AHLA---MP	SY---GSVRC	GHYYWP-----	-----
	-----	SCSHSQQCQS	CR-PKGDASR	PYYTWI-----	-----
	-----	VDPACPQSLP	CF-EASDCK-	-----	-----
	-----	VGLPGPHSLP	SSEEASNSG-	-----	-----

+ RAxML Black Box

RAxML BlackBox

Data

Paste your sequence alignment

Must be in **Phyip** format, or convertible by
[reduced](#)

or upload a file

ファイルを選択 ファイル未選択

Gamma model of rate heterogeneity

If not checked, CAT model will be used

Protein sequences

If not checked, dna sequences will be assumed

Outgroup

The name of one or more (comma-separated) outgroup(s) can be specified

Constraint

ファイルを選択 ファイル未選択

This option allows you to specify an incomplete or comprehensive multifurcating constraint tree in NEWICK format. More help [here](#).

Binary backbone

ファイルを選択 ファイル未選択

This option allows you to pass a binary/bifurcating constraint/backbone tree in NEWICK format. More help [here](#).

Phyip形式のアライメントを入力

タンパク配列を用いる場合にチェック

Substitution Matrix

Substitution matrix

WAG

タンパク配列を用いる場合に進化モデルを選択

Use empirical base frequencies

Model

Use a mixed/partitioned model?

ファイルを選択 ファイル未選択

You may provide an alignment partition file. More help [here](#)

Maximum likelihood search

Will search for best-scoring tree after the bootstraps

最尤系統樹を出力したいのでチェック

Notification

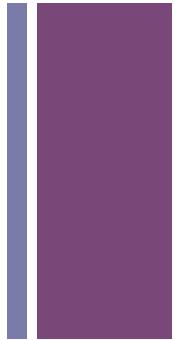
Email

Enter your email if you want to be notified upon job completion

データが大きく時間がかかるときには E-mailでアラートしてもらうと便利

Run

+ RAxML Black Box



RAxML BlackBox

100 bootstrapped trees

[Download](#), [View](#).

Consensus tree

[Download \(MRE\)](#), [View \(MRE\)](#).

[Download \(MR\)](#), [View \(MR\)](#).

[Download \(STR\)](#), [View \(Strict\)](#).

Best-scoring ML tree

Tree with branch lengths only: [Download](#), [View](#).

Tree with branch lengths and support values: [Download](#), [View](#).

Job report

[Download](#), [View](#).

ビュワーは使い物にならないので、系統樹データをダウンロードして
スタンドアロンのArchaeopteryxで表示させるのがよい。

+

課題

- 別ページの課題 2 を行います。

+

もっと簡単に相同遺伝子群を作りたい！

- オーソログデータベース
 - あらかじめ同定されたオーソログ群を格納したデータベース
 - 遺伝子名検索や相同性検索により、目的の相同遺伝子群を見つけられる
 - アラインメントや分子系統樹も表示してくれることが多い
 - 遺伝子データは最新のバージョンではないことがほとんどで、生物種も限られる
 - 自動化アルゴリズムでオーソログを同定しているので、正解ではないことがある
- オーソログデータベース
 - あらゆる生物種: PhylomeDB, EggNog, OrthoMCLDB
 - 動物: Ensembl、Evola
 - 微生物: MBGD
 - etc.

Evola (<http://hinv.jp/evola/>)

■ 脊椎動物を対象にしたオーソログデータベース

The screenshot displays the Evola database interface, which includes:

- Left Panel (Ortholog):** Shows orthologous gene alignments across various species. A green box highlights the entry for Human (Homo sapiens) with ID HIT000000011 (AB002303) ZFYVE16, which encodes a Zinc finger FYVE domain-containing protein 16 (Endofin) (Endosome-associated FYVE domain protein). A red box highlights the alignment details.
- Middle Panel (Similarity Search Tool):** Provides a search interface for orthologs and gene families. A green box highlights the Ortholog section, showing alignments for Chimp, Macaque, Mouse, and Rat. A pink box highlights the Similarity Search Tool section.
- Top Right Panel (ATV | A Tree Viewer):** Displays a phylogenetic tree of the Zinc finger FYVE domain-containing protein 16 orthologs. The tree shows relationships between Human, Chimpanzee, Macaque, Mouse, Rat, Dog, Horse, Cow, and Zebrafish. Node numbers (e.g., 1000, 498, 713, 992, 721) are indicated next to branches. A pink arrow points from the tree viewer towards the sequence alignment panel.
- Bottom Right Panel (Sequence Alignment):** Shows the amino acid sequence alignment for the Zinc finger FYVE domain-containing protein 16 orthologs. The sequence is shown in FASTA format with gaps indicated by dashes. A pink box highlights the sequence alignment area.

+

他のマルチプルアラインメント・分子系統解析方法は？

- グラフィカルなソフトウェア
 - MEGA
- コマンドライン (主にMac OSX, Linux)
 - アラインメント法も系統樹推定法も色々ある
 - それらを組み合わせて試したい条件ごとに行うことが可能
 - 少しのプログラミング技術があれば、ゲノムプロジェクトのような数千～数万のオーソログ群を用いた分子系統推定も自動的に実行可能

MEGA

WindowsやMac上で動作する
視覚的に分かりやすい
GUIインターフェース
(コマンドを打つ必要がない)

多彩なオプション

インターネットに接続せずに
クローズドな環境で利用可能。
未公開配列の解析に便利
(配列相同性検索を除く)



<http://www.megasoftware.net/>

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011)
MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood,
Evolutionary Distance, and Maximum Parsimony Methods.
Molecular Biology and Evolution 28: 2731-2739.

遺伝子レベル、タンパク質
モジュールレベルの
配列解析に便利

無料
(ダウンロードの際に
メールアドレスと名前を入力)

1993年にVersion1が公開
されて以来、バージョンアップ
継続中。

強力な機能を持つ統合配列解析環境のGUIソフトウェア

+ 気をつけること

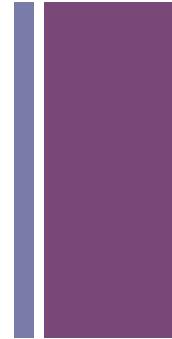
- いろいろ便利なツールはあるが、最終的に重要なのはユーザの生物学的感性
- データから生物学的に新規なことを見いだすことも、データ自体のおかしなところを見いだすことも必要
- 解析ソフトウェアはおかしなデータからも「とりあえず」結果を出すので、常に目視で確認してデータがまともか確認すること

+ おかしなアラインメントにはウラがある（ことが多い）

1塩基挿入によるフレームシフト

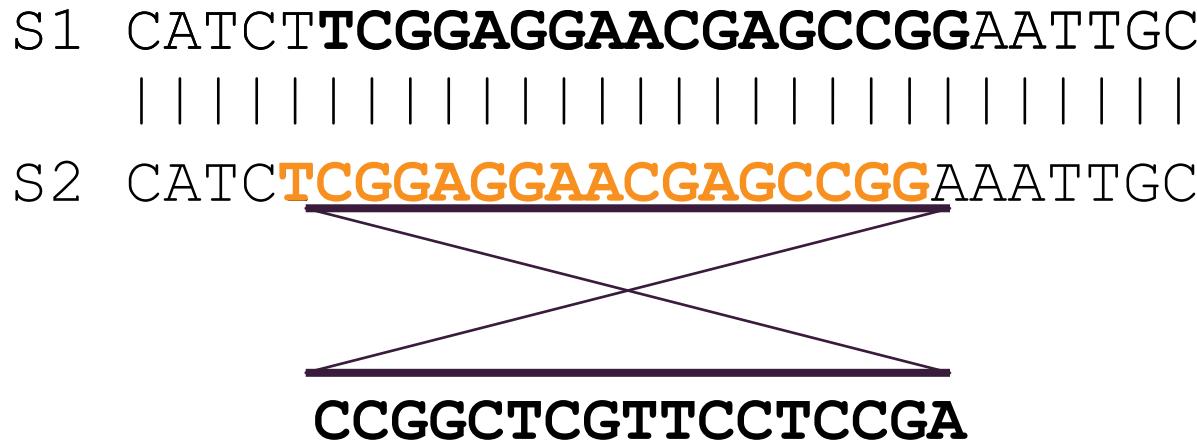


S1	CTCTTTCTGTGAGACCTGGAGCATTAAAGCTCTGGCCTCTCATTCTCCTTGT S2	CTCTTTCTGTGAGACCTGGAGCATTAAAGCTCTGGCCTCTCATTCTCCTTGT
S1	GAAAGGAAGGTTGTACTTGTATTGTAAGATATCTCACTGTTCTTATATACACAAAA S2	GAAAGGAAGGTTGTACTTGTATTGTAAGATATCTCAC-GTTCTTATATACACAAAA
S1	TGAGGGATTCTCTAACATCTCAGTTACTTACCCATGAGACTCTCGAAC S2	TGAGGGATTCTCTAACATCTCAGTTACTTACCCATGAGACTCTCGAAC
S1	TTACTCTGTGATTGGAAAAGAGTTCACACTTGCTCGCCTCAAGCATGTTAGCTCAGT S2	TTACTCTGTGATTGGAAAAGAGTTCACACTTGCTCGCCTCAC-GCATGTTAGCTCAGT
S1	CTTTATCAGCAACTCTCTTTGGAGCAAATTCACTGGTCCAGATTGCCTGTTACGCC S2	CTTTATCAGCAACTCTCTTTGGAGCAAATTCACTGGTCCAGATTGCCTGTTACGCC
S1	CAGGCAGGAAAGCTTATTCATCTATGTGAGTGATAAGGATCTAGAAAGAGTTCTGA S2	CAGGCAGGAAAGCTTATTCATCTATGTGAGTGATAAGGATCTAGAAAGAGTTCTGA
S1	TTGGTGTGCTTGTAACTGCGGCAGGATTCCATTACCACTTAGTCAGCTGGTAATGA S2	TTGGTGTGCTTGTAACTGCGGCAGGATTCCATTACCACTTAGTCAGCTGGTAATGA
S1	GGGTCTCTGTACCAGGCAACTGTGGCTGCCATCTGGCGGCCAGCACACGCCCGCCC S2	GGGTCTCTGTACCAGGCAACTGTGGCTGCCATCTGGCAAGGCCAGCACACGCCCGCCC
S1	GGCAGACAGAACCAATTAGTCGGTCCCCACACCTCCCACAAGGCCTTCCTGTTGGTGG S2	GGCAGACAGAACCAATTAGTCGGTCCCCACACCTCCCACAAGGCCTTCCTGTTGGTGG
S1	CAGAGTGTAGCCCCGCTGAGCCTCCACTAGAAACCTGTTGAGTTCTGGGAGGGGAG S2	CAGAGTGTAGCCCCGCTGAGCCTCCACTAGAAACCTGTTGAGTTCTGGGAGGGGAG





Ultramicro inversion (超微小逆位)



- Such small-scale inversions (**Ultramicro Inversions**) may be erroneously forwardly aligned in local alignments.
- These inversions have not been focused though they have the potential for clarifying evolution of genome structures at minimum level.

RESEARCH ARTICLE

Open Access

Abundance of Ultramicro Inversions within Local Alignments between Human and Chimpanzee Genomes

Yuichiro Hara and Tadashi Imanishi*



Grants
JSPS KAKENHI Grant
Number 24657168

科 研 費
KAKENHI



最後に

- 多くの生物学研究には、あまり気づかれないですが、進化学的視点が含まれています
 - 相同性検索
 - モデル生物、種間比較
- 分子系統樹を用いた解析は、研究に内在する進化学的視点を鮮明にし、思考に時間軸という幅を持たせてくれます
- 時間に質問できなかった方は後ほどにでもお越しください
- あるいは@Amiacalvaまでメンションください