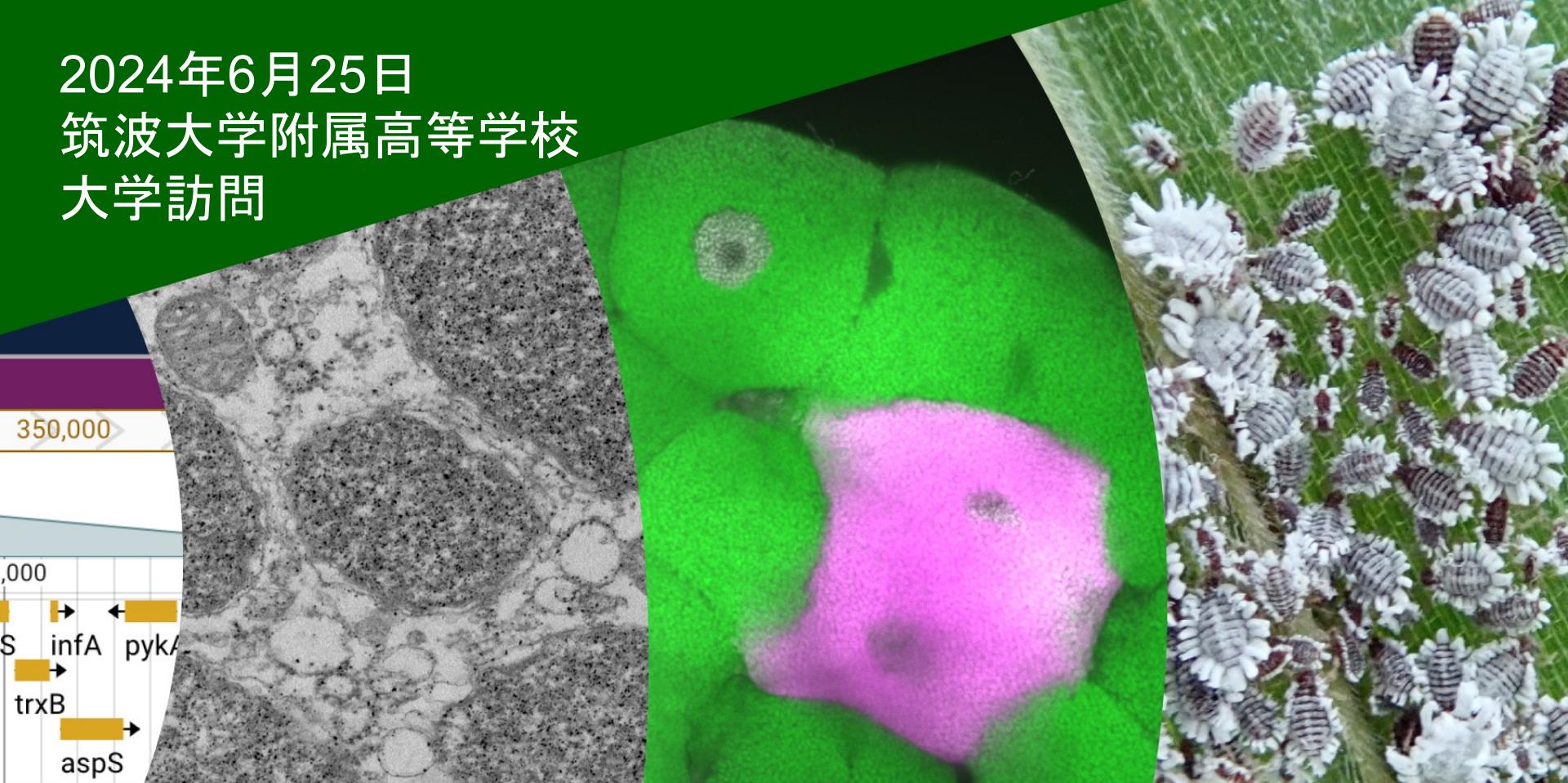


2024年6月25日
筑波大学附属高等学校
大学訪問



コンピューターを用いた生物研究

筑波大学 生存ダイナミクス研究センター
賴本 隼汰

本講義を始めるにあたって

- ・ わからないことがあったら、途中でも遠慮なく質問してください。
- ・ 高校生物「生物と遺伝子」分野を学びましたか？
 - ✓ゲノム
 - ✓セントラルドグマ
 - ✓ポリメラーゼ連鎖反応(PCR)
 - ✓次世代シークエンサー(NGS)
 - ✓etc...
- ・ 実習ではインターネットに接続できるデバイスが必要です。スマートフォンでも操作できますが、画面が大きいほど操作しやすいため、パソコンやタブレットがあるならそちらを使うことをお勧めします。

本講義を始めるにあたって

講義資料や演習課題は特設ホームページに公開されています。
講義終了後もアクセスできます。



<https://github.com/ShuntaYorimoto/2024Jun-BioinfoHandson/wiki>

生物の研究と聞いて思い浮かぶのは？

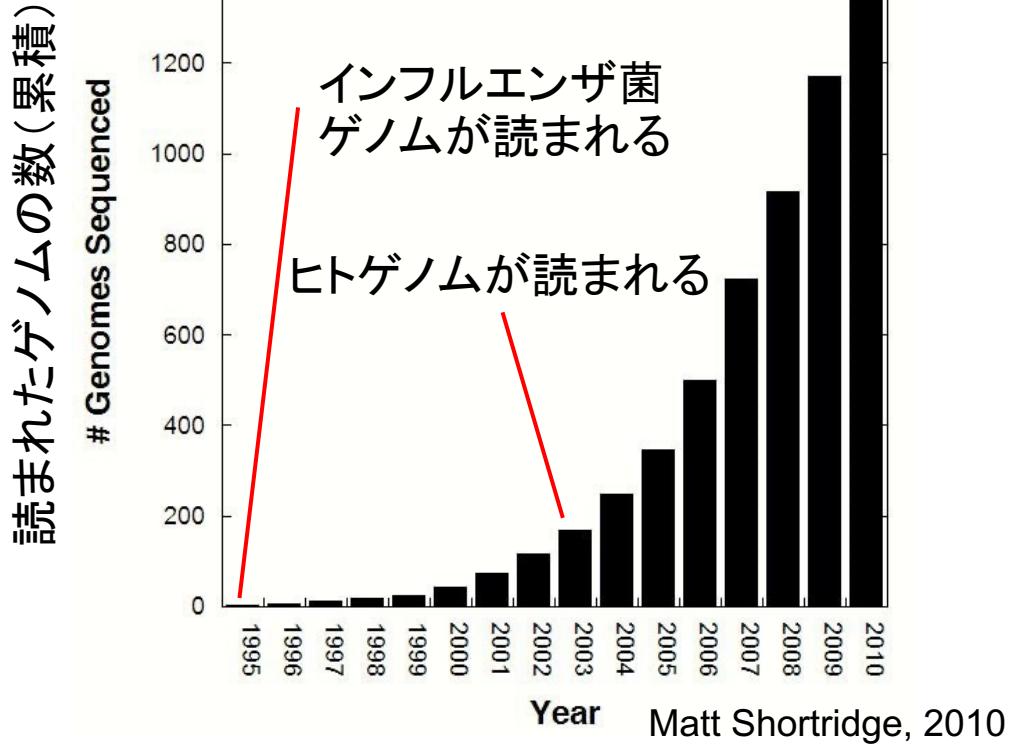


ゲノムとバイオインフォマティクス

基礎用語

- DNA(デオキシリボ核酸)
生物を構成する遺伝情報を保持する物質
- ゲノム
生物を構成する全遺伝情報
- 遺伝子
遺伝形質を規定する因子
- セントラルドグマ
DNA—(転写)→RNA—(翻訳)→タンパク質
という遺伝情報の流れは一方向的であるという生物則

続々と報告されているゲノム



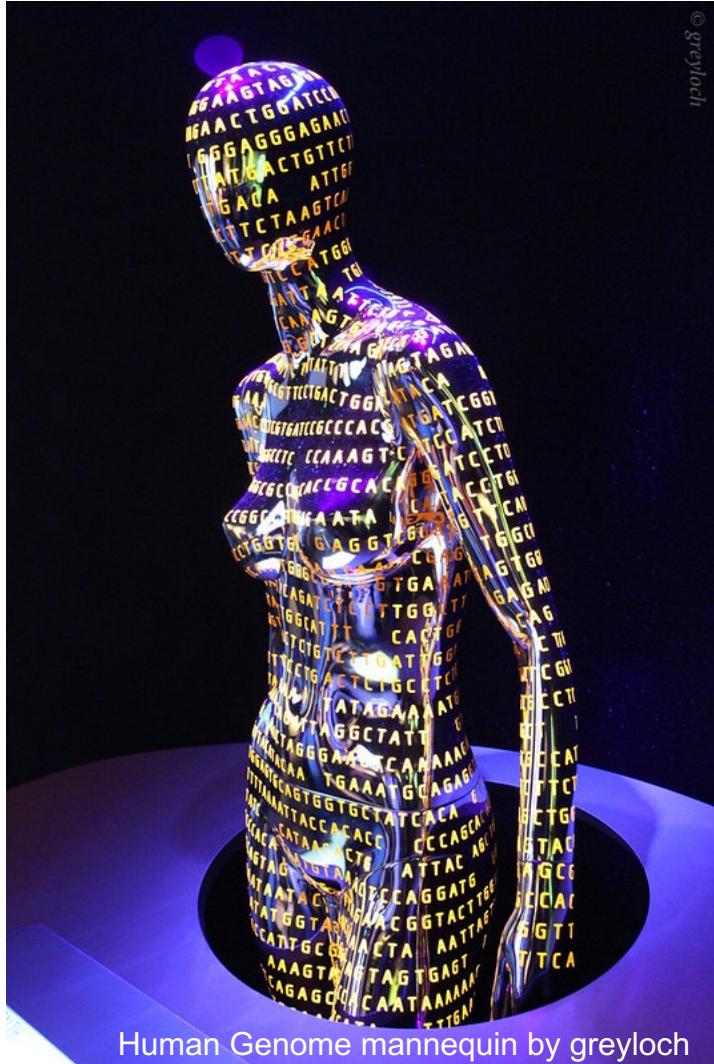
2024年6月22日現在

真核生物: 36,964

原核生物: 707,463 計: 約81万!

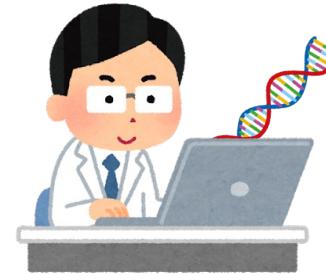
ウイルス: 75,745

<https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/>



バイオインフォマティクス(生物情報学)

- 生命科学と情報科学を融合した分野
- 生物学的データを計算機で解析する



次世代シーケンサー(NGS)



Illumina MiSeq

https://www.nibb.ac.jp/analyins/jp/?page_id=456



PacBio
Sequel IIe

https://www.nibb.ac.jp/analyins/jp/?page_id=6029

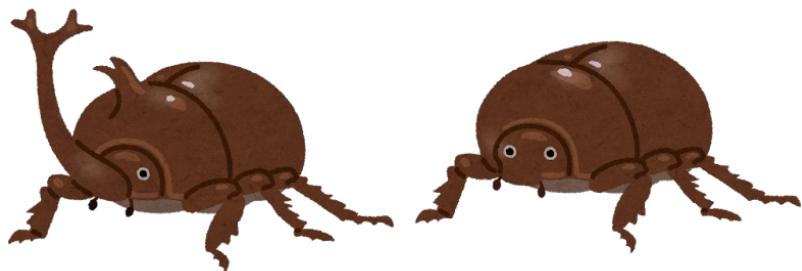


NANOPORE MinION

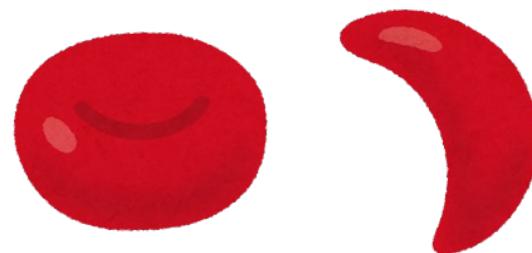
<https://nanoporetech.com/ja/products/sequence/minion>

ゲノムや遺伝子を調べると何が分かるのか

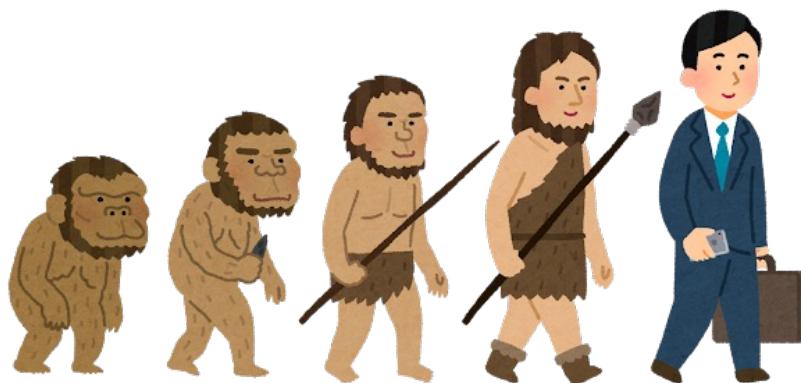
遺伝子の機能と形質



遺伝子の変異と病気



進化・系統



個体差



バイオインフォマティクスの応用例

- 基礎生物学
- 医学・薬学
- 農林水産・畜産
- 食品科学
- 環境科学
- etc...

生物配列解析(実習)

Pairwise sequence alignment

二つの塩基配列またはアミノ酸配列同士を比較し、最適な一致を見つけ、その類似性を調べる手法

例：

DNA

配列1: GACTCTTACCT

配列2: GGC-CCTGCC

* * * * **

protein

配列1: QSDGNFLDED

配列2: QTDGFLDDD

* : * * * * * : *

"-": gap 空白

"*": match 一致

": conserved substitutions アミノ酸の性質が似ている

" ": mismatch 不一致

どのようにアライメントを評価するのか

スコアルール 例1:

一致 = +1

不一致 = -1

空白 = -1

スコアルール 例2:

一致 = +1

不一致 = -1

空白 = 0

(A)

配列1: GACTCTTACCT

配列2: GGC-CCTGCC

* * * * **

1

(A)

配列1: GACTCTTACCT

配列2: GGC-CCTGCC

* * * * **

2

(B)

配列1: GACTCTTA-CCT

配列2: GGC-CCTGCC-

* * * * **

0

(B)

配列1: GACTCTTA-CCT

配列2: GGC-CCTGCC-

* * * * **

3

アミノ酸配列のアライメントはややこしい

BLOSUM62 matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1												
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-1	-1	-3										
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1												
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1												
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2												
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1												
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3												
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-1	-4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

Self-match scores

- Common amino acids (e.g., A,I,L,V) have lower scores
- Rare amino acids (e.g., W) and those with special roles (e.g., P) have higher scores

Frequently seen substitutions get positive scores

Rarely seen substitutions get negative scores

Substitutions expected by chance get a zero

練習問題1: pairwise sequence alignment



<https://github.com/ShuntaYorimoto/2024Jun-BioinfoHandson/wiki/exercise-1>

GitHub wikiの操作方法

The screenshot shows a GitHub wiki page for the HBB gene. At the top, there's a navigation bar with icons for signal, Wi-Fi, battery level (80%), and the URL 'github.com'. The time '13:46' is also displayed. Below the navigation, there's a sidebar with a 'Tool' section containing a link to 'water'. A red arrow points from the text '解析ツールのリンクを長押しし、「新しいタブで開く」で開いてください。' to this link.

Tool

- [water](#)

**1-1. 鎌状赤血球症の原因遺伝子β
グロビンにおける変異の検出**

鎌状赤血球症は、βグロビン遺伝子に変異が生じることで鎌状の奇形赤血球が形成され、貧血などの症状が見られる遺伝性疾患です。

正常な配列と変異のある配列を比較してみましょう。

正常

```
>HBB_normal NP_000509.1 HBB [organ ✓ H  
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVYPWT  
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLG  
ALAHKYH
```

変異

```
>HBB_mutated NP_000509.1 HBB [orga □ =  
MVHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVYPWT  
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLG  
AI AHKYH
```

解析ツールのリンクを長押しし、「新しいタブで開く」で開いてください。

このボタンを押すと、枠内の配列情報をコピーできます。

waterの操作方法

13:56
ebi.ac.uk
79%

Input sequence ⓘ
Sequence type
 Protein DNA

Paste your sequence here - or use the example sequence

```
>HBB_normal NP_000509.1 HBB [organism=Homo sapiens] [GeneID=3043] MVHLTPEEKSAVTALWGKVNDEVG GEALGRLLVYPWTQRFESFGDLST PDAVMGNPKVKAHGKKVLG AFSDGLAHLDNLKGTFATLSELHCDK
```

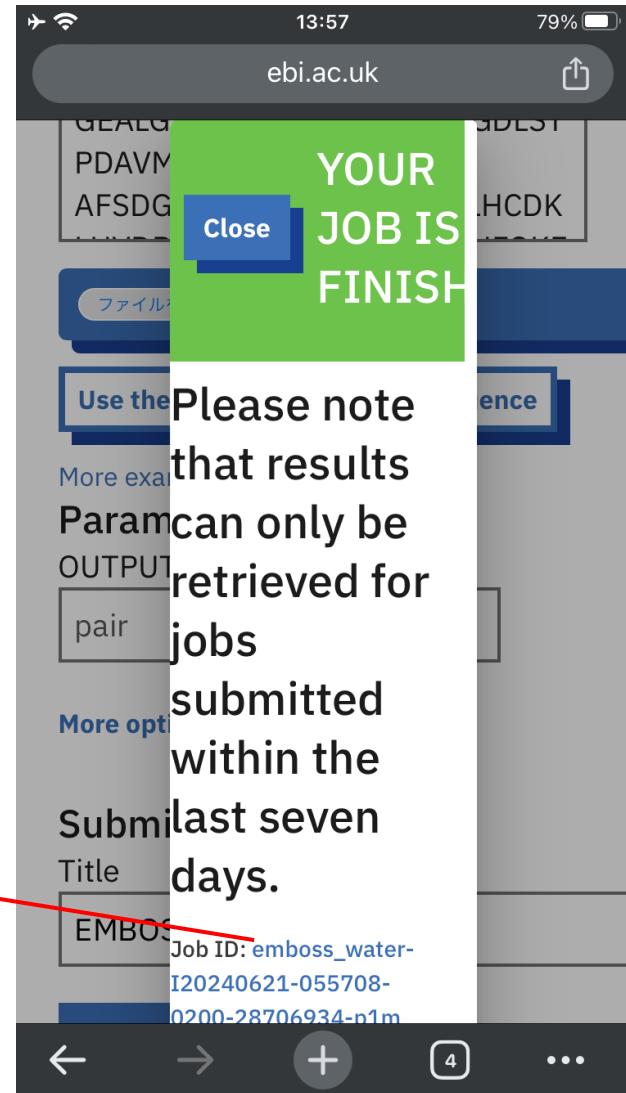
Submit
Title
EMBOSS Water's job

Submit

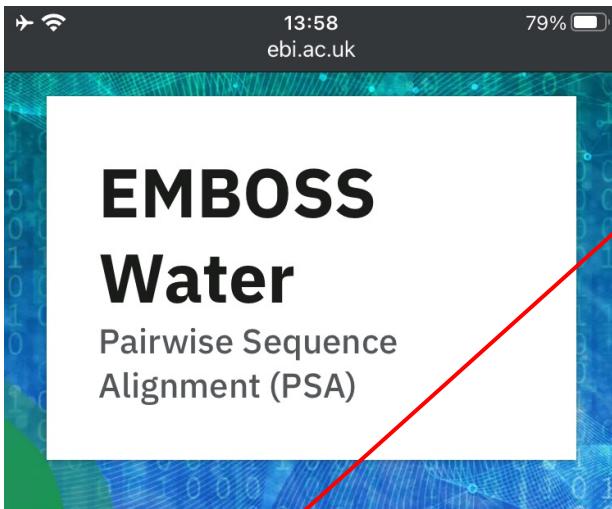
入力する配列に合わせてどちらか選択してください。

入力が終わったら、「Submit」を押してください。

「YOUR JOB IS FINISHED」と表示されたら、下のリンクを押してください。



もし「Close」を押してしまったら



解析結果は数日間
保存されます。
「Your Jobs」を押して
ください。

先ほどの「Job ID」を
押せば、結果が
見れます。

EMBOSS Water uses the Smith-Waterman algorithm (modified for speed enhancements) to calculate the local alignment of two sequences.

A screenshot of the Job Dispatcher website. At the top, it shows the time (13:58), battery level (79%), and URL (ebi.ac.uk). The main title is "Your Jobs" and there is a "Feedback" button. A yellow feedback banner says "Welcome to the new Job Dispatcher website. We'd love to hear your feedback about the new webpages!" with a close button. Below this is a section titled "Your Recent Jobs" with a "Filter by tool name" dropdown set to "EMBOSS WATER". There is also a "Delete All" button. A table lists recent jobs with columns: "Job ID (ALL)", "Status", and "Last update". The first job listed is "emboss_water-I20240621-055708-0200-28706934-p1m" with a status of "✓" and "8 hours ago". A red arrow points from the text above to the "Job ID" column of the first table row.

1-1. 鎌状赤血球症の原因遺伝子βグロビンにおける変異の検出

Q1. どこに変異が起きているでしょうか？

HBB_normal	1	MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVYPWTQRFFESFGDLS .	50
HBB_mutated	1	MVHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVYPWTQRFFESFGDLS	50
HBB_normal	51	TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD 	100
HBB_mutated	51	TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD	100
HBB_normal	101	PENFRLLGTVLVCVLAHHFGKEFTPPVQAAYQKVAGVANALAHKYH 	147
HBB_mutated	101	PENFRLLGTVLVCVLAHHFGKEFTPPVQAAYQKVAGVANALAHKYH	147

7番目のグルタミン酸(E)がバリン(V)に
変わっている！

1-2. ミトコンドリアCOI遺伝子の種間比較

Q1. ヒトとチンパンジー間のCOI遺伝子の類似性は何%ですか？

Q2. ヒトとマウス間のCOI遺伝子の類似性は何%ですか？

Q3. Q1, Q2の結果から考えられることは何ですか？

14:40
ebi.ac.uk
75%

```
#=====
#  
# Aligned_sequences: 2  
# 1: COI_human  
# 2: COI_chimpanzee  
# Matrix: EDNAFULL  
# Gap_penalty: 10.0  
# Extend_penalty: 0.5  
#  
# Length: 1542  
# Identity: 1408/1542 (91.3%)  
# Similarity: 1408/1542 (91.3%)  
# Gaps: 0/1542 ( 0.0%)  
# Score: 6504.0  
#  
#=====
```

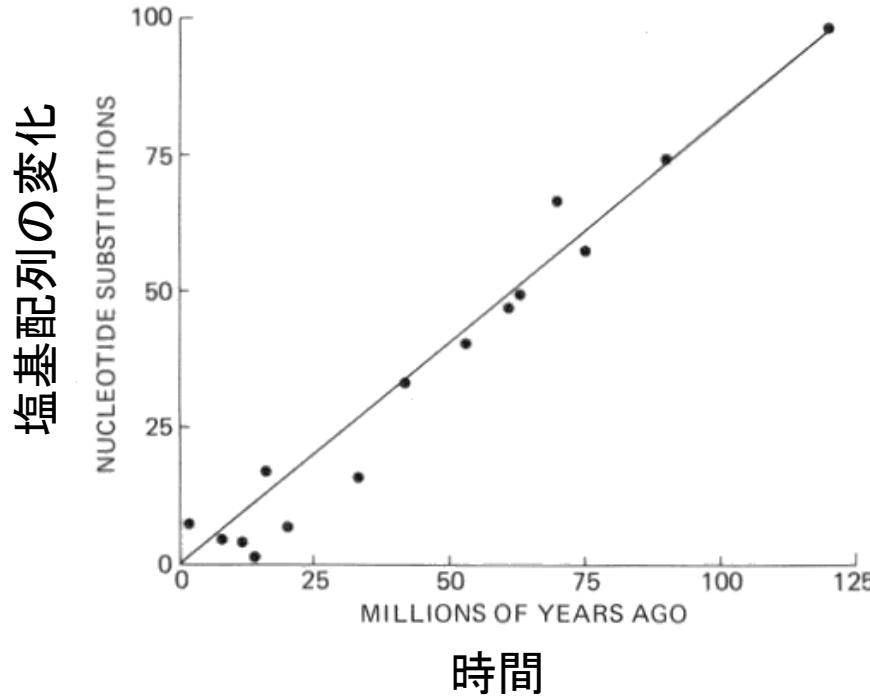
14:42
ebi.ac.uk
75%

```
#=====  
#  
# Aligned_sequences: 2  
# 1: COI_human  
# 2: COI_mouse  
# Matrix: EDNAFULL  
# Gap_penalty: 10.0  
# Extend_penalty: 0.5  
#  
# Length: 1550  
# Identity: 1210/1550 (78.1%)  
# Similarity: 1210/1550 (78.1%)  
# Gaps: 20/1550 ( 1.3%)  
# Score: 4674.5  
#  
#=====
```

分子時計

DNA/タンパク質配列の変化が、時間と共に一定の速度で起こる。この変化速度を基に生物種間の分岐時期を推定できる。

Zuckerkandl and Pauling, 1965

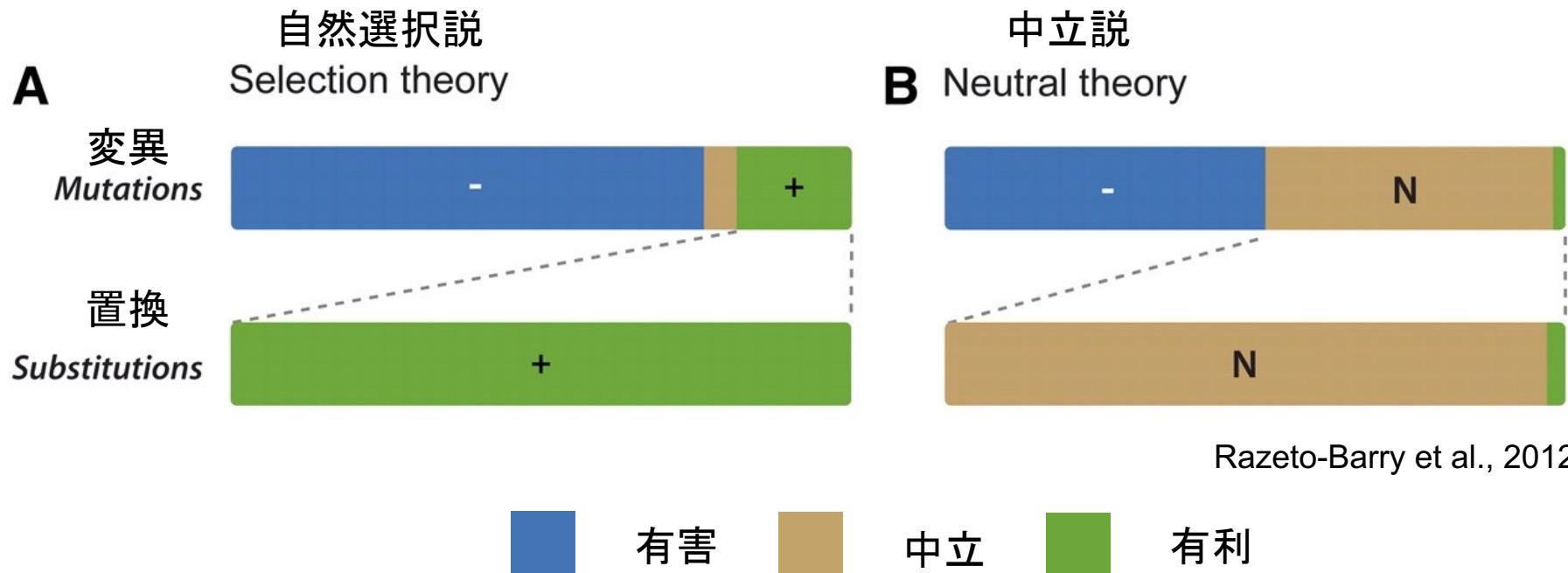


https://www.mun.ca/biology/scarr/Molecular_Clock.html

分子進化の中立説

自然選択に対して良くも悪くもない突然変異が偶然生じる。

木村資生, 1968



BLAST (Basic Local Alignment Search Tool)

Pairwise sequence alignment: 1 vs 1

BLAST検索: 1 vs database

- 手元にある塩基配列やアミノ酸配列(クエリ配列)を、大規模な配列データベースに対して検索し、類似する配列を見つけるためのアルゴリズムまたはツール
- バイオインフォマティクス解析で広く使われ、配列の機能予測や進化的関係の解明などに利用される

練習問題2: BLAST



<https://github.com/ShuntaYorimoto/2024Jun-BioinfoHandson/wiki/exercise-2>

BLASTの操作方法1

blast.ncbi.nlm.nih.gov

BLAST®

Home Recent Results Saved Strategies Help

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

NEWS

BLAST+ 2.15.0 is here!

We have included two exciting new features in the latest BLAST+ release

Tue, 28 Nov 2023

[More BLAST news...](#)

Web BLAST

blastn

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

blastp

Protein BLAST
protein ▶ protein

The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with links for Home, Recent Results, Saved Strategies, and Help. Below that, a banner for 'BLAST+ 2.15.0 is here!' is displayed, featuring a green 'NEWS' box and a white box with text and a date (Tue, 28 Nov 2023). The main content area is titled 'Basic Local Alignment Search Tool'. It describes what BLAST does and includes a 'Learn more' link. Below this, there are four search tools: 'blastn' (Nucleotide BLAST), 'blastx' (translated nucleotide ▶ protein), 'tblastn' (protein ▶ translated nucleotide), and 'blastp' (Protein BLAST). Each tool has a corresponding graphic: a DNA double helix for blastn, arrows for blastx/tblastn, and a protein ribbon for blastp.

入力する配列とデータベースの配列が塩基配列かアミノ酸配列かによって使うアルゴリズムを選択する。

BLASTの操作方法2

blast.ncbi.nlm.nih.gov

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

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

>unknown_DNA
AAATTGAAGAGTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCA
ACACATGCAAGTCGAACGGT
AACAGGAAGAACGTTGCTCTTGCTGACGAGTGGCGGACGGGTGAGTA

Query subrange [?](#)

From
To

Reset page Bookmark

Or, upload file [ファイルを選択](#) [ファイル未選択](#)

Job Title unknown_DNA

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

Align two or more sequences [?](#)

Choose Search Set

Database Standard databases (nr etc.) rRNA/ITS databases Genomic + transcript databases Betacoronavirus
 Experimental databases

Nucleotide collection (nr/nt) [?](#)

Organism [Optional](#)

Enter organism name or id—completions will be suggested exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Models (XM/XP) Uncultured/environmental sample sequences

Sequences from type material

Entrez Query [Optional](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn)
Choose a BLAST algorithm [?](#)

BLAST

Search database nt using Megablast (Optimize for highly similar sequences)

Show results in a new window

+ Algorithm parameters

ここに配列を入力してください。

データベースを選択してください。

指示がある場合は、検索対象の生物を指定してください。

ここにチェックをつけてから「BLAST」を押すと、結果が新しいタブで表示されます。

2-1. 食中毒の原因菌(BLASTN)

Q1. この配列はどの細菌に由来するでしょうか？

大腸菌(*Escherichia coli*) O-157株の16S ribosomal RNA gene

blast.ncbi.nlm.nih.gov

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download ▾

Select columns ▾ Show 100 ▾ ?

select all 100 sequences selected GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Escherichia sp. strain Esraa 4 16S ribosom... Escherichia ...	2848	2848	100%	0.0	100.00%	1542	MT647245	
<input checked="" type="checkbox"/>	Escherichia coli strain Dam/Dcm chromoso... Escherichia ...	2848	19547	100%	0.0	100.00%	4628041	CP053603	
<input checked="" type="checkbox"/>	Escherichia coli strain NEB10-beta chromo... Escherichia ...	2848	19530	100%	0.0	100.00%	4667764	CP053604	
<input checked="" type="checkbox"/>	Escherichia coli strain NEB_Turbo chromos... Escherichia ...	2848	19547	100%	0.0	100.00%	4527032	CP053605	
	Escherichia coli strain pK19EC119 chromo... Escherichia ...	2848	19502	100%	0.0	100.00%	1696649	CP050280	

Feedback

2-2. mRNAワクチンの配列(BLASTX)

Q1. この配列はどのウイルスを標的としますか？

Q2. この配列が標的とするタンパク質は何ですか？

blast.ncbi.nlm.nih.gov

select all 74 sequences selected

[GenPept](#) [Graphics](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	surface glycoprotein [Severe acute respir...	Severe ...	2547	2547	94%	0.0	99.84%	1273	YP_009724390.1
<input checked="" type="checkbox"/>	spike glycoprotein [SARS coronavirus Tor2]	SARS c...	1966	1966	93%	0.0	76.80%	1255	YP_009825051.1
<input checked="" type="checkbox"/>	spike protein [Bat coronavirus BM48-31/B...	Bat coro...	1818	1818	94%	0.0	72.02%	1259	YP_003858584.1
<input checked="" type="checkbox"/>	spike protein [Bat Hp-betacoronavirus/Zh...	Bat Hp...	828	828	89%	0.0	40.27%	1317	YP_009072440.1

surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]

Sequence ID: [YP_009724390.1](#) Length: 1273 Number of Matches: 1

Range 1: 12 to 1273 [GenPept](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Related Information

Score	Expect	Method	Identities	Positives	Gene - associated	Gaps	Frame
2547 bits(6601)	0.0	Compositional matrix adjust.	1260/1262(99%)	1260/1262(99%)	0/1262(0%)	+1	

Query 91 SSQCVNLTRTQLPPAYTNSFTRGVYYPDVKFRSSVLHSTC
Sbjct 12 SSQCVNLTRTQLPPAYTNSFTRGVYYPDVKFRSSVLHSTC

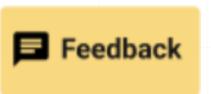
2-3. 生物間で保存されている遺伝子(BLASTP)

Q1. ショウジョウバエのeyeless遺伝子と最も類似しているヒトの遺伝子は何ですか？

blast.ncbi.nlm.nih.gov

select all 100 sequences selected

GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform a [Homo ...	Homo s...	264	437	30%	7e-79	93.94%	422	NP_000271.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform c [Homo ...	Homo s...	263	434	30%	1e-78	93.94%	401	NP_001297088.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform h [Homo ...	Homo s...	263	434	30%	8e-78	93.94%	473	NP_001355844.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform e [Homo ...	Homo s...	263	436	30%	1e-77	93.94%	503	NP_001355839.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform b [Homo ...	Homo s...	254	428	30%	4e-75	84.93%	436	NP_001245391.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform g [Homo ...	Homo s...	253	425	30%	4e-74	84.93%	487	NP_001355841.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform f [Homo ...	Homo s...	253	425	30%	5e-74	84.93%	488	NP_001355840.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform j [Homo ...	Homo s...	251	424	30%	1e-73	78.98%	447	NP_001355849.1
<input checked="" type="checkbox"/>	paired box protein Pax-6 isoform i [Homo ...	Homo s...	245	419	30%	2e-71	72.51%	461	NP_001355847.1
	paired box protein Pax-5 isoform 5 [Homo ...	Homo s...	214	214	19%		Feedback	295	NP_001267479.1
	paired box protein Pax-5 isoform 4 [Homo ...	Homo s...	214	214	19%			324	NP_001267478.1

2-3. 生物間で保存されている遺伝子(BLASTP)

Q2. 最も類似していたヒトの遺伝子の機能は何ですか？

ncbi.nlm.nih.gov

COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from [Z95332.1](#), [M93650.1](#) and [Z83307.1](#).

Summary: This gene encodes paired box protein Pax-6, one of many human homologs of the *Drosophila melanogaster* gene *prd*. In addition to a conserved paired box domain, a hallmark feature of this gene family, the encoded protein also contains a homeobox domain. Both domains are known to bind DNA and function as regulators of gene transcription. Activity of this protein is key in the development of neural tissues, particularly the eye. This gene is regulated by multiple enhancers located up to hundreds of kilobases distant from this locus. Mutations in this gene or in the enhancer regions can cause ocular disorders such as aniridia and Peter's anomaly. Use of alternate promoters and alternative splicing results in multiple transcript variants encoding different isoforms. Interestingly, inclusion of a particular alternate coding exon has been shown to increase the length of the paired box domain and alter its DNA binding specificity. Consequently, isoforms that carry the shorter paired box domain regulate a different set of genes compared to the isoforms carrying the longer paired box domain. [provided by RefSeq, Mar 2019].

Identical Structures
OMIM
Protein (UniProtKB)
PubMed (RefSeq)
PubMed (Weighted)
Related Structures (List)
Related Structures (Summary)
SNP

LinkOut to external resources ▾
Dryad Digital Repository
[Dryad Digital Repository]
NP_000271

このタンパク質の活性は、神経組織、特に眼の発達において重要である。

ショウジョウバエのeyeless遺伝子と似た機能を有している！

Multiple sequence alignment

- 複数の塩基配列またはアミノ酸配列同士を比較し、一致する領域や保存された領域を見つける手法
- 配列の構造や進化的関係を推定することができる

RLAO_METVA	--MIDAKSEHKIAPWKIEEVNALKELLKSANVIALIDMMEVPAVQLQETRDK
RLAO_METJA	---METKVKAHVAPWKIEEVKTLKGLIKSKPVVVAIVDMMDVPAPQLQETRDK
RLAO_PYRAB	-----MAHVAEWKKKEVEELANLIIKSYPVIALVDVSSSMPAYPLSQMRRL
RLAO_PYRHO	-----MAHVAEWKKKEVEELAKLIIKSYPVIALVDVSSSMPAYPLSQMRRL
RLAO_PYRFU	-----MAHVAEWKKKEVEELANLIIKSYPVVALVDVSSSMPAYPLSQMRRL
RLAO_PYRKO	-----MAHVAEWKKKEVEELANIICKSYPVIALVDVAGVPAYPLSKMRDK
RLAO_HALMA	MSAEESERKTETIPEWKQEEVDAIVEEMIESYESVGVVNIAGIPS RQLQDMRRD
RLAO_HALVO	MSESEVRQTEVIPQWKREEVDELVDFIESYESVGVVGVAGIPS RQLQSMRRE
RLAO_HALSA	MSAEEQRTTEEVPPEWKRQEVAELVDLLETYDSVGVVNVTGIPS KQLQDMRRG
RLAO_THEAC	-----MKEVSQQKKELVNEITQRRIKASRSVAIVDTAGIRT RQIQDIRGK
RLAO_THEVO	-----MRKINPKKKEIVSELAQDITKSKAVAIVDIKGVRT RQM QDIRAK
RLAO_PICTO	-----MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNEFQKIRNS

<https://bioinformaticsreview.com/20151018/multiple-sequence-alignment/>

練習問題3: multiple sequence alignment



<https://github.com/ShuntaYorimoto/2024Jun-BioinfoHandson/wiki/exercise-3>

3-1. アブラムシのBCR遺伝子

Q1. よく保存されている領域はどこでしょうか。

Q2. BCRの名前の由来は、"Bacteriocyte-specific Cystein-Rich protein"です。何個のシステイン残基が保存されていますか。

※スマートフォン (iPhone 8) だと段ズレを起こしていたため、下記はPCのスクショ。

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Apis_BCR4	MRLLYGFLIIMLTIYLSVQDFDPTEFKGPFPPTIEICSKYC--AVVCNYTSPRCYCVEAAK
Apis_BCR5	MRLLYGFLIIMLTIHLSVQDIDPNTLRGPYPTKEICSKYCEYNVVC GASLP-CICVQDAR
Apis_BCR1	MKLLHGFLIIMLTMHLSIQ----YAYGGPFLT KYLC DRVC--HKLCGDEFV-CSCIQYKS
Apis_BCR2	MKLLYGFLIIMLTIHLSVQ----Y-FESPFETKYNCDTHC--NKLCGKIDH-CSCIQYHS
	.:*****:*****:***:*
	. *: * *. * :*. * *::
Apis_BCR4	ERDQWFPYCYD-----
Apis_BCR5	QLDHWFACCYDGGPEMLM-----
Apis_BCR1	LKG LWFPHCPTGKASVVLHNFLTSP
Apis_BCR2	MEGLWFPHCRTGSAAQMLHDFLSNP
	. **. *

3-2. 遺伝情報を用いた系統関係の推定

- Q1. 系統樹の結果から、ヒトと最も近縁な生物はどれですか。
Q2. 系統樹の結果から、ヒトと最も遠縁な生物はどれですか。

16:01
66%
ebi.ac.uk

Results for Job ID
muscle-l20240621-075834-0052-
4392208-p1m **Copy**

< ments Phylogenetic Tree >

Result画面の”Phylogenetic Tree”を選択してください。

Phylogram

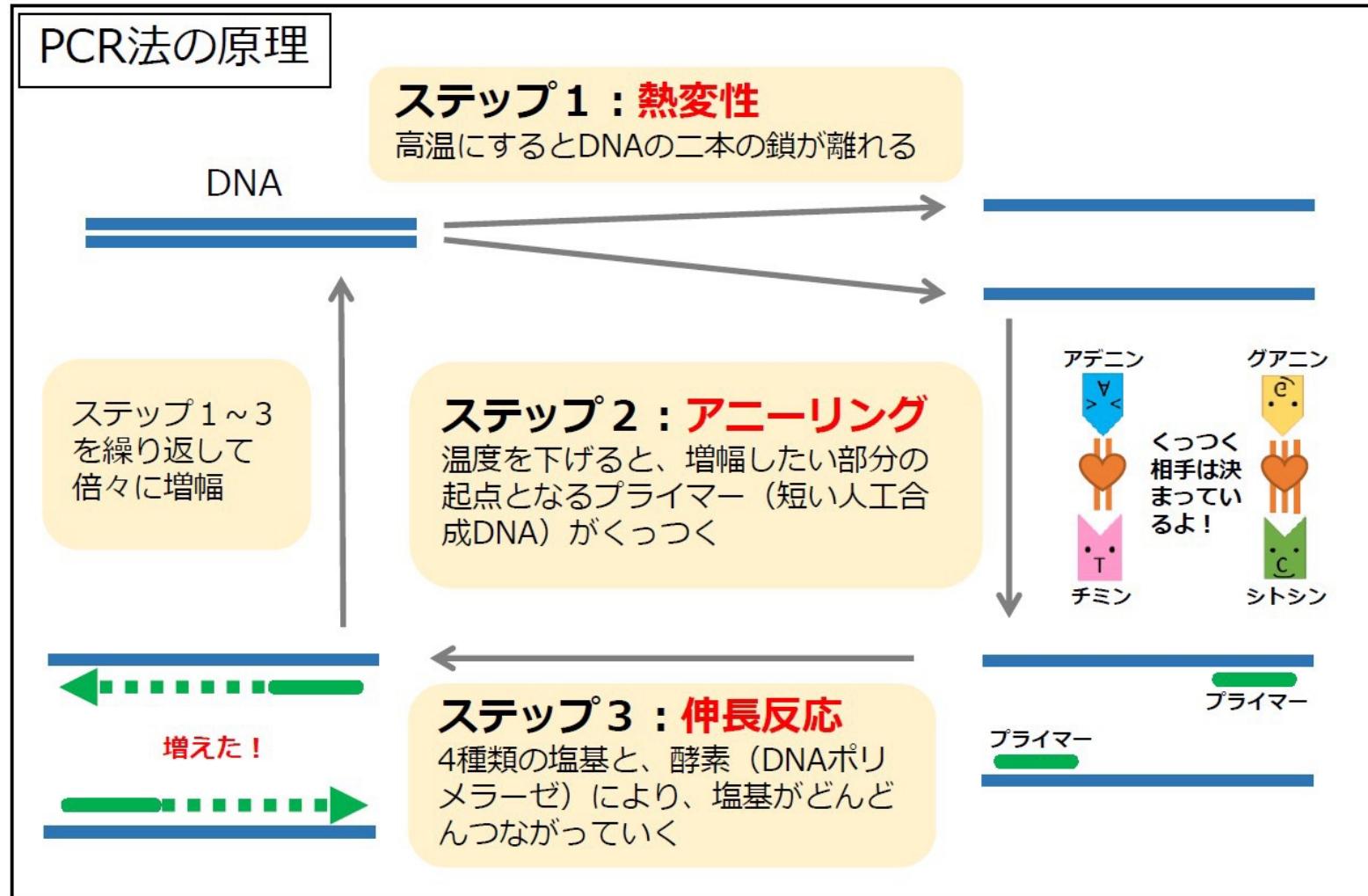


Radial

Selected 0 branches with current label



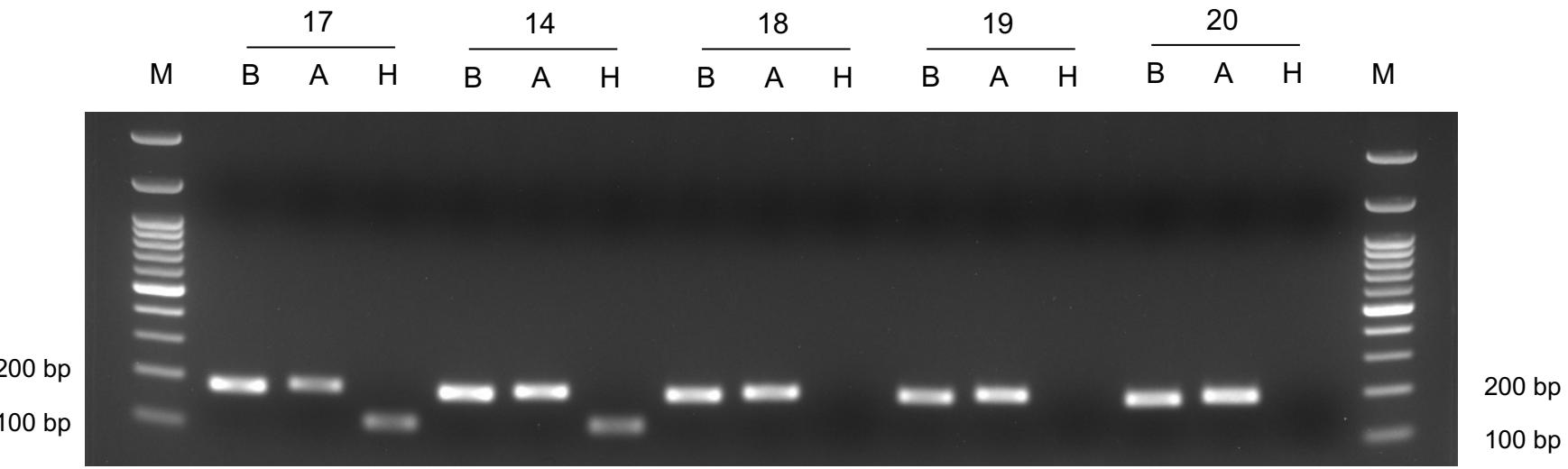
PCR (Polymerase Chain Reaction)



<https://www.huhp.hokudai.ac.jp/covid-19/pcr/>

電気泳動

- ・ ゲルを用いて電場をかけることで、DNA、RNA、タンパク質などの分子を大きさで分離する技術
- ・ PCR産物を電気泳動で流すと、目的の遺伝子が増幅されたか確認できる



Yorimoto et al., 2022

練習問題4: design primers for PCR



<https://github.com/ShuntaYorimoto/2024Jun-BioinfoHandson/wiki/exercise-4>

4-1. 新型コロナウイルス(COVID-19)のPCR検査に向けて

Q1. 何種類の組み合わせが提案されましたか。

primer3plus.com

Primer3Plus
pick primers from a DNA sequence

Load server settings: Default
Activate Settings

Task: generic

Main General Settings Advanced Settings Internal Oligo Penalties Advanced Seq.

Sequence Id:

Paste template sequence below or upload sequence file: ファイルを選択 ファイル未選択

Load Example

ATGACTCTCGTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACTCTTTCTTGCTTCG
TGGTATTCTGCTAGTACACTAGCCATCCTACTGCGCTCGATTGTTGCGTACTGCTGCAATATTG
TAACCGTAGCTGTAACCTCTTTACGTTACTCTCGTGTAAAATCTGAATTCTTAGAGTT
CCTGATCTCTGGTCTAA

More... Source C Help About

Pair 1: Primer

Left Primer 1: TCGTTTCGGAAGAGACAGGT
Start: 11 Length: 20 bp Tm: 58.7°C R: 91.710% GC: 50.0% Any: 0.0 End: 0.0 TB: 6.0 HP: 44.7 Y_Stab: 4.0 Penalty: 1.3

Hairpin Secondary Structure:
Tm: 44.7°C dG: -648 cal/mol dH: -26600 cal/mol dS: -84 cal/mol
5' TCGTTTCGG-
|| |
3' TGGACAGAGAA-

Right Primer 1: AGACCAGAAGATCGGAACTCT
Start: 226 Length: 22 bp Tm: 57.9°C R: 90.720% GC: 45.5% Any: 0.0 End: 0.0 TB: 6.0 HP: 34.6 Y_Stab: 3.2 Penalty: 4.0

Hairpin Secondary Structure:
Tm: 34.6°C dG: 55 cal/mol dH: -7000 cal/mol dS: -23 cal/mol*K
5' AGACCAGAAGATCGG-
| | A
3' TCTCA-

Pair: Product Size: 216 bp Tm: 81.6°C Any: 0.0 End: 0.0 TB: 11.0 Penalty: 5.3

First Primer Pair Previous Primer Pair Next Primer Pair Last Primer Pair

1	ATGACTCTCAT	TCGTTTCGGA	AGAGACAGGT	ACGTTAACAG	TTAATAGCGT
51	ACTTCTTTT	CTTGCTTTCG	TGGTATTCTT	GCTAGTACAA	CTAGCCATCC
101	TTACTGCGCT	TCGATTGTTG	GCGTACTGCT	GCAATATTGT	TAACGTGAGT
151	CTTGTAAAAC	CTTCTTTTA	C GTTTACTCT	CGTGTAAAAA	ATCTGAATTG
201	TTCTAGAGTT	CCTGATCTTC	TGGTCTAA		

Link to the results (valid 3 days): <https://www.primer3plus.com/index.html?UUID=1e22e838-8787-4f09-8d48-279c4e387200>

Statistics:

Left Primer:	considered 1280, low tm 598, high tm 176, ok 506
Right Primer:	considered 1280, GC content failed 2, low tm 743, high tm 157, high hairpin stability 11, ok 367
Primer Pair:	considered 2111222, unacceptable product size 2111208, primer in pair overlaps a primer in a better pair 28471, ok 14

ここに配列を入力してください。
“Pick Primers”を押してください。

4-1. 新型コロナウイルス(COVID-19)のPCR検査に向けて

Q2. “Pair1”のプライマーは適切か評価しましょう。

Pair 1:

Left Primer 1: TCGTTTCGGAAAGAGACAGGT

Start: 11 Length: 20 bp Tm: 58.7 C B: 91.710 % GC: 50.0 % Any: 0.0 End: 0.0 TB: 6.0 HP: 44.7 3' Stab: 4.0 Penalty: 1.3

Hairpin Secondary Structure:

Tm: 44.7°C dG: -648 cal/mol dH: -26600 cal/mol dS: -84 cal/mol

5' TCGTTTCGG
 || ||
3' TGGACAGAGAA

Right Primer 1: AGACCAGAAGATCAGGAACCT

Start: 226 Length: 22 bp Tm: 57.9 C B: 90.720 % GC: 45.5 % Any: 0.0 End: 0.0 TB: 6.0 HP: 34.6 3' Stab: 3.2 Penalty: 4.0

Hairpin Secondary Structure:

Tm: 34.6°C dG: 55 cal/mol dH: -7000 cal/mol dS: -23 cal/mol*K

5' AGACCAGAAGATCAGG
 || A
3' TCTCA

Pair: Product Size: 216 bp Tm: 81.6 C Any: 0.0 End: 0.0 TB: 11.0 Penalty: 5.3

今回着目するのは:

- 増幅サイズ
- プライマーの長さ
- GC含量
- Tm値

データベース

- ・大量のデータを効率的に管理、検索、取得するためのシステム
- ・遺伝子配列、タンパク質の配列や構造、遺伝子発現データなどが公開されている
- ・本実習で扱ってきた塩基配列やタンパク質配列は、いずれも公開データベースからダウンロードしたもの
- ・生命科学研究の基盤として多くの研究者に利用されている

練習問題5: explore database



<https://github.com/ShuntaYorimoto/2024Jun-BioinfoHandson/wiki/exercise-5>

5-1. NCBI

ゲノム、遺伝子、タンパク質などのデータベースやBLASTなどの分析ツールを提供

ncbi.nlm.nih.gov

 National Library of Medicine
National Center for Biotechnology Information

All Databases  aldh2 homo sapiens  

NCBI Home

Resource List (A-Z)

- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology

Welcome to NCBI

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

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

Submit
Deposit data or manuscripts into NCBI databases 

Download
Transfer NCBI data to your computer 

Learn
Find help documents, attend a class or watch a tutorial 

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

5-1. NCBI

Q1. ヒトのALDH2遺伝子の機能は何ですか。



Results found in 20 databases (1 error)

GENE

ALDH2 – aldehyde dehydrogenase 2 family member
Homo sapiens (human)

Also known as: ALDH-E2, ALDHI, ALDM
Gene ID: 217

RefSeq transcripts (2) RefSeq proteins (2)
RefSeqGene (1) PubMed (794)

Orthologs Genome Data Viewer
BLAST

Was this helpful?

RefSeq Sequences +

Summary This protein belongs to the aldehyde dehydrogenase family of proteins. Aldehyde dehydrogenase is the second enzyme of the major oxidative pathway of alcohol metabolism. Two major liver

このタンパク質はアルデヒド脱水素酵素ファミリーに属する。アルデヒド脱水素酵素はアルコール代謝の主要な酸化経路の第二の酵素である。

国税庁「あなたはお酒が強い人？弱い人？」

<https://www.nta.go.jp/about/organization/tokyo/sake/seminar/h19/02/02.htm>

5-1. NCBI

Q2. ID: ABB48431.1 に対応する生物と遺伝子の名前は何ですか。

The screenshot shows the NCBI protein search interface. The search term "pikachurin" is entered in the search bar. The results page displays the entry for "pikachurin [Mus musculus]". Key details shown include:

- GenBank ID: ABB48431.1
- Identical Proteins, FASTA, Graphics links
- LOCUS: ABB48431, 1017 aa
- DEFINITION: pikachurin [Mus musculus].
- ACCESSION: ABB48431
- VERSION: ABB48431.1
- DBSOURCE: accession D0223720.1
- KEYWORDS: .
- SOURCE: Mus musculus (house mouse)
- Post-translational modifications: linear, ROD, 24-OCT-2008

Text on the right side of the page:

光を受けて電気に換える細胞のイメージ
から、ピカチュウの名前を援用した遺伝子
古川貴久教授 インタビュー
https://resou.osaka-u.ac.jp/ja/story/2012/121201_2

5-2. UniProt

タンパク質の配列および機能に関する包括的なデータベース

The screenshot shows the UniProt homepage with the URL uniprot.org in the browser address bar. The main search bar contains the identifier ACYPI009125. Below the search bar, there are links for UniProtKB, Advanced search, and a blue 'Search' button. To the right of the search bar, there is a 'Feedback' link. The main content area features a large blue banner with the text 'Find your protein'. Below the banner, examples of search terms like 'Insulin', 'APP', 'Human', 'P05067', and 'organism_id:9606' are listed. A descriptive text about UniProt follows, mentioning it is the world's leading high-quality, comprehensive, and freely accessible resource of protein sequence and functional information, with a link to 'Cite UniProt'. At the bottom, there are four navigation links: 'Proteins UniProt' (blue), 'Species Proteomes' (red), 'Protein Clusters' (orange), and 'Sequence Archive' (green).

5-2. UniProt

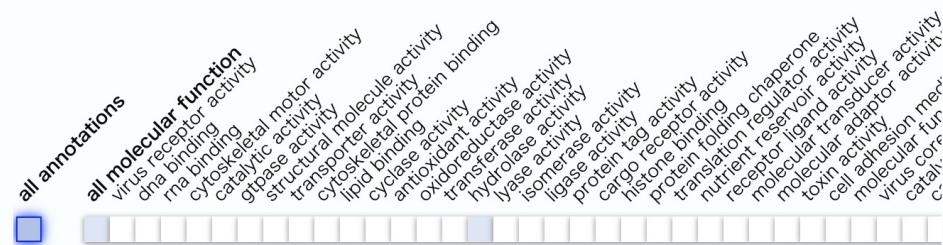
Q1. このタンパク質の機能は何ですか。

GO annotationsⁱ

Access the complete set of GO annotations on QuickGO [↗](#)

Slimming set:

generic



ASPECT	TERM
Molecular Function	lysozyme activity <small>↗ Source:UniProtKB-EC</small>
Biological Process	defense response to bacterium <small>↗ Source:UniProtKB-KW</small>
Biological Process	killing of cells of another organism <small>↗ Source:UniProtKB-KW</small>
Biological Process	metabolic process <small>↗ Source:UniProtKB-KW</small>

細菌への防御反応
他の生物の細胞を殺す

5-2. UniProt

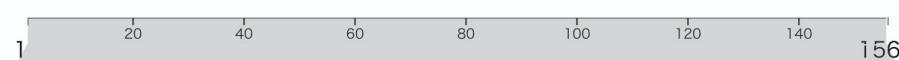
Q2. このタンパク質はどのようなドメインを持ちますか。

Q3. Q2を踏まえ、どのドメインがどの立体構造と対応しますか。

PTM/Processingⁱ

Features

Showing features for signalⁱ, chainⁱ, disulfide bondⁱ.



TYPE	ID	POSITION(S)	DESCRIPTION
------	----	-------------	-------------

-- Select --

▶ Signal		1-19	Automatic Annotation
▶ Chain	PRO_5010960718	20-156	lysozyme Automatic Annotation
▶ Disulfide bond		35↔123	Automatic Annotation
▶ Disulfide bond		40↔46	Automatic Annotation
▶ Disulfide bond		52↔61	Automatic Annotation
▶ Disulfide bond		93↔99	Automatic Annotation

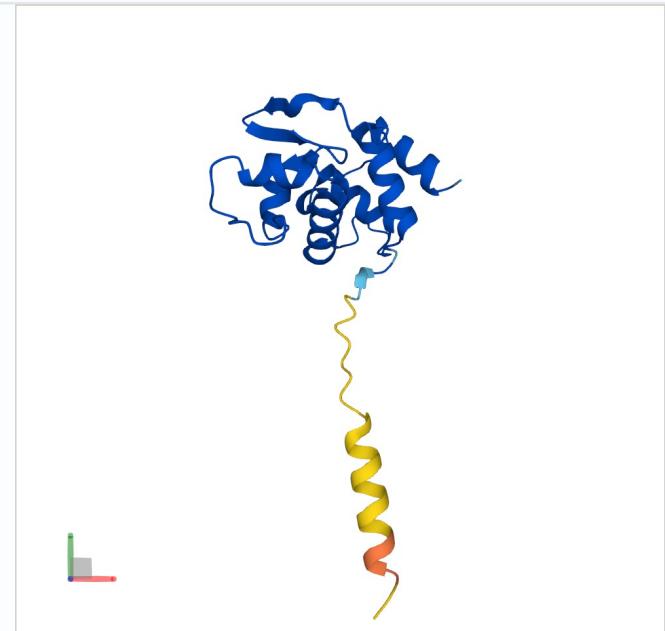
Structureⁱ

Model

Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.



他にもまだあるバイオインフォマティクス解析

- ゲノム解析
ゲノムアセンブリ、遺伝子予測、ゲノム比較など
- トランスクriプトーム解析
遺伝子発現パターンの解析(bulk, visium, single-cell, etc...)
- プロテオミクス
タンパク質の構造や機能、タンパク質相互作用など
- メタゲノミクス
環境中の微生物群のDNA解析
- エピゲノミクス
DNAメチル化やヒストン修飾など遺伝子発現調節機構の解析
- システム生物学
生命をシステムとして理解する(ネットワーク解析、モデル化)
- etc...

質疑応答

本講義に関すること、大学受験について、研究室生活について等、何でも気軽に質問して下さい。

本講義のアンケートにご協力ください

2024年6月25日「コンピューターを用いた生物研究」アンケート



https://docs.google.com/forms/d/e/1FAIpQLSfpArWuawq9E7caBctvciEVF6fFLMWRwUmtjFZQJeutGJVXpw/viewform?usp=sf_link