

# 今回を始める前に ～ 準備

今回は少しややこしいので  
よく注意を聞いてステップごとに

ファイルをダウンロードしよう

①下記のホームページ

<http://pepper.is.sci.toho-u.ac.jp/DL>  
から、デスクトップへ、mb-10.zip をダウンロード

②デスクトップで、mb-10.zipをダブルクリックすると、解凍される

⇒ デスクトップにフォルダmb-10ができる

⇒ その中のファイルを1つずつ違う所へコピーする (以下に指示)

慣れない学生に対して  
TAはサポートをお願いします

<Jupyter Notebook>を停止しておく

③mb-10の中のうちで、  
フォルダ clustralw 全体を  
D(USB):¥Wpy-3662 の下(中)にコピー

④mb-10の中のうちで、  
ファイル env.bat を  
D(USB):¥Wpy-3662¥scripts の下(中)  
にコピー (置き換え)

⑤mb-10の中のうちで、  
フォルダ try-clustral 全体を  
D(USB):¥Wpy-3662¥notebooksの  
下(中)にコピー

⑥テスト: Jupyter Notebookを起動  
フォルダ try-clustral へ移動し  
clustalw.ipynb を開き、上段をrun  
OKなら下段をrun ⇒ 樹形図?

## 配列の整列と距離

配列の整列 = (微妙に異なる配列の) 位置を合せる

イメージ:

MVLSADDKTN IKNCWGKIGG  
MVLSAADKSN VKAAWGKVGG

距離 = 違いの個数

違いが少ない ⇒ 変異が少ない

⇒ 進化上の距離が小さい

違いが多い ⇒ 変異が多い

⇒ 進化上の距離が大きい

抜けがあるときのイメージ:

MVLSADDKTN MVLSADDKTN  
MVLAADKSNV MVL-AADKSN

抜けの可能性を含めて、最もそれらしい位置を探す

違いの重み付け

塩基・アミノ酸の種類によって

違い度合 (距離) を変えるべき

(変わりやすさが違うはず)

⇒ いくつかのモデル

⇒ <https://www.ddbj.nig.ac.jp/clustalw-help.html>

あたりから始めるとよいのでは?

処理の結果:

整列された配列位置 \*\*\*.aln

配列間の距離 \*\*\*.dnd

配列間の距離から、

(クラスター化の手続きによって)

樹が描ける ⇒ 系統樹

# 配列の整列と距離

やってみよう

入力として用意するもの：

遺伝子かアミノ酸配列

いくつかの種の、同じタンパクの  
遺伝子orアミノ酸列

何がいいか？ ⇒ 試してみたら？  
(私に聞かないで！)

どこから拾ってくるか？

「相同性検索」で探すのが一手

**BLAST**

⇒ [https://www.ddbj.nig.ac.jp/column/blast\\_doc.html](https://www.ddbj.nig.ac.jp/column/blast_doc.html)

⇒ <http://blast.ddbj.nig.ac.jp/blastn?lang=ja>

(11/22までお休みらしい)

さしあたって、お試し用に

den.test.fasta ヘモグロビン 6 個

den.mio.fasta ミオグロビン

den.hemo.fasta ヘモグロビン20余個

仕組として用意したもの：

BioPython

Pythonでバイオインフォマティクス  
を扱うためのパッケージ

<https://biopython.org/>

<https://biopython-tutorial-ja.readthedocs.io/en/latest/index.html>

<https://biopython.org/wiki/Documentation>

ClustalW (Windows版)

多重整列プログラム

<https://www.ebi.ac.uk/Tools/msa/clustalw2/>

<https://www.ddbj.nig.ac.jp/clustalw-help.html>

<https://ja.wikipedia.org/wiki/Clustal>

3

# ClustalWの操作

やってみよう

入力として用意するもの：

遺伝子かアミノ酸配列

いくつかの種の、同じタンパクの  
遺伝子orアミノ酸列

何がいいか？ ⇒ 試してみたら？  
(私に聞かないで！)

どこから拾ってくるか？

「相同性検索」で探すのが一手

**BLAST**

⇒ [https://www.ddbj.nig.ac.jp/column/blast\\_doc.html](https://www.ddbj.nig.ac.jp/column/blast_doc.html)

⇒ <http://blast.ddbj.nig.ac.jp/blastp?lang=ja>

<http://blast.ddbj.nig.ac.jp/blastx?lang=ja>

さしあたって、お試し用に

den.test.fasta ヘモグロビン 6 個

den.mio.fasta ミオグロビン

den.hemo.fasta ヘモグロビン20余個

仕組として用意したもの：

BioPython

Pythonでバイオインフォマティクス  
を扱うためのパッケージ

<https://biopython.org/>

<https://biopython-tutorial-ja.readthedocs.io/en/latest/index.html>

<https://biopython.org/wiki/Documentation>

ClustalW (Windows版)

多重整列プログラム

<https://www.ebi.ac.uk/Tools/msa/clustalw2/>

<https://www.ddbj.nig.ac.jp/clustalw-help.html>

<https://ja.wikipedia.org/wiki/Clustal>

4

# 入力を自分で探してみよう①

アミノ酸列：たとえば Cytochrome c

1つ見つけて、それに似た配列をBLAST（相同検索）で検索する方針

最初の1つは適当に検索：タンパク質データベース UniProtKB

https://www.uniprot.org/uniprot/ ⇒ Cytochrome c mouseで検索

BLAST Align Download Add to basket Columns

Quote terms: "cytochrome c"

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9CR68	UCRI_MOUSE	Cytochrome b-c1 complex subunit Rie...	Uqcrfs1	Mus musculus (Mouse)	274
Q9D0M3	CY1_MOUSE	Cytochrome c1, heme protein, mitochond...	Cyc1	Mus musculus (Mouse)	325
P00158	CYB_MOUSE	Cytochrome b	Mt-Cyb Cob, Cytb, mt-Cytb, Mtcyb	Mus musculus (Mouse)	381
P00397	COX1_MOUSE	Cytochrome c oxidase subunit 1	Mtco1 COI, mt-Co1	Mus musculus (Mouse)	514
Q64429	CP1B1_MOUSE	Cytochrome P450 1B1	Cyp1b1 Cyp1-b1	Mus musculus (Mouse)	543
Q91X77	CY250_MOUSE	Cytochrome P450 2C50	Cyp2c50	Mus musculus (Mouse)	490
Q811W2	CP26B_MOUSE	Cytochrome P450 26B1	Cyp26b1	Mus musculus (Mouse)	512
P00184	CP1A1_MOUSE	Cytochrome P450 1A1	Cyp1a1 Cyp1a-1	Mus musculus (Mouse)	524

ページ内検索でCYC\_MOUSEを検索 ⇒ P62897 をクリック

5

# 入力を自分で探してみよう②

アミノ酸列：たとえば Cytochrome c

ページ内検索でCYC\_MOUSEを検索 ⇒ P62897 をクリック ⇒ 下の方に Sequences

## Sequence<sup>1</sup>

Sequence status<sup>1</sup>: Complete.

Sequence processing<sup>1</sup>: The displayed sequence is further processed into a mature form.

P62897-1 [UniParc] FASTA Add to basket

« Hide

```
10      20      30      40      50
MGDVEKGGKI FVQKCAQCHT VEKGGKHKTG PNLHGLFGRK TGQAAGFSYT
60      70      80      90     100
DANKNKGITW GEDTLMLEYL NPKKYIPGTK MIFAGIKKKG ERADLIAYLK
KATNE
```

Length: 105  
Mass (Da): 11,605  
Last modified: January 23, 2007 - v2  
Checksum:<sup>1</sup> B5BCA779BCE40492

BLAST GO

これがアミノ酸配列  
ClustalWに掛けるには  
FASTAでダウンロード  
してコピペで結合する

右の BLAST を選んでいるところで GO をクリック

⇒ 更に次の画面で、下の方 Run BLAST をクリック

⇒ Job status: RUNNING の画面が出る

⇒ 検索に時間がかかる（10分ぐらい？）が、次ページのような結果

6

# 入力を自分で探してみよう③

アミノ酸列：たとえば Cytochrome c

BLASTの結果 (BLASTは相同性検索 ~ 似た配列の検索)

## Overview

Show all 250

Entry	Protein names	Match hit
P62898	 <b>Cytochrome c, somatic</b> (Rattus norvegicus)	<div style="width: 100%;"><div style="width: 100%;"></div></div>
P62897	 <b>Cytochrome c, somatic</b> (Mus musculus)	<div style="width: 100%;"><div style="width: 100%;"></div></div>
G4XX57	 Cytochrome c (Rattus sp.)	<div style="width: 100%;"><div style="width: 100%;"></div></div>
Q56A15	 Cytochrome c (Mus musculus)	<div style="width: 100%;"><div style="width: 100%;"></div></div>

下に引くと  
沢山の候補

## Alignments

Alignment overview	Info
<input type="checkbox"/> Query: sp P62897 CYC_MOUSE B20181126A7FB2F8316FC890C270F4DE55871329902B9237	
<input type="checkbox"/> CYC_RAT - Cytochrome c, somatic - Rattus norvegicu... - View alignment	E-value: 1.2e-73 Score: 565 Ident.: 100.0%
<input type="checkbox"/> CYC_MOUSE - Cytochrome c, somatic Mus musculus (Mouse) - View alignment	E-value: 1.2e-73 Score: 565 Ident.: 100.0%
<input type="checkbox"/> G4XX57_MURI - Cytochrome c (Rattus sp.) - View alignment	E-value: 1.2e-73

1つ1つの  
候補の詳細

ブラウザのページ内検索で、種を探してみたら? (黄色いマーク: Reviewed)

homo sapiens, gorilla, Anas platyrhynchos(マガモ), Alligator mississippiensis, ...

## 期末レポート課題

課題を1つ選んでレポートを提出

11月28日現在

### 1) 英語の文字頻度

いくつかの英文について文字頻度を比較し  
各英文を区別する特徴となるか検討せよ  
(英文のソースの例)

<http://www.gutenberg.org/>

<https://www.whitehouse.gov/briefings-statements/the-inaugural-address/>

### 2) 英語の単語頻度

1) を単語の頻度について行ってみよ

### 3) 英語の文の長さ (文字数・単語数)

1) 同様に文章中の文の長さの分布を比較せよ

### 4) 日本語の単語頻度

幾人かの作家について単語頻度を比較せよ

### 5) 日本語の文の長さ (文字数・単語数)

幾人かの作家について単語頻度を比較せよ  
(日本語の文学作品のソースの例)

<http://www.aozora.gr.jp> 漢字文字コードはS-JIS

### 6) 系統樹

タンパク質を2~3種類選び、7~8種の  
動物の系統樹を作って、比較してみよ

### 7) 系統樹の問題点の指摘

取り上げた方法はごく簡略・模式的なもので  
多くの問題点を含む。分子進化学の参考書等  
を参照して、問題点を検討せよ

(入門書) ブルーボックス B-1849

「分子から見た生物進化」宮田隆 講談社

締切り: 2019年 1月16日

提出場所: 2号館2212室