

公開されているプログラムのコンパイル (clustalw, dendromat)

0. 作業用ディレクトリに移動

```
cd UNIXEX
mkdir 14
cd 14
```

1. ファイルをダウンロード

```
git clone https://github.com/NGS-maps/PracticeTargz
cd PracticeTargz
```

2. ダウンロードしたclustalw2を解凍・コンパイル

```
ls
clustalw-2.1.tar.gz      dendromat06.tar.gz      hemo.fasta              : ダウンロードしたファイルを確認
gunzip clustalw-2.1.tar.gz      : 圧縮ファイルを解凍

ls
clustalw-2.1.tar      dendromat06.tar.gz      hemo.fasta
tar -xf clustalw-2.1.tar      : tar アーカイブを解凍

ls
clustalw-2.1      clustalw-2.1.tar      dendromat06.tar.gz      hemo.fasta
cd clustalw-2.1      : 解凍したディレクトリ内に移動

ls
COPYING      Makefile.am      aclocal.m4      config.log      configure      install-sh      src
COPYING.LESSER      Makefile.in      clustalw_help      config.status      configure.ac      m4
Makefile README      config.guess      config.sub      depcomp      missing

: 沢山ファイルとディレクトリがある

./configure      : マシンの状況に合わせたmakefileを作成
複数行出力

make      : make により自動コンパイル
複数行出力

cd src      : src というディレクトリに移動
ls -lh | grep clustalw2      : ここにclustalw2と言うファイルが
-rwxr-xr-x  1 kenske  staff  944360  5 10 15:43 clustalw2      : 出来ていることを確認
cp clustalw2 ~/bin      : clustalw2を~/binに移動
clustalw2      : 実行できることを確認

*****
***** CLUSTAL 2.1 Multiple Sequence Alignments *****
*****
1. Sequence Input From Disc
2. Multiple Alignments
3. Profile / Structure Alignments
4. Phylogenetic trees

S. Execute a system command
H. HELP
X. EXIT (leave program)

x で抜ける
```

3. dendromat のソースを展開・コンパイル

```
cd ~/UNIXEX/14      : 作業ディレクトリに移動して
gunzip dendromat06.tar.gz      : 圧縮ファイルを解凍
tar -xf dendromat06.tar      : tarアーカイブを解凍
cd dendromat06      : 解凍したディレクトリ内に移動
make      : make でコンパイル
複数行出力

ls -l | grep dendromat      : dendromatというファイルが
-rwxr-xr-x@ 1 kenske  staff   53K  2 13  2009 dendromat      : 出来ていることを確認
cp dendromat ~/bin      : dendromatを~/binに移動
dendromat      : 実行できることを確認
ERROR      : ERRORとなっているがOK

Usage:  dendromat [-options] filename (input file is newick format, typically generated by
clustalW or T_coffee .dnd, .phb etc)
options: -fs      8      (Font size, default 10)
         -fn      courier      (font name, default Helvetica)
         -cs      1      (caption style, 1: Sideway, 2: Radial, 3: Radial Split, default 2)
         -mt      1      (map type, 0: Cladogram, 1: Phylogram, 2: Dendrogram, default 2)
         -lw      6      (line width, *0.1 point, default 10 = 1 point)
```

4. アラインメントを実行

```
cd -/UNIXEX/14
ls
clustalw-2.1      clustalw-2.1.tar dendromat06      dendromat06.tar hemo.fasta

clustalw2 hemo.fasta

    複数行出力

ls
clustalw-2.1      clustalw-2.1.tar dendromat06      dendromat06.tar
hemo.aln          hem.dnd          hemo.fasta
more hemo.aln
CLUSTAL 2.1 multiple sequence alignment

HITO              MVLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPFHDLSHGSQAQVKGHG
BUTA              -VLSAADKANVKAAGKVGQAGAHGAEALERMF LGFPTTKTYFPFHNLSHGSQDVKAHG
TOGARINEZUMI      -VLSPGDKSNIKAAWGKIGGQAPQYGAEALERMF LSFPPTTKTYFPFDMSHGSAIQAHG
NIWATORI          MVLSAADKNNVKGI FTKIAGHAE EYGAE TLERMF TTYPTTKTYFPFHDLSHGSQAQIKHG
HATO              MVLSANDKSNVKA VFAKIGGQAGDLGGEALERL FITYPQT KTYFPFHDLSHGSQAQIKHG
WANI_KUROKODAIRU  -VLSDDDKCNVKA VWSKVAGHLEEYGAEALERMF CAYPQTKI YFPFHDLSHGSQAQIRAHG
                  ***. ** *:. : :.:. : *.*:***:* : * * *****:**** *:. **

HITO              KKVADALTNVAHVHDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTF
BUTA              QKVADALTKAVGHLLDLPALGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNP
TOGARINEZUMI      KKVADALSTAVGHLLDLPALGALSALSDLHAHKLRVDPANFKLLSHC ILVTLACHHPGDFNP
NIWATORI          KKVVAALIEAANH IDDIAGT LSKLSDLHAHKLRVDPVNFKLLGQCFLVVAIHHPAALTP
HATO              KKVAEALVEAANH IDDIAGALS KLSDLHAQKLRVDPVNFKLLGHCFV VVAHVHFPSSLTP
WANI_KUROKODAIRU  KKVFAALHEAVNH IDDLPGALCRLSELHAHSLRVDPVNFKFLAQCVLVVAIHHPGSLTP
                  :** ** *. *:.:. :.*. **:***:*****.***:*. :.***: * * :.*

HITO              AVHASLDKFLASVSTVLTISKYR
BUTA              SVHASLDKFLANVSTVLTISKYR
TOGARINEZUMI      EIHASLDKFLANVSTVLTISKYR
NIWATORI          EVHASLDKFLCAVGTVLTAKYR
HATO              EVHASLDKFLVAVGTVLTAKYR
WANI_KUROKODAIRU  EVHASLDKFLCAVSSVLTISKYR
                  :*****: *. :*****
```

5. 系統樹を作成 *dendromat*

```
dendromat -mt 1 hemo.dnd
```

複数行出力

```
ls
clustalw-2.1      clustalw-2.1.tar dendromat06      dendromat06.tar
hemo.aln         hemo.dnd         hemo.dnd.ps      hemo.fasta
evince hemo.dnd.ps
```

描画ウィンドウが開く

課題1 : .psファイルを添付して提出 012Yamada14.ps (012を学生番号下3桁、yamadaを自分の苗字で置き換え)
 メールタイトルは 山田_u14
 hemo.fastaから上記のように作った.psを提出しても良いが、次ページ以降の方法で、自分で選んだ生物種で作成した方がよい
 注) 厳密には、.dndファイルはガイドTreeと呼ばれるもので系統樹ではない。

6. マルチファスタファイルを作成

別途配布するスイスプロット生物種略号forヘモグロビンから、10種程度の生物種を選択する。選択した生物種の左側の記号をHBA_ の後につなげたものがその生物種のヘモグロビンのUNIPROT IDとなる。例えばヒトならHBA_HUMAN, リスト最初のトガリネズミならHBA_TUPGL である。NCBIサイトで検索窓左のAll DatabaseとなっているメニューバーをProteinsとし検索窓にHBA_HUMANなどと打ち込み検索を実行し、FASTA と書かれたリンクをクリックするとアミノ酸配列のfasta フォーマットが示される。

NCBI Resources How To

NCBI National Center for Biotechnology Information

All Databases Search

Protein HBA_HUMAN Search

Protein Search Advanced Help

NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! Read more...

GenPept Send to: Change region shown

RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain

UniProtKB/Swiss-Prot: P69905.2

Identical Proteins FASTA Graphics

Go to: Customize view

LOCUS HBA_HUMAN 142 aa linear PRI 13-APR-2016

NCBI Resources How To

Protein Search Advanced Help

NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! Read more...

FASTA Send to: Change region shown

RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain

UniProtKB/Swiss-Prot: P69905.2

GenPept Identical Proteins Graphics

>gi|57013850|sp|P69905.2|HBA_HUMAN RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain

MVLSPADKTNVKAAGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA

VAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLT

YR

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

コピー&ペースト

```
>gi|57013850|sp|P69905.2|HBA_HUMAN RecName: Full=Hemoglobin subunit alpha; AltName:
MVLSPADKTNVKAAGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA
VAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLT
YR
```

コピーペーストによりこの部分を選択した10種程度の生物種についてつなげ合わせてゆく。この時それぞれの1行目の ">" に続く文字列を、以下のように「ローマ字」または英語名で表記したものに置き換えておく。この表記が系統樹作成時のラベルとして表示されるのでこの作業を忘れると、意味のわからない結果となる。

```
>HITO
MVLSPADKTNVKAAGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA
VAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLT
YR
```

同様に、ブタ、トガリネズミ、などをつなぎ合わせて行き、選んだ生物種で以下の例のようなマルチファスタファイル(hemo.fasta)を作成する

```
>HITO
MVLSPADKTNVKAAGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFT
PAVHASLDKFLASVSTVLTskyr
>BUTA
VLSAADKANVKAAGKVGAGHAGEYGAELERMFLGFPTTKTYFPHFNLHGSDQVKAHQ
KVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHHPDFNPS
VHASLDKFLANVSTVLTskyr
>TOGARINEZUMI
VLSPGDKSNIAAGKIGGQAPQYGAELERMFLSFPTTKTYFPHFDMSHGSAQIQAHGK
KVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHHPDFNPS
VHASLDKFLANVSTVLTskyr
>NIWATORI
MVLSPADKTNVKAAGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQIKGHG
KKVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHHPDFNPS
VHASLDKFLANVSTVLTskyr
>HATO
MVLSPADKTNVKAAGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQIKGHG
KKVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHHPDFNPS
VHASLDKFLANVSTVLTskyr
>WANI
VLSAADKANVKAAGKVGAGHAGEYGAELERMFLGFPTTKTYFPHFNLHGSDQVKAHQ
KVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHHPDFNPS
VHASLDKFLANVSTVLTskyr
```