

- 分子生物学関連データベース, 配列解析ツールなどの網羅的コレクション.
- ライフサイエンス統合データベースセンター (<https://dbcls.rois.ac.jp/>)
ライフサイエンス分野におけるデータベース統合化の拠点.
- 京都大学大学院 農学研究科 (<http://www.kais.kyoto-u.ac.jp>)
分子生物学実験担当研究室 → <http://www.biochemistry.kais.kyoto-u.ac.jp>
- 京都大学大学院 生命科学研究科 (<http://www.lif.kyoto-u.ac.jp>)
分子生物学実験担当研究室 → <http://www.lif.kyoto-u.ac.jp/labs/plantmb/>

参考書籍

- 「バイオデータベースとウェブツールの手とり足とり活用法 今日からできるバイオインフォマティクス はじめの一步」, 中村保一, 磯合 敦, 石川 淳／編, 羊土社, ISBN 4-89706-357-4
- ・「初心者でもわかる! バイオインフォマティクス入門—やさしい UNIX 操作から遺伝子・タンパク質解析まで—」, 坊農秀雅／著, 羊土社, ISBN 4-89706-290-X
- ・「実践 バイオインフォマティクス」, Cynthia Gibas, Per Jambeck／著 水島 洋／監訳, オライリー・ジャパン, ISBN 4-87311-068-8
- ・「バイオインフォマティクス ゲノム配列から機能解析へ」, David W. Mount／著 岡崎康司, 坊農秀雅／監訳, メディカル・サイエンス・インターナショナル, ISBN 4-89592-307-x

1.9.2. アミノ酸配列の解析

1.9.2.1. ドメイン構造の検索

– 応用生命科学科のホームページ(<http://www.biochemistry.kais.kyoto-u.ac.jp/>)に入って以下の実習を行なう –

蛋白質は通常一つのドメインからなるのではなく, 多くのドメインが集まったモジュール構造をとっている. 各ドメインはドメイン単位で機能を持つと共に, 蛋白質内の他のドメイン構造と関係し, 互いの機能を制御しあって初めて蛋白質として機能することができる. この実習では, 蛋白質のドメイン(モジュール)構造を見つけだし, 他の蛋白質と比較する.

モジュール(ドメイン)の例

酵素活性ドメイン	-- キナーゼ, ATPase など
蛋白質-蛋白質相互作用	-- SH(src homology)3 domain, SH2 domain, PDZ domain など
蛋白質-DNA 相互作用	-- homeobox, helix-turn-helix など
蛋白質-脂質相互作用	-- PH (Pleckstrin homology) domain

モジュール構造の利点

- 単にドメイン(モジュール)を入れ替えるだけで、異なる機能を持つ蛋白質ができる。たとえば一部のモジュールを入れ換えることにより、同じ酵素活性を持っていても、細胞内局在が異なる蛋白質や異なる高度複合体に属する蛋白質を簡単に作ることができる。
 - 蛋白質のバリエーションを簡単に増やすことができる。
- モジュールの微調整により同じようなモジュールで異なる機能(異なる結合相手)を持たせることができる。
 - 蛋白質のバリエーションを簡単に増やすことができる。
- 一つの機能ドメインだけでは細胞内で機能しないことが多い。
 - 基質, 局在, 高度複合体

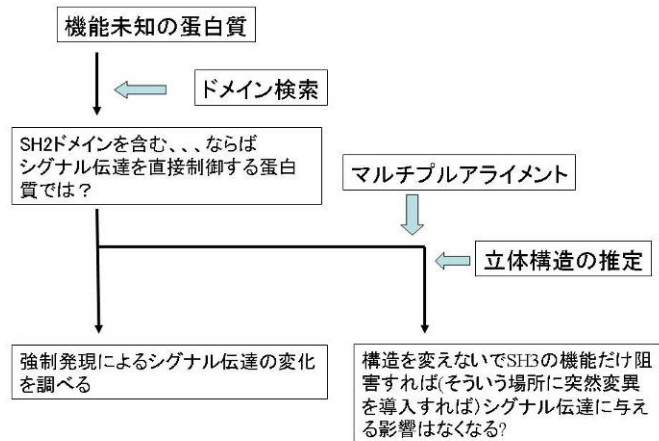


図 10

各モジュール(ドメイン)の役割の例

酵素活性

- | | | |
|------------|----|---|
| 細胞内の局在を決める | -- | 細胞内のどこで機能する必要がある? 核内?
細胞膜直下? |
| 高度複合体を形成する | -- | 何ステップかに必要な酵素や基質が複合体を作り、特定の反応が非常に効率良く進む。 |

実習の目的

なんらかの方法で「遺伝子をクローニングした」とする。

機能的スクリーニング, 抗体スクリーニング, ハイブリダイゼーション, PCR, two-hybrid, プロットオーバーレイ法, ゲノムプロジェクト

↓

どのような蛋白質なのかをドメイン, モジュールの観点からコンピューターを使って調べる。

- (実習 1) どのようなモジュールが含まれているか?
(含まれているモジュールの特徴からタンパク質の機能を類推できないか?)

↓

- (実習 2) 似たドメイン-モジュールを持つ蛋白質を並べて比較する(マルチプルアライメント).
(どのアミノ酸が特に保存されている? どのアミノ酸に変異を入れればこのドメインの機能を破壊できる?)

↓

- ・ (実習 3) 未知蛋白質は膜貫通ドメインを持っているか？
(細胞内の局在場所を推定することによってタンパク質の機能を類推できないか？)

実習 (1) ドメイン構造をさがす

それぞれのドメインは機能を果たすために、あるいはその構造を維持するために、進化上特定のアミノ酸配列が保存され、個々のドメインを特徴付けている。現在これらの保存されたアミノ酸配列(モチーフ)がデータベース化され、それを検索することによって未知の蛋白質の配列中にどのようなドメインが含まれているかを推定することができる。

- ① 本実習では下記例 1 配列, および未知配列を用いてドメインの検索を行なう。

>例 1(実習 1.8.2 未知融合プラスミドにコードされているタンパク質の一部)

```
DSMLGSLQSDLNKLGVATVAKGVCGACKKPIAGQVVTAMGKTWHPEHFVCTHCQEEIGS
RNF FERDGPYCEKDYHNLFSRRCYCCNGPILDKVVTALDRTWHPEHFFCAQC GAFFGP
EGFHEKDGKAYCRKDYDFMFAFKCGGCARAI LENYISALNTLWHPECFVCRECFTPFVN
GSFFEHDGQPYCEVHYHERRGSLCSGCQK PITGRCITAMAKKFHPEHFVCAFLKQLNK
GTFKEQNDKPYCQNCFLKLF
```

>未知配列

出席番号 1,4,7,10,13,16,19,22,25,28,31,34,37,40,43,46,49,52,55

```
MNNMNSLQDAEWYWGDISREEVNEKLRDTADGTF LVRDASTKMHGDYTLTLRKGGN
NKL I K I F H R D G K Y G F S D P L T F S S V V E L I N H Y R N E S L A Q Y N P K L D V K L L Y P V S K Y Q Q D Q
V V K E D N I E A V G K K L H E Y N T Q F Q E K S R E Y D R L Y E E Y T R T S Q E I Q M K R T A I E A F N E T I K I
F E E Q C Q T Q E R Y S K E Y I E K F K R E G N E K E I Q R I M H N Y D K L K S R I S E I I D S R R R L E E D L K K
Q A A E Y R E I D K R M N S I K P D L I Q L R K T R D Q Y L M W L T Q K G V R Q K K L N E W L G N E N T E D Q Y S L
V E D D E D L P H H D E K T W N V G S S N R N K A E N L L R G K R D G T F L V R E S S K Q G C Y A C S V V V D G E V
K H C V I N K T A T G Y G F A E P Y N L Y S S L K E L V L H Y Q H T S L V Q H N D S L N V T L A Y P V Y A Q Q R R
```

出席番号 2,5,8,11,14,17,20,23,26,29,32,35,38,41,44,47,50,53,56

```
DGGSPFLGRDRFVYPSSTRDPSASNGGGSPARRE EKKRKAARLKDFDFAQSPKELTL
QKGDIVYIHKVDKNWLEGEHHGRLGIFPANYVEVLP ADEIPKPIKPPTYQVLEYGAV
AQYTFKGDLEVELSFRKGEHICLIRKVNENWYEGRITGTGRQGFIPASYVQVSREPRLR
LCDDGPQLPTSPRLTAAARSARHPSSPSALRSPADPTDLGGQTSRRTGFSFPTQEP RP
QTQNFGTGPGALSHSRGSPHPLDLGTSSPNTSQIHWT PYRAMYQYRPNQNEDELELREGD
RVDVMQQCDDGWFVGVSRRTQKFGTFPGNYVAPV
```

出席番号 3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57

```
VNTDTLEAPGYELQVNGTEGEMEYEEITLERGNSGLGFSIAGGTDNPHIGDDPSIFI
TKIIPGGAAAQDGR LRVNDSILFVNEVDVREVTHSAAVEALKEAGSIVRLVYMRKPPA
EKVMEIKLIKPGKLGFSIAGGVGNQHIPGDNSIYVTKIEGGAHKDGR LQIGDKILA
VNSVGLVEDVMHEDAVAALKNTYDVVYLKVAKPSNAYLSDSYAPPDITTSYSQHLDNEIS
HSSYLGTDYPTAMPTSPRRYSPVAKDLLGEEDIPREPRRIVIRHGSTGLGFNIVGGED
GEGIFISFILAGGPADLSGELRKGQILSVNGVDLRNASHEQAALKNAGQTVTIIAQ
YKPEEYSRFEAKIHDLREQLMNSSLGS GTASLRSNPKRGFYIRALFDYDKTKDCGFLS
```

- ② SMART <http://smart.embl-heidelberg.de/>

(または NCBI CD search <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)

InterPro (<http://www.ebi.ac.uk/InterProScan/>)

を用いてドメインを検索する。

- ③ 配列の中で見つかったドメイン名(例1の場合 LIM)が表示される。蛋白質によってはいくつものドメインが見つかることもあるし、蛋白質によっては既知のドメインが見つからないこともある。
- ④ ドメインの機能について、a)ドメイン検索に用いたサイトを利用して b)PubMed など文献を利用して、調べてレポートに記す。(例えば、LIMドメインは2つの Znフィンガーを含む構造をしており、転写調節や細胞骨格の調節に働き、,,,))
- ⑤ 見つかったドメインが未知配列のおよそ何番目から何番目までアミノ酸に相当したか。

質問

- ① 例1と未知配列にはそれぞれどのようなドメインが含まれていたか？また、例1と未知配列のそれぞれ何番目のアミノ酸から何番目のアミノ酸までが該当のドメインであったか？
- ② 見つかったドメインはどのような機能を持っているか。
- ③ 未知配列の等電点と分子量を以下のサイトの「Compute pI/Mw」を利用して予測しなさい。
<http://www.expasy.org/tools/>

上記課題について文献を調べる場合は

文献検索 <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

このサイトでキーワードとして例えば“SH3 Review”といれて検索すると、SH3 という言葉を含んだ Review(総説, 解説)がリストアップされる。Review は普通の論文に比べ、解説が多い。

図書館・オンラインジャーナル

京都大学附属図書館 <http://www3.kulib.kyoto-u.ac.jp/>

農学部図書室 <http://www.agril.kais.kyoto-u.ac.jp/>

医学図書館 <http://www.lib.med.kyoto-u.ac.jp/>

実習(2) 蛋白質のマルチプルアライメント

未知配列中に見つけたドメインがどのような蛋白質ととくに相同性が高いのか、また、ドメイン中のどの配列がとくに保存されているのかを知るために、未知配列を含む複数の蛋白質のアミノ酸配列を並べる(マルチプルアライメント)。

そのために、まず BLASTP を用いて類似配列を集め、その後 ClustalW を用いてマルチプルアライメントを作成する。

① 実習1で各自に割り振られたアミノ酸配列のうち、見出したドメインに相当する領域を1つ(例えば例1の場合、4番目 LIM 領域に相当する195番目から最後までのアミノ酸配列)を copy&paste で切り取っておく。

② ①の配列を用いて、BlastP を使い興味のある類似配列を4つ集める

- BlastP は <http://www.ncbi.nlm.nih.gov/BLAST/> (または <http://blast.genome.jp/>) のサイトを用いる。
- データベースは SwissProt を選ぶ。(より多くものを検索するためにはデータベースに nr-aa を選択する方が良いが、SwissProt は注釈等が詳しいので今回はこのデータベースに限る)
- リストアップされたタンパク質のうち、最も相同性の高かった配列を一つと、それ以外に自分が興味を持ったタンパク質を3つ選択する。
- 選択したタンパク質の DEFINITION(蛋白質名), SOURCE(種)を記録する。

③ マルチプルアライメント検索用ファイルを作成する。

- アミノ酸配列の中から、目的とするドメインに相当する部分を COPY する。下図左にあるように、BlastP の結果のページで Query に用いた配列がヒットした配列の何番目のアミノ酸から何番目のアミノ酸と相同性があったのかをメモする。その領域に相当する部分が予想通り(実習1)で見出したドメインになっていることを、タンパク質のページの REGION の項目で確認する(図 11 右)。

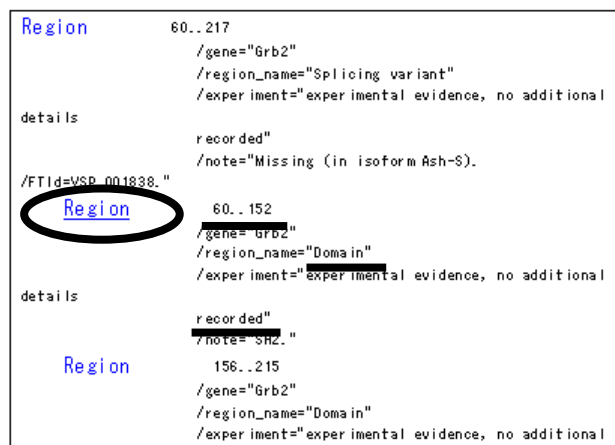
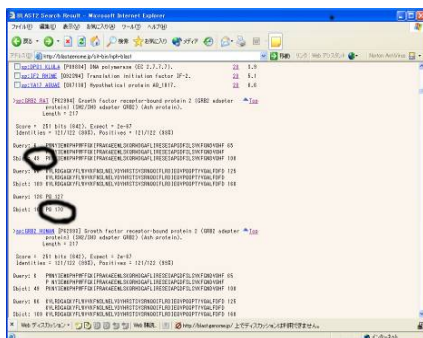


図 11

- ドメインに相当する領域を Copy&paste で Word あるいはメモ帳ファイルにコピーして、その前に ">TEST" 等、蛋白質あるいは遺伝子の名前を入れる。(このとき、名前の中の全角、スペース、ハイフンは避ける。配列中のスペースは問題なし。)

- 上記を選んだ4つすべてのタンパク質についてくり返す。下図のようなファイルができるはずである。

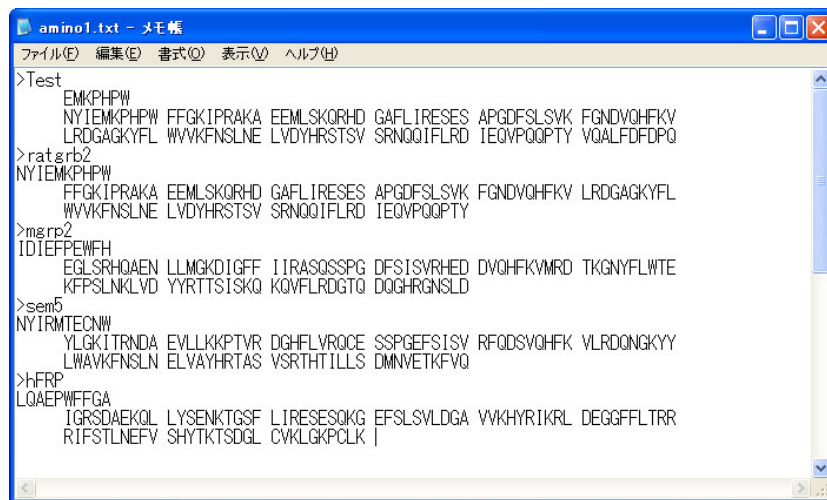


図 12

④ マルチプルアライメントを作成する。

- <http://clustalw.ddbj.nig.ac.jp/top-j.html> (または <http://www.genome.jp/tools/clustalw/>) で
③で作成したファイルを送信する。

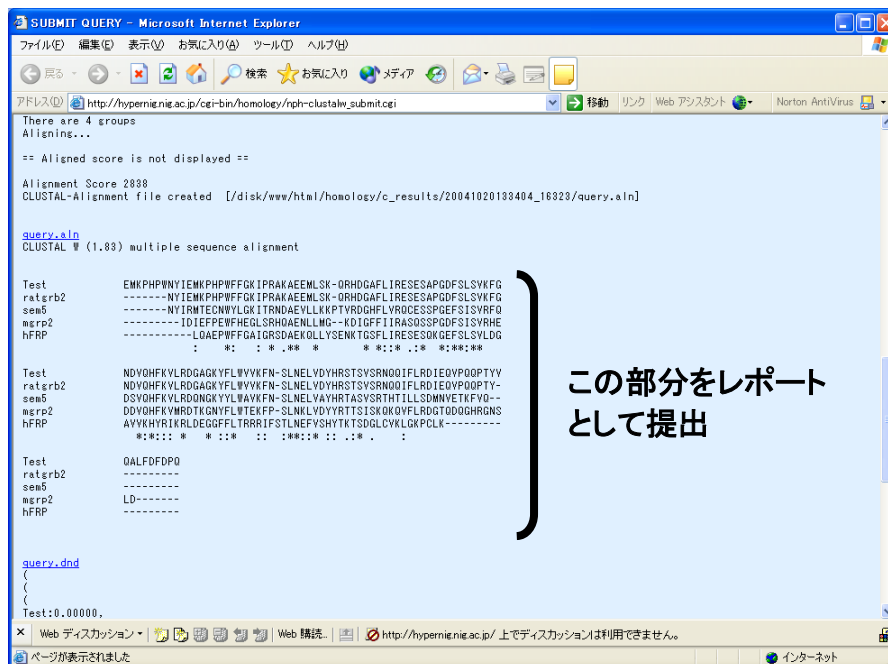


図 13

- 5つの配列がアライメントされたものが表示される。"*" は5つの配列すべてで同じであったアミノ酸, ":" は性質の似たアミノ酸に置換されているが非常によく保存されているアミノ酸, "."は ":" 程ではないが一応保存されたアミノ酸, を示す。よく保存されているアミノ

酸はそのドメインが機能するために重要な役割を果たしていると考えられる。

- ⑤ アライメントされている領域を Word の書類に Copy&Paste 後保存し、レポートに貼り付けて提出する。(もし、アライメントがずれるなら左右のマージンを変更したり、フォントサイズを変更したり、ページ設定で紙を横にしてみる。どうしてもずれる場合は表示されている Web のページを"名前をつけて保存"を選び、"ファイルの種類"を"テキストファイル"と選択した後、適当な名前で保存する。このファイルを"Word"で開き、マルチプルアライメント以外の部分を削除し、保存し直す。)

実習(3) 膜貫通ドメインを探す

実習1-2で検索してきたような機能ドメイン以外に、細胞内局在を決定するドメインを知ることは未知蛋白質の機能を推定する上で非常に重要である。今回は膜貫通ドメインに限って検索し、未知蛋白質が膜蛋白質かどうか SOSUI システムを用いて判定する。膜貫通領域は疎水性が高く、またヘリックス構造をとるので、これを指標に膜貫通領域を推定する。

「SOSui は、美宅成樹教授を中心とする研究グループによって開発された、膜タンパク質やダンベル型タンパク質の判別と2次構造予測、およびシグナルペプチドの予測を行うソフトウェアです。

既知の配列との相同性比較やニューラルネットワークなどを用いた予測法と異なり、新規に定義したものを含む複数の指標を用いることで、従来の手法よりも高速に予測・判別を行うことが可能です。たとえば、Kyte-Doolittle の疎水性指標と、新規に定義した両親媒性指標ならびに非電荷指標を用いることにより、従来の手法に比べて高い成績での膜タンパク質と水溶性タンパク質の判別を実現しています。」(三井情報株式会社), だそうです。

- ① 未知配列(unknown_membrane)を含むファイルを開く。

```
>unknown_membrane
malrgfcsad gsdplwdwnv twntsnpdf t kcfqntvlw vpcfylwacf pfyflylsrh
drgyiqmtp1 nktktalglf lwivcwadlf ysfwersrgi flapvflvsp tllgitmlla
tfliqlerrk gvqssgimlt fwlvalvcal ailrskimta lkedaqvdlf rditfyvyfs
llliqlvlsc fsdrspflse tihdnpncpe ssasflsrit fwwitglivr gyrqplegsd
lwslnkedts eqvvpvlvkn wkkecaktrk qpvkvvyssk dpaqpkessk vdaneeveal
ivkspqkewn pslfkvlykt fgpyflmsff fkaihdlmmf sgpqilklli kfvndtkapd
wqgyfytvll fvtaclqtlv lhqyfhicfv sgmriktavi gavyrkalvi tnsarksstv
geivnlmsvd aqrfmdlaty inmiwsaplq vilalylwl nlgpsvlagv avmvlmvpvn
avmamtkty qvahmkskdn riklmneiln gikvlklyaw elafkdkvla irqeelkvlk
ksaylsavgt ftwvctpflv alctfavyvt idennildaq tafvslalfn ilrfplnilp
mvisivqas vslkririfl sheelepdsi errpvkdggg tnsitvrnat ftwarsdppt
lgitfsipe galvavvgqv gcgkssllsa llaemdkveg hvaikgsvay vpqqawiqnd
slrenilfgc qleepyysv iqacallpdl eilpsgdrte igekgvnlsg gqkqrslar
avsynadiyl fddplsavda hvgkhifenv igpkgmlknk trilvthsms ylpqvdviiv
msggkisemg syqellardg afaeflrtya steqeadae ngvtgvsppg keakqmengm
lvtdsagkql qrqlsssssy sgdisrhhs taelqkaeak keetwklmea dkaqtgqvkl
svywdymkai glfisflsif lfmchnvsal asnywlslwt ddpivngtqe htkvrlsvgy
```

algisqgiav fgysmavsig gilasrcclhv dllhsilrsp msffertpsg nlnvrfskel
dtvdsmipev ikmfmgslfn vigacivill atpiaaiiip plgliyffvq rfyvassrql
krlesvsrsp vyshfnetll gvsvirafee qerfihqsd1 kvdenqkayy psivanrwla
vrlecvgnci vlfaalfavi srhslsaglv glsvsyslvq ttylnwlvrm ssemetniva
verlkeyset ekeapwqiqe tappsswpqv grvefrnycl ryredldfvl rhinvtingg
ekvgivgrtg agkssl1t1gl frinesaege iidginiak iglhdlrfki tiipqdpvlf
sgslrmnldp fsqysdeevw tslelahlkd fvsalpdkld hecaeggenl svgqrqlvcl
arallrktki lvldeataav dletddliqs tirtqfedct vltiahr1nt imdytrvivi
dkgeiqeyga psdllqqr1l fysmakdagl v

- ② <http://harrier.nagahama-i-bio.ac.jp/sosui/> を開く. または
http://www.ch.embnet.org/software/TMPRED_form.html を開く
- ③ SOSUI を選択した後, 未知配列(配列名: membrane)の膜貫通領域を検索する.
- ④ 予想される膜貫通ドメインが表と図で表示される. 表中 Primary は膜貫通の性質が強いタイプ,
Secondary は Primary が存在する状況下で膜貫通する可能性の高い領域を示す.

質問

- ① 未知配列(unknown_membrane)は膜タンパク質と予想されたか? そうだとしたら何回膜貫通タンパク質であったか?
- ② 何番目のアミノ酸から何番目のアミノ酸が膜貫通領域と推定されたか?
- ③ ハイドロパシープロファイルとは何か? 説明せよ.
- ④ Option(未知配列はどのような機能をもつタンパク質であったか?)

実習(4) 自分たちが増幅した塩基配列がヒト染色体のどこにあるかを調べる

1.8.1 で増幅した繰り返し配列をヒトゲノム配列に対して BLAST 検索を行い, PCR にて増幅した配列が第何番目染色体由来であったかを調べる. 本年度は塩基配列決定実習をクローン化した大腸菌染色体で行なうため, 以下に過去に決定した配列を示すので, これを用いて実習を行なうこと.

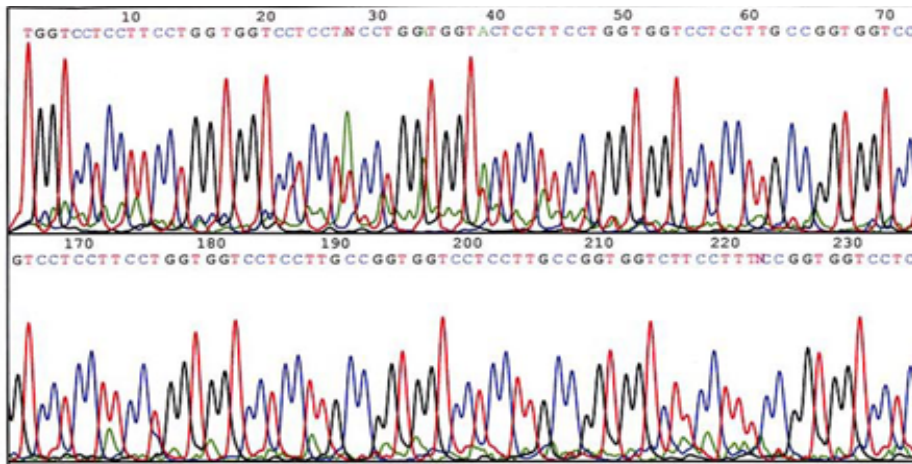
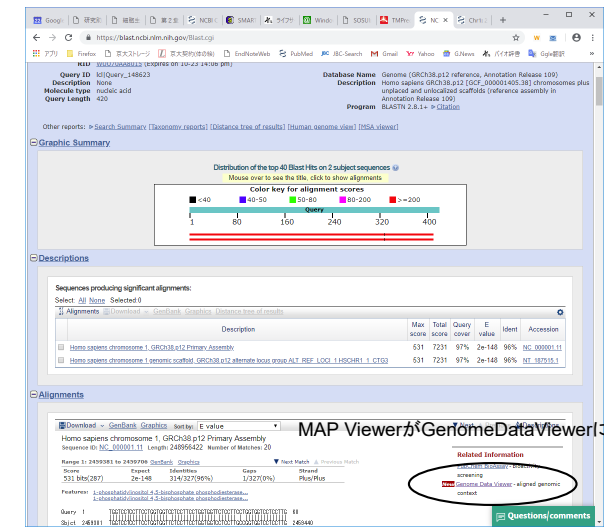


図 14

>Sequence

TGGTCTCCTTCCTGGTGGTCTCCTTCCTG
 GTGGTTCTCCTTCCTGGTGGTCTCCTTG
 CCGGTGGTCTCCTTCCTGGTGGTTCCTT
 CCTGGTGGTCTCCTTCCTGGTGGTCTC
 CTTCTGGTGGTCTCCTTCCTGGTGGTCT
 CCTCCTGGTGGTCTCCTTCCTGGTGGT
 CCTCCTTCCTGGTGGTCTCCTTCCTGGTGG
 TCTTCCTTTCCGGTGGTCTCCTTCCTG
 TGGTCTCCTTCCTGGTGGTCTCCTTCCTG
 GTGGTCTTCCTTCCTGGTGGTCTCCTTC
 CCGGTGGTCTCCTATGCCTGTGGTCTGTCT
 ATGCCTTGGNCTGNCTCTGACCCAGCCC
 TTACTCTNANTGACTCCTNTCTCATGGNGA
 TTCCGCTCCTTAGTGATCCNANGCTGGA
 TNGTCCGGCCGTGGGACNNCGGCNANCNGT
 NNNGGCNAAGCNNNNCTNNTNNNNCTCGG
 CCTCGNCCCTNNNTNCCTNNCCCTCCTCCC
 TCCTTCCTNNTCTCTCCTCTCCTCCNC
 TCTCCTNCTCTTCTCTCTC

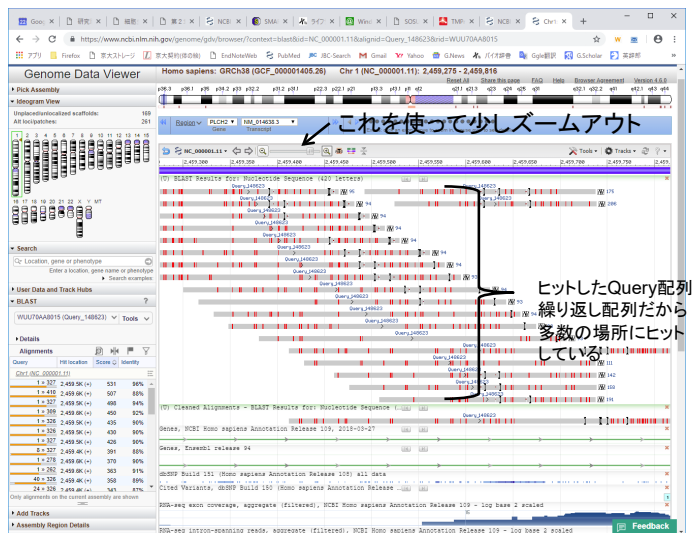
Page 60 ⑥ Map ViewerがGenome Data Viewerに変更されてる



① <http://www.ncbi.nlm.nih.gov/BLAST/>
 を開く

② Genomes の中の human を選択す
 る

③ 決定した塩基配列のうち信頼性の
 高い領域 (N を殆ど含まない領域)
 を COPY&PASTE で Query 欄に入



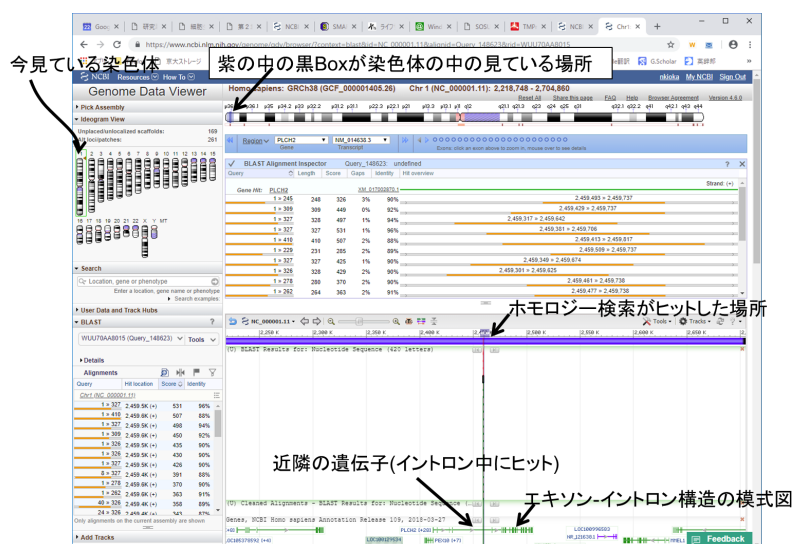
れ, Algorithm Parameters の Filter のうち Low complexity regions の✓をはずす(繰り返し配列なので, 通常のフィルターをかけると除外されてしまう) . 検索を開始する

- ④ いろいろな情報を獲得することができるが, ここでは Alignments 欄の右 Map Viewer(図15上)をクリックして情報を得る。
- ⑤ PCR で増幅した断片は何番目の染色体にあったか? また, その近傍にはどのような遺伝子があったか? 近傍の遺伝子は染色体上のどの領域(8p21.3 など)に位置するものであったか?. (ちなみに p は短腕, q は長腕を意味する)

図 15

本日のコンピューター実習のレポートには少なくとも以下の事項を記載すること

- ① 例 1, および未知配列から見つけた「ドメインの機能」をコンピューター上で調べ, 説明せよ。
- ② 未知配列を用いて作成したマルチプルアライメント
- ③ マルチプルアライメントに用いた4つの蛋白質の機能はそれぞれ何か.(簡単に)
- ④ Membrane 配列中にはいくつの膜貫通ドメインが含まれているとよそうされたか(グラフィックを打ち出したものも貼付, あるいは, 模式図を描く)? これは文献に出ている膜貫通領域の数と一致しているか?



参考図書 できるバイオインフォマティクス(中山書店)

文献検索 <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

図書館・オンラインジャーナル

京都大学附属図書館 <http://www.kulib.kyoto-u.ac.jp>

農学部図書室 <http://www.agril.kais.kyoto-u.ac.jp>

医学図書館 <http://www.lib.med.kyoto-u.ac.jp>