生物情報工学 *BioInformatics*

4

遺伝子解析のためのツール

遺伝子解析のためのツール

各種データベースの活用 (2015年度版)

第2回目（10/14）文献データベースと特許データベース

第3回目（10/21）遺伝子データベース

第4回目（10/28）遺伝子解析のためのツール

第5回目（11/4）ゲノムデータベース

リンク集

データベース検索：

1. PubMed: 論文検索
2. NCBI databases：総合データベース
3. Google Scholar：文献データベース
4. 特許情報プラットフォーム ：特許データベース

ホモロジー検索：

1. BLAST [GenomeNET]
2. FASTA [GenomeNET]

配列解析：

1. DNA → AA ：DNA配列をアミノ酸配列に変換
2. GENSCAN：スプライシングの予測（新）

今日のメニュー

- 制限酵素切断部位の検索
- 塩基配列をアミノ酸配列に変換する
- Open reading frameの検索
- スプライシングの予測
- 転写因子結合部の予測

演習：制限酵素切断部位の検索とマップ作成

- リンク集:制限酵素マップのNEB Cutterを使う。
- データはプラスミドpUC18を使おう。

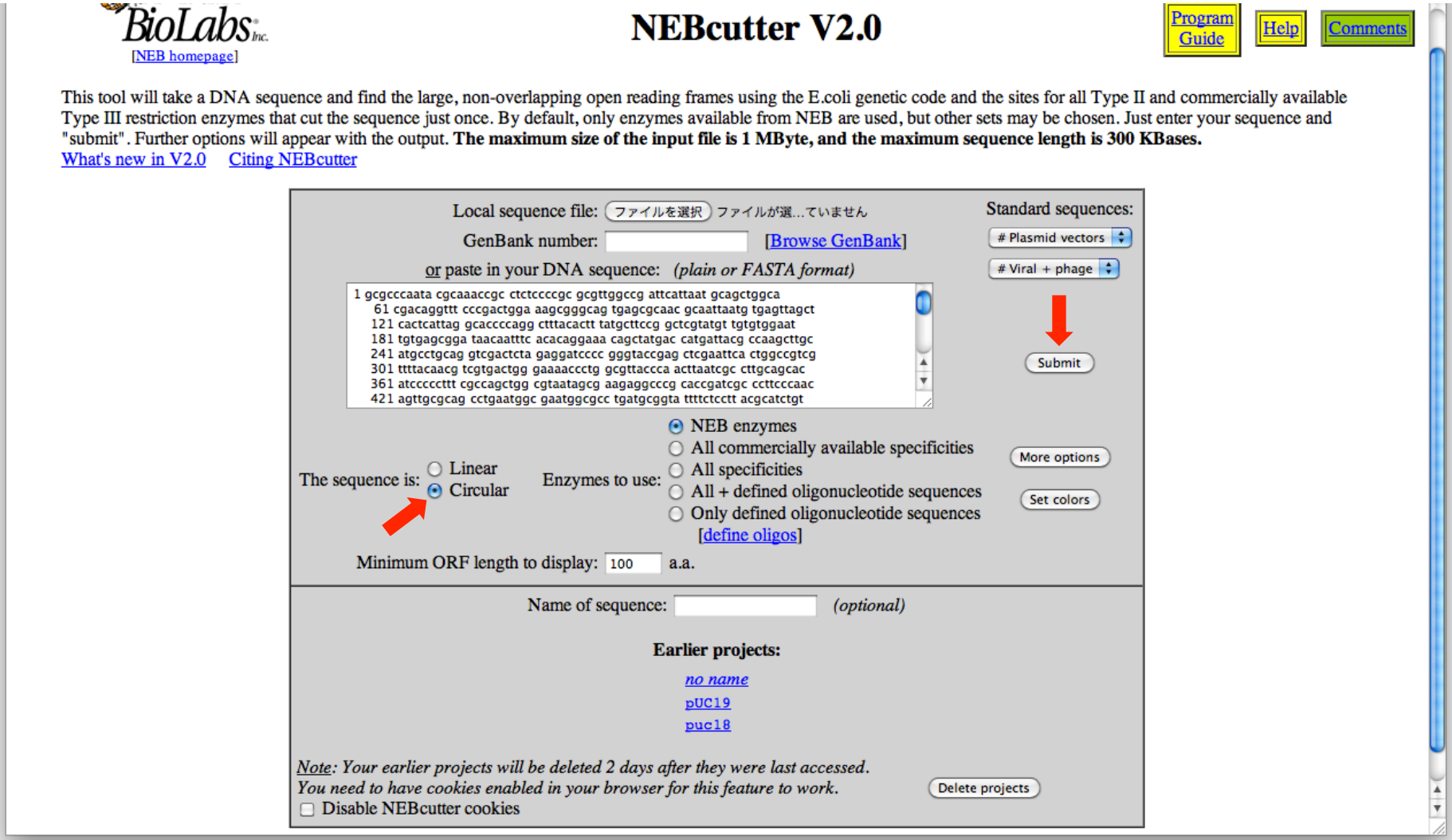
まずは、先週テキスト保存したpUC18の情報を開く

```
                1- 229  1069-1297      Lac-Operon
              230- 286     1-  57      polylinker of M13mp18
              289- 447  1303-1461      Lac-Operon
              448- 547  2351-2252  (c) pBR322
              548- 684  2210-2074  (c) pBR322
              685-2686  4355-2354  (c) pBR322
            Conflict (cfl) and Mutations (mut):
                 pUC18  source
            mut 1128 T  C 3912 (c) pBR322
            mut 1429 A  G 3611 (c) pBR322
         FEATURE
            952-1740 1-789 Ap-R; b-lactamase
         POLYLINKER  EcoRI-SacI-KpnI-SmaI-BamHI-XbaI-SalI-PstI-SphI-HindIII
         SELECTION
            #resistance  Ap
            #indicator beta-galactosidase
         SUMMARY  pUC18    #length 2686   #checksum 5464.
FEATURES             Location/Qualifiers
     source          1..2686
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN
        1 gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat gcagctggca
       61 cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct
      121 cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat
      181 tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg aattcgagct
      241 cggtacccgg ggatcctcta gagtcgacct gcaggcatgc aagcttggca ctggccgtcg
      301 ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc cttgcagcac
      361 atccccctttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac
      421 agttgcgcag cctgaatggc gaatggcgcc tgatgcggta ttttctcctt acgcatctgt
      481 gcggtatttc acaccgcata tggtgcactc tcagtacaat ctgctctgat gccgcatagt
      541 taagccagcc ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct tgtctgctcc
      601 cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt
      661 caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgccta tttttatagg
      721 ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaaatgtgc
      781 gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac
      841 aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt
      901 tccgtgtcgc ccttattccc ttttttgcgg catttgcct tcctgttttt gctcacccag
```

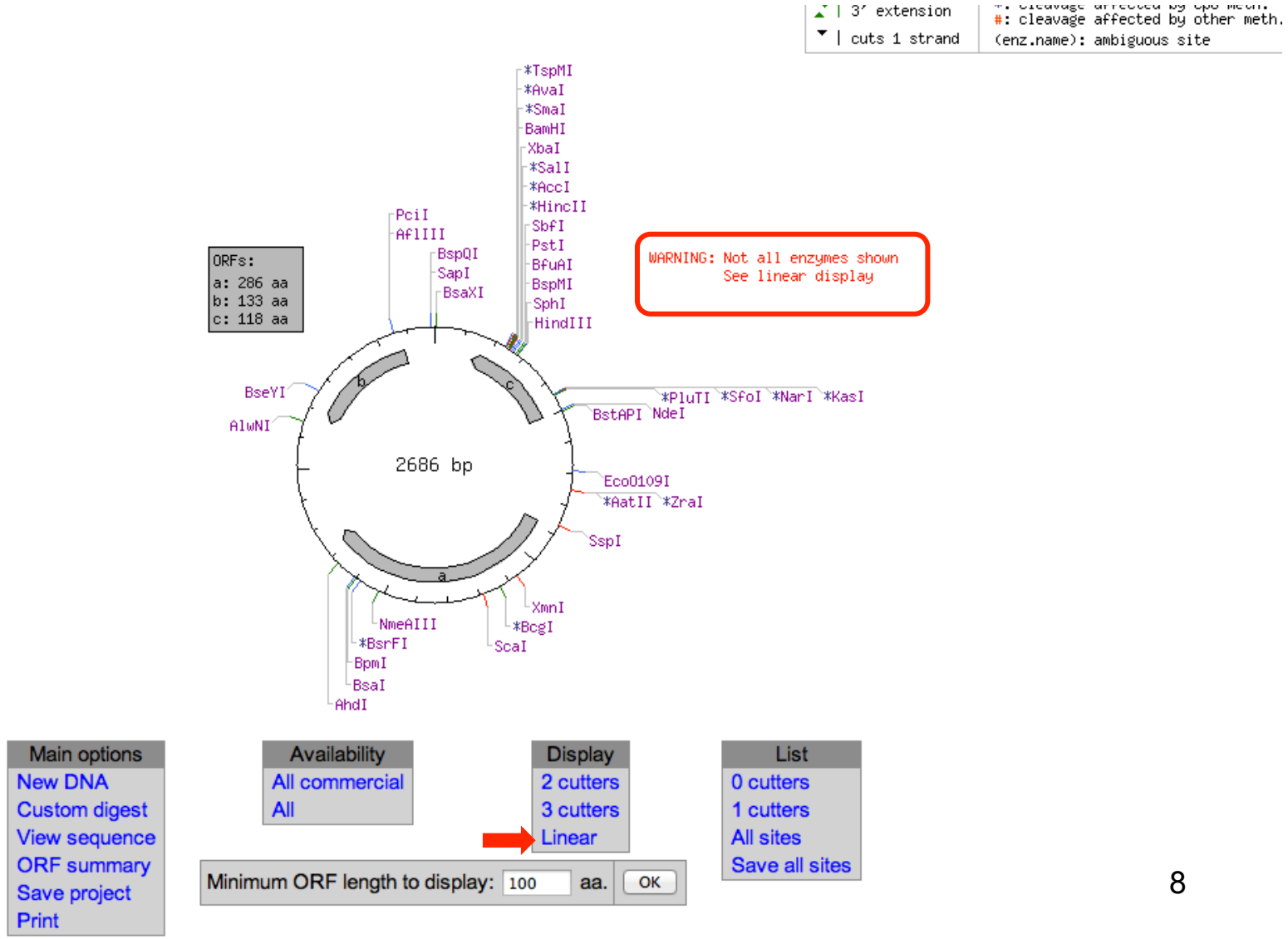
配列部分をコピーしておく

```
                448- 547  2351-2252  (c) pBR322
                548- 684  2210-2074  (c) pBR322
                685-2686  4355-2354  (c) pBR322
              Conflict (cfl) and Mutations (mut):
                   pUC18  source
              mut 1128 T  C 3912 (c) pBR322
              mut 1429 A  G 3611 (c) pBR322
            FEATURE
              952-1740 1-789 Ap-R; b-lactamase
            POLYLINKER  EcoRI-SacI-KpnI-SmaI-BamHI-XbaI-SalI-PstI-SphI-HindIII
            SELECTION
              #resistance  Ap
              #indicator beta-galactosidase
            SUMMARY  pUC18    #length 2686   #checksum 5464.
FEATURES             Location/Qualifiers
     source          1..2686
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN
        1 gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat gcagctggca
       61 cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct
      121 cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat
      181 tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg aattcgagct
      241 cggtacccgg ggatcctcta gagtcgacct gcaggcatgc aagcttggca ctggccgtcg
      301 ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc cttgcagcac
      361 atccccctttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac
      421 agttgcgcag cctgaatggc gaatggcgcc tgatgcggta ttttctcctt acgcatctgt
      481 gcggtatttc acaccgcata tggtgcactc tcagtacaat ctgctctgat gccgcatagt
      541 taagccagcc ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct tgtctgctcc
      601 cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt
      661 caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgccta tttttatagg
      721 ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaaatgtgc
      781 gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac
      841 aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt
      901 tccgtgtcgc ccttattccc ttttttgcgg catttgcct tcctgttttt gctcacccag
      961 aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg ggttacatcg
     1021 aactggatct caacagcggt aagatccttg agagttttcg ccccgaagaa cgttttccaa
     1081 tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt atcccgtatt gacgccgggc
```

NEB cutterを開き、配列をペーストする



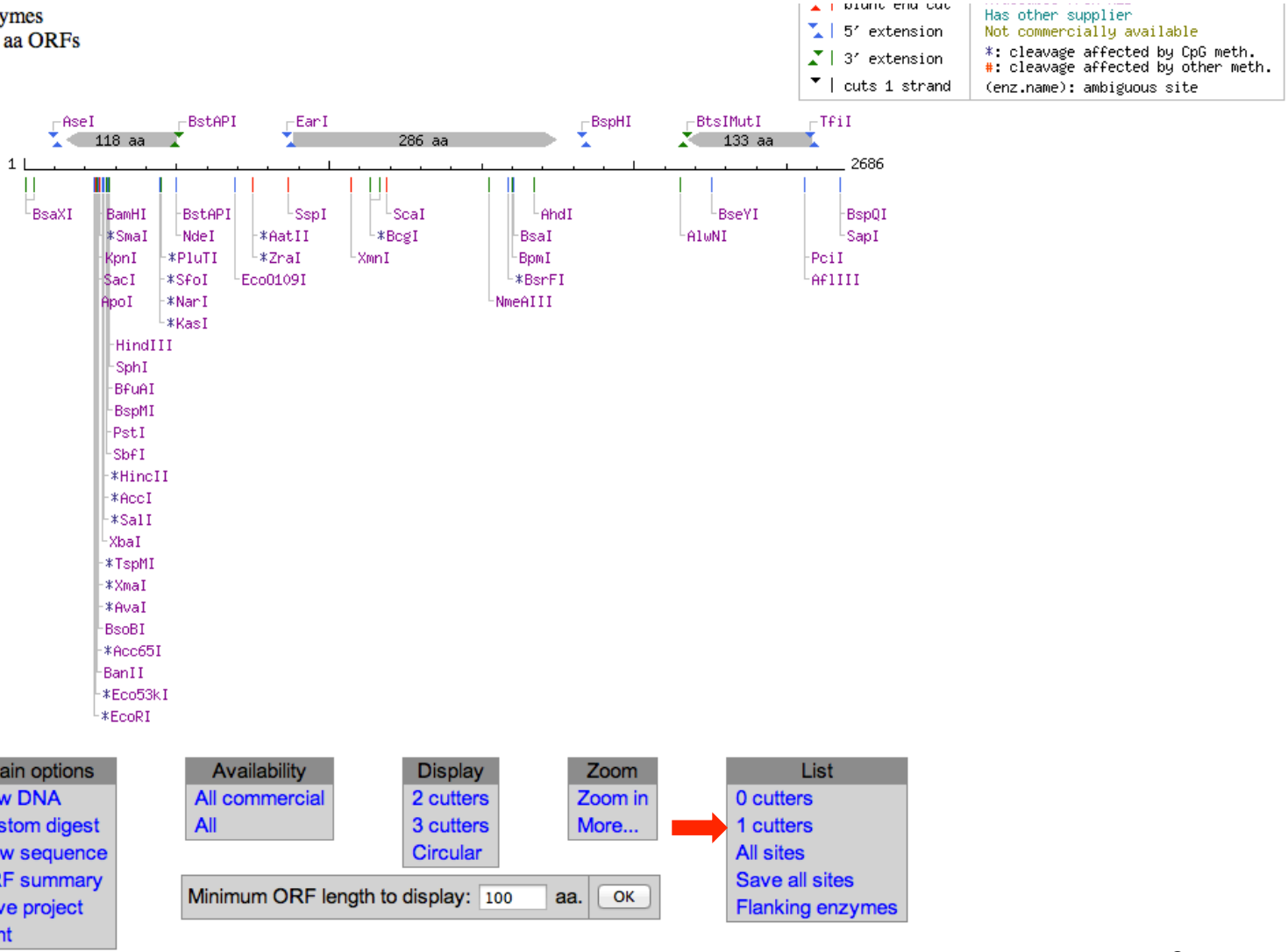
結果(1)


3′ extension
cuts 1 strand
#: cleavage affected by other meth.
(enz.name): ambiguous site
ORFs:
a: 286 aa
b: 133 aa
c: 118 aa
WARNING: Not all enzymes shown
See linear display
*TspMI
*AvaI
*SmaI
BamHI
XbaI
*SalI
*AccI
*HincII
SbfI
PstI
BfuAI
BspMI
SphI
HindIII
PciI
AflIII
BspQI
SapI
BsaXI
BseYI
AlwNI
*PluTI
*SfoI
*NarI
*KasI
BstAPI
NdeI
2686 bp
Eco0109I
*AatII
*ZraI
SspI
XmnI
*BcgI
ScaI
NmeAIII
*BsrFI
BpmI
BsaI
AhdI
a
b
c
Main options
New DNA
Custom digest
View sequence
ORF summary
Save project
Print
Availability
All commercial
All
Display
2 cutters
3 cutters
Linear
List
0 cutters
1 cutters
All sites
Save all sites
Minimum ORF length to display:
100
aa.
OK

結果（2）

Display: - NEB single cutter restriction enzymes
- Main non-overlapping, min. 100 aa ORFs

GC=51%, AT=49%



結果(3)

NEW ENGLAND BioLabs Inc. NEBcutter

[Back to main display]

Single cutters

unnamed sequence

Help | Comments

Number of cuts = 1 OK Sort order: Alphabetical Save as text file

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AatII	G_ACGT^C	list	*753/749
2	Acc65I	G^GTAC_C	list	*242/246
3	AccI	GT^MK_AC	list	*264/266
4	AflIII	A^CRYG_T	list	2560/2564
5	AhdI	GACNN_N^NNGTC	list	1672/1671
6	AlwNI	CAG_NNN^CTG	list	2151/2148
7	ApoI	R^AATT_Y	list	230/234
8	AvaI	C^YCGR_G	list	*246/250
9	BamHI	G^GATC_C	list	251/255
10	BanII	G_RGCY^C	list	240/236
11	BcgI	_NN^(N)$_{10}$CGA(N)$_6$TGC(N)$_{10}$_NN^	list	*1134/1132+1168/1166
12	BfuAI	ACCTGCNNNN^NNNN_	list	276/280
13	BpmI	CTGGAG(N)$_{14}$_NN^	list	1603/1601
14	BsaI	GGTCTCN^NNNN_	list	1606/1610
15	BsaXI	_NNN^(N)$_9$AC(N)$_5$CTCC(N)$_7$_NNN^	list	6/3+36/33
16	BseYI	C^CCAG_C	list	2256/2260
17	BsoBI	C^YCGR_G	list	246/250
18	BspMI	ACCTGCNNNN^NNNN_	list	276/280
19	BspQI	GCTCTTCN^NNN_	list	2677/2680
20	BsrFI	R^CCGG_Y	list	*1587/1591

結果(4)

NEW ENGLAND BioLabs Inc. NEBcutter

[Back to main display]

BamHI

Help | Comments

```
5'... G▼G A T C C ... 3'
3'... C C T A G▲G ... 5'
```

突出末端 (protruding end)
または
粘着末端(sticky end)

Available from NEB, Catalog # R0136

View product page | 5 minute Time-Saver | REBASE enzyme page | Methylation Sensitivity

Buffer name: NEBuffer 3.1
Salt: 100 mM NaCl
Main: 50 mM Tris-HCl
pH: 7.9
Mg: 10 mM MgCl2
BSA: 100

Reaction temperature: 37 °C

Overlapping methylation:

NOT ANALYZED

Neoschizomers:

Isoschizomers:

結果(5)

NEW ENGLAND BioLabs Inc. NEBcutter

[Back to main display]

SmaI

```
5'... C C C▼G G G ... 3'
3'... G G G▲C C C ... 5'
```

Help | Comments

平滑末端 (blunt end)

Available from NEB, Catalog # R0141

View product page | 5 minute Time-Saver | REBASE enzyme page | Methylation Sensitivity

Buffer name: CutSmart Buffer

Salt: 50 mM KOAc
Main: 20 mM Tris-OAc
pH: 7.9
Mg: 10 mM MgOAc
BSA: 100

Reaction temperature: 25 °C

Overlapping methylation:

NOT ANALYZED

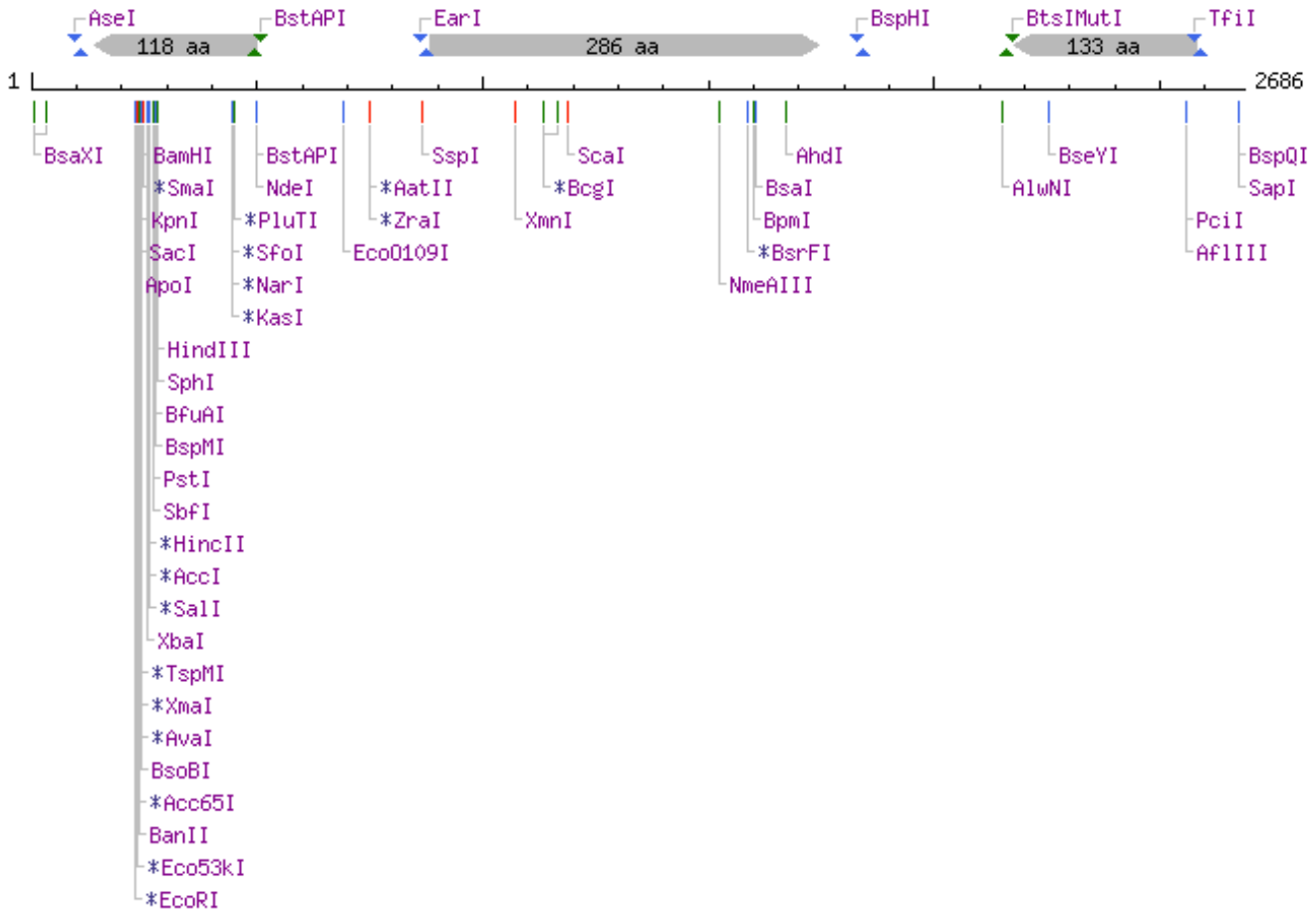
Neoschizomers:

	5'... C▼C C G G G ... 3'

Isoschizomers:

パターンが違う場合はリンク集からコピーする

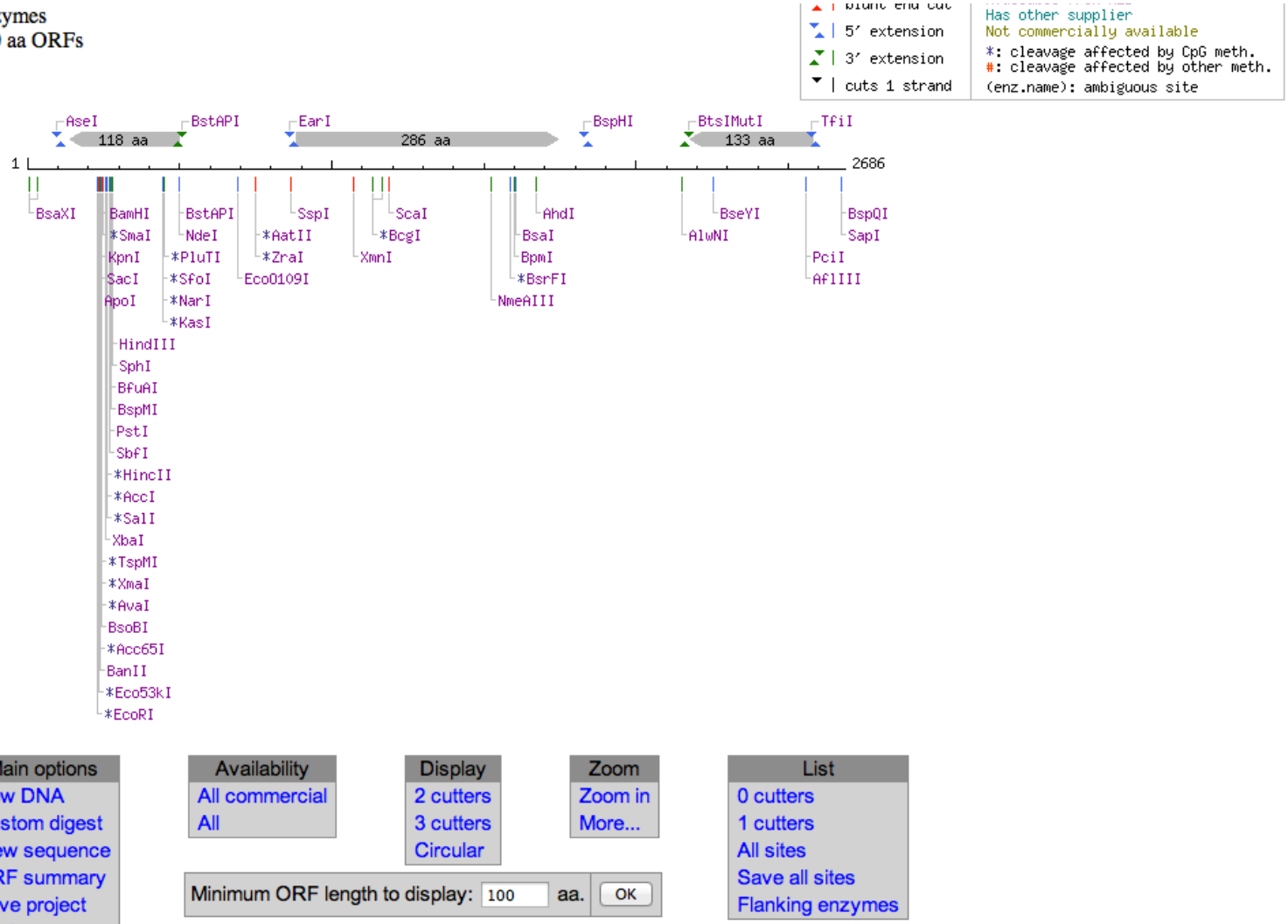
GCGCCCで始まる配列



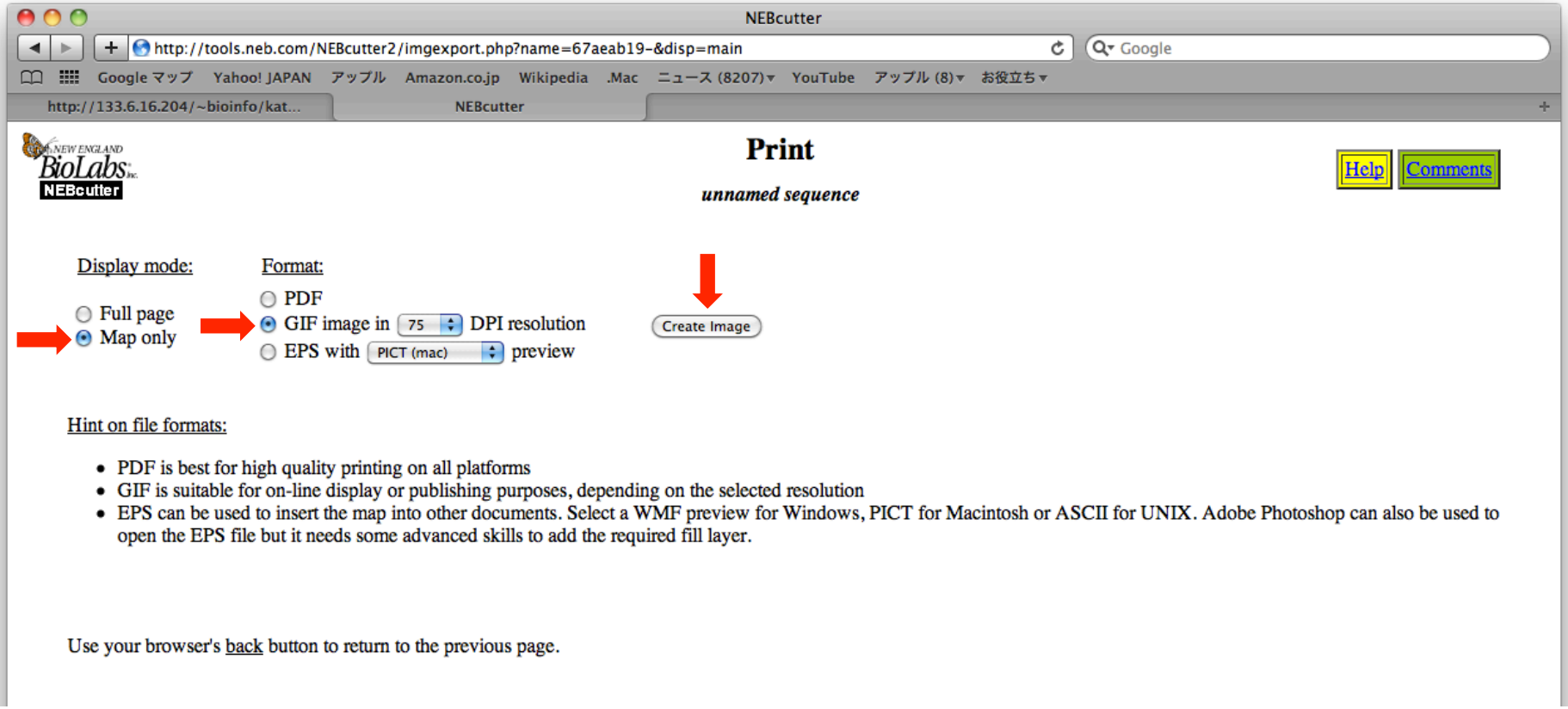
画像ファイルをダウンロードする（1）

Display: - NEB single cutter restriction enzymes
- Main non-overlapping, min. 100 aa ORFs

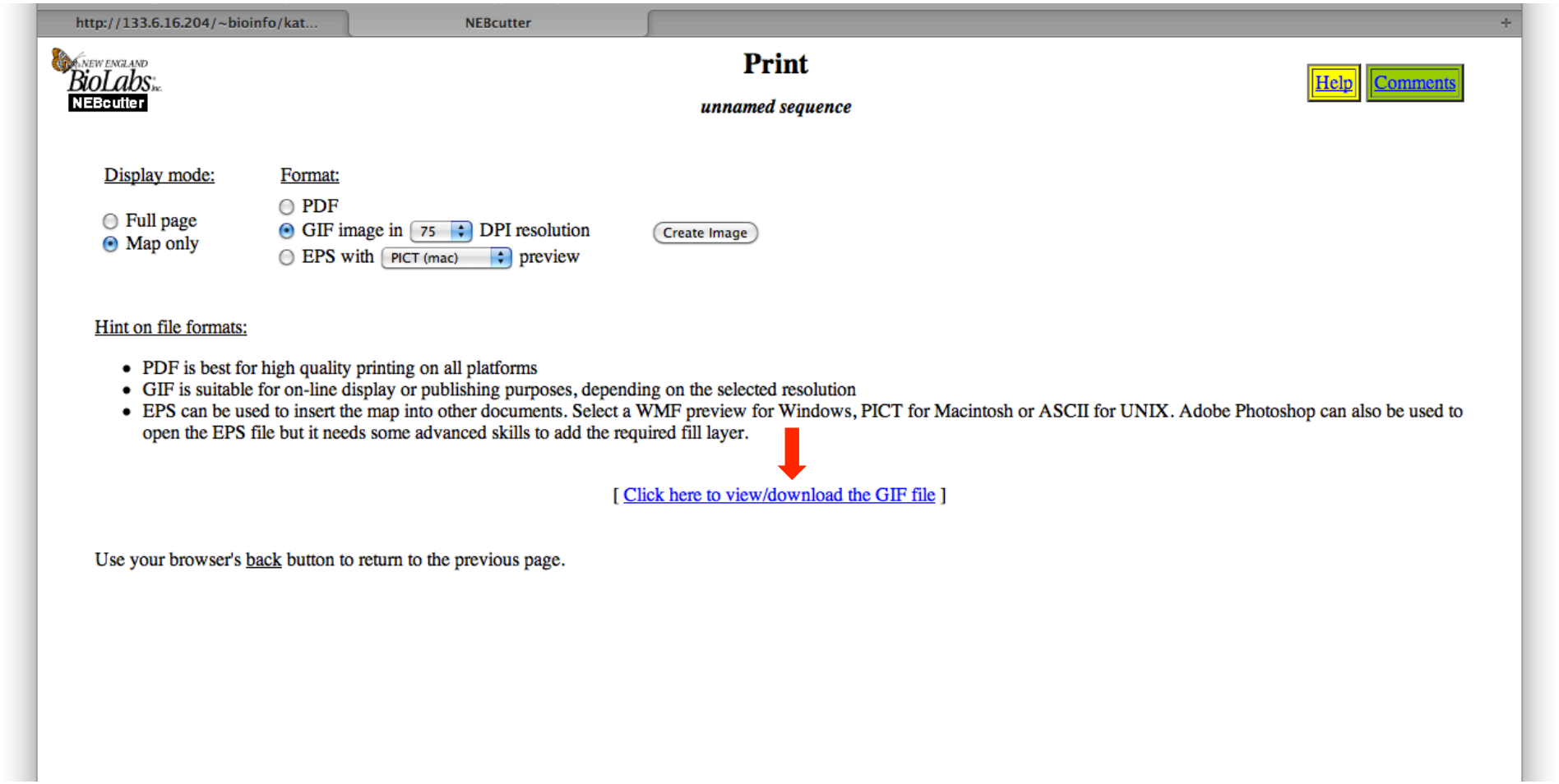
GC=51%, AT=49%



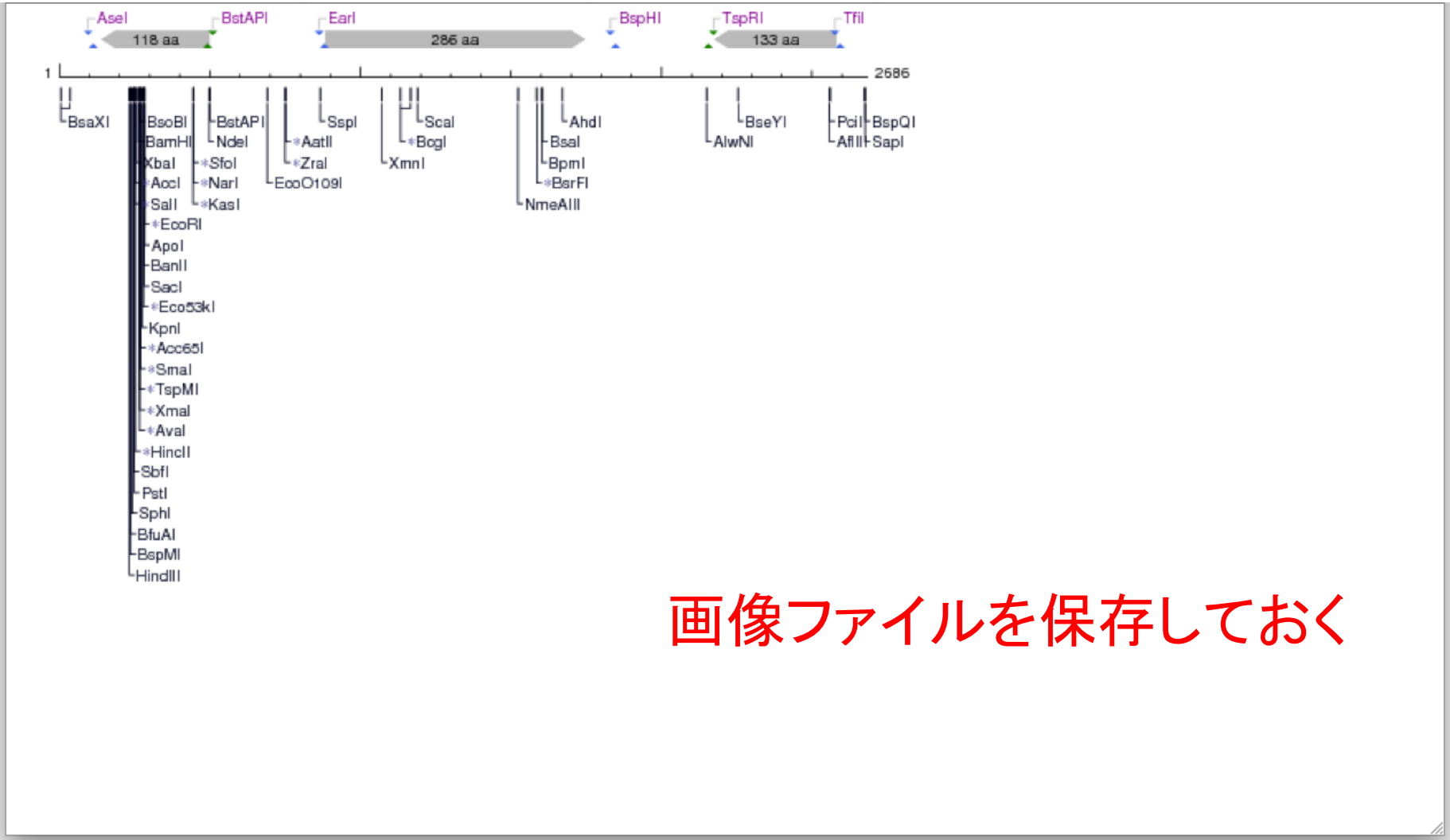
画像ファイルをダウンロードする(2)



画像ファイルをダウンロードする(3)



結果(4)



画像ファイルを保存しておく

課題

- pUC19の情報を取得し、制限酵素地図のイメージ(linear)を作成、画像ファイルを保存する。
- pUC18とpUC19の制限酵素地図をパワーポイントのスライドに並べてに貼り付け、スライドを作成して下さい。
- PDFファイルに変換し、メールに添付する。
- 件名は「講義４課題１」とする。
- pUC18とpUC19の違いに気がついた人は、メールの本文に記載して下さい。

まずは、pUC１９の情報をINSDICから取得して塩基配列をコピーする

・リンク集の“ゲノムネットWWWサーバー”へ

・DBGET searchへ

・検索対象のデータベースを”INSDIC”に設定

・pUC19を検索（必要に応じてテキストを追加

（例：cloning vectorなど）

・検索結果からpUC19の完全長配列を含むものとを選抜

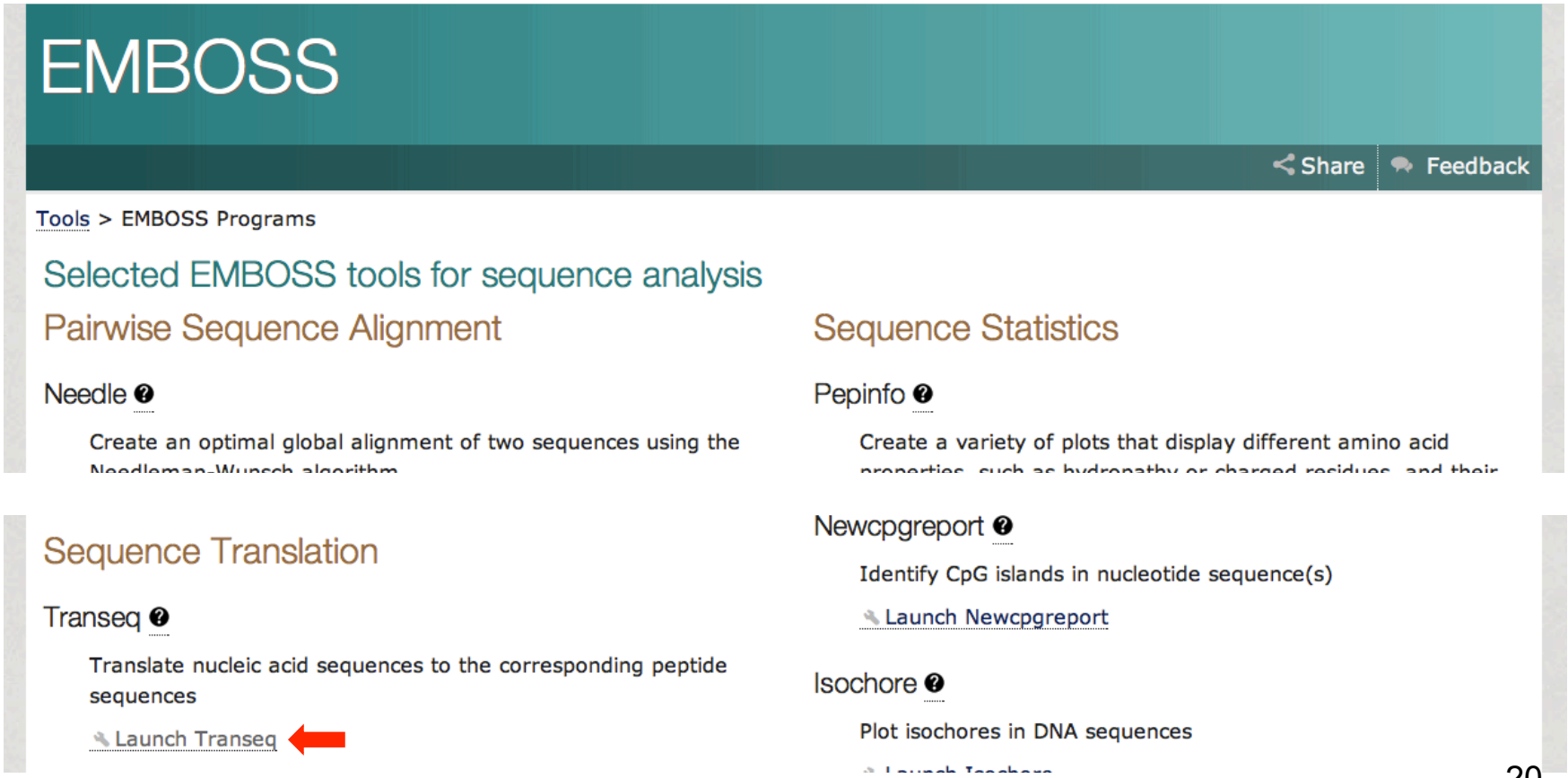
・配列部分（GCGCCCで始まる配列）をコピー。

演習：塩基配列をアミノ酸配列に変換する

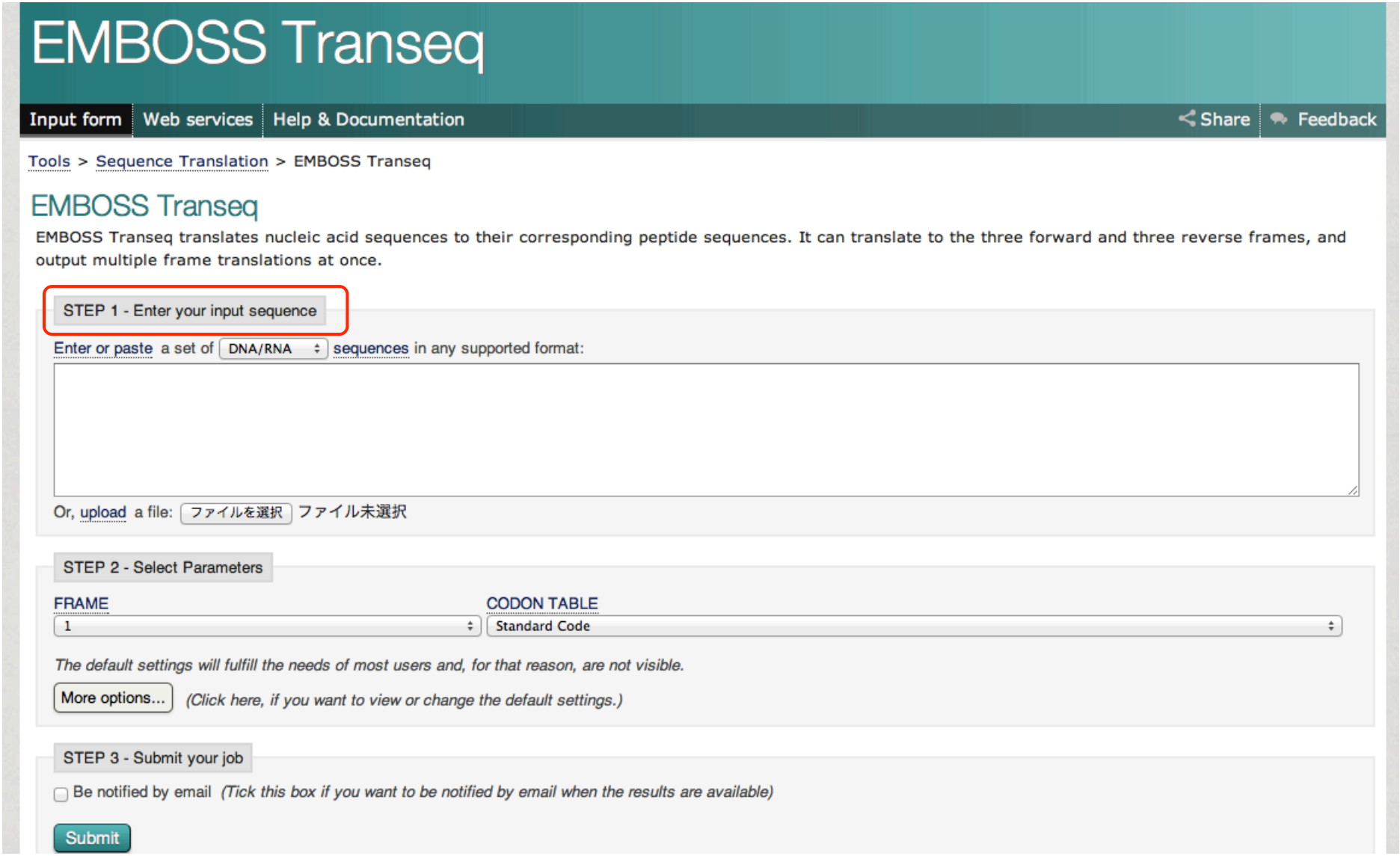
リンク集のDNA→AAを使う

– EMBL-EBI EMBOSS Transeq

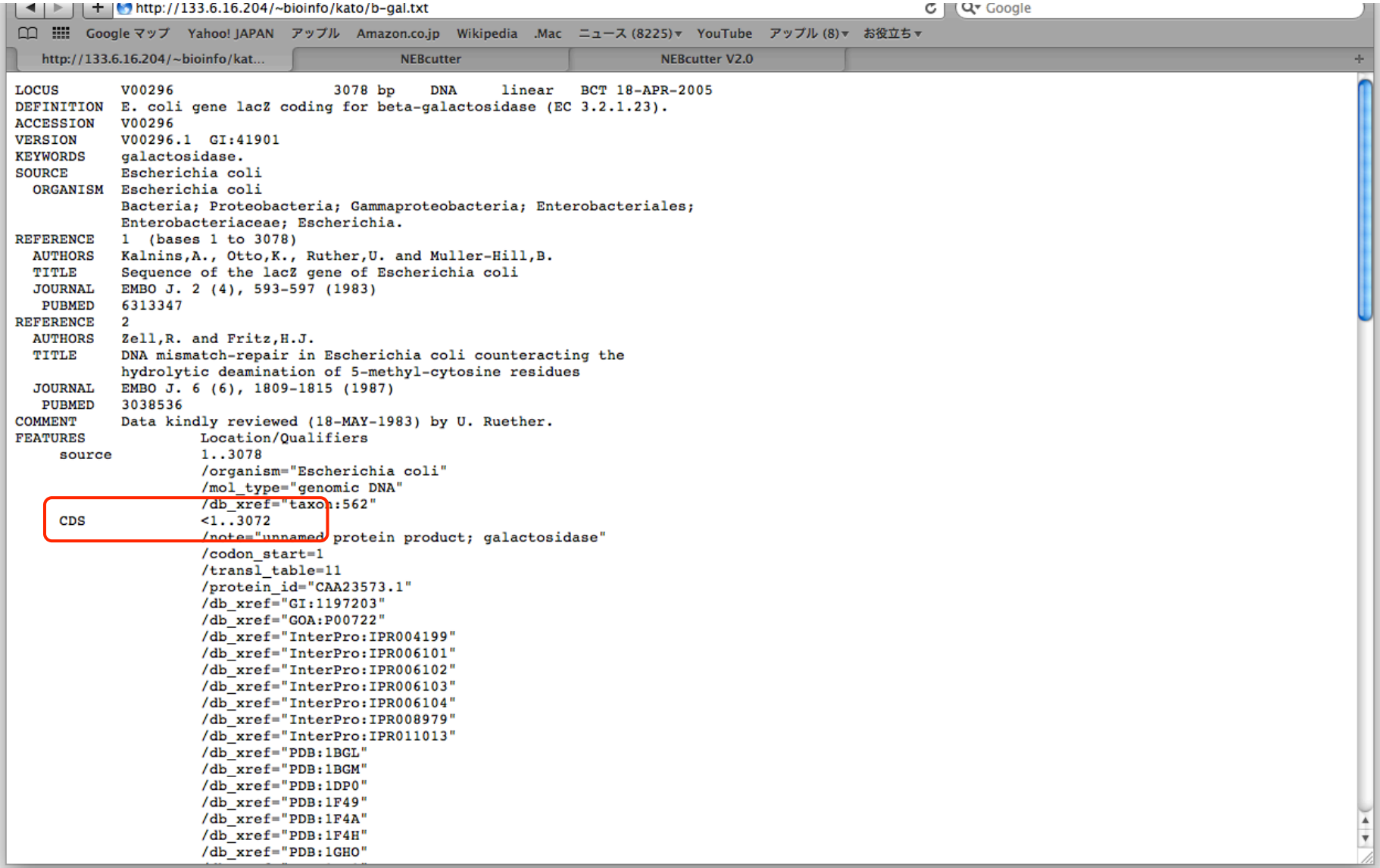
使用するのはβガラクトシダーゼ遺伝子



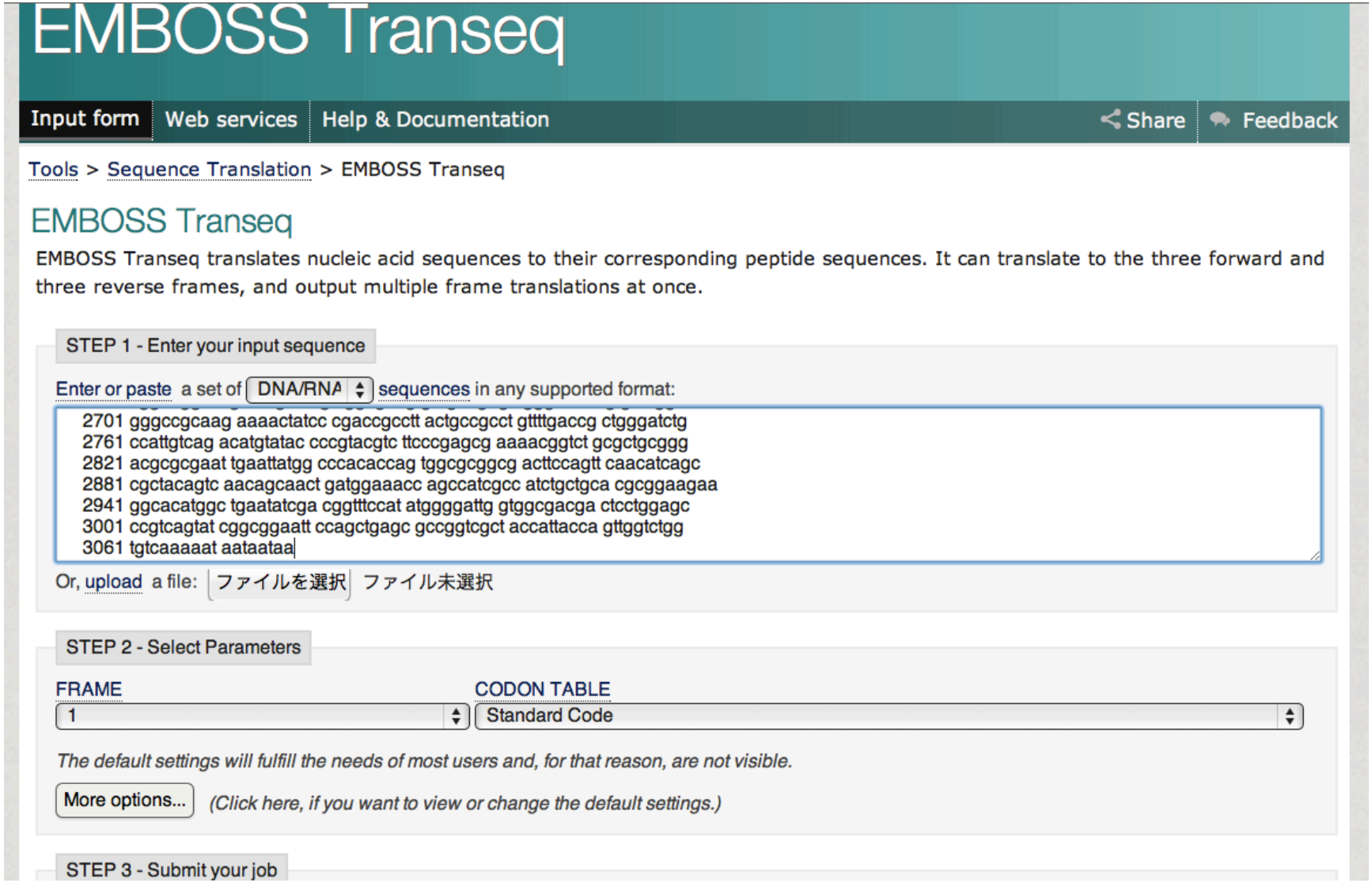
EMBL-EBI EMBOSS Transeq



βガラクトシダーゼ遺伝子のテキストファイルを開く



ボックスの中に配列をペーストする



結果

EMBOSS Transeq

Input form | Web services | Help & Documentation | Share | Feedback

Tools > Sequence Translation > EMBOSS Transeq

Results for job emboss_transeq-I20131022-094451-0222-62149969-oy

Tool Output | Submission Details

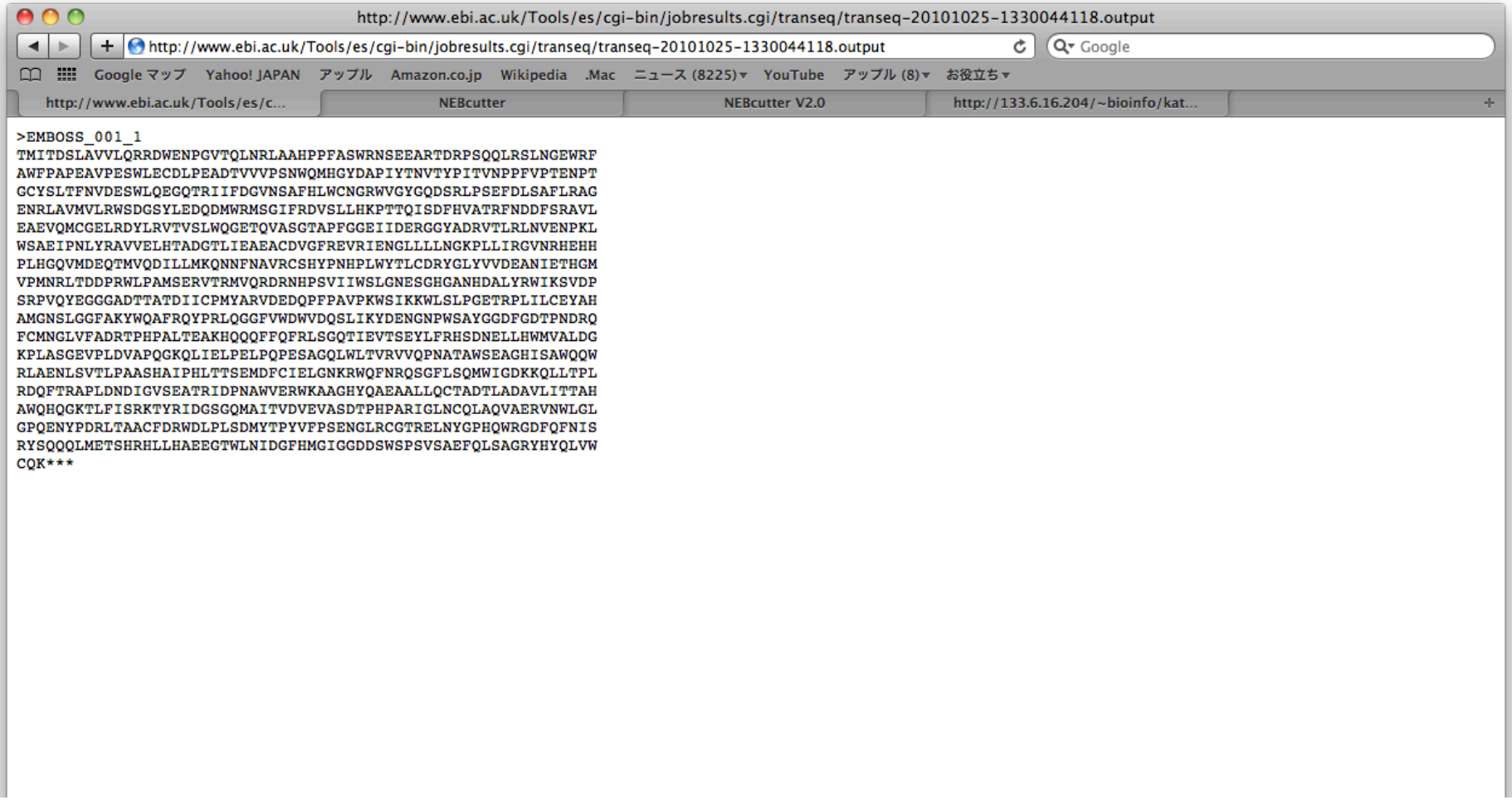
Download | Show Colors

```
>EMBOSS_001_1
TMITDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWRF
AWFPAPEAVPESWLECDLPEADTVVVPSNWQMHGYDAPIYTNVTYPITVNPPFVPTENPT
GCYSLTFNVDESWLQEGQTRIIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRAG
ENRLAVMVLRWSDGSYLEDQDMWRMSGIFRDVSLLHKPTTQISDFHVATRFNDDFSRAVL
EAEVQMCGELRDYLRVTVSLWQGETQVASGTAPFGGEIIDERGGYADRVTLRLNVENPKL
WSAEIPNLYRAVVELHTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLIRGVNRHEHH
PLHGQVMDEQTMVQDILLMKQNNFNAVRCSHYPNHPLWYTLCDRYGLYVVDEANIETHGM
VPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVIIWSLGNESGHGANHDALYRWIKSVDP
SRPVQYEGGGADTTATDIICPMYARVDEDQPFPAVPKWSIKKWLSLPGETRPLILCEYAH
AMGNSLGGFAKYWQAFRQYPRLQGGFVWDWVDQSLIKYDENGNPWSAYGGDFGDTPNDRQ
FCMNGLVFADRTPHPALTEAKHQQQFFQFRLSGQTIEVTSEYLFRHSDNELLHWMVALDG
KPLASGEVPLDVAPQGKQLIELPELPQPESAGQLWLTVRVVQPNATAWSEAGHISAWQQW
RLAENLSVTLPAASHAIPHLTTSEMDFCIELGNKRWQFNRQSGFLSQMWIGDKKQLLTPL
RDQFTRAPLDNDIGVSEATRIDPNAWVERWKAAGHYQAEAALLQCTADTLADAVLITTAH
AWQHQGKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAERVNWLGL
GPQENYPDRLTAACFDRWDLPLSDMYTPYVFPSENGLRCGTRELNYGPHQWRGDFQFNIS
RYSQQQLMETSHRHLLHAEEGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLVW
CQK***
```

演習

- フレームをFに変えてやってみる。読み枠以外の配列も出てくる。3通りの読み枠で出力する。はじめと終わりが分からないときに便利。
- フレームを6にしてみる。反対鎖もふくめ全ての読み枠を調べる方法。
- 結果を表示しているページの「ダウンロード」をクリックしてみよう。

ファスタ形式で出力される



課題

- 未知mRNA X（解析用配列にリンクがある）のコードするアミノ酸配列をファスタ形式で提出
- 講義４課題２

  - 塩基配列をボックスにペースト
  - フレームを F に設定してRUN
  - M（メチオニン）から始まり*（終始コドン）で終わる配列のうち、最も長いものを探す
  - 書式を整える（Seqretをつかう：次ページ）
  - メールの本分に貼り付けて、提出

配列情報（塩基配列・アミノ酸配列）のフォーマット変換

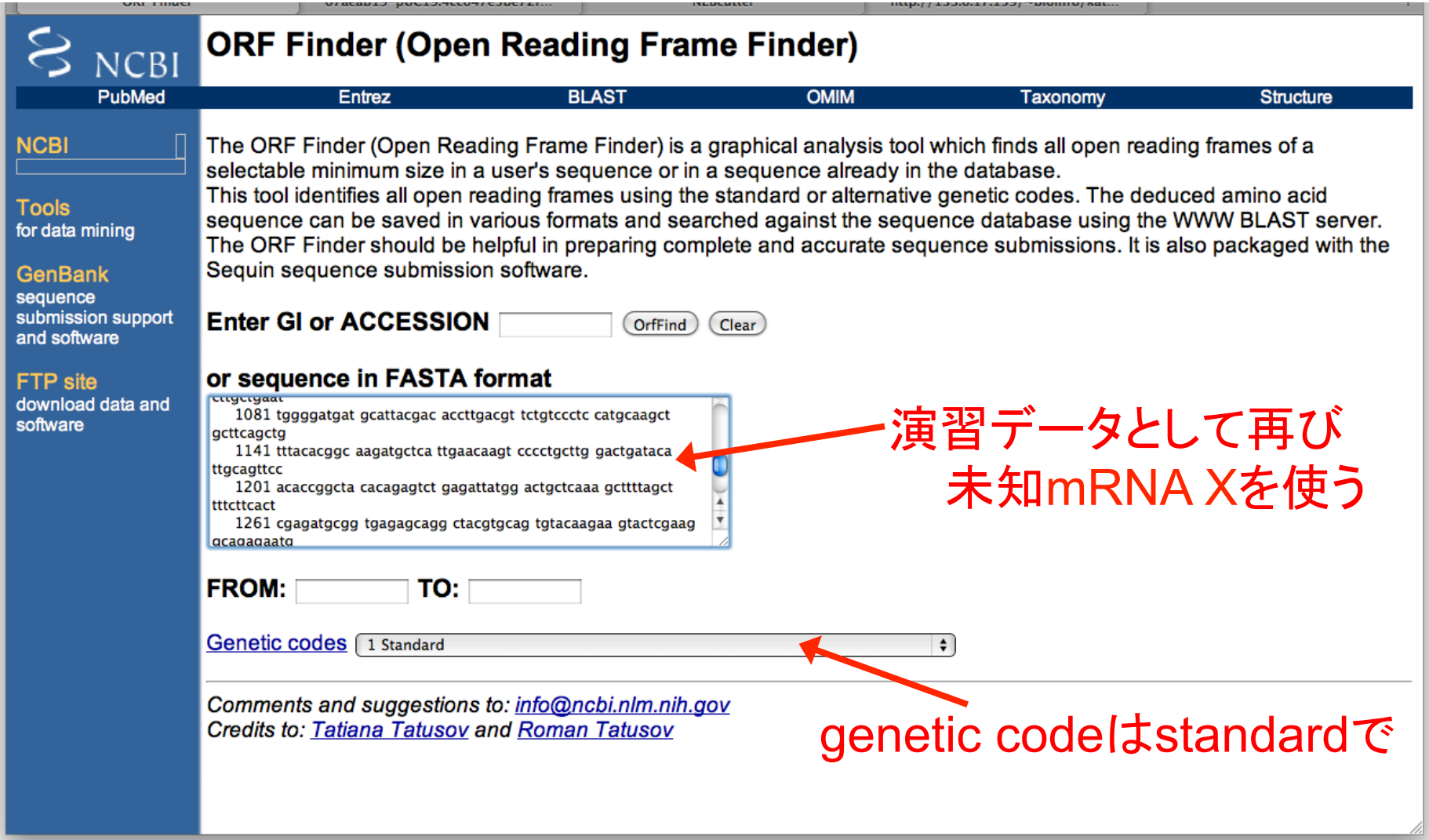
EMBOSS Seqret にアクセス
(http://www.ebi.ac.uk/Tools/sfc/emboss_seqret)

- ボックスに貼り付ける
- PROTEIN/DNA/RNAを選択
- OUTPUT FORMATをPlain Textに設定
- Submitボタンで開始

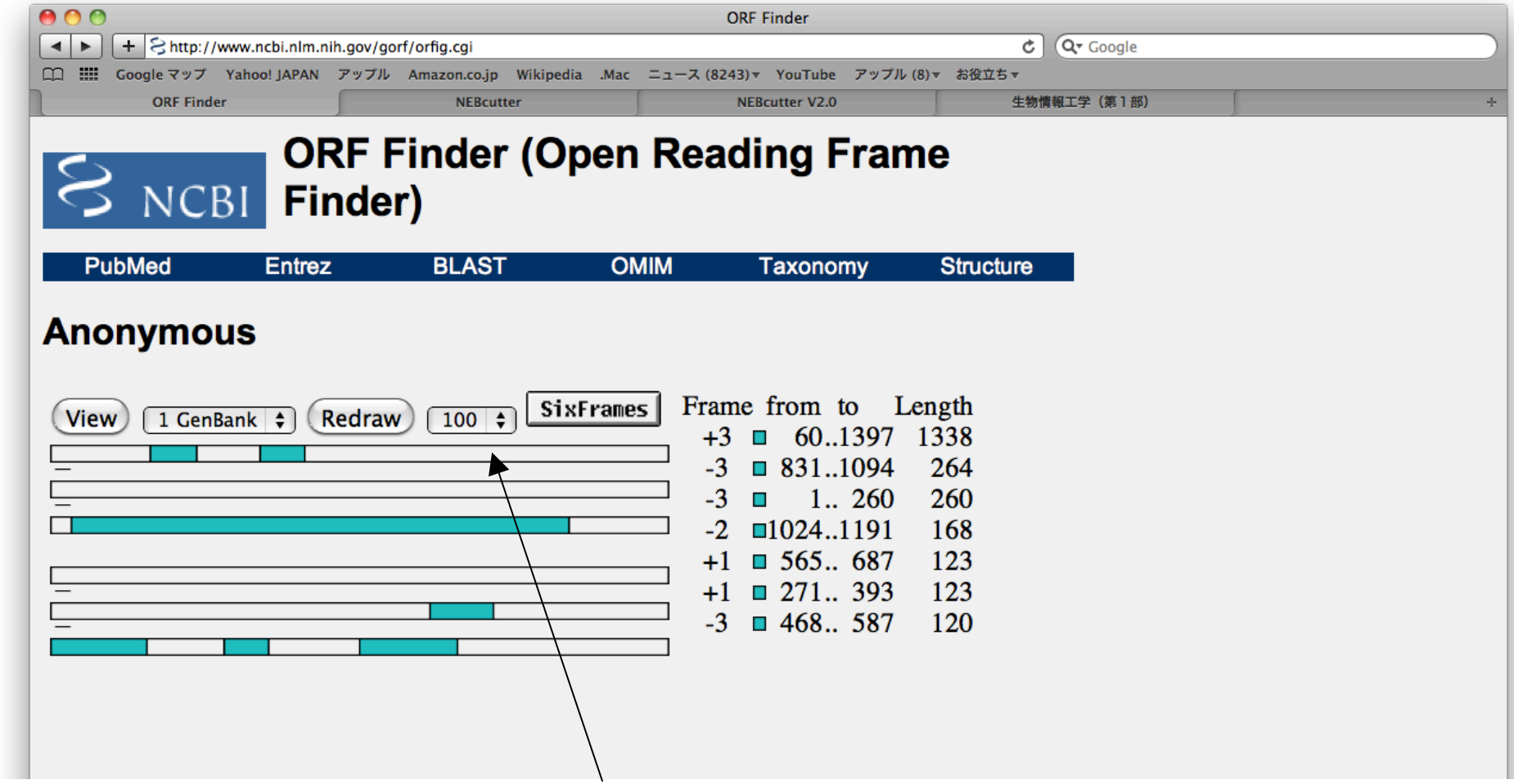
Open Reading Frame の検索

- DNA配列から蛋白質をコードしている部分（CDS)を探す。
- 長いORF (open reading frame)がCDSの有力な候補。
- リンク集のORF Finder (NCBI)を使う。

ORF Finderにアクセスする

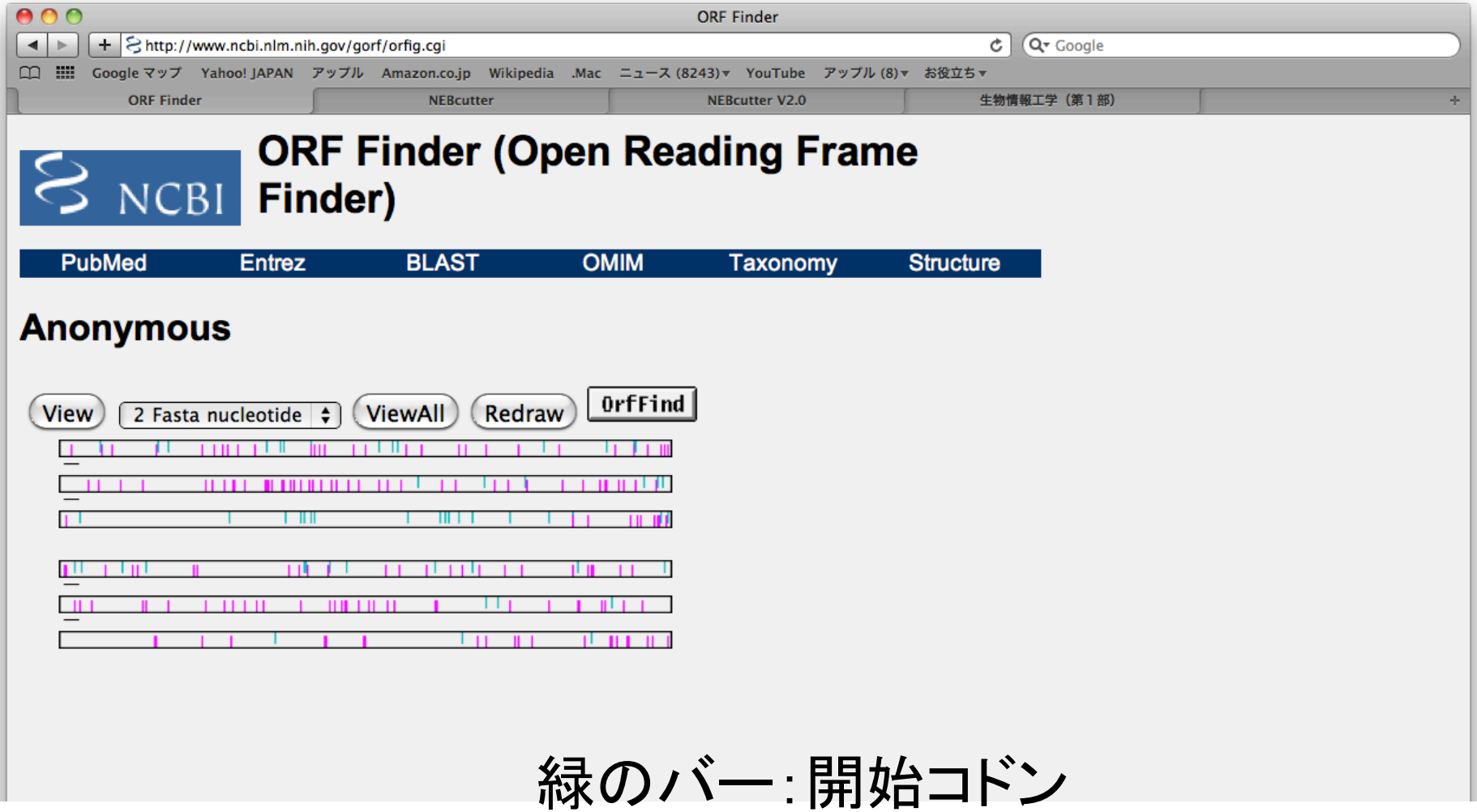


結果の表示（1）



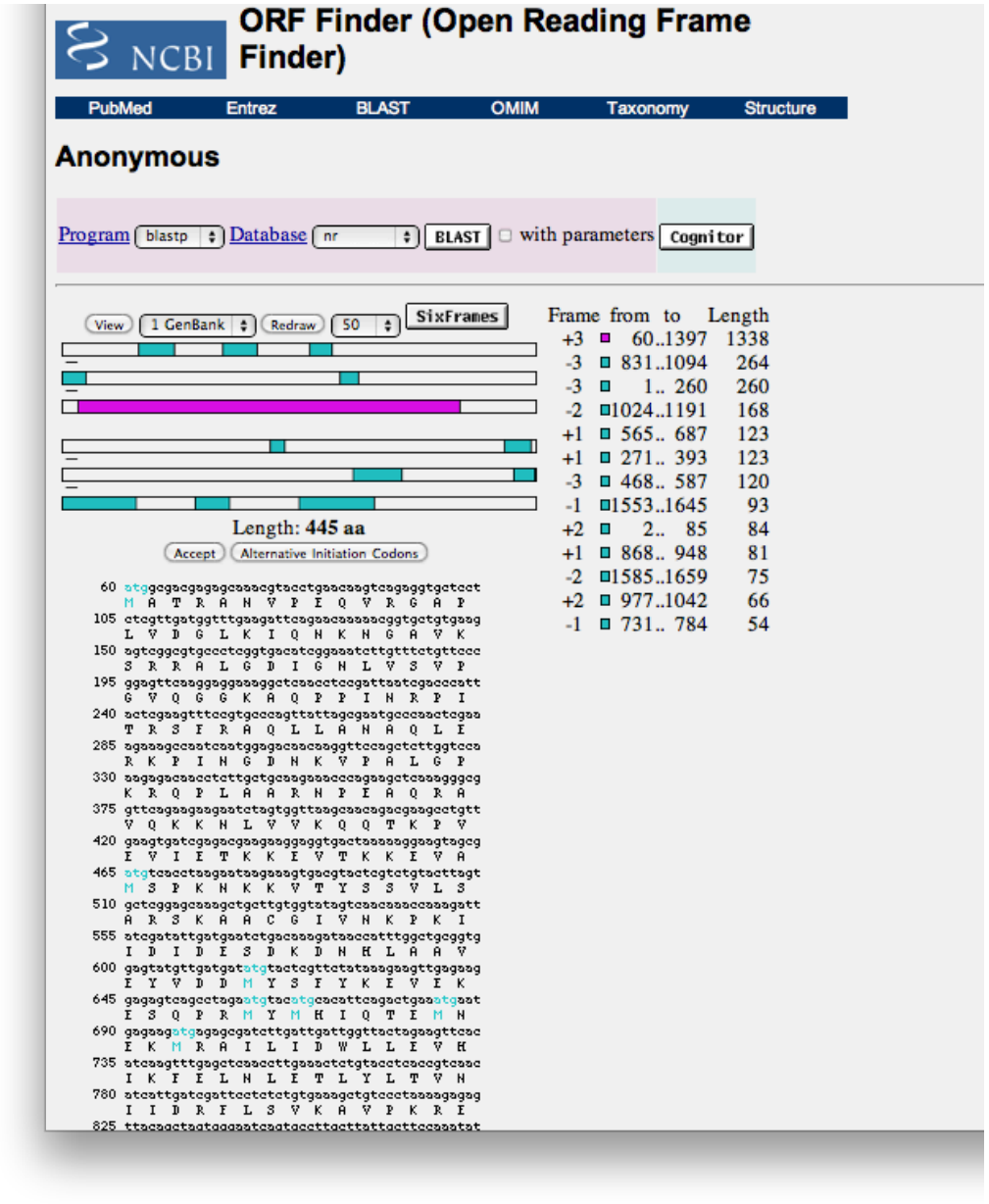
100 base以上のORFを表示している

結果の表示(2)



緑のバー：開始コドン
ピンクのバー：終始コドン

結果の表示（3）


NCBI
ORF Finder (Open Reading Frame Finder)
PubMed Entrez BLAST OMIM Taxonomy Structure
Anonymous
Program blastp Database nr BLAST with parameters Cognitor
View 1 GenBank Redraw 50 SixFrames
Length: 445 aa
Accept Alternative Initiation Codons
Frame from to Length
+3 60..1397 1338
-3 831..1094 264
-3 1.. 260 260
-2 1024..1191 168
+1 565.. 687 123
+1 271.. 393 123
-3 468.. 587 120
-1 1553..1645 93
+2 2.. 85 84
+1 868.. 948 81
-2 1585..1659 75
+2 977..1042 66
-1 731.. 784 54

大腸菌ラクトースオペロンの構造

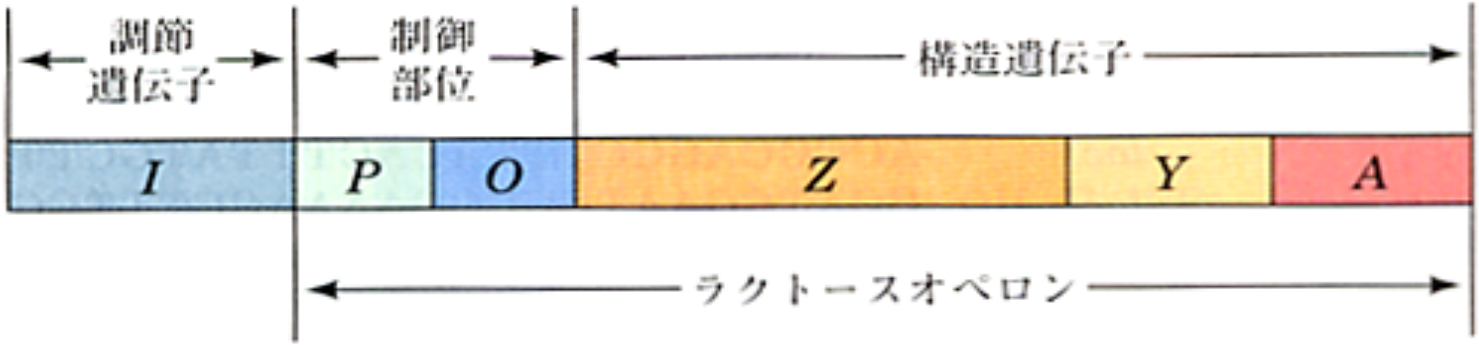
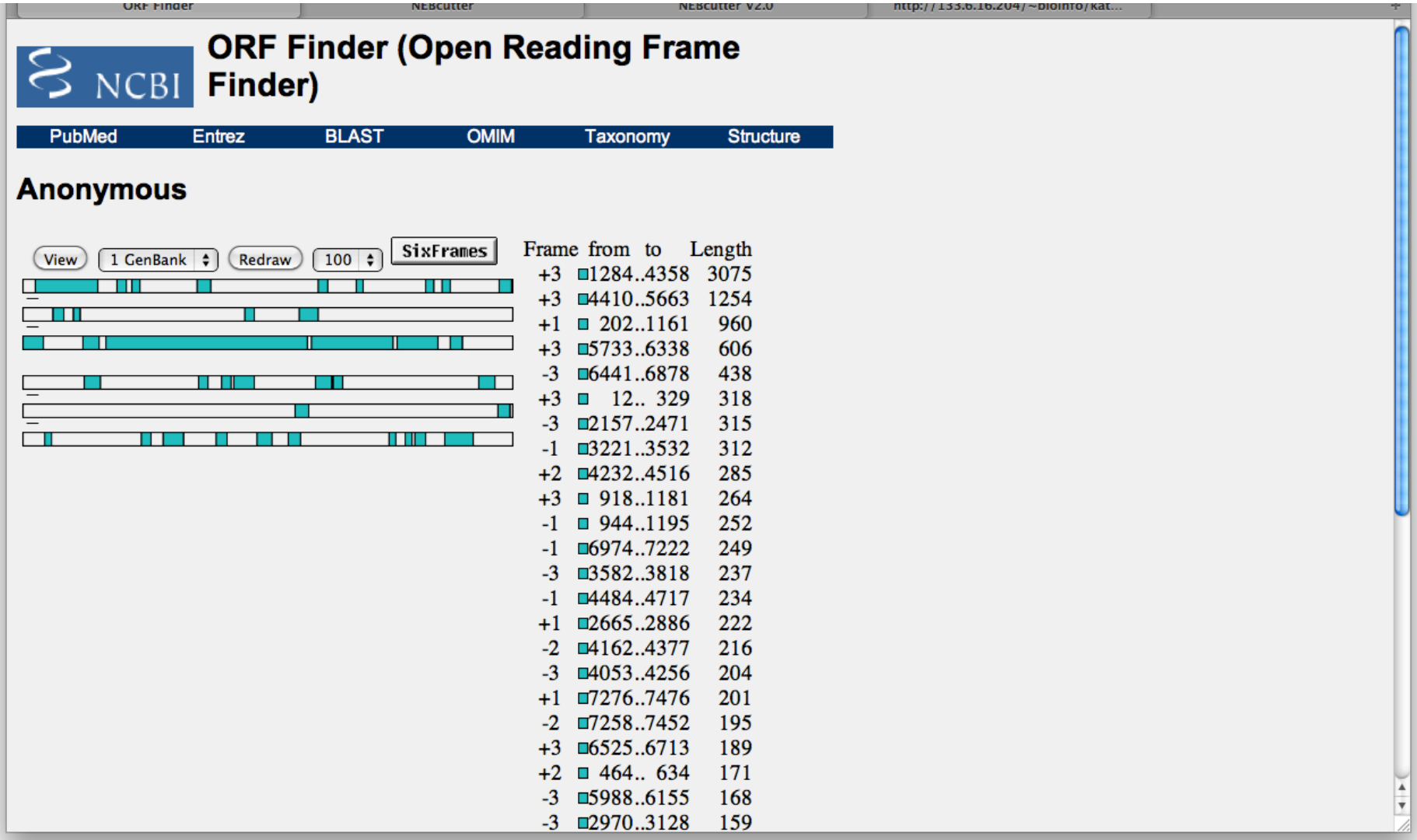


図 25・4　大腸菌 *lac* オペロン　このDNAはラクトース代謝に関与するタンパクをコードする遺伝子と，その発現を制御する遺伝子部位を含む．Z遺伝子は**β-ガラクトシダーゼ**，Y遺伝子は**ガラクトシドパーミアーゼ**，A遺伝子は**チオガラクトシドアセチルトランスフェラーゼ**をコードする．すぐ近くにある*I*遺伝子は*lac*オペロンには属さないが*lac*オペロンの転写を抑制するリプレッサーをコードする調節遺伝子である．

課題

- lactose operonの配列全部（7kb以上）選択し、コピーする。ボックスにペーストし、ORFを検索。
- 4つのORFを探し出し、塩基の番号（最初と終わり）および、それぞれがコードするタンパク質の名称を示せ。
- 講義4課題3

結果



ORF Finder (Open Reading Frame Finder)

PubMed Entrez BLAST OMIM Taxonomy Structure

Anonymous

View 1 GenBank Redraw 100 SixFrames

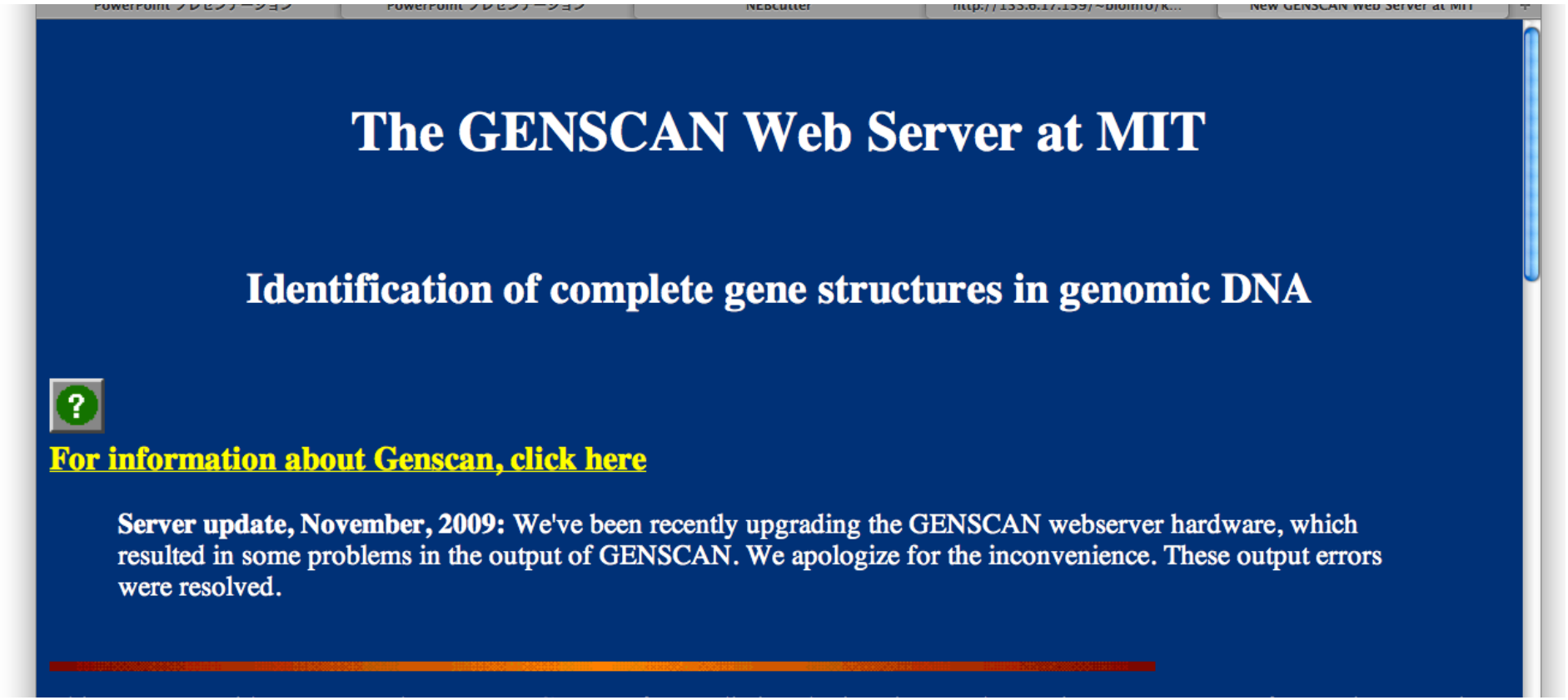
Frame	from	to	Length
+3	1284	4358	3075
+3	4410	5663	1254
+1	202	1161	960
+3	5733	6338	606
-3	6441	6878	438
+3	12	329	318
-3	2157	2471	315
-1	3221	3532	312
+2	4232	4516	285
+3	918	1181	264
-1	944	1195	252
-1	6974	7222	249
-3	3582	3818	237
-1	4484	4717	234
+1	2665	2886	222
-2	4162	4377	216
-3	4053	4256	204
+1	7276	7476	201
-2	7258	7452	195
+3	6525	6713	189
+2	464	634	171
-3	5988	6155	168
-3	2970	3128	159

演習

- 出てきたアミノ酸配列と、DNAデーターベース中に書いてある情報と比べてみよう。

スプライシングの予測

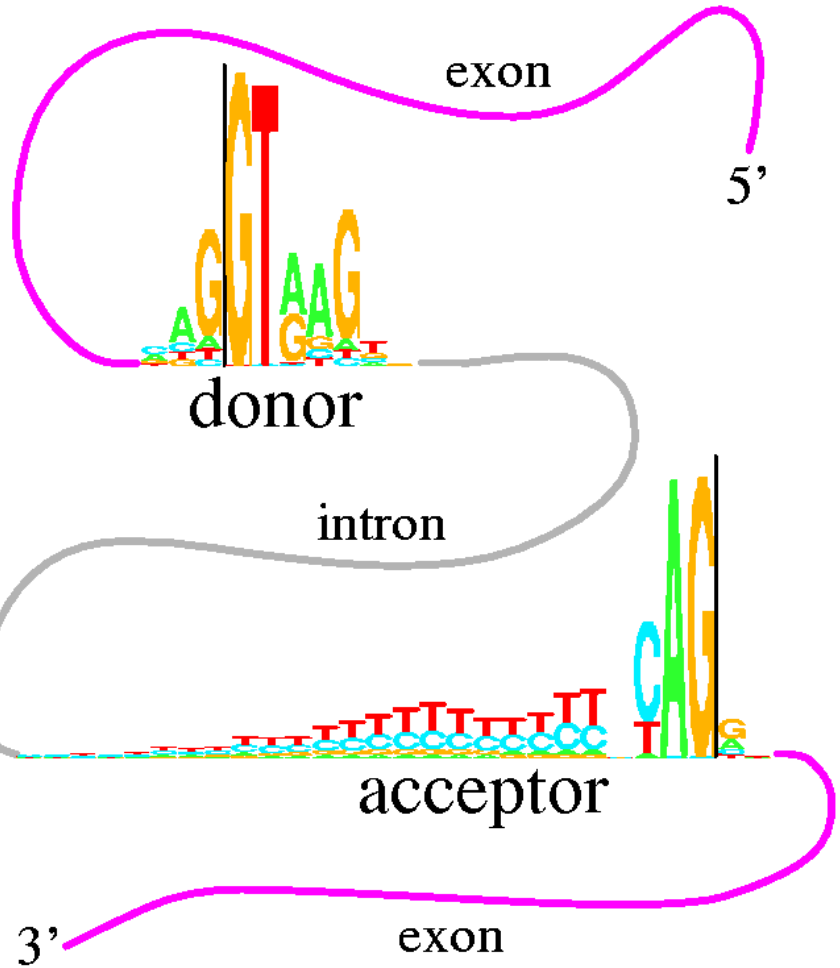
- リンク集のGENESCANを使う。
- 配列はovalbuminを使う。



イントロンとエキソンを塩基配列から推定する

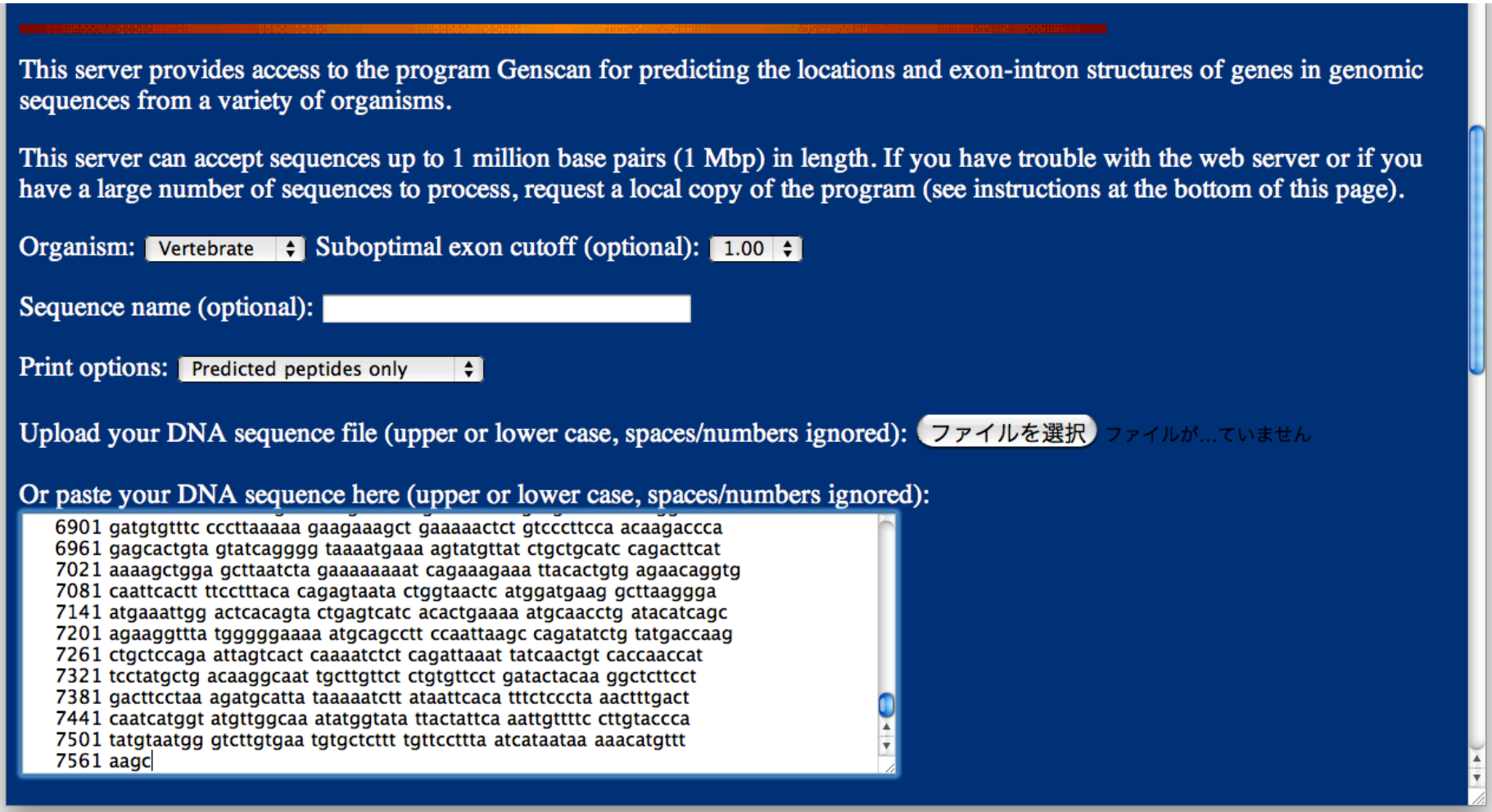
・エキソン／イントロン境界部位の塩基配列

・エキソンにはタンパク質をコードする読み枠がある



https://schneider.ncifcrf.gov/sequencelogo.html (2016/7/26)

ovalbumin遺伝子の配列をボックスにペースト



結果の表示（1）

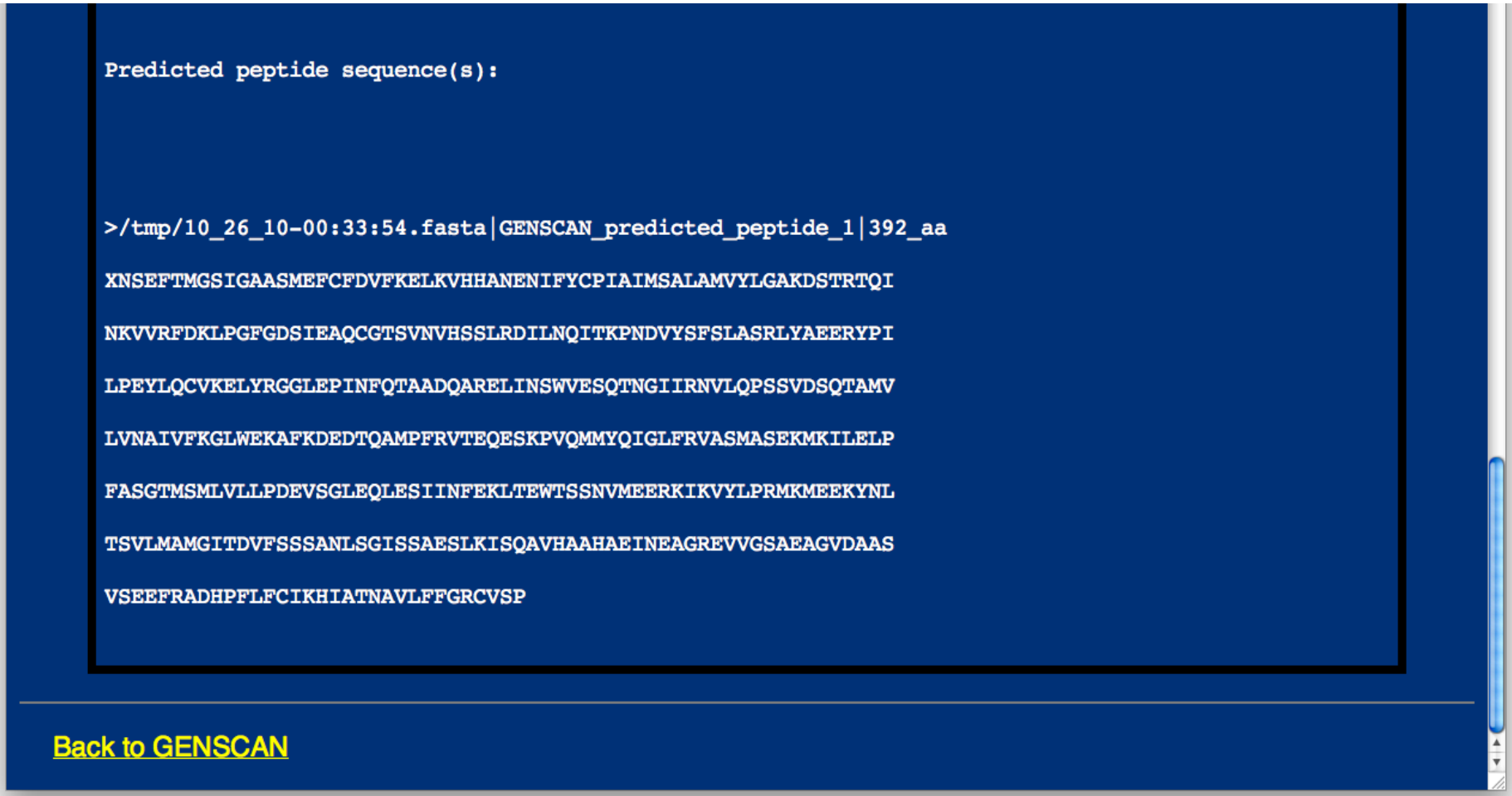
```
Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------

 1.01 Intr +   1637   1821  185  0  2   81  105   164 0.576  15.99
 1.02 Intr +   2073   2123   51  2  0   83   92    91 0.993   7.19
 1.03 Intr +   2705   2833  129  1  0   70  100    80 0.991   7.37
 1.04 Intr +   3234   3351  118  2  1   80   91   102 0.975   8.82
 1.05 Intr +   4310   4452  143  0  2   78   61   157 0.998  11.15
 1.06 Intr +   4784   4939  156  1  0   92   80   121 0.998  10.99
 1.07 Term +   6522   6917  396  2  0   69   42   244 0.974  11.99
 1.08 PlyA +   7546   7551    6                              1.05
```

結果の表示（2）



Predicted peptide sequence(s):

>/tmp/10_26_10-00:33:54.fasta|GENSCAN_predicted_peptide_1|392_aa

XNSEFTMGSIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSALAMVYLGAKDSTRTQI

NKVVRFDKLPGFGDSIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSLASRLYAEERYPI

LPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLQPSSVDSQTAMV

LVNAIVFKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIGLFRVASMASEKMKILELP

FASGTMSMLVLLPDEVSGLEQLESIINFEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNL

TSVLMAMGITDVFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGVDAAS

VSEEFRADHPFLFCIKHIATNAVLFFGRCVSP

Back to GENSCAN

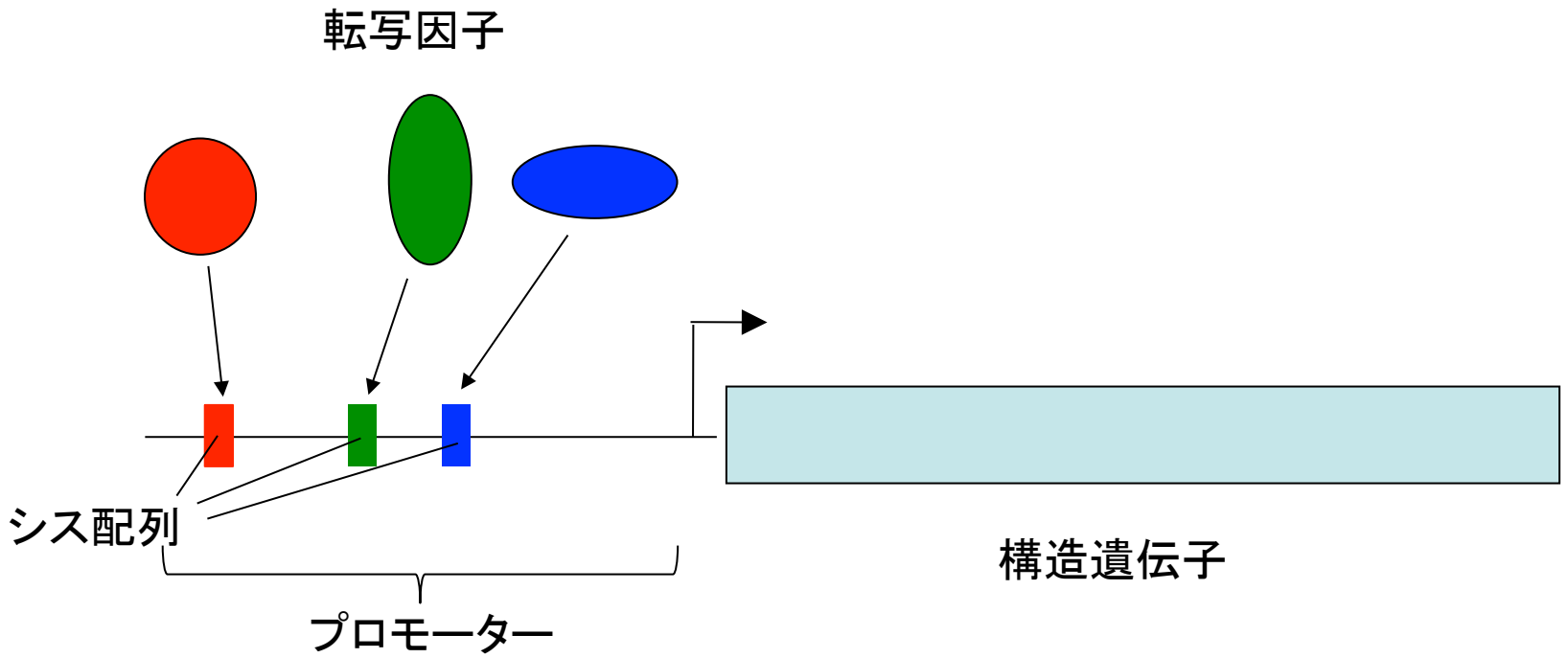
演習

- 解析結果を見てみよ。
- 実際のイントロン、エキソンと比べ考察せよ。
  - DNAデータの中に記述がある。

```
FEATURES             Location/Qualifiers
     source          1..7564
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9031"
     exon            1..47
                     /number=1
     intron          48..1636
                     /number=1
     exon            1637..1821
                     /number=2
     CDS             join(1654..1821,2073..2123,2705..2833,3234..3351,
                     4310..4452,4784..4939,6522..6917)
                     /codon_start=1
                     /product="ovalbumin"
                     /protein_id="CAA23716.1"
                     /db_xref="GI:808974"
                     /db_xref="GOA:P01012"
                     /db_xref="PDB:1JTI"
                     /db_xref="PDB:1OVA"
                     /db_xref="PDB:1P1Z"
                     /db_xref="PDB:1UHG"
                     /db_xref="PDB:1VAC"
                     /db_xref="UniProtKB/Swiss-Prot:P01012"
                     /translation="MGSIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSALAMVYL
                     GAKDSTRTQINKVVRFDKLPGFGDSIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSL
                     ASRLYAEERYPILPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRN
                     VLQPSSVDSQTAMVLVNAIVFKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIGLF
```

転写因子結合部位の予測

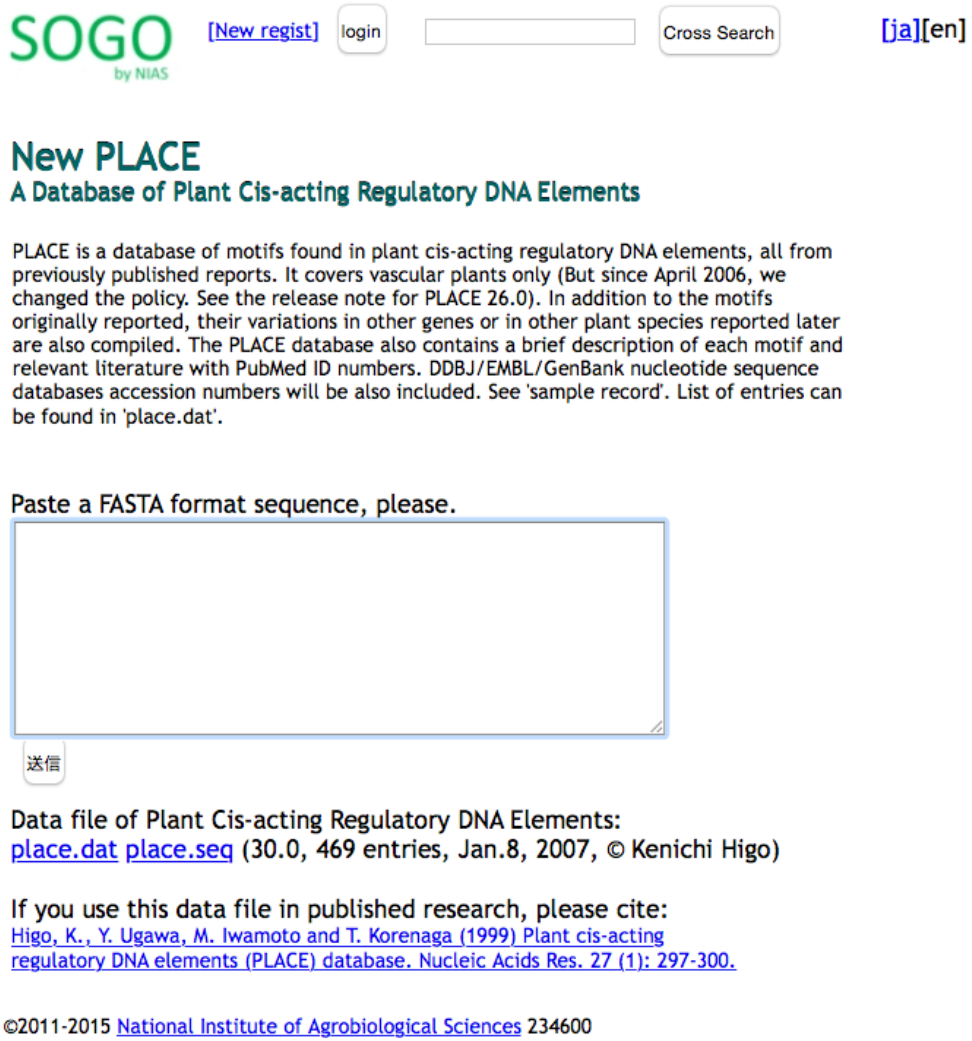
- PLACEを使って、プロモーターのシス配列を探そう。
- シス配列：シス配列：プロモーター中で発現に影響を与える短い塩基配列



PLACEにアクセスする

(A Database of Plant Cis-acting Regulatory DNA Elements)

http://www.dna.affrc.go.jp/PLACE/

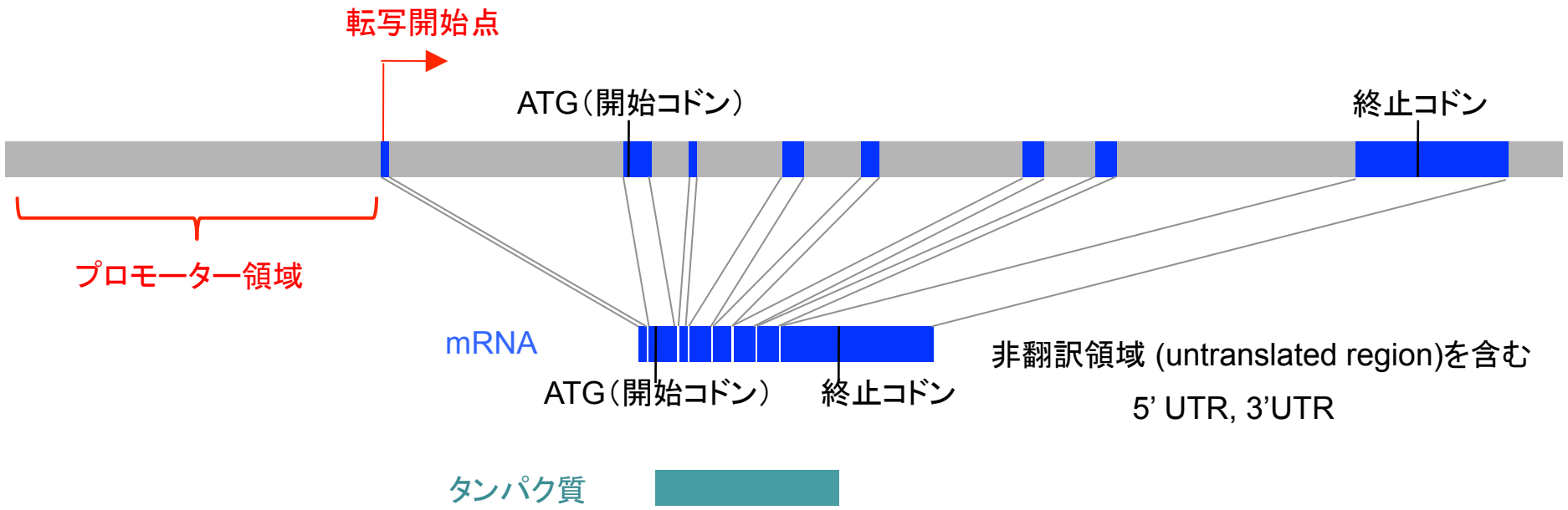


演習：使い方

- Wheat histone H4 geneのプロモーターを調べる
- DNAデータよりプロモーター部分を選択する。
- Featuresをみると、mRNAが669からとなっている。→1－668をプロモーターとする。
- 1から668の配列をコピーする。
- DNA配列をボックスにペーストする。
- submitボタンで検索開始。

遺伝子の構造

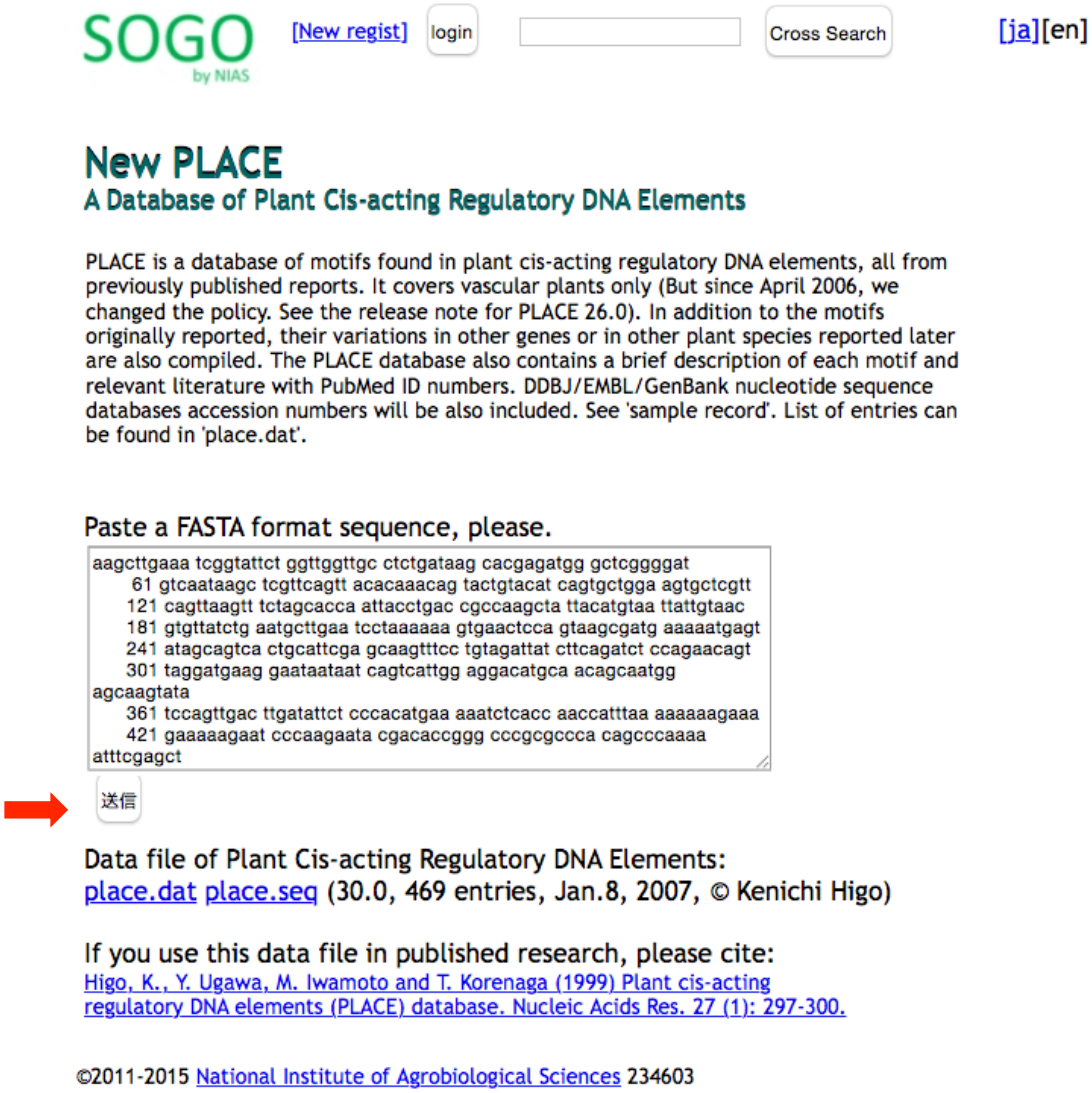
転写を制御するプロモーター領域



コムギのヒストンH4遺伝子のプロモーター領域の配列をコピー

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4565"
     mRNA            669..>1200
                     /product="messenger RNA"
     CDS             736..1047
                     /note="unnamed protein product; histone H4"
                     /codon_start=1
                     /protein_id="CAA24924.1"
                     /db_xref="GI:21795"
                     /db_xref="GOA:P62785"
                     /db_xref="InterPro:IPR001951"
                     /db_xref="InterPro:IPR007125"
                     /db_xref="InterPro:IPR009072"
                     /db_xref="InterPro:IPR019809"
                     /db_xref="UniProtKB/Swiss-Prot:P62785"
                     /translation="MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGV
                     KRISGLIYEETRGVLKIFLENVIRDAVTYTEHARRKTVTAMDVVYALKRQGRTLYGFG
                     G"
ORIGIN
        1 aagcttgaaa tcggtattct ggttggttgc ctctgataag cacgagatgg gctcggggat
       61 gtcaataagc tcgttcagtt acacaaacag tactgtacat cagtgctgga agtgctcgtt
      121 cagttaagtt tctagcacca attacctgac cgccaagcta ttacatgtaa ttattgtaac
      181 gtgttatctg aatgcttgaa tcctaaaaaa gtgaactcca gtaagcgatg aaaaatgagt
      241 atagcagtca ctgcattcga gcaagtttcc tgtagattat cttcagatct ccagaacagt
      301 taggatgaag gaataataat cagtcattgg aggacatgca acagcaatgg agcaagtata
      361 tccagttgac ttgatattct cccacatgaa aaatctcacc aaccatttaa aaaaaggaaa
      421 gaaaaaggaat cccaagaata cgacaccggg cccgcgccca cagcccaaaa atttcgagct
      481 gccgccacag ctctgaccca gcccgccaaa ccaccggtcc aatctctcgg ccacgtcacc
      541 gatccgcggc atctctcccc cggatcgccg tctcgaccgt ccactccatc cgcatccaac
      601 ggcagccaca cgcctcctcc aacctctcga ccccttaaag acgcccttcg ccccacccag
      661 caaatcacag caccagacgc cacccaccac cgttcctccc atcccacact cgctcgcagc
      721 tcgagatcgt cggccatgtc cgggcgcggc aaggggaggca agggcctagg caagggcggc
      781 gccaagcgcc accggaaggt cctccgcgat aacatccagg gcatcaccaa gccggcgatc
      841 cggcggctgg cgcggcgggg cggcgtgaag cgcatctcgg ggctcatcta cgaggagacc
      901 cgcggcgtgc tcaagatctt cctcgagaac gtcatccgcg atgccgtcac ctacaccgag
      961 cacgcccgcc gcaagaccgt caccgccatg gacgtcgtct acgcgctcaa gcgccagggc
     1021 cgcaccctct acggcttcgg cggctaaggg ccggccggcc gacgggagtc actctttgtc
     1081 gccgcctgca gattccagaa gcctgatgaa gccccgactt gtttagttcg ctatttcctc
     1141 tgtagtttga actcaatcgt ggaacaaagt tattcgcata tattgttgga aatgaagctt
```

Search boxにペースト



結果の表示(1)

```
668 base pairs

 (+) = Current Strand
 (-) = Opposite Strand

1       AAGCTTGAAATCGGTATTCTGGTTGGTTGCCTCTGATAAGCACGAGATGG
                (-) ARR1AT S000454 9 NGATT
                     (+) -10PEHVPSBD S000392 15 TATTCT
                           (-) MYBPLANT S000167 21 MACCWAMC
                           (-) MYBPZM S000179 21 CCWACC
                           (-) BOXLCOREDCPAL S000492 21 ACCWWCC
                             (-) REALPHALGLHCB21 S000362 23 AACCAA
                                       (+) GATABOX S000039 35 GATA
                                       (+) IBOX S000124 35 GATAAG
                                       (+) IBOXCORE S000199 35 GATAA
                                            (+) RHERPATEXPA7 S000512 40 KCACGW
                                                  (+) SITEIIATCYTC S000474 48 TGGGCY

51      GCTCGGGGATGTCAATAAGCTCGTTCAGTTACACAAACAGTACTGTACAT
                 (+) BIHD1OS S000498 60 TGTCA
                  (-) WBOXATNPR1 S000390 61 TTGAC
                  (-) WRKY71OS S000447 61 TGAC
                    (+) CAATBOX1 S000028 63 CAAT
                              (-) MYB2AT S000177 76 TAACTG
                              (-) MYB2CONSENSUSAT S000409 76 YAACKG
                              (+) MYBCORE S000176 76 CNGTTR
                                        (-) CACTFTPPCA1 S000449 89 YACT
                                         (-) CURECORECR S000493 90 GTAC
                                         (+) CURECORECR S000493 90 GTAC
                                          (+) CACTFTPPCA1 S000449 91 YACT
                                              (-) CURECORECR S000493 95 GTAC
                                              (+) CURECORECR S000493 95 GTAC

101     CAGTGCTGGAAGTGCTCGTTCAGTTAAGTTTCTAGCACCAATTACCTGAC
         (-) CACTFTPPCA1 S000449 102 YACT
                  (-) CACTFTPPCA1 S000449 111 YACT
                            (-) MYB2AT S000177 121 TAACTG
                            (-) MYB2CONSENSUSAT S000409 121 YAACKG
                            (+) MYBCORE S000176 121 CNGTTR
                                    (-) POLLEN1LELAT52 S000245 129 AGAAA
                                              (+) CCAATBOX1 S000030 138 CCAAT
                                               (+) CAATBOX1 S000028 139 CAAT
                                                 (-) GT1CONSENSUS S000198 141 GRWAAW
                                                      (+) WBOXNTCHN48 S000508 146 CTGACY
                                                       (+) WRKY71OS S000447 147 TGAC
                                                       (+) WBOXNTERF3 S000457 147 TGACY

151     CGCCAAGCTATTACATGTAATTATTGTAACGTGTTATCTGAATGCTTGAA
```

結果の表示(2)

Factor or Site Name	Loc.(Str.)	Signal Sequence	SITE #
ARR1AT	9 (-)	NGATT	S000454
-10PEHVPSBD	15 (+)	TATTCT	S000392
MYBPLANT	21 (-)	MACCWAMC	S000167
MYBPZM	21 (-)	CCWACC	S000179
BOXLCOREDCPAL	21 (-)	ACCWWCC	S000492
REALPHALGLHCB21	23 (-)	AACCAA	S000362
GATABOX	35 (+)	GATA	S000039
IBOX	35 (+)	GATAAG	S000124
IBOXCORE	35 (+)	GATAA	S000199
RHERPATEXPA7	40 (+)	KCACGW	S000512
SITEIIATCYTC	48 (+)	TGGGCY	S000474
BIHD1OS	60 (+)	TGTCA	S000498
WBOXATNPR1	61 (-)	TTGAC	S000390
WRKY71OS	61 (-)	TGAC	S000447
CAATBOX1	63 (+)	CAAT	S000028
MYB2AT	76 (-)	TAACTG	S000177
MYB2CONSENSUSAT	76 (-)	YAACKG	S000409
MYBCORE	76 (+)	CNGTTR	S000176
CACTFTPPCA1	89 (-)	YACT	S000449
CURECORECR	90 (-)	GTAC	S000493
CURECORECR	90 (+)	GTAC	S000493
CACTFTPPCA1	91 (+)	YACT	S000449
CURECORECR	95 (-)	GTAC	S000493
CURECORECR	95 (+)	GTAC	S000493
CACTFTPPCA1	102 (-)	YACT	S000449
CACTFTPPCA1	111 (-)	YACT	S000449
MYB2AT	121 (-)	TAACTG	S000177
MYB2CONSENSUSAT	121 (-)	YAACKG	S000409
MYBCORE	121 (+)	CNGTTR	S000176
POLLEN1LELAT52	129 (-)	AGAAA	S000245
CCAATBOX1	138 (+)	CCAAT	S000030
CAATBOX1	139 (+)	CAAT	S000028
GT1CONSENSUS	141 (-)	GRWAAW	S000198
WBOXNTCHN48	146 (+)	CTGACY	S000508
WRKY71OS	147 (+)	TGAC	S000447

結果の表示(3)

SOGO by NIAS [新規登録] ログイン 横断検索 [ja][en]

```
ID   MYB2CONSENSUSAT
XX
AC   S000409
XX
DT   03-Jun-2003 (last modified) kehi
XX
DE   MYB recognition site found in the promoters of the
DE   dehydration-responsive gene rd22 and many other genes in
DE   Arabidopsis; Y=C/T; K=G/T; See S000177 (MYB2), S000175
DE   (MYBATRD22);
XX
KW   MYB; rd22BP1; ABA; leaf; seed; stress;
XX
OS   Arabidopsis thaliana
XX
RA   Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki
RA   K.
RT   Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as
RT   transcriptional activators in abscisic acid signaling.
RL   Plant Cell 15: 63-78 (2003)
RD   PubMed: 12509522;
XX
SQ
     YAACKG コンセンサス配列が標示される
//
```

国際塩基配列データベースで使用する核酸コード

シンボル	意味	説明
a	a	adenine
c	c	cytosine
g	g	guanine
t	t	thymine in DNA; uracil in RNA
m	a or c	amino
r	a or g	purine
w	a or t	
s	c or g	
y	c or t	pyrimidine
k	g or t	keto
v	a or c or g	not t
h	a or c or t	not g
d	a or g or t	not c
b	c or g or t	not a
n	a or c or g or t	any

課題

- シロイヌナズナ（*Arabidopsis thaliana*）の*ACT2*遺伝子のプロモーター配列を調べてみる。

  （配列はリンク集にあります。）

- 推定されるTATA boxのうち、転写開始点に最も近いものの配列とポジション、ストランド（+/-）についてメールの本文にまとめて提出。

- 講義４課題４