

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
source Location/Qualifiers
1..2686
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
1 gcgccaata cgcaaaccgc ctctcccgc gcggtggccg attcattaat gcagctggca
61 cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct
121 cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat
181 tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg aattcgagct
241 cggtagccgg ggatcctcta gagtcgacct gcagggcatgc aagcttggca ctggccgctc
301 ttttacaacg tcgtgactgg gaaaaccctg gcggtaccca acttaatcgc cttgcagcac
361 atcccccttt cgcagctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac
421 agttgcgag cctgaatggc gaatggcgcc tgatgcggtg ttttctcctt acgcatctgt
481 gcggtatttc acaccgata tgggtcactc tcagtacaat ctgctctgat gccgcatagt
541 taagccagcc ccgacaccgc ccaacaccgc 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 cttattccc ttttttgccg cattttgcct tcctgttttt gctcaccag
```

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
  source Location/Qualifiers
          1..2686
          /organism="synthetic construct"
          /mol_type="genomic DNA"
          /db_xref="taxon:32630"
ORIGIN
1 ggcgccaata cgcaaaccgc ctctccccgc gcgttgccg attcattaat gcagctggca
61 cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct
121 cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat
181 tgtgagcgga taacaatttc acacagaaa cagctatgac catgattacg aattcgagct
241 cggtaaccgg ggatcctcta gagtcgacct gcaggcatgc aagcttggca ctggccgctc
301 ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc cttgcagcac
361 atcccccttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ctttcccaac
421 agttgcgag cctgaatggc gaatggcgcc tgatgcggtt ttttctcctt acgcatctgt
481 gcggtatttc acaccgcata tggtgacctc tcagtacaat ctgctctgat gccgcatagt
541 taagccagcc cgcacaccgc ccaacaccgc ctgacgcgcc ctgacgggct tgtctgctcc
601 cggcatcgcg ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt
661 caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgcta tttttatagg
721 ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttogg ggaaatgtgc
781 gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac
841 aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt
901 tccgtgtcgc ccttattccc ttttttgcgg cattttgcct tcctgttttt gctcaccag
961 aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg ggttacatcg
1021 aactggatct caacagcggg 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



NEBcutter V2.0



This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. **The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.**

[What's new in V2.0](#) [Citing NEBcutter](#)

Local sequence file: ファイルが選...ていません

GenBank number: [\[Browse GenBank\]](#)

Standard sequences:
Plasmid vectors
Viral + phage

or paste in your DNA sequence: (plain or FASTA format)

```
1 gcgcccaata cgaaaacgc ctctcccgc gcgttgccg attcattaat gcagctggca
61 cgacaggttt cccgactgga aagcgggag tgagcgaac gcaattaatg tgagttagct
121 cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat
181 tgtgagcggg taacaattc acacaggaaa cagctatgac catgattacg ccaagcttgc
241 atgcctgcag gtcgacteta gaggatcccc gggtaaccgag ctggaattca ctggccgctg
301 tttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaacgcg cttgcagcac
361 atccccctt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc cttcccaac
421 agttgcgag cctgaatggc gaatggcgc tgatgcgga tttctcctt acgcatctgt
```

The sequence is: Linear Circular

Enzymes to use: NEB enzymes
 All commercially available specificities
 All specificities
 All + defined oligonucleotide sequences
 Only defined oligonucleotide sequences
[\[define oligos\]](#)

Minimum ORF length to display: a.a.

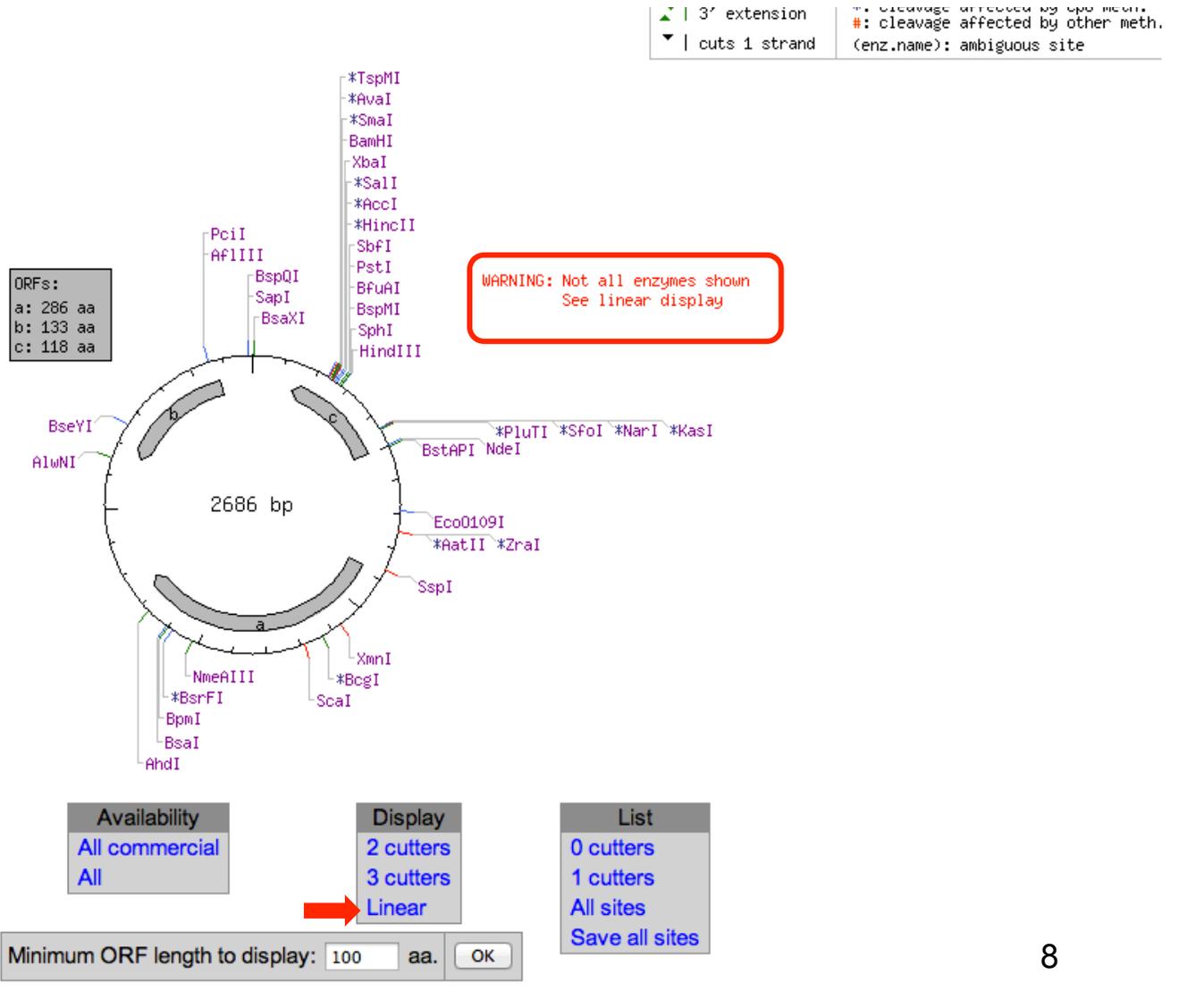
Name of sequence: (optional)

Earlier projects:
[no name](#)
[pUC19](#)
[puc18](#)

*Note: Your earlier projects will be deleted 2 days after they were last accessed.
You need to have cookies enabled in your browser for this feature to work.*

Disable NEBcutter cookies

結果(1)

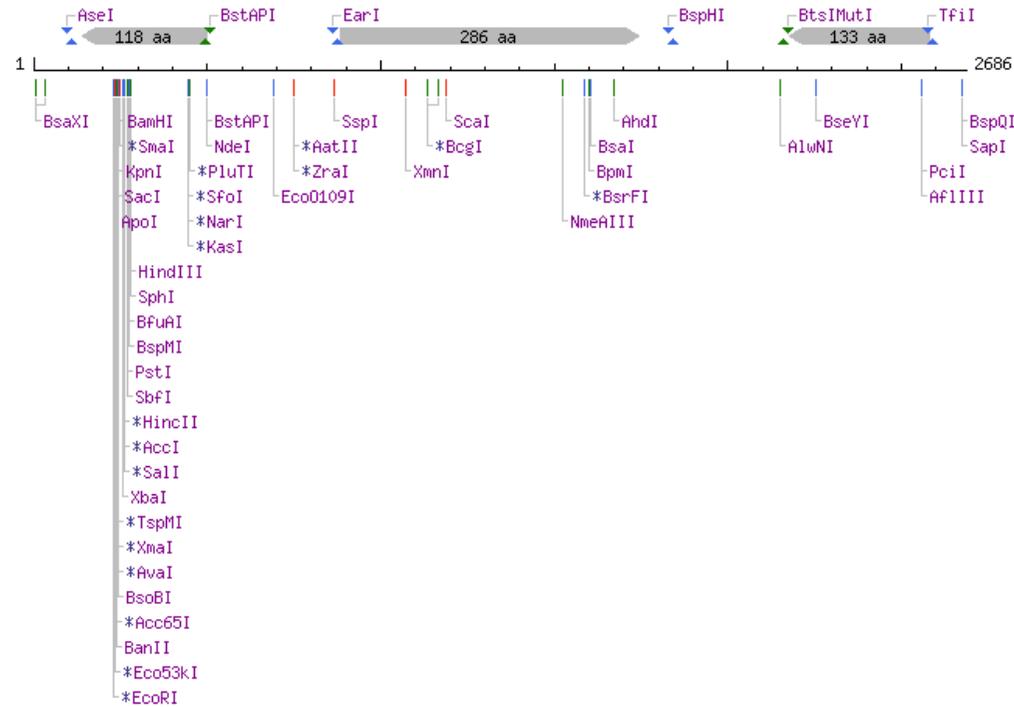


結果(2)

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

GC=51%, AT=49%

▲ 1 strand cut	Has other supplier
▼ 5' extension	Not commercially available
▲ 3' extension	*: cleavage affected by CpG meth.
▼ cuts 1 strand	#: cleavage affected by other meth.
	(enz.name): ambiguous site



Main options New DNA Custom digest View sequence ORF summary Save project Print	Availability All commercial All	Display 2 cutters 3 cutters Circular	Zoom Zoom in More...	List 0 cutters 1 cutters All sites Save all sites Flanking enzymes
--	--	--	-----------------------------------	--

Minimum ORF length to display: 100 aa.

結果 (3)



[\[Back to main display\]](#)

Single cutters

unnamed sequence

[Help](#) [Comments](#)

Number of cuts Sort order:

#	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	GCTCTCN ⁺ NNN _↓	list	2677/2680
20	BsrFI	R ⁺ CCGG _↓ Y	list	*1587/1591

結果(4)



[\[Back to main display\]](#)

BamHI

[Help](#)

[Comments](#)

```
5' ... GvG 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

Overlapping methylation:

NOT ANALYZED

Reaction temperature: 37 °C

Neoschizomers:

Isoschizomers:

結果 (5)



[\[Back to main display\]](#)

SmaI

[Help](#)

[Comments](#)

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

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

Overlapping methylation:

NOT ANALYZED

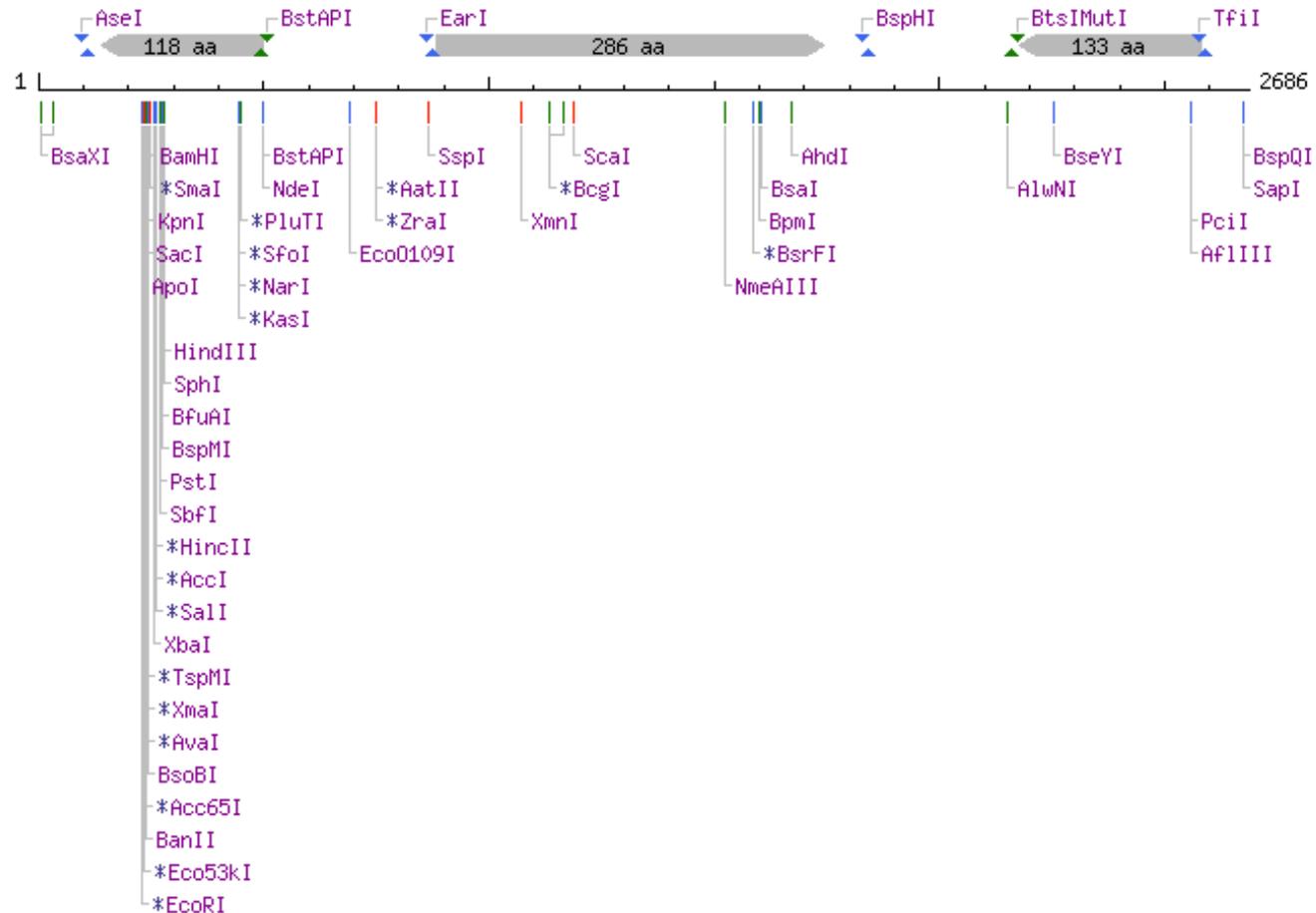
Reaction temperature: 25 °C

Neoschizomers:

```
5' ... C▽C C G G G ... 3'
```

Isoschizomers:

パターンが違えばリンク集からコピーする GCGCCCで始まる配列

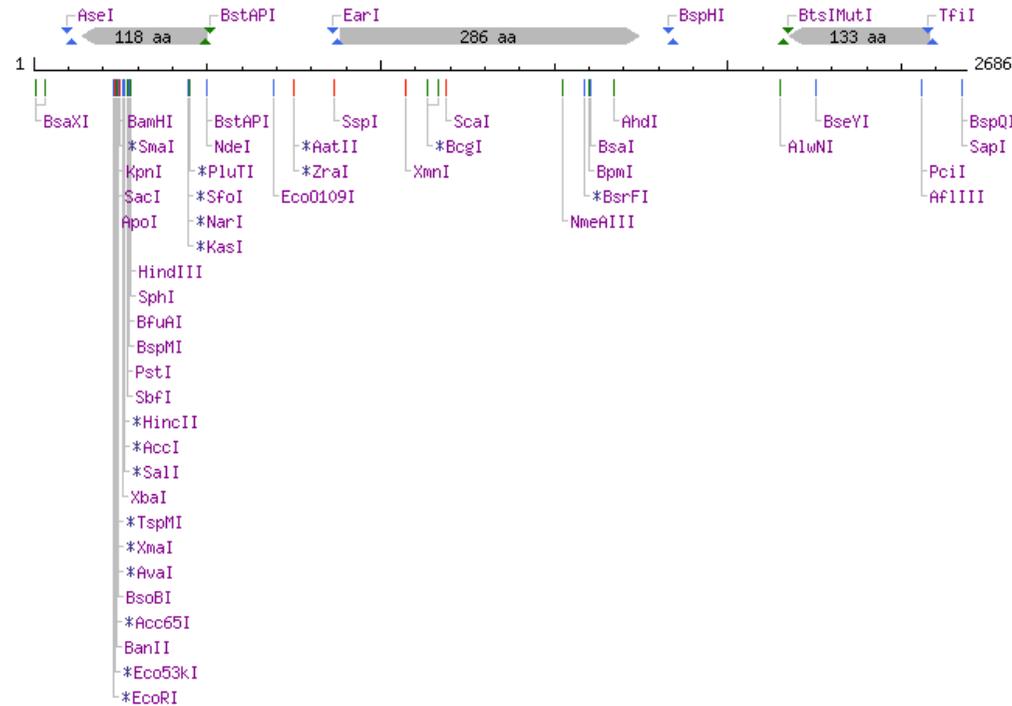


画像ファイルをダウンロードする(1)

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

GC=51%, AT=49%

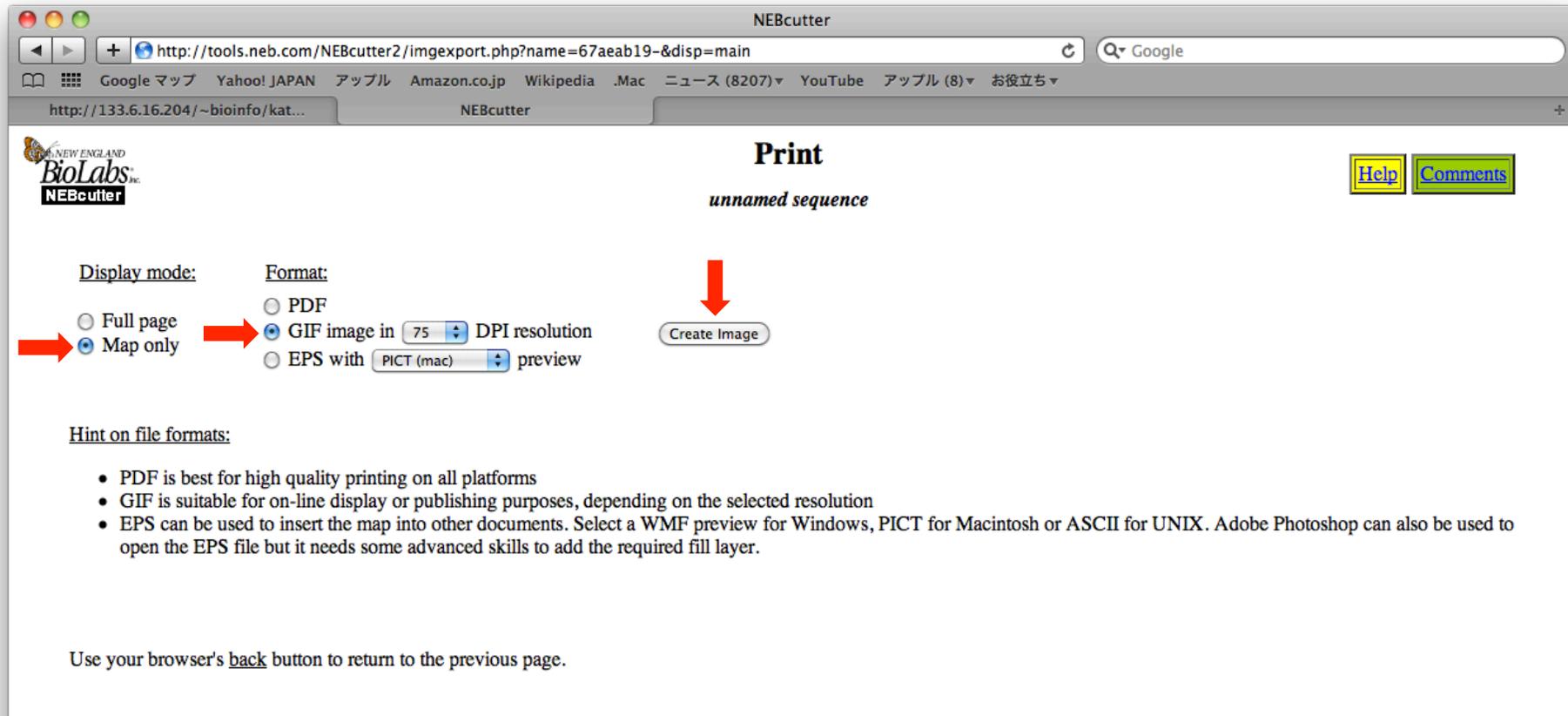
▲ 1 strand cut	Has other supplier
▼ 5' extension	Not commercially available
▼ 3' extension	*: cleavage affected by CpG meth.
▼ cuts 1 strand	#: cleavage affected by other meth.
	(enz.name): ambiguous site



Main options New DNA Custom digest View sequence ORF summary Save project Save project Print	Availability All commercial All	Display 2 cutters 3 cutters Circular	Zoom Zoom in More...	List 0 cutters 1 cutters All sites Save all sites Flanking enzymes
--	--	--	-----------------------------------	--

Minimum ORF length to display: 100 aa.

画像ファイルをダウンロードする(2)



NEBcutter

http://tools.neb.com/NEBcutter2/imgexport.php?name=67aeab19-&disp=main

Google

Google マップ Yahoo! JAPAN アップル Amazon.co.jp Wikipedia .Mac ニュース (8207) YouTube アップル (8) お役立ち

http://133.6.16.204/~bioinfo/kat... NEBcutter

Print
unnamed sequence

Help Comments

NEW ENGLAND
BioLabs
NEBcutter

Display mode: Format:

Full page Map only

PDF
 GIF image in 75 DPI resolution
 EPS with PICT (mac) preview

Create Image

Hint on file formats:

- PDF is best for high quality printing on all platforms
- GIF is suitable for on-line display or publishing purposes, depending on the selected resolution
- EPS can be used to insert the map into other documents. Select a WMF preview for Windows, PICT for Macintosh or ASCII for UNIX. Adobe Photoshop can also be used to open the EPS file but it needs some advanced skills to add the required fill layer.

Use your browser's back button to return to the previous page.

画像ファイルをダウンロードする(3)

http://133.6.16.204/~bioinfo/kat... NEBcutter

Print
unnamed sequence

NEW ENGLAND
BioLabs
NEBcutter

Help Comments

Display mode: Format:

Full page PDF

Map only GIF image in 75 DPI resolution

EPS with PICT (mac) preview

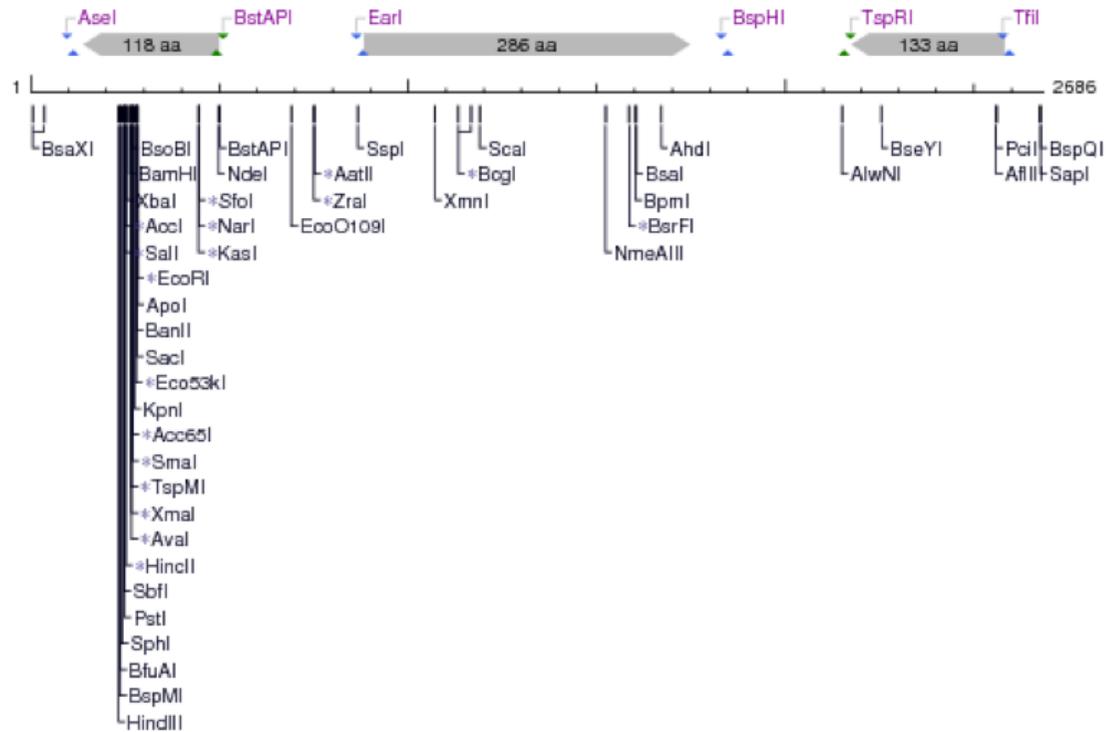
Hint on file formats:

- PDF is best for high quality printing on all platforms
- GIF is suitable for on-line display or publishing purposes, depending on the selected resolution
- EPS can be used to insert the map into other documents. Select a WMF preview for Windows, PICT for Macintosh or ASCII for UNIX. Adobe Photoshop can also be used to open the EPS file but it needs some advanced skills to add the required fill layer.

[\[Click here to view/download the GIF file \]](#)

Use your browser's back button to return to the previous page.

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

EMBOSS

[Share](#) [Feedback](#)

[Tools](#) > [EMBOSS Programs](#)

Selected EMBOSS tools for sequence analysis

Pairwise Sequence Alignment

Needle [?](#)

Create an optimal global alignment of two sequences using the Needleman-Wunsch algorithm

Sequence Translation

Transeq [?](#)

Translate nucleic acid sequences to the corresponding peptide sequences

[Launch Transeq](#) 

Sequence Statistics

Pepinfo [?](#)

Create a variety of plots that display different amino acid properties, such as hydrophobicity or charged residues, and their

Newcpgreport [?](#)

Identify CpG islands in nucleotide sequence(s)

[Launch Newcpgreport](#)

Isochore [?](#)

Plot isochores in DNA sequences

[Launch Isochore](#)

EMBL-EBI EMBOSS Transeq

EMBOSS Transeq

[Input form](#) | [Web services](#) | [Help & Documentation](#)

[Share](#) | [Feedback](#)

[Tools](#) > [Sequence Translation](#) > [EMBOSS Transeq](#)

EMBOSS Transeq

EMBOSS Transeq translates nucleic acid sequences to their corresponding peptide sequences. It can translate to the three forward and three reverse frames, and output multiple frame translations at once.

STEP 1 - Enter your input sequence

Enter or paste a set of DNA/RNA sequences in any supported format:

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

STEP 2 - Select Parameters

FRAME

1

CODON TABLE

Standard Code

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

[Submit](#)

βガラクトシダーゼ遺伝子のテキストファイルを開く

```
http://133.6.16.204/~bioinfo/kato/b-gal.txt
Google マップ Yahoo! JAPAN アップル Amazon.co.jp Wikipedia .Mac ニュース (8225) YouTube アップル (8) お役立ち
http://133.6.16.204/~bioinfo/kat... NEBcutter NEBcutter V2.0
LOCUS V00296 3078 bp DNA linear BCT 18-APR-2005
DEFINITION E. coli gene lacZ coding for beta-galactosidase (EC 3.2.1.23).
ACCESSION V00296
VERSION V00296.1 GI:41901
KEYWORDS galactosidase.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 3078)
AUTHORS Kalnins,A., Otto,K., Ruther,U. and Muller-Hill,B.
TITLE Sequence of the lacZ gene of Escherichia coli
JOURNAL EMBO J. 2 (4), 593-597 (1983)
PUBMED 6313347
REFERENCE 2
AUTHORS Zell,R. and Fritz,H.J.
TITLE DNA mismatch-repair in Escherichia coli counteracting the
hydrolytic deamination of 5-methyl-cytosine residues
JOURNAL EMBO J. 6 (6), 1809-1815 (1987)
PUBMED 3038536
COMMENT Data kindly reviewed (18-MAY-1983) by U. Ruether.
FEATURES
source Location/Qualifiers
1..3078
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
CDS <1..3072
/note="unnamed protein product; galactosidase"
/codon_start=1
/transl_table=11
/protein_id="CAA23573.1"
/db_xref="GI:1197203"
/db_xref="GOA:P00722"
/db_xref="InterPro:IPR004199"
/db_xref="InterPro:IPR006101"
/db_xref="InterPro:IPR006102"
/db_xref="InterPro:IPR006103"
/db_xref="InterPro:IPR006104"
/db_xref="InterPro:IPR008979"
/db_xref="InterPro:IPR011013"
/db_xref="PDB:1BGL"
/db_xref="PDB:1BGM"
/db_xref="PDB:1DP0"
/db_xref="PDB:1F49"
/db_xref="PDB:1F4A"
/db_xref="PDB:1F4H"
/db_xref="PDB:1GHO"
```

ボックスの中に配列をペーストする

EMBOSS Transeq

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

Tools > Sequence Translation > EMBOSS Transeq

EMBOSS Transeq

EMBOSS Transeq translates nucleic acid sequences to their corresponding peptide sequences. It can translate to the three forward and three reverse frames, and output multiple frame translations at once.

STEP 1 - Enter your input sequence

Enter or paste a set of **DNA/RNA** sequences in any supported format:

```
2701 gggccgcaag aaaactatcc cgaccgcctt actgccgcct gtttgaccg ctgggatctg
2761 ccattgtcag acatgtatac cccgtacgtc ttcccgagcg aaaacggctt gcgctgcggg
2821 acgcgcgaat tgaattatgg cccacaccag tggcgcggcg acttccagtt caacatcagc
2881 cgctacagtc aacagcaact gatggaaacc agccatgcc atctgctgca cgcggaagaa
2941 ggcacatggc tgaatatcga cggtttccat atggggattg gtggcgacga ctctggagc
3001 ccgtcagtat cggcgggaatt ccagctgagc gccggctcgt accattacca gttggtctgg
3061 tgtcaaaaat aataataa|
```

Or, upload a file: ファイル未選択

STEP 2 - Select Parameters

FRAME **CODON TABLE**

The default settings will fulfill the needs of most users and, for that reason, are not visible.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

結果

EMBOSS transeq

[Input form](#) | [Web services](#) | [Help & Documentation](#)

[Share](#) | [Feedback](#)

[Tools](#) > [Sequence Translation](#) > [EMBOSS Transeq](#)

Results for job `emboss_transeq-l20131022-094451-0222-62149969-oy`

[Tool Output](#) | [Submission Details](#)

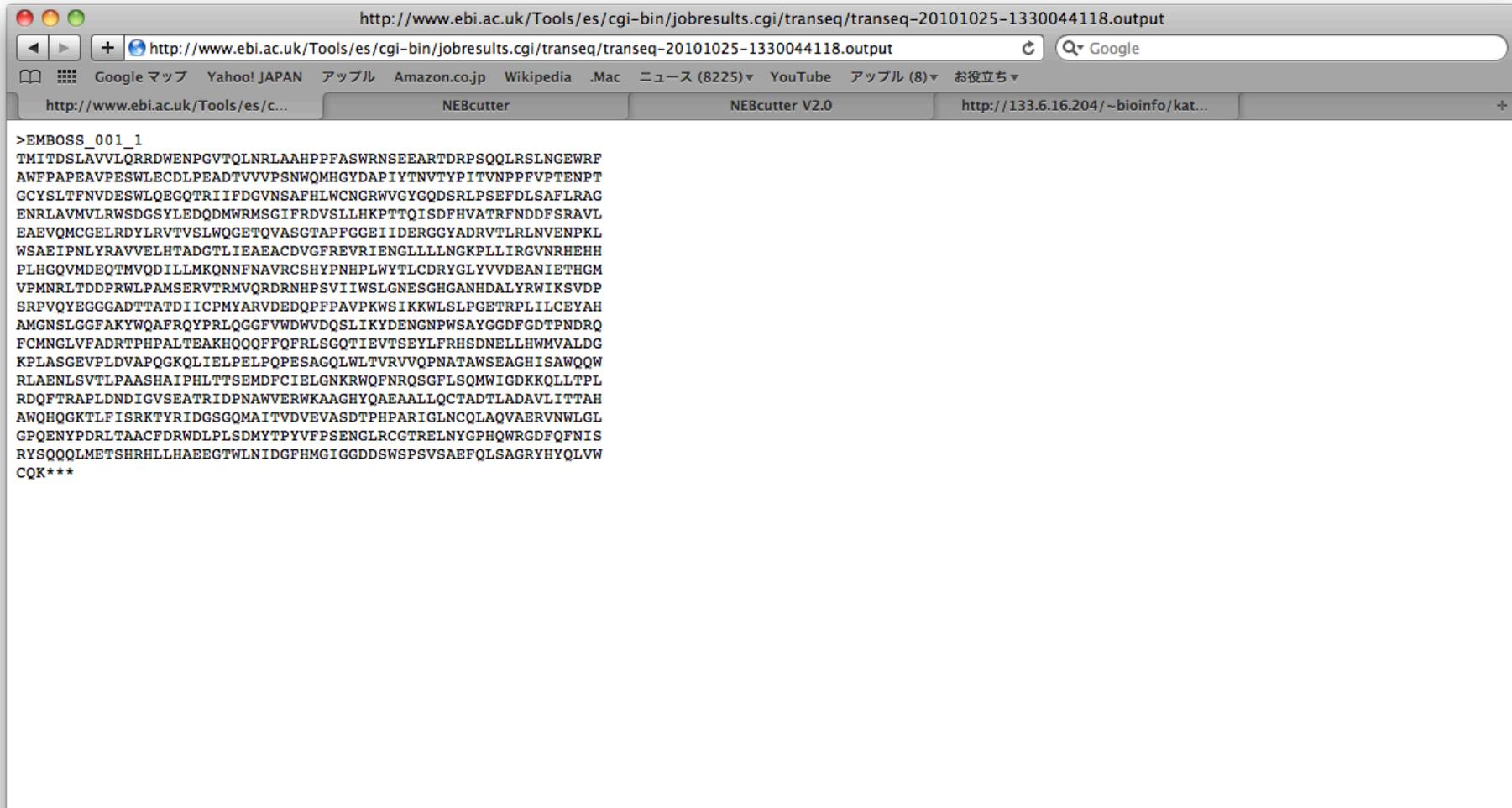
[Download](#) | [Show Colors](#)

```
>EMBOSS_001_1
TMITDSLAVVLQRRDVENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR
AWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPIYTNVTYPI TVNPPFVPTENPT
GCYSLTFNVDESWLQEGQTRII FDGVNSAFHLWCNCRWVGYGQDSRLPSEFDLSAFLRAG
ENRLAVMVLRWSDGSYLEQDMWRMSGIFRDVSL LHKPTTQISDFHVATR FNDDFSRAVL
EAEVQMC GELRDYLRVTVSLWQGETQVASGTAPFGGEI I DERGGYADRVT LRLNVENPKL
WSAEIPNLYRAVVELHTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLIRGVNRHEHH
PLHGQVMDEQTMVQDILLMKQNNFNAVRC SHYPNHPLWYTLCDRYGLYVVDEANIETHGM
VPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSV I IWSLGNESGHGANHDALYRWIKSVDP
SRPVQYEGGGADTTATDIIICPMYARVDEDQFP PAVPKWSIKKWSLPGETRPLILCEYAH
AMGNSLGGFAKYWQAFRQYPRLQGGFVWDWVDQSLIKYDENGNPWSAYGGDFGDTPNDRQ
FCMNGLVFADRTPHPALTEAKHQQQFFQFRLSGQTIEVTSEYLF R HSDNELLHWMVALDG
KPLASGEVPLDVAPQ GKQLIELPELPQPE SAGQLWLTVRVVQPNATAWSEAGHISAWQQW
RLAENLSVTLPAASHAIPHLT TSEMDFCIELGNKRWFNRQSGFLSQMWIGDKKQLLTPL
RDQFTRAPLDNDIGVSEATR IDPNAWVERWKAAGHYQAE AALLQCTADTLADAVLITTAH
AWQH QGKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQAERVNWGLG
GPQENYPDRLTAACFDRWDLPLSDMYTPYVFPSENGLR CGTRELNYGPHQWRGDFQFNIS
RYSQQQLMETSHRLLHAE EGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLVW
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



The screenshot shows a web browser window with the address bar containing the URL: `http://www.ebi.ac.uk/Tools/es/cgi-bin/jobresults.cgi/transeq/transeq-20101025-1330044118.output`. The browser tabs include "NEBcutter" and "NEBcutter V2.0". The main content area displays a FASTA format sequence for "EMBOSS_001_1".

```
>EMBOSS_001_1
TMITDSLAVVLQRRDWNPGVTQLNRLAAHPPFASWRNSEARTDRPSQQLRSLNGEWRP
AWFPAPEAVPESWLECDLPEADTVVVPNSWQMHGYDAPITYTNVTYPI TVNPPFVPTENPT
GCYSLTFNVDESWLQEGQTRII FDGVNSAFHLWCNCRWVGYGQDSRLPSEFDLSAFLRAG
ENRLAVMVLRWSDGSYLEDDQDMWRMSGIFRDVSLHKPTTQISDFHVATR FNDDFSRAVL
EAEVQMCGELRDYLRVTVSLWQGETQVASGTAPFGGEI IDERGGYADRVTLRNLNVPKL
WSAEIPNLYRAVVELHTADGTLIEAEACDVGFREVR IENGLLLLNGKPLLIRGVNRHEHH
PLHGQVMDEQTMVQDILLMKQNNFNVAVRC SHYPNHPWYTLCDRYGLYVVDEANIETHGM
VPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVI IWSLGNESGHGANHDALYRWIKSVDP
SRPVQYEGGGADTTATDI ICPMYARVDEDEDQFP PAVPKWSIKKWL SLPGETRPLILCEYAH
AMGNSLGGFAKYWQAFRQYPRLQGGFVWDVVDQSLIKYDENGPNWSAYGGDFGDTPNDRQ
FCMNGLVFADRTPHPALTEAKHQQFFQFRLSGQTIEVTSEYLF R HSDNELLHWMVALDG
KPLASGEVPLDVAPQKQLIELPELPQ PESAGQLWLT VVRVVPNATAWSEAGHISAWQQW
RLAENLSVTLPAASHAI PHLT TSEMDFCIELGNKRWFNRQSGF LSQMWIGDKKQLLTPL
RDQFTRAPLDNDIGVSEATR IDPNAWVERWKAAGHYQAEALLQCTADTLADAVLITTAH
AWQHGGKTLFISRKTYRIDGSGQMAITVDVEVASDTPH PARI GLNCQLAQVAERNWLGL
GPQENY PDRLTAACFDRWDLPLSDMYTPVFPSENGLCGTRELNYPHQRGDFQFNIS
RYSQQQLMETSHRLLHAEECTWLNIDGFHMGIGDDSWSPSVSAEFQLSAGRYHYQLVW
CQK***
```

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

The screenshot shows the NCBI ORF Finder web interface. The page title is "ORF Finder (Open Reading Frame Finder)". The navigation bar includes links for PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. The main content area contains a description of the tool and a form for inputting a sequence. The form has two sections: "Enter GI or ACCESSION" and "or sequence in FASTA format". The "or sequence in FASTA format" section contains a text area with the following sequence:

```
cttgcgaac
1081 tggggatgat gcattacgac accttgacgt tctgtccctc catgcaagct
gcttcagctg
1141 ttacacggc aagatgctca tgaacaagt cccctgcttg gactgataca
ttgcagttcc
1201 acaccggcta cacagagtct gagattatgg actgctcaaa gcttttagct
tttcttcaact
1261 cgagatgcgg tgagagcagg ctacgtgcag tgtacaagaa gtaactcgaag
acaaaaata
```

Below the text area are fields for "FROM:" and "TO:", and a dropdown menu for "Genetic codes" set to "1 Standard".

Red annotations with arrows point to the FASTA sequence and the "Genetic codes" dropdown. The first annotation says "演習データとして再び未知mRNA Xを使う" (Use unknown mRNA X again as exercise data). The second annotation says "genetic codeはstandardで" (genetic code is standard).

結果の表示(1)

ORF Finder (Open Reading Frame Finder)

PubMed Entrez BLAST OMIM Taxonomy Structure

Anonymous

View 1 GenBank Redraw 100 SixFrames

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

100 base以上のORFを表示している

結果の表示(2)

The screenshot shows the NCBI ORF Finder web interface. At the top, the browser address bar displays the URL <http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi>. The page title is "ORF Finder (Open Reading Frame Finder)". Below the title, there is a navigation bar with links to PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. The user is logged in as "Anonymous".

Below the navigation bar, there are several buttons: "View", "2 Fasta nucleotide", "ViewAll", "Redraw", and "OrfFind". The main content area displays a sequence analysis result as a series of horizontal bars. Each bar represents a different reading frame. The bars are color-coded: green bars indicate the start of an open reading frame (start codon), and pink bars indicate the end of an open reading frame (stop codon).

緑のバー: 開始コドン
ピンクのバー: 終始コドン

大腸菌ラクトースオペロンの構造

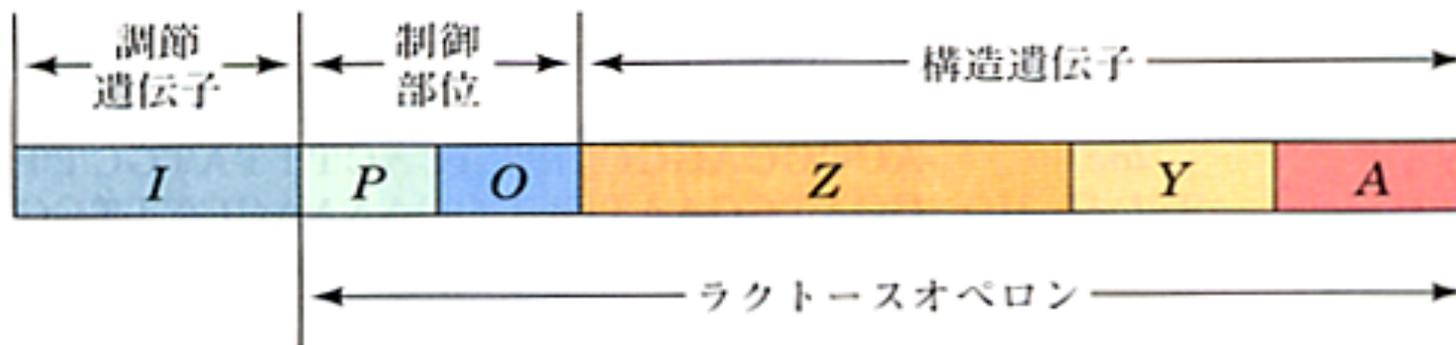
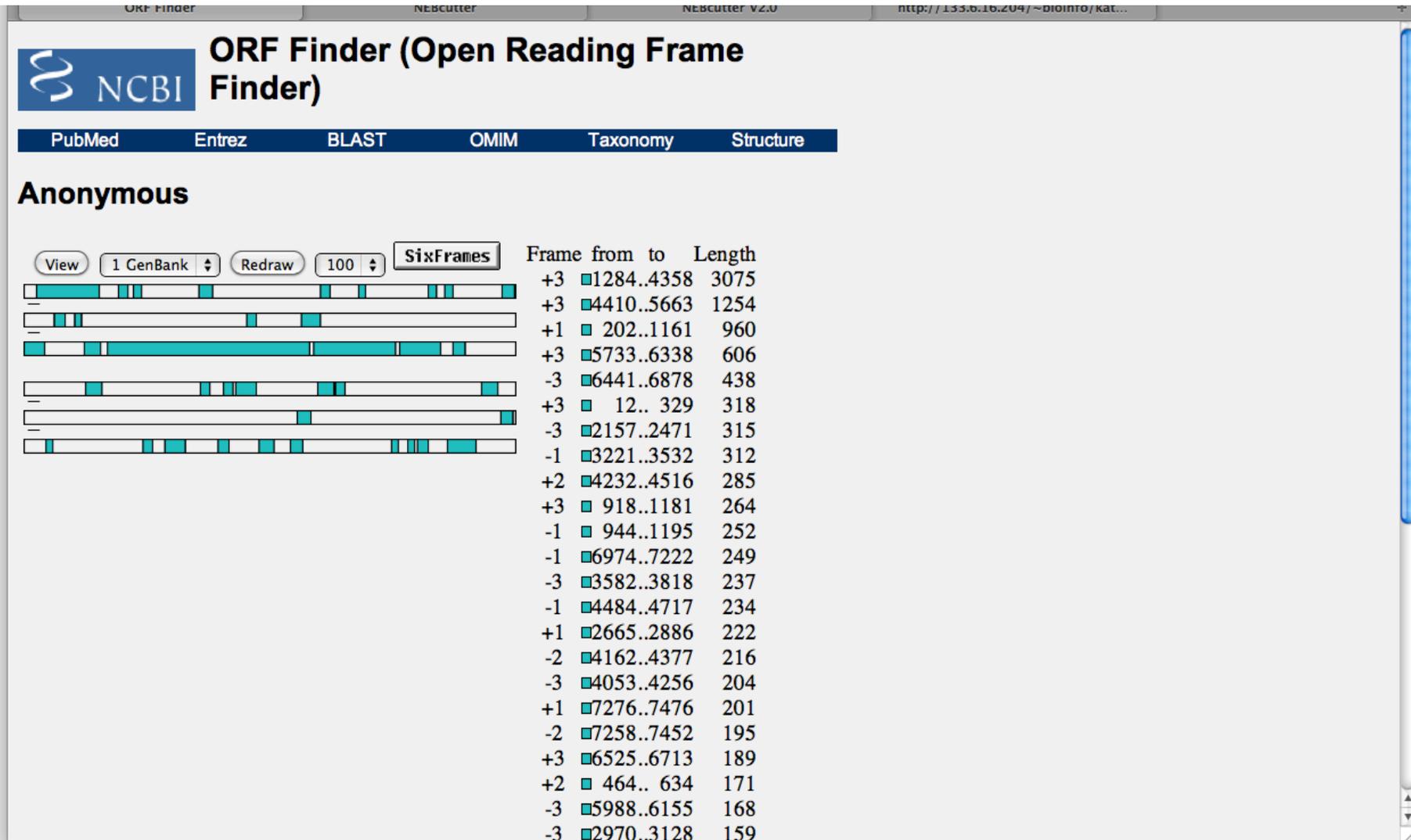


図 25・4 大腸菌 *lac* オペロン このDNAはラクトース代謝に関与するタンパクをコードする遺伝子と、その発現を制御する遺伝子部位を含む。Z遺伝子は β -ガラクトシダーゼ、Y遺伝子はガラクトシドパーミアーゼ、A遺伝子はチオガラクトシドアセチルトランスフェラーゼをコードする。すぐ近くにあるI遺伝子は*lac* オペロンには属さないが*lac* オペロンの転写を抑制するリプレッサーをコードする調節遺伝子である。

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



演習

- 出てきたアミノ酸配列と、DNAデータ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。

The GENSCAN Web Server at MIT

Identification of complete gene structures in genomic DNA

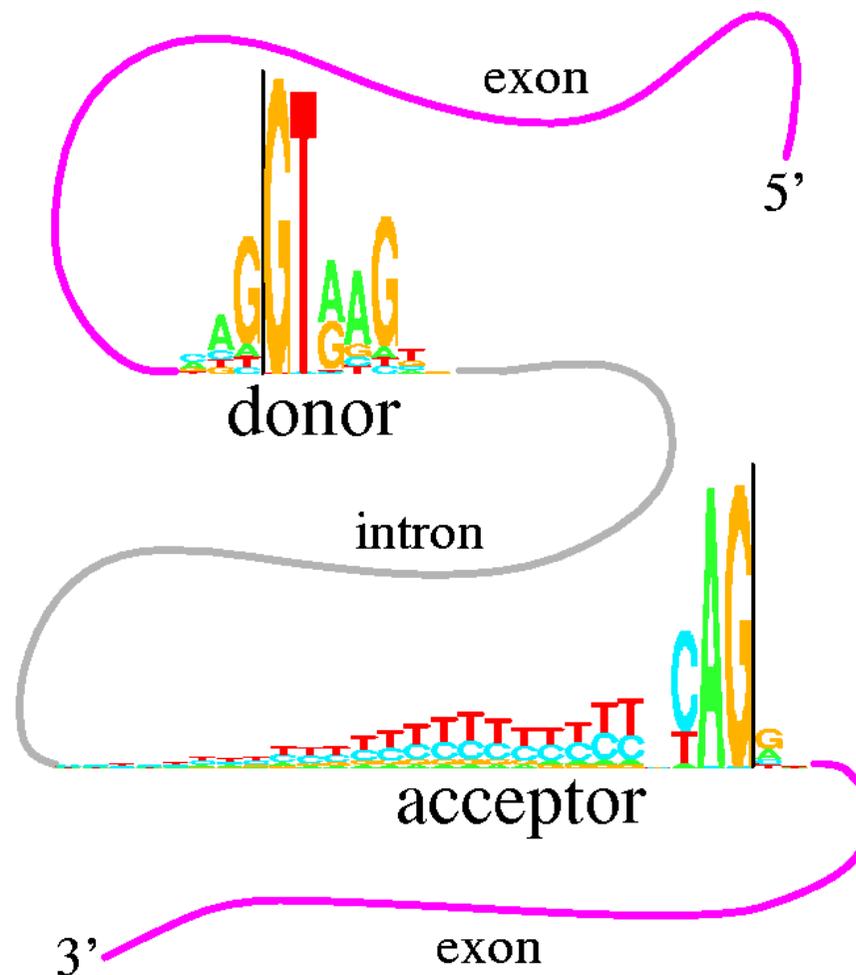


[For information about Genscan, click here](#)

Server update, November, 2009: We've been recently upgrading the GENSCAN webserver hardware, which resulted in some problems in the output of GENSCAN. We apologize for the inconvenience. These output errors were resolved.

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

This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms.

This server can accept sequences up to 1 million base pairs (1 Mbp) in length. If you have trouble with the web server or if you have a large number of sequences to process, request a local copy of the program (see instructions at the bottom of this page).

Organism: Suboptimal exon cutoff (optional):

Sequence name (optional):

Print options:

Upload your DNA sequence file (upper or lower case, spaces/numbers ignored): ファイルが...ていません

Or paste your DNA sequence here (upper or lower case, spaces/numbers ignored):

```
6901 gatgtgttc ccctaaaaa gaagaaagct gaaaaactct gtccttcca acaagacca
6961 gagcactgta gtatcagggg taaaatgaaa agtatgttat ctgctgcatc cagacttcat
7021 aaaagctgga gcttaatcta gaaaaaaaaat cagaaagaaa ttactctgtg agaacaggtg
7081 caattcactt ttccttaca cagagtaata ctggttaactc atggatgaag gcttaagggg
7141 atgaaattgg actcacagta ctgagtcac acactgaaaa atgcaacctg atacatcagc
7201 agaaggttta tgggggaaaa atgcagcctt ccaattaagc cagatatctg tatgaccaag
7261 ctgctccaga attagtcact caaaatctct cagattaaat tatcaactgt caccaacat
7321 tcctatgctg acaaggcaat tgcttgttct ctgtgttct gatactacaa ggctcttct
7381 gacttctaa agatgcatta taaaatctt ataattcaca tttctccta aacttgact
7441 caatcatggt atgttgcaa atatggtata ttactattca aattgtttc cttgtacca
7501 tatgtaatgg gtcttgtaa tgtgctcttt tgttcttta atcataataa aaacatggtt
7561 aagc
```

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
```

```
XNSEFTMGSIGAASMEFCFDVFKELKVHHANENIFYCPIATMSALAMVYLGAKDSTRTQI
```

```
NKVVRFDKLPGFGDSIEAQCSTSVNVHSSLRDILNQITKPNVYVYSLASRLYAEERYPI
```

```
LPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLQPSVDSQTAMV
```

```
LVNAIVFKGLWEKAFKDEDQAMPFRVTEQESKPVQMMYQIGLFRVASMASEKMKILELP
```

```
FASGTMSMLVLLPDEVSGLEQLESIIINFEKLTWTSNVMEERKIKVYLPRMKMEKYNL
```

```
TSVLMAMGITDVFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAGVDAAS
```

```
VSEEFRADHPFLFCIKHIATNAVLFFGRCVSP
```

[Back to GENSCAN](#)

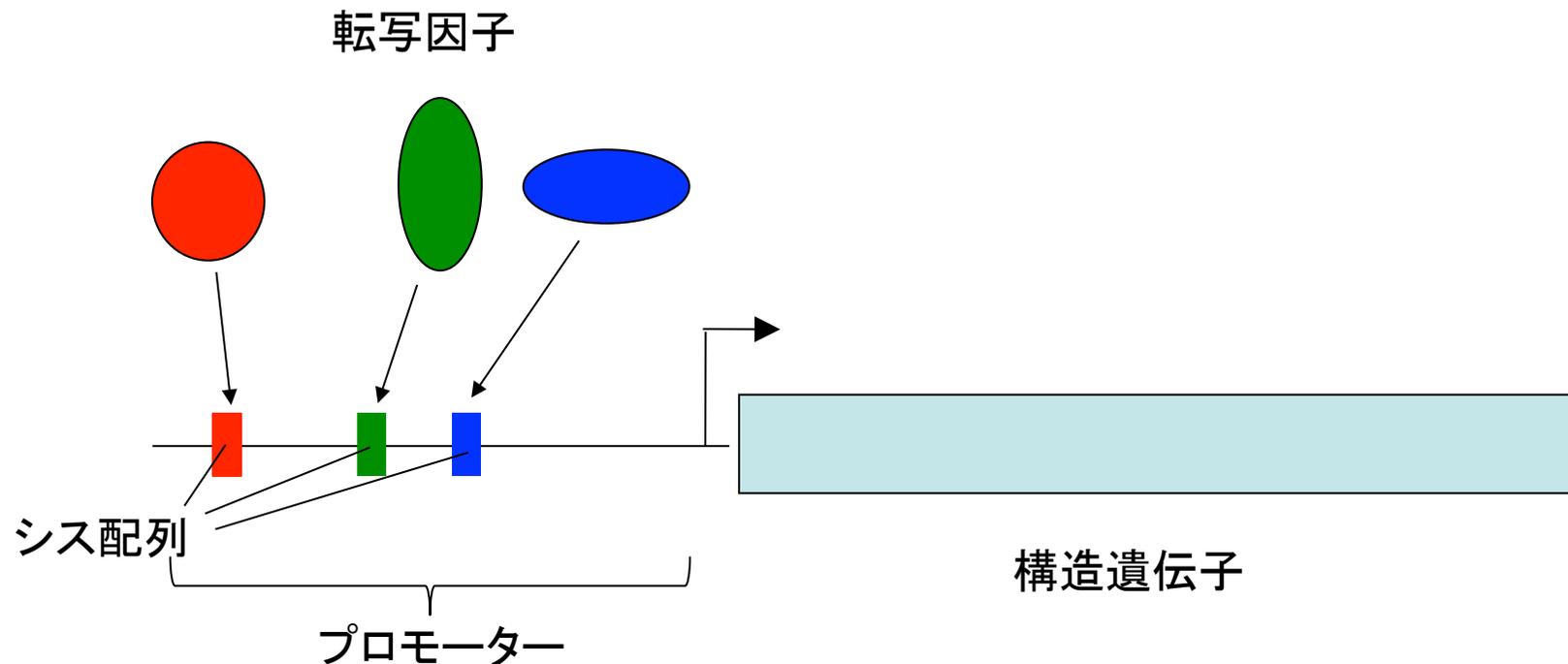
演習

- 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
GAKDSTRQINKVVRFDKLPFGDSIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSL
ASRLYAERYPIPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRN
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/>



New PLACE

A Database of Plant Cis-acting Regulatory DNA Elements

PLACE is a database of motifs found in plant cis-acting regulatory DNA elements, all from previously published reports. It covers vascular plants only (But since April 2006, we changed the policy. See the release note for PLACE 26.0). In addition to the motifs originally reported, their variations in other genes or in other plant species reported later are also compiled. The PLACE database also contains a brief description of each motif and relevant literature with PubMed ID numbers. DDBJ/EMBL/GenBank nucleotide sequence databases accession numbers will be also included. See 'sample record'. List of entries can be found in 'place.dat'.

Paste a FASTA format sequence, please.

送信

Data file of Plant Cis-acting Regulatory DNA Elements:
[place.dat](#) [place.seq](#) (30.0, 469 entries, Jan.8, 2007, © Kenichi Higo)

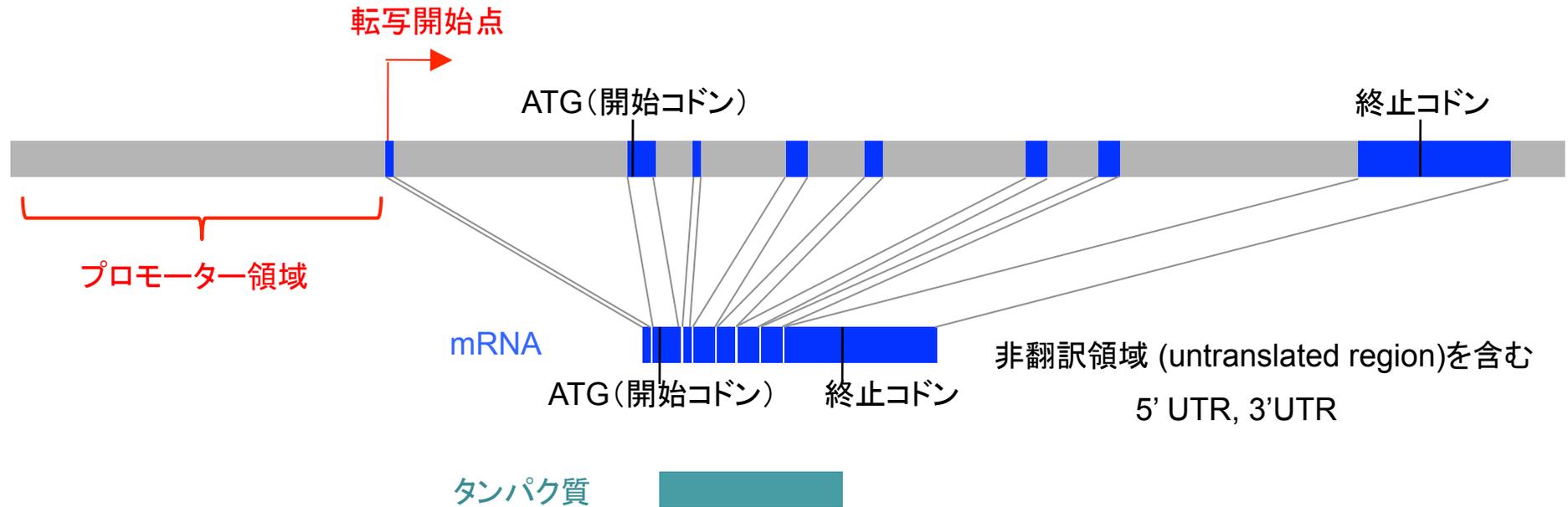
If you use this data file in published research, please cite:
[Higo, K., Y. Ugawa, M. Iwamoto and T. Korenaga \(1999\) Plant cis-acting regulatory DNA elements \(PLACE\) database. Nucleic Acids Res. 27 \(1\): 297-300.](#)

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
KRISGLIYEETRGLKIFLENVIRDAVTYTEHARRKTVTAMDVVYALKROGRTLYGFG
G"

```

ORIGIN

```

1 aagcttgaaa tcggtattct gattggttgc ctctgataag cacgagatga gctcggggat
61 gtcaataagc tcgttcagtt acacaaacag tactgtacat cagtgcctga agtgctcgtt
121 cagttaagtt tctagcacca attacctgac cgccaagcta ttacatgtaa ttattgtaac
181 gtgttatctg aatgcttgaa tcctaaaaaa gtgaactcca gtaagcgatg aaaaatgagt
241 atagcagtcg ctgcattcga gcaagtttcc ttagattat cttcagatct ccagaacagt
301 taggatgaag gaataataat cagtcattgg aggacatgca acagcaatgg agcaagtata
361 tccagttgac ttgatattct cccacatgaa aaatctcacc aaccatttaa aaaaaagaaa
421 gaaaaagaat cccaagaata cgacaccggg cccgcgcccc cagcccaaaa atttcgagct
481 gccgccacag ctctgacctc gcccgccaaa ccaccggtcc aatctctcga ccacgtcacc
541 gatccgcggc atctctcccc cggatcgccg tctcgaccgt ccactccatc cgcacccaac
601 ggcagccaca cgctctctcc aacctctcga cccttttaag acgcccttcg ccccaaccag
661 caaatcacag caccagagcg caccaccacg cgttcctccc atcccacact cgtcgcagc
721 tcgagatcgt cggccatgct cgggcgcggc aaggaaggca agggcctagg caagggcggc
781 gccaaagcgc accggaaggt cctccgcgat aacatccagg gcacaccaa gccggcgatc
841 cggcggctgg cgcggcgggg cggcgtgaag cgcacatcgg ggctcatcta cagggagacc
901 cgcggcgtgc tcaagatctt cctcgagaaac gtcatccgcg atgccgtcac ctacaccgag
961 cacgcccgcg gcaagaccgt caccgccatg gacgtcgtct acgcctcaa gcgccagggc
1021 cgcaccctct acggcttcgg cggctaaggg ccggccggcc gacgggagtc actctttgtc
1081 gccgcctgca gattccagaa gcctgatgaa gccccgactt gtttagttcg ctatttcctc
1141 tgtagtttga actcaatcgt ggaacaaagt tattcgata tattgttggg aatgaagctt

```

Search box(にペースト



[\[New regist\]](#)

login

Cross Search

[\[ja\]](#)[\[en\]](#)

New PLACE

A Database of Plant Cis-acting Regulatory DNA Elements

PLACE is a database of motifs found in plant cis-acting regulatory DNA elements, all from previously published reports. It covers vascular plants only (But since April 2006, we changed the policy. See the release note for PLACE 26.0). In addition to the motifs originally reported, their variations in other genes or in other plant species reported later are also compiled. The PLACE database also contains a brief description of each motif and relevant literature with PubMed ID numbers. DDBJ/EMBL/GenBank nucleotide sequence databases accession numbers will be also included. See 'sample record'. List of entries can be found in 'place.dat'.

Paste a FASTA format sequence, please.

```
aagcttgaaa tcggtattct ggttggtgc ctctgataag cagcagatgg gctcggggat
61 gtcaataaagc tcgttcagtt acacaaacag tactgtacat cagtgcctgga agtgctcgtt
121 cagttaagtt tctagcacca attacctgac cgccaagcta ttacatgiaa ttattgtaac
181 gtgttatctg aatgcttgaa tcctaaaaaa gtgaactcca gtaagcgatg aaaaatgagt
241 atagcagtcg ctgcattcga gcaagtttcc thtagattat cttcagatct ccagaacagt
301 taggatgaag gaataataat cagtcattgg aggacatgca acagcaatgg
agcaagtata
361 tccagttgac ttgatattct cccacatgaa aaatctacc aaccatttaa aaaaaagaaa
421 gaaaaagaat cccaagaata cgacaccggg cccgcccga cagcccaaaa
atttcgagct
```



送信

Data file of Plant Cis-acting Regulatory DNA Elements:

[place.dat](#) [place.seq](#) (30.0, 469 entries, Jan.8, 2007, © Kenichi Higo)

If you use this data file in published research, please cite:

[Higo, K., Y. Ugawa, M. Iwamoto and T. Korenaga \(1999\) Plant cis-acting regulatory DNA elements \(PLACE\) database. Nucleic Acids Res. 27 \(1\): 297-300.](#)

結果の表示(1)

668 base pairs

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

```
1      AAGCTTGAAATCGGTATTCTGGTTGGTTGCCTCTGATAAGCACGAGATGG
      (-) ARRIAT S000454 9 NGATT
      (+) -10PEHVPSBD S000392 15 TATTCT
      (-) MYBPLANT S000167 21 MACCWAMC
      (-) MYBPZM S000179 21 CCWACC
      (-) BOXLCOREDPCAL 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 TGTC
      (-) WBOXATNPR1 S000390 61 TTGAC
      (-) WRKY71OS S000447 61 TGAC
      (+) CAATBOX1 S000028 63 CAAT
      (-) MYB2AT S000177 76 TAACTG
      (-) MYB2CONSENSUSAT S000409 76 YAACKG
      (+) MYBCORE S000176 76 CNGTTR
      (-) CACTFTPPCAL S000449 89 YACT
      (-) CURECORECR S000493 90 GTAC
      (+) CURECORECR S000493 90 GTAC
      (+) CACTFTPPCAL S000449 91 YACT
      (-) CURECORECR S000493 95 GTAC
      (+) CURECORECR S000493 95 GTAC

101    CAGTGCTGGAAGTGCTCGTTCAGTTAAGTTTCTAGCACCAATTACCTGAC
      (-) CACTFTPPCAL S000449 102 YACT
      (-) CACTFTPPCAL S000449 111 YACT
      (-) MYB2AT S000177 121 TAACTG
      (-) MYB2CONSENSUSAT S000409 121 YAACKG
      (+) MYBCORE S000176 121 CNGTTR
      (-) POLLENILELAT52 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    CGCCAAGCTATTACATGTAATTATTGTAACGTGTATCTGAATGCTTGAA
```

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 (-)	ACCWCC	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
BIHD10S	60 (+)	TGTCA	S000498
WBOXATNPR1	61 (-)	TTGAC	S000390
WRKY710S	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
WRKY710S	147 (+)	TGAC	S000447

結果の表示(3)



[\[新規登録\]](#)

[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