

1. GeneMarkのホームページへ。(http://opal.biology.gatech.edu/GeneMark/)

GeneMark™ - Free gene prediction software - Windows Internet Explorer の提供元: MSN Japan

http://opal.biology.gatech.edu/GeneMark/

Google 検索

GeneMark™ - Free gene prediction soft..

GeneMark
A family of gene prediction programs developed at
Georgia Institute of Technology, Atlanta, Georgia, USA.

What's New:
Gene identification with updated heuristic models in **metagenomes**.

Supported by NIH

Gene Prediction in Bacteria, Archaea and Metagenomes

For bacterial and archaeal gene prediction we recommend to use a parallel combination of [GeneMark-P*](#) and [GeneMark.hmm-P](#) with pre-computed models.

A novel genome can be analyzed either by the program with [Heuristic models](#) (if the sequence is shorter than 200 kb) or by the self-training program [GeneMarkS*](#) (aka GeneMark.hmm-PS).

Multiple contig metagenomic sequences can be analyzed by our [new program](#) with updated heuristic models.

Gene Prediction in Eukaryotes

For eukaryotic gene prediction you can use the parallel combination of [GeneMark-E*](#) and [GeneMark.hmm-E](#).

For a novel genome (the one whose name is not in the list of available models) you can run [GeneMark.hmm-ES](#), the self-training program (just 10MB sequence is needed for training).

Gene Prediction in Viruses, Phages and Plasmids

For novel virus, phage and plasmid gene prediction you can use either the [Heuristic approach](#) (if the sequence is shorter than 50 kb) or the self-training program [GeneMarkS](#) (aka GeneMark.hmm-PS). Both options will run the parallel combination of GeneMark and GeneMark.hmm.

Gene Prediction in EST and cDNA

To analyze ESTs and cDNAs you can use [GeneMark-F](#).

Powered by IBM

Borodovsky Group

Gene Prediction Programs

- [GeneMark](#)
- [GeneMark.hmm](#)
- [GeneMarkS](#)
- [Heuristic models](#)
- [Metagenomes](#)
- [Mirror site](#)

Information

- [Background](#)
- [References](#)
- [In GenBank](#)
- [FAQ](#)
- [Contact](#)

Databases of predicted genes

- [Prokaryotes](#)
- [Viruses/Phages \(VIOLIN\)](#)

Models for Gene Prediction

大腸菌ゲノムなのでここをクリック。

2. データを入力。

GeneMark.hmm for Prokaryotes – Windows Internet Explorer の提供元: MSN Japan

http://opal.biology.gatech.edu/GeneMark/gmhm2_prok.cgi

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

Google 検索 ブックマーク ブロック数: 491 チェック 次へ送信 設定

GeneMark.hmm for Prokaryotes

GeneMark.hmm for Prokaryotes (Version 2.4) [\(Reload this page\)](#)

Reference: Lukashin A. and Borodovsky M., *GeneMark.hmm: new solutions for gene finding*, *NAR*, 1998, Vol. 26, No. 4, pp. 1107-1115.
[\[Download PDF \]](#)

- 1) This page has been updated to run version 2.4 of GeneMark.hmm, as well as version 2.5 of GeneMark.
- 2) Processing speed: 1 million nucleotides in 15 seconds.
- 3) Prediction results for sequences longer than 5 MB are sent by e-mail.

UPDATE (November 8, 2005):
Prediction models have been pre-computed for a [265](#) completely sequenced prokaryotic genomes from the NCBI RefSeq database.
Gene predictions made for these genomes are available in the [GeneMark prokaryotic database](#).

Input Sequence

Title (optional):

Sequence Text:

Sequence File upload:

Species:

塩基配列を入力。

学習セットを入力。
実習では大腸菌K12を入力。

3. 出力オプションを入力。

GeneMark.hmm for Prokaryotes – Windows Internet Explorer の提供元: MSN Japan

http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

Google 検索 検索 検索 ブックマーク 491 チェック 次へ送信 設定

GeneMark.hmm for Prokaryotes

Sequence Text:

```
ACACCATCGAATGGCGCAAAACCTTCGCGGTATGGCATGATAGCGCCG
GAAGAGAGTCAATTGCGGTGGTGAATGTGAAACAGTAACGTTATACGA
TGTCGCAGAGTATGCGGTGTCTTATCAGACCGTTCCCGCGTGGTGA
ACGAGGCCAGCCAGTTTCTGCGAAACGCGGGAAGTGAAGCGGCG
ATGGCGGAGCTGAATTACATCCCAACCGCGTGGCACAACAACCTGGCGG
CAACAGTCGTTGCTGATTGGCGTGGCACCTCCAGTCTGGCCCTGCACG
CGCCGTCGCAATTGTGCGGCGGATTAATCTCGCGCGATCAACTGGGT
GCCAGCGTGGTGGTGTGATGGTAGAACGAAGCGCGTCCGAAGCCTGTAA
AGCGGCGGTGCACAATCTCTCGCGCAACGCGTCAGTGGGCTGATCATT
ACTATCGCTGGATGACAGGATGCCATTGCTGTGAAGCTGCCTGCACT
```

Sequence File upload:

参照

Species:

Escherichia_coli_K12

☒ Use RBS model, if available

Output Options

E-Mail Address (required for graphical output or sequences longer than 5000000 bp)

☐ Generate PDF graphics (screen)

☐ Generate PostScript graphics (email)

☐ Print GeneMark 2.4 predictions in addition to GeneMark.hmm predictions

☒ Translate predicted genes into proteins

☒ Sequences of predicted genes

Run

Start GeneMark.hmm Default

Web pages maintained by [GeneMark administrator, wenhan@amber.gatech.edu](mailto:wenhan@amber.gatech.edu)
Please send any suggestions for improvements or problems to the web page maintainer.

読枠ごとの予測ORFを出力。

予測ORFのアミノ酸配列を出力。

予測ORFの塩基配列を出力。

入力がすべて完了したら
ここをクリック。

4. GeneMarkの出力。

GeneMarkhmm for Prokaryotes - Windows Internet Explorer の提供元: MSN Japan

http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi

View PDF Graphical Output

Gene Predictions in Text Format

Information on input sequence

Sequence title: reidai
Length: 6409 bp
G+C Content: 52.58 %

Parse predicted by GeneMark.hmm 2.4

GeneMark.hmm PROKARYOTIC (Version 2.6q)
Model organism: Escherichia_coli_K12
Mon Jun 16 22:00:24 2008

← +は順方向、-は逆方向。

← 予測ORF。
1番目は短いので無視する。

Gene #	Strand	LeftEnd	RightEnd	Gene Length	Class
1	+	<3	41	39	1
2	+	78	1160	1083	1
3	+	1283	4357	3075	1
4	+	4409	5662	1254	1
5	+	5779	6339	561	1

Protein translations of predicted genes

>Translation: 78..1160 (direct), 361 amino acids
VKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTRKVEAAMAEALNYIPNRVAQQLAGKQ
SLIGVATSSALHAPSQIVAAIKSRADQLGASVVVSMVERSGVEACKAAVHNLQAQRVS
GLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIIFSHEDGTRLGVEHLVALGHQ
QIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMLNEGIVPT
AMLVANDQMALGAMRAITESGLRVGADISVVGYYDDTSSCYIPPLTTIKQDFRLGQTS
VDRLLQLSQGQAVKGNQLLPSVSLVKRKTTLAPNTQTASPRALADSLMLARQVSRLESGQ
*

← 1番目(上の表では2番目)の
予測ORFのアミノ酸配列。

>Translation: 1283..4357 (direct), 1025 amino acids

インターネット 100%

GeneMark.hmm for Prokaryotes - Windows Internet Explorer の提供元: MSN Japan

http://opal.biology.gatech.edu/GeneMark/gmnhmm2_prok.cgi

Protein translations of predicted genes

>Translation: 78..1160 (direct), 361 amino acids
VKFVTLVDVAEYAGVSYQTIVSRVNVQASHVSAKTRKVEAAMAEALNYIPNRVAQQLAGKQ
SLLIQVATSSSLALHAPSQIVAAIKSRADQLGASVVVSMVERSGVEACKAAVHNLQAQVRS
GLIINYPLDDQDAIAVEAACTNVFALFLDVSQTPINSIIFSHEDGTRLGVEHLVALGHQ
QIALLAGPLSSSARLRLAGWHKYLTRNQPFIAREEGDWSAMSGFQTMQMLNEGIVPT
AMLVANDQMALGAMRAITESGLRVGADISVVGYYDDTEDSSCYIFPLTTIKQDFRLLGQTS
VDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNTQTASPRALADSLMQLARQVSRLESQ
*

>Translation: 1283..4357 (direct), 1025 amino acids
MTMITDSLAVVLQRRDWNPGVTQLNRLAAHPFASWRNSEEARTDRPSQQLRSLNGEWR
FAWFPAPAEVPEWLECDLPEADTVVPSNWQMHGYDAPIYTNVTYPTVNPFFVPTENP
TGCYSLTFNVDESWLQEGQTRIIIDGVNSAFHLWCNWRVVGQDSRLPSEFDLSAFLRA
GENRLAVMLRWSDGSYLEQDMMWRMSGIFRDVSLHLKFTTQISDFHVATRFNDDFSRAV
LEAEVQMCCELDRDYLRVTVSLWQGETQVAGSTAPFGGEIIDERGGYADRVTLRNLVENPK
LWSAEITPNLYRAVVELHTADGTLEAEACDVGFREVRTEIENGLLNGKPELLIRGVNRHEH
HPLHGQVMDEQTMVDIILLMKQNNFNVRCSHYPNHPLWYTLCDRYGLYVVDEANIEHGH
MVMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVIIWISLGNESGHHANHDALYRWIKSVD
PSRPFVYEGGGADTTATDIIICPMYARVDEDQPPFAVPKWSIKKWSLPGETRPLILCEYA
HAMGNSLGGFAKYWQAFRQYPRLQGGFVWDWVDSLIKIDENGPNWSAYGGDFGDTFNDR
QFCMNGLVFADRTPHPALTEAKHQQFFQFRLSGQTIEVTSEYLFRHSDNELLHWMVALD
GKPLASGEVPLDVAPQKQLIELPELPQESAGQLWLTVRVVPNATAWSEAGHISAWQQ
WRLAENLSVTLPAASHAIPHLLTSEMDFCIELGNKRWFQNRQSGFSLQMWIGDKQLLTP
LRDQFTRAPLNDIGVSEATRIPNAWVERWKAAGHYQAEAAQLQCTADTLADAVLITTA
HAWQHGGKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAERNWLG
LGPQENYPRDLTAACFDRWDLPLSDMYTPYVFPSENGLRGCTRELNYGFHQWRGDFQFNI
SRYSQQLMETSHRLLHAEBSGTLNIDGFHMGIGGDSWSPVSAEFQLSAGRYHYQLV
WCQK*

>Translation: 4409..5662 (direct), 418 amino acids
MYLKNNTNFWMFLGFFFFYFFIMGAYFPFFPIWLHDINHISKSDTGIIFAAISLSFLLFQ
PLFGLSLDKLGLRKYLLWITGMLVMFAPFFIFIFGPLLQYNILVGSIVGGIYLGFCFNA
GAPAVEAFIEKVSRRSNFEFGARMFGCVGWALCASIVGIMFTINNQFVFWLGSQCALIL
AVLLFFAKTDAPSSATVANAVGANHSFSLKALELFRQPKLWFLSLYVIGVSTYDVDF
QQFANFFTSFFATGEQGRVFEVYVTMGELNLASIMFFAPLIINRIGGKNALLAGTIMS
VRIIGSSFATSALLEVILKTLHMFVFPFLVGCFFKYITSQFEVFRSATIYLVCFCEFFKQL
AMIFMSVLAGNMYESIGFQGAYLVGLVALGFTLISVFTLSGPGPLSLRRQVNEVA*

>Translation: 5779..6339 (direct), 187 amino acids
MCEGLPEKRLRGKTLMEFNSHSPSEVEKRESLIKEMFATVGENAWVEPPYFYSYGSNIH
IGRNFYANFNLTIVDDYTVTIGDNVLIAPNVTLSVTGHPVHHELRKNGEMYSFPITIGNN
VWIGSHVIVNPGVTIGDNSVIGAGSIVTKDIPNVVAAGVPCRVIREINDRKHYFKDY
KVESSV*

1番目(上の表では2番目)の
予測ORFのアミノ酸配列。

2番目の予測ORFのアミノ酸配列。

3番目の予測ORFのアミノ酸配列。

4番目の予測ORFのアミノ酸配列。

インターネット 100%

GeneMark.hmm for Prokaryotes - Windows Internet Explorer の提供元: MSN Japan

http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi

Google 検索

GeneMark.hmm for Prokaryotes

```
GGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCG
GCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGA
TTAGGGCCGCAAGAAAATATCCCGACCGCCTTACTGCCGCTGTTTGGACCGCTGGGAT
CTGCCATTGTGACACATGTATACCCCGTACGTCTTCCCGAGCGAAACGGTCTGCGCTGC
GGGACGCGGAATTGAATTATGGCCCAACACAGTGGCGCGGCGACTTCCAGTTCAACATC
AGCCGCTACAGTCAACAGCAACTGATGGAACAGCCATCGCCATCTGCTGCACGCGGAA
GAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGG
AGCCCGTCAGTATCGGCGGAATCCAGCTGAGCGCGGTCGCTACCATTACCAGTTGGTC
TGGTGTCAAAAATAA
```

>Nucleotides: 4409..5662 (direct), 1254 nucleotides

```
ATGTAATTTTAAAAACACAACTTTTGGATGTTTCGTTTATCTTTTCTTTTACTTT
TTTATCATGGGAGCCTACTTCCCGTTTTTCCCGATTGGGTACATGACATCAACCATATC
AGCAAAAGTGATACGGGTATTTTGGCGCTATTCTCTGTTCTCGCTATTATCCAA
CCGCTGTTTGGTCTGCTTTCTGACAACTCGGGCTGCGCAAATACCTGCTGGGATTAT
ACCGCATGTTAGTGATGTTTGGCGCTTCTTTATTTTATCTCGGGCCACTGTTACAA
TACAACATTTTAGTAGGATCGATTGTTGGTGGTATTTATCTAGGCTTTTGGTTTAAACCC
GGTGGCGCAGCAGTAGAGGCTTTATTTGAGAAAGTCAGCGTCGCGAGTAATTTGCAATTT
GGTCGCGCGCGGATGTTTGGCTGTGTTGGCTGGGCGCTGTGCTCGCTCGATTGTCGCAATC
ATGTTACCATCAATAATCAGTTTGTCTGGCTGGGCTCTGGCTGTGCACTCATCTC
GCCGTTTTTACTCTTTTTCGCCAAACGAGTGGCGCTCTTCTGCCACGGTGGCAATGCG
GTAGGTGCCAACCATTCGGCATTAGCCTTAAGCTGGCACTGGAACGTGTCAGACAGCCA
AAACTGTGGTTTTGTCACTGTATGTTATTTGGCGTTTCTGCACCTACGATGTTTTGAC
CAACAGTTTGCTAATTTCTTACTTCTGTTCTTTGCTACCGGTGAACAGGTCACGCGGTA
TTTGGCTACGTAAACGACAATGGGGAATTAACGCTCGATTATGTTCTTTGCGCCA
CTGATCATTAATCGCATCGTGGGAAAAACGCCCTGCTGCTGGCTGGCACTATTATGCTC
GTACGTATTATTGGCTCATCGTTCCGCCCTCAGCGCTGGAAGTGGTTATTCTGAAAACG
CTGCATATGTTGAAGTACCGTTCTGCTGGTGGGCTGCTTTAAATATATTACAGCCAG
TTTGAAGTGGCTTTTTCAGCGACGATTTATCTGGTCTGTTCTGCTTCTTAAAGCAACTG
GCGATGATTTTATGCTGTACTGGCGGCAATATGATGAAAGCATCGGTTTCCAGGGC
GCTTATCTGGTGTGGGTCTGGTGGCGCTGGGCTTACCTTAATTTCCGTGTTACGCTT
AGCGGCCCGGCCCGCTTTCCCTGCTGCTGCTCAGGTGAATGAAGTCGCTTAA
```

>Nucleotides: 5779..6339 (direct), 561 nucleotides

```
ATGTGCGAAGGCTTACCGGAAAAAGACTTCGTGGGAAAAACGTTAATGTATGAGTTTAA
CACTCGCATCCATCAGAAAGTTGAAAAAGAGAAAGCCTGATTAAAGAAATGTTTGCCAG
GTAGGGGAAAAACGCTGGGTAGAACCGCTGTCTATTTCTCTTACGGTTCCAACATCCAT
ATAGGCGCAATTTTATGCAAAATTTCAATTTAACCATTTGTCGATGACTACACGGTAACA
ATCGGTGATAACGTACTGATTGACCCACGTTACTCTTTCCGTTACGGGACACCTGTA
ACCATGAATTGAGAAAAAACCGCGAGATGTACTCTTTCCGATAACGATTGGCAATAAC
GTCGGATCGGAAGTCATGTGGTTATTAATCCAGGCGTCACCATCGGGGATAATTCTGTT
ATTGGCGGGGTAGTATCGTCAAAAAAGACATTCACCAAACGTCGTGGCGGCTGGCGTT
CCTTGTGGGTTATTCGCGAAATAAACGACGGGATAAGCACTATTATTTCAAAGATTAT
AAAGTTGAATCGTCAGTTTAA
```

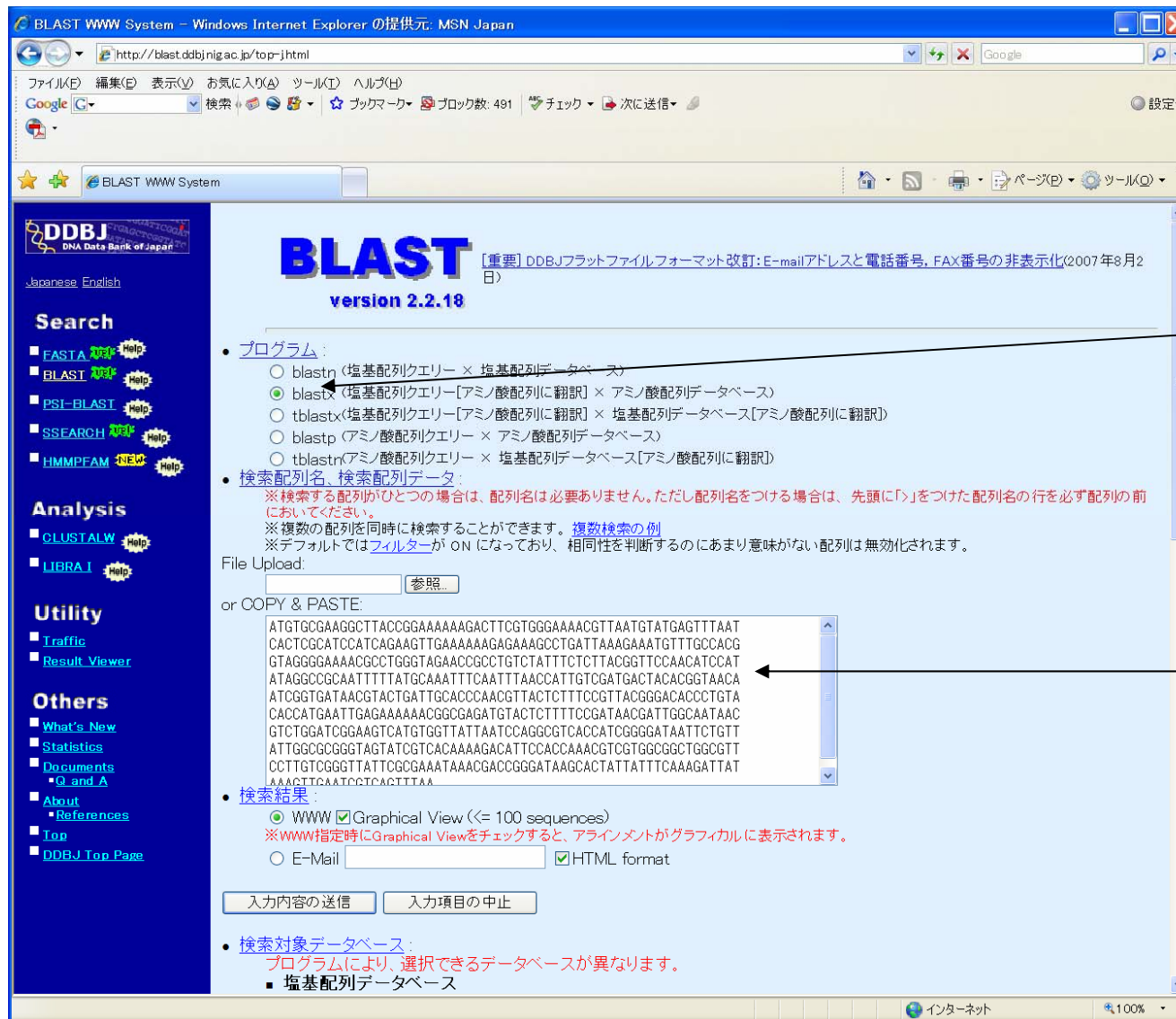
← 3番目の予測ORFの塩基配列。

← 4番目の予測ORFの塩基配列。

ページが表示されました

インターネット 100%

5. BLASTの入力。予測されたORFの確認。



アミノ酸配列でも塩基配列でも
どちらでもよいが、ここでは塩基
配列で入力する。

—例として4番目の予測ORFの塩基配列。

5. BLASTの出力。予測されたORFの確認。

RESULT OF BLAST - Windows Internet Explorer の提供元: MSN Japan

http://blast.ddbj.nig.ac.jp/blast/result

Google 検索

RESULT OF BLAST

受付番号は、【20080617112048_29366】です。

[CLUSTALW SETUP (Graphical View(<= 100 sequences) | Text View(any number of sequences))]

BLASTX 2.2.18 [Mar-02-2008]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= query
(561 letters)

Databasc: PROTEIN: UniProt/Swiss-Prot + TREMBL + PDBSH + PRF protein sequence database [Last update Jun/17/2008]
7,301,011 sequences; 2,374,778,426 total letters

Searching.....done

Sequences producing significant alignments:

	Score	E
	(bits)	Value
prf 1111315A transacetylase,thiogalactoside	387	e-106
pdb 1KRV_C 2.80 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRV_B 2.80 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRV_A 2.80 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRU_C 2.80 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRU_B 2.80 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRU_A 2.80 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRR_C 2.50 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRR_B 2.50 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KRR_A 2.50 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KQA_C 3.20 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KQA_B 3.20 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
pdb 1KQA_A 3.20 3 203 NIMOL_ID: 1; MOLECULE: GALACTOSIDE O-ACE...	387	e-106
tr B1XBJO B1XBJO_EC0DH Thiogalactoside acetyltransferase.	387	e-106
tr B1J0T7 B1J0T7_EC0LC Transferase hexapeptide repeat containing...	387	e-106
sp P07464 THGA_ECOLI Galactoside O-acetyltransferase (EC 2.3.1.1...	387	e-106
prf 3217355UK galactoside O-acetyltransferase	380	e-104

インターネット 100%

ここらあたりは同じスコアなので
同じタンパク質。

RESULT OF BLAST - Windows Internet Explorer の提供元: MSN Japan

http://blast.ddbj.nig.ac.jp/blast/result

RESULT OF BLAST

Rank	Accession	Protein Name	Score
175	spIP529841TH6A_LACLA	Galactoside O-acetyltransferase (EC 2.3.1.18) (GAT) (Thiogal...	1e-42
174	prf13221374EKU1	galactoside O-acetyltransferase	3e-42
174	prf12887262A1	thiogalactoside transacetylase	3e-42
174	trIQ93LI51Q93L_I5_LACDL	Thiogalactoside transacetylase,	3e-42
174	trIQ04A591Q04A59_LACDB	Galactoside O-acetyltransferase,	3e-42
174	trIQ65VI61Q65VI6_HANSH	MbbJ protein,	4e-42
173	prf13219351AUC1	maltose O-acetyltransferase	5e-42
173	trIQ08TR51Q08TR5_LACPL	Galactoside O-acetyltransferase (EC2.3.1.18),	5e-42
173	trIQ18A661Q18A66_CL006	Maltose O-acetyltransferase (EC 2.3.1.79),	5e-42
173	trIA5N1K01A5N1K0_CL0K5	Predicted acetyltransferase (EC 2.3.1.-),	7e-42

GO TO CLUSTALW

Go to top

>prf1111315A|transacetylase,thiogalactoside
Length = 203

Score = 387 bits (994), Expect = e-106
Identities = 186/186 (100%), Positives = 186/186 (100%)
Frame = +1

Query: 1 MCEGLPEKRLRGKTLMYEFNHSHPSEVEKRESLIKEMFATVGENAWVEPPVYFSGSNIH 180
MCEGLPEKRLRGKTLMYEFNHSHPSEVEKRESLIKEMFATVGENAWVEPPVYFSGSNIH
Sbjct: 18 MCEGLPEKRLRGKTLMYEFNHSHPSEVEKRESLIKEMFATVGENAWVEPPVYFSGSNIH 77

Query: 181 IGRNFYANFLTIVDDYTVTIGDNVLIAPNVTLSVTGHPVHHELKNGEMYSFPITIGNN 360
IGRNFYANFLTIVDDYTVTIGDNVLIAPNVTLSVTGHPVHHELKNGEMYSFPITIGNN
Sbjct: 78 IGRNFYANFLTIVDDYTVTIGDNVLIAPNVTLSVTGHPVHHELKNGEMYSFPITIGNN 137

Query: 361 VWIGSHVVINPGVTIGDNSVIGAGSIVTKDIPPNVVAAGVPCRVIREINDRDKHYFVKDY 540
VWIGSHVVINPGVTIGDNSVIGAGSIVTKDIPPNVVAAGVPCRVIREINDRDKHYFVKDY
Sbjct: 138 VWIGSHVVINPGVTIGDNSVIGAGSIVTKDIPPNVVAAGVPCRVIREINDRDKHYFVKDY 197

Query: 541 KVESSV 558
KVESSV
Sbjct: 198 KVESSV 203

Go to top

>pdb11KRY_C|2.80| 3| 203|NMOL_ID: 1; MOLECULE: GALACTOSIDE
O-ACETYLTRANSFERASE; C
Length = 203

Score = 387 bits (994), Expect = e-106
Identities = 186/186 (100%), Positives = 186/186 (100%)

当然、全く同じタンパク質が見つかった。しかしよく見ると、
入力した塩基配列は、18番目のアミノ酸からのものである。

検索されたタンパク質は203アミノ酸なので、17アミノ酸少ない。

確認するため、ここをクリック。

DDBJ sequence search by accession number - Windows Internet Explorer の提供元: MSN Japan

http://getentry.ddbj.nig.ac.jp/search/get_entry?mode=view&type=flatfile&database=prf&accnumber=1111315A

Number = [1111315A]

>ref:1111315A
NAME transacetylase,thiogalactoside
determine gene
SOURCE Escherichia coli
taxon Bacteria;Proteobacteria;gamma subdivision;Enterobacteriaceae
JOURNAL Proc.Natl.Acad.Sci.U.S.A., 82(19),6414-6418(1985)
AUTHOR Hediger,M.A., Johnson,D.F., Nierlich,D.P., Zabin,I.
TITLE DNA Sequence of the lactose operon. The lacA gene and the transcriptional termination region.
KEYWORD lacA Gene Lactose Operon E.coli Plasmid Construction
Seq Determination 1820bp 203AAs Lac mRNA 4410-6230
Thiogalactoside Transacetylase
ORF of 234AAs on Complementary Strand
COMMENT Completion of lac Operon Seq UUG as Translation Initiation Codon
gene lacA EC2.3.1.18
CROSSREF PIR=XXECTG
SEQUENCE
MNMPMTIRIR AGKLFTDMCE GLPEKRLRGK TLMYEFNHSH PSEVEKRESL IKEMFATVGE
NAWVEPPVYF SYGSNIHIGR NFYANFNLT I YDDYTYTIGD NVLIAPNVTL SVTGHVPVHHE
LRKNGEMYSF PITIGNNVWI GSHVVINPGV TIGDNSVIGA GSIIVTKDIPP NWVAAGVPCR
VIREINDRDK HYYFKDYKVE SSV
END

ページが表示されました インターネット 100%

入力した塩基配列は、18番目のアミノ酸(MCEGL...)からのものであった。

塩基配列を17アミノ酸(17×3=51塩基)さかのぼってみる。

塩基配列を17アミノ酸(17×3=51塩基)さかのぼってみるとTTGを開始コドンとしてMNMPMT・・・と続き、検索されたタンパク質のORFであることがわかる。つまりGeneMarkの予測は、今回、開始コドンを間違ったことになる。しかしながら、BLASTによるホモロジー検索で正しいORFを導くことができた。ちなみにGeneMarkは、他の3つの予測ORFについては正しく予測している。

正しい開始コドン GeneMarkが予測した開始コドン

CCAACATATCATAACGGAGTGATCGCA**TTGAACATGCCAATGACCGAAAG**
AATAAGAGCAGGCAAGCTATTTACCGATATGTGCGAAGGCTTACCGGAAA
AAAGACTTCGTGGGAAAACGTTAATGTATGAGTTTAATCACTCGCATCCA
TCAGAAGTTGAAAAAAGAGAAAGCCTGATTAAAGAAATGTTTGCCACGGT
AGGGGAAAACGCCTGGGTAGAACCGCCTGTCTATTTCTCTTACGGTTCCA
ACATCCATATAGGCCGCAATTTTTATGCAAATTTCAATTTAACCATTGTC
6000
GATGACTACACGGTAACAATCGGTGATAACGTACTGATTGCACCCAACGT
TACTCTTTCCGTTACGGGACACCCTGTACACCATGAATTGAGAAAAAACG
GCGAGATGTACTCTTTTCCGATAACGATTGGCAATAACGTCTGGATCGGA
AGTCATGTGGTTATTAATCCAGGCGTCACCATCGGGGATAATTCTGTTAT
TGGCGCGGGTAGTATCGTCACAAAAGACATTCCACCAAACGTCGTGGCGG
CTGGCGTTCCCTTGTCGGGTTATTCGCGAAATAAACGACCGGGATAAGCAC
TATTATTTCAAAGATTATAAAGTTGAATCGTCAGTTTAAATTATAAAAAT
TGCCTGATACGCTGCGCTTATCAGGCCTACAAGTTCAGCGATCTACATTA
GCCGCATCC