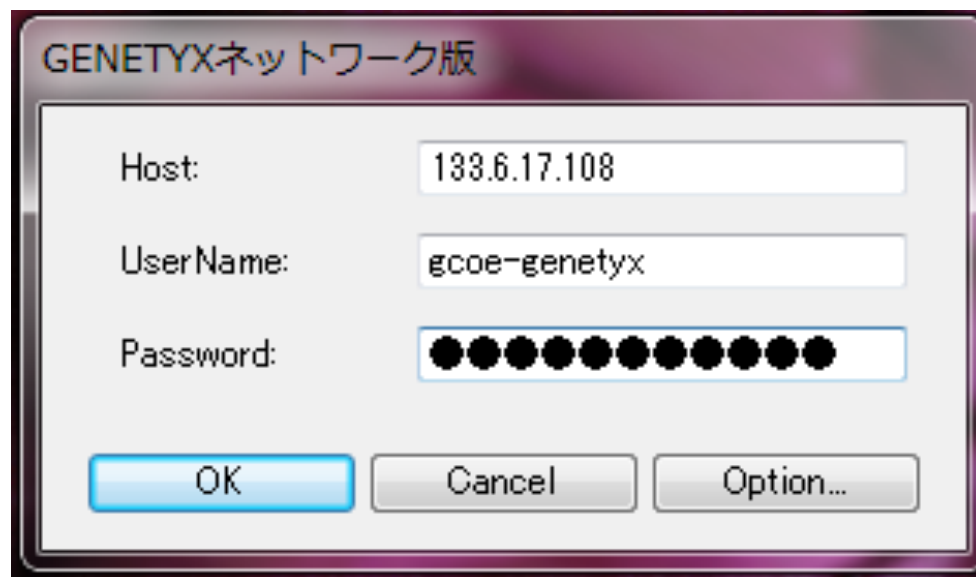


11月25日

内容

配列解析ソフト”Genetyx”を使ってみる

## Genetyx ネットワーク版の起動



The image shows a login dialog box titled "GENETYXネットワーク版". It contains three input fields: "Host:" with the value "133.6.17.108", "UserName:" with the value "gcoe-genetyx", and "Password:" which is obscured by ten black dots. At the bottom, there are three buttons: "OK" (highlighted in blue), "Cancel", and "Option...".

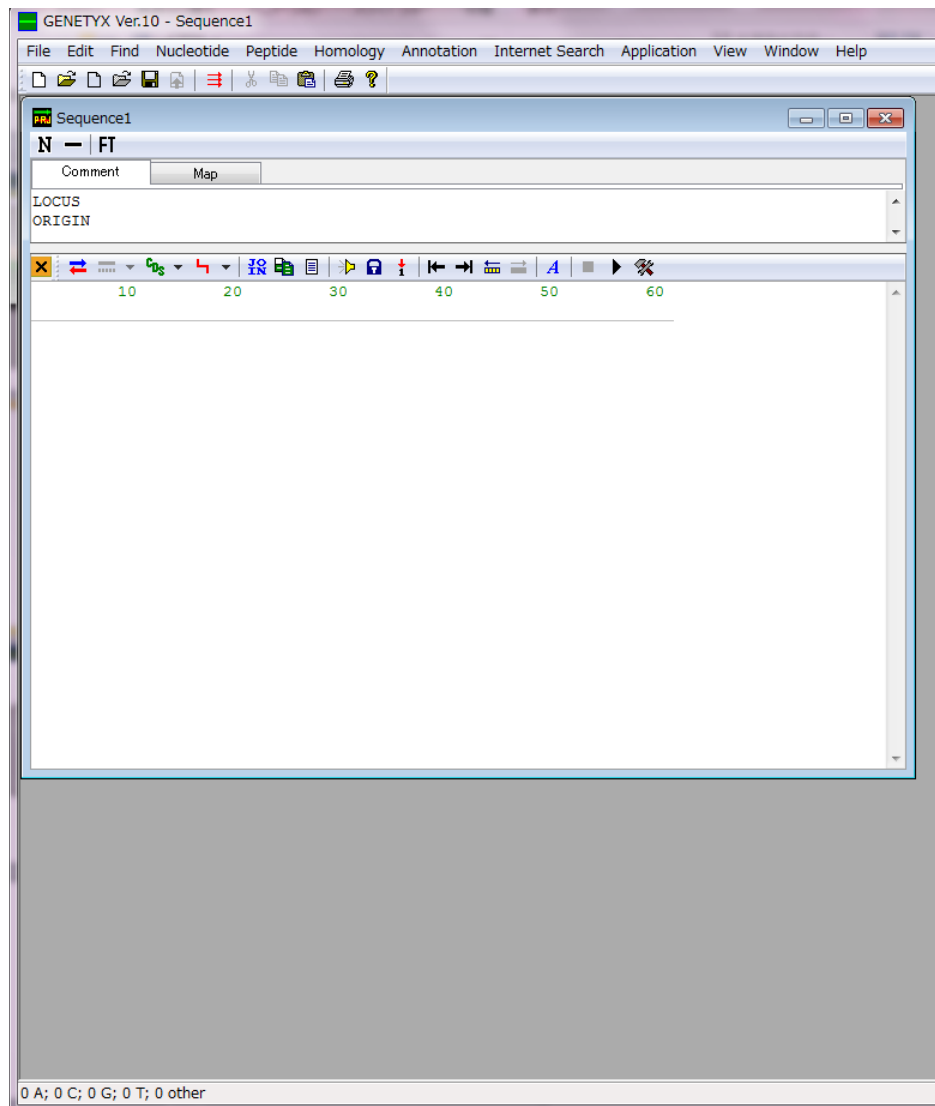
授業HPからHost, User Name, Passwordをコピー&ペースト

# Genetyx ネットワーク版の起動



# 復習: 未知配列から、そこに含まれる遺伝子を予測する

① 授業HPよりGenome seqをコピー&ペースト



GENETYX Ver.10 - Sequence1

File Edit Find Nucleotide Peptide Homology Annotation Internet Search Application View Window Help

Sequence1

N — FT

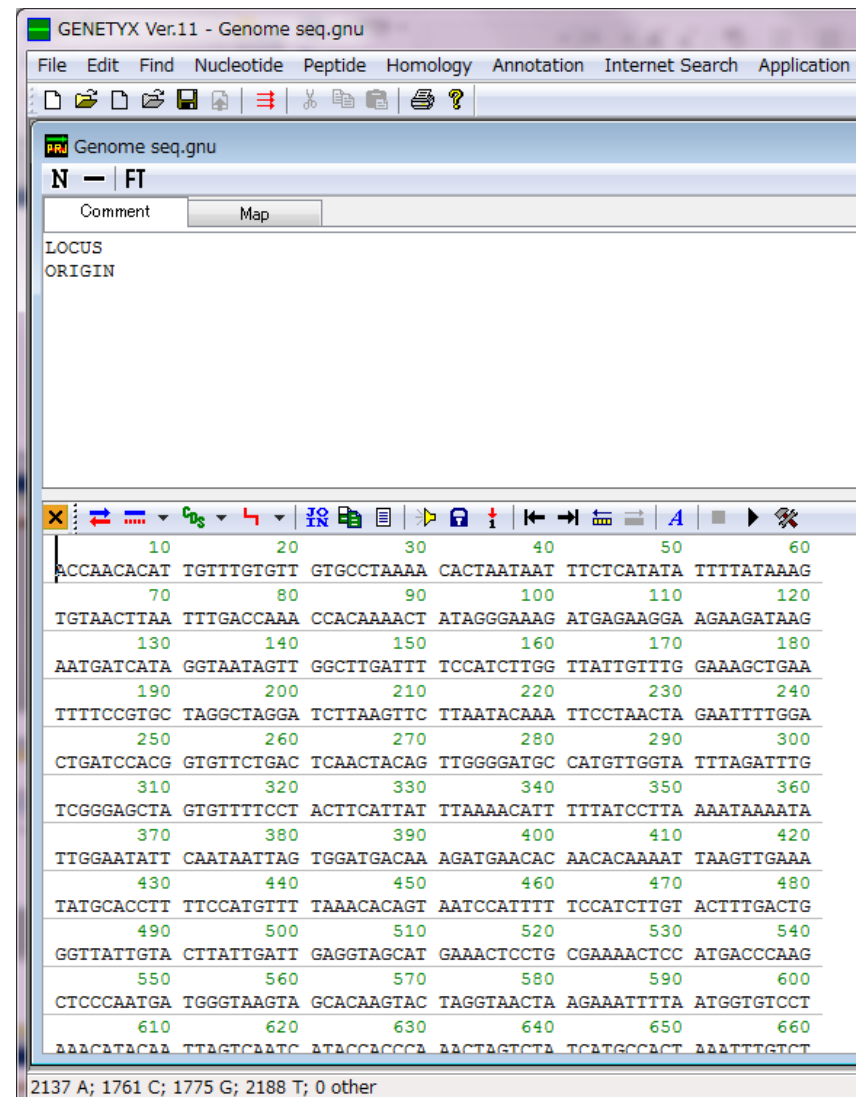
Comment Map

LOCUS  
ORIGIN

10 20 30 40 50 60

0 A; 0 C; 0 G; 0 T; 0 other

② Ctrl + Sで配列を"Genome seq"として保存



GENETYX Ver.11 - Genome seq.gnu

File Edit Find Nucleotide Peptide Homology Annotation Internet Search Application

Genome seq.gnu

N — FT

Comment Map

LOCUS  
ORIGIN

10 20 30 40 50 60  
ACCAACACAT TGTTTGTTT GGCCTAAAA CACTAATAAT TTCTCATATA TTTTATAAAG  
70 80 90 100 110 120  
TGTAACITAA TTGACCAAA CCACAAAAC ATAGGGAAAG ATGAGAAGGA AGAAGATAAG  
130 140 150 160 170 180  
AATGATCAIA GGTAATAGTT GGCTTGATTI TCCATCTTGG TTATTGTTTG GAAAGCTGAA  
190 200 210 220 230 240  
TTTCCGTGC TAGGCTAGGA TCTTAAGTTC TTAATACAAA TTCCTAATA GAATTTTGGA  
250 260 270 280 290 300  
CTGATCCACG GTGTTCTGAC TCAACTACAG TTGGGGATGC CATGTTGGTA TTTAGATTG  
310 320 330 340 350 360  
TCGGGAGCTA GTGTTTTCCT ACTTCATTAT TTAACAACATT TTTATCCTTA AAATAAAATA  
370 380 390 400 410 420  
TTGGAATATT CAATAATTAG TGGATGACAA AGATGAACAC AACACAAAAT TAAGTTGAAA  
430 440 450 460 470 480  
TATGCACCTT TTCCATGTTT TAAACACAGT AATCCATTTT TCCATCTTGT ACTTTGACTG  
490 500 510 520 530 540  
GGTTATTGTA CTTATTGATT GAGGTAGCAT GAAACTCCTG CGAAAACCTC ATGACCCAAG  
550 560 570 580 590 600  
CTCCCAATGA TGGGTAAGTA GCACAAGTAC TAGGTAATA AGAAATTTTA ATGGTGTCCCT  
610 620 630 640 650 660  
AAACATACAA TTAGTCATTC ATACCACCCA AACTAGTCTA TCATGCCACT AATTGTGTCT

2137 A; 1761 C; 1775 G; 2188 T; 0 other



# 復習: 未知配列から、そこに含まれる遺伝子を予測する

Google

fgenesh softberry ① fgenesh softberryのキーワードで検索

ウェブ 地図 ニュース 動画 ショッピング もっと見る ▼ 検索ツール

約 7,580 件 (0.39 秒)

## Softberry Home Page

[www.softberry.com/](http://www.softberry.com/) ▼ このページを訳す

Softberry Home Page. ... Annotation of Animal Genomes. Gene identification, Fgenesh gene finder and Fgenesh++ genome annotation pipeline, ~50 custom made gene-finders ...

### Fgenesh

FGENESH is the fastest (50-100 times faster than GenScan) and ...

### Annotation of Bacterial Geno...

Annotation of Bacterial Genomes and Community Sequences ...

### Softberry

Softberry developed genefinding parameters for 30 new ...

② Softberryをクリック

This is a partial list of Softberry products. All programs available ...

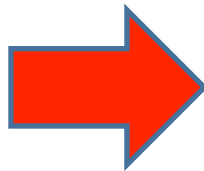
### FGENESH HELP

FGENESH - Program for predicting multiple genes in genomic DNA ...

### Annotation of Plant Genomes

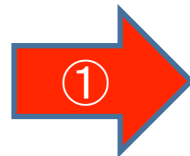
Annotation of Plant Genomes: Genes, Promoters, Functional ...

[softberry.com](http://softberry.com) からの検索結果 »



# 復習: 未知配列から、そこに含まれる遺伝子を予測する

① カーソルを"GENE FINDING in Eularyota"に合わせる



SoftBerry

TEST ON LINE

- GENE FINDING in Eukaryota
- GENE FINDING WITH SIMILARITY
- OPERON AND GENE FINDING IN BACTERIA
- GENE FINDING IN VIRUSES
- NEXT GENERATION
- ALIGNMENT /Sequences&genomes
- GenomeSequence EXPLORER/Infogene
- SEARCH FOR MOTIFS /promoters&functional
- PROTEIN LOCATION /patterns/Epitops
- RNA STRUCTURE COMPUTING
- PROTEIN STRUCTURE
- PROTEIN / DNA 3D-Visual Works
- SEQMAN
- MULTIPLE ALIGNMENTS
- ANALYSIS OF EXPRESSION DATA
- PLANT PROMOTERS DATABASE
- REPEATS /find&map repeats

FGENESH HMM based Gene structure prediction

FGENES Pattern based Human Gene structure

FGENES-M Multiple variants of Gene structure

FGENESH-M Prediction of multiple variants potential genes in genomic DNA

FGENESH\_GC (with possible donor GC) HMM based Human Gene structure prediction

BESTORF Finding coding fragment EST/mRNA

FEX Finding potential 5', internal and 3'-coding exons

SPL Search for potential splice sites

SPLM Search for non-standard splice sites using weight matrices

RNA SPL Search for exon-exon junction positions in cDNA

FSPLICE Find splice sites in genomic DNA

pipelines on line.

②

②一番上の"FGENESH"をクリック

- Annotation of Bacterial Genomes and Community Sequences: Genes, Operons, Promoters, Terminators, Protein Sub-Cellular Localization.
- Analyze RNASeq Next Generation Sequencing Data:
  - Accurate alignment of high-throughput RNA-seq data to a reference genome (ReadsMap);
  - De novo transcriptome reads assembly into RNA transcripts (TransSeq);
- Transomics - pipeline to map RNAseq data, assemble them into transcripts and quantify the abundance of these transcripts in particular datasets.
- Analyze Genomic Next Generation Sequencing Data:
  - De novo reconstruction (assembling) of genomic sequence,
  - Reconstruction of sequences using reference genome,
  - Mutation profiling and SNP discovery (OligoZip Assembler);
  - Functional analysis of SNP (SNP-effect);

<http://linux1.softberry.com/berry.phtml?topic=index&group=programs&subgroup=gfind>

# 復習: 未知配列から、そこに含まれる遺伝子を予測する

The screenshot shows the SoftBerry FGENESH web interface. The page has a navigation menu at the top with links like HOME, ALL SOFTWARE, PRODUCTS, etc. On the left, there is a 'TEST ON LINE' section with various analysis options. The main content area features the FGENESH logo and a reference citation: 'Reference: Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 2006,7, Suppl 1: P. 10.1-10.12.' Below this is a description of the HMM-based gene structure prediction tool. The interface includes a text input field for 'Paste nucleotide sequence here:' containing a DNA sequence, a 'Local file name:' field, and a dropdown menu for 'Select organism specific gene-finding parameters:'. The dropdown menu is open, showing a list of organisms under 'Monocot plants', with 'Monocot plants (generic, Corn, Rice, Wheat, Barley)' selected. A red arrow labeled '1' points to the text input field, and another red arrow labeled '2' points to the organism dropdown menu. A third red arrow labeled '3' points to the 'SEARCH' button at the bottom right. Three orange callout boxes provide instructions: '1 Genome seqの配列をコピー&ペースト', '2 Organism -> Monocot plants -> Monocot plants (generic, Corn, Rice, Wheat, Barley) を選択', and '3 SEARCHをクリック'.

① Genome seqの配列をコピー&ペースト

② Organism  
→ Monocot plants  
→ Monocot plants (generic, Corn, Rice, Wheat, Barley) を選択

③ SEARCHをクリック

# 復習: 未知配列から、そこに含まれる遺伝子を予測する

## 検索結果

```
Show picture of predicted genes in PDF file

FGENESH 2.6 Prediction of potential genes in Monocot genomic DNA
Time      :   Sat Nov 21 09:40:42 2015
Seq name:  test sequence
Length of sequence: 7861
Number of predicted genes 2: in +chain 1, in -chain 1.
Number of predicted exons 9: in +chain 2, in -chain 7.
Positions of predicted genes and exons: Variant 1 from 1, Score:156.457642
  G Str  Feature  Start      End      Score      ORF      Len
  1 +    TSS      2926             -5.48
  1 +    1 CDSf     3001 -          3154    19.41     3001 -      3153    153
  1 +    2 CDS1     3255 -          3886    72.94     3257 -      3886    630
  1 +    PolA      4490             -1.06

Predicted protein(s):
>FGENESH:[mRNA] 1 2 exon (s) 3001 - 3886 786 bp, chain +
ATGGCCGGCTCGTCAGCCGCAACGTCGTGTGCTCGGTTCCCTGGCATTGCTGGCGACATGC
CTCCTCTGGAACGAGGCCGCATCGTTCACGGCGTCCGGCTGGAACAAGGCGTTCCGCCACC
TTCTACGGCGGCGAGTGACGCTTCAGGAACGATGGGTGGGGCGTGTGGGTACGGGGACCTG
TACTCGACGGGGTACGGGACGAACACGGCGGCGCTGAGCACGGTGTGTTCAACGACGGG
GCGTCGTGCGGGCAGTGTACCCGGATCATGTGCGACTACCAAGCGGACAGGCGGTCTGTC
ATCTCCGGCACGTCGGTGACCATCACGGCGACCAACCTCTGCCCGCCGAACACTACGGCTC
CCCAACGACGCGCGGGTGGTGCACCCCGCGCGGCGAGCACTTCGACATGGCCGAACCG
GCGTGGCTCAAGATCGGCGTCTACGTGCGGGCATCGTGCCGGTGTATGATACCAGCGGGTG
CCGTGCGCCAAGCAGGGCGGGTGGAGTTCACCATCAACGGCAGGACTACTTCGAGCTG
GTGTCGTCTCCAACGTCGGCGGCGTCCGCTCCATCCAGTCCGGTGTGATCAAGGGGTCCG
AGGACCGGGTGGATGGCCATGTCCAGGAATTGGGGCGTCAACTGGCAGTCCCAACGCCTAC
CTCGACGGCCAGACCTGTGTTCAAGGTCACCGAGCAGCGCCAGACGCTCACCTTC
CTCGACGTCGCCCCGCGGGCTGGACGTTCCGGCCAGACCTTCTCGACCTCACAGCAGTTC
TCTTAA

>FGENESH: 1 2 exon (s) 3001 - 3886 261 aa, chain +
MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL
YSTGYGINTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPNYAL
PNDAGGWCNPPRQHFDMAPAWLKIGVYVGGIVPMYQRPVCAKQGGVRFITINGRDYFEL
VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGVNWQSNAYLDGQSLSFKVTSSDGQTLTF
LDVAPAGWTFGQTFSTSQQFS
```

遺伝子①

→1つの遺伝子の存在が予想された

アミノ酸配列をもとにBLAST検索を行い、  
どんな遺伝子なのかを調べる

①CDSの配列をコピー

# 復習: 未知配列から、そこに含まれる遺伝子を予測する

Genetyxでアミノ酸に変換する

The screenshot shows the GENETYX Ver.11 interface. A red arrow points to the 'New' button (labeled ①) in the top-left corner. The main window displays a DNA sequence with a 'Comment' field and a 'Map' button. The sequence is as follows:

```
LOCUS
ORIGIN
10      20      30      40      50      60
ATGGCCGGCT CGTCAGCCGC AACGTCGTGT GCTCGGTTCC TGGCATTGCT GCGGACATGC
70      80      90     100     110     120
CTCCTCTGGA ACGAGGCCGC ATCGTTACAG GCGTCCGGCT GGAACAAGGC GTTCGCCACC
130     140     150     160     170     180
TTCTACGGCG GCAGTGACGC TTCAGGAACG ATGGGTGGGG CGTGTGGGTA CGGGGACCTG
190     200     210     220     230     240
TACTCGACGG GGTACGGGAC GAACACGGGC GCGCTGAGCA CCGTGCTGTT CAACGACGGG
250     260     270     280     290     300
GCGTCGTGCG GGCAGTGCTA CCGGATCATG TCGCACTACC AGGCGGACAG GCGGTTCTGC
310     320     330     340     350     360
ATCTCCGGCA CGTCGGTGAC CATCACGGCG ACCAACCTCT GCCCGCCGAA CTACGCGCTC
370     380     390     400     410     420
CCCAACGACG CCGGCGGGTG GTGCAACCCG CCGCGGCAGC ACTTCGACAT GGCCGAACCG
430     440     450     460     470     480
GCGTGGCTCA AGATCGGCGT CTACGTCGGC GGCATCGTGC CCGTGATGTA CCAGCGGGTG
490     500     510     520     530     540
CCGTGCGCCA AGCAGGGCGG GGTGAGGTTT ACCATCAACG GCAGGGACTA CTTTCGAGCTG
550     560     570     580     590     600
GTGCTCGTCT CCAACGTCGG CGGCGTCGGC TCCATCCAGT CCGTGTCGAT CAAGGGGTCG
```

At the bottom of the window, the sequence composition is shown: 131 A; 252 C; 267 G; 136 T; 0 other. The GenBank logo is visible in the bottom right corner.

①新規作成

②配列をペースト

131 A; 252 C; 267 G; 136 T; 0 other

GenBank

L

# 復習: 未知配列から、そこに含まれる遺伝子を予測する

① Nucleotideをクリック

② Translate to AAをクリック

核酸からアミノ酸に

0	50	60
C	TGGCATTGCT	GGCGACATGC
0	110	120
T	GGAACAAGGC	GTTCGCCACC
0	170	180
G	CGTGTGGGTA	CGGGGACCTG
0	230	240
A	CGGTGCTGTT	CAACGACGGG
0	290	300
C	AGGCGGACAG	GCGGTTCTGC
0	350	360
T	GCCCCCGGAA	CTACGCGCTC
0	410	420
C	ACTTCGACAT	GGCCGAACCG
0	470	480
C	CGGTGATGTA	CCAGCGGGTG
0	530	540
G	GCAGGGACTA	CTTCGAGCTG
0	590	600
T	CGGTGTCGAT	CAAGGGGTCG



# 復習: 未知配列から、そこに含まれる遺伝子を予測する

Translate to AA for Thesis

Nuc. Seq. Position ( 1 - 786 ) : 1 - 786

1 - 786

Translation Position : 1 - 786

Add Remove

Genetic Code : The Standard Code

A.A. Format :  1-Left  1-Center  1-Right  3-Letter

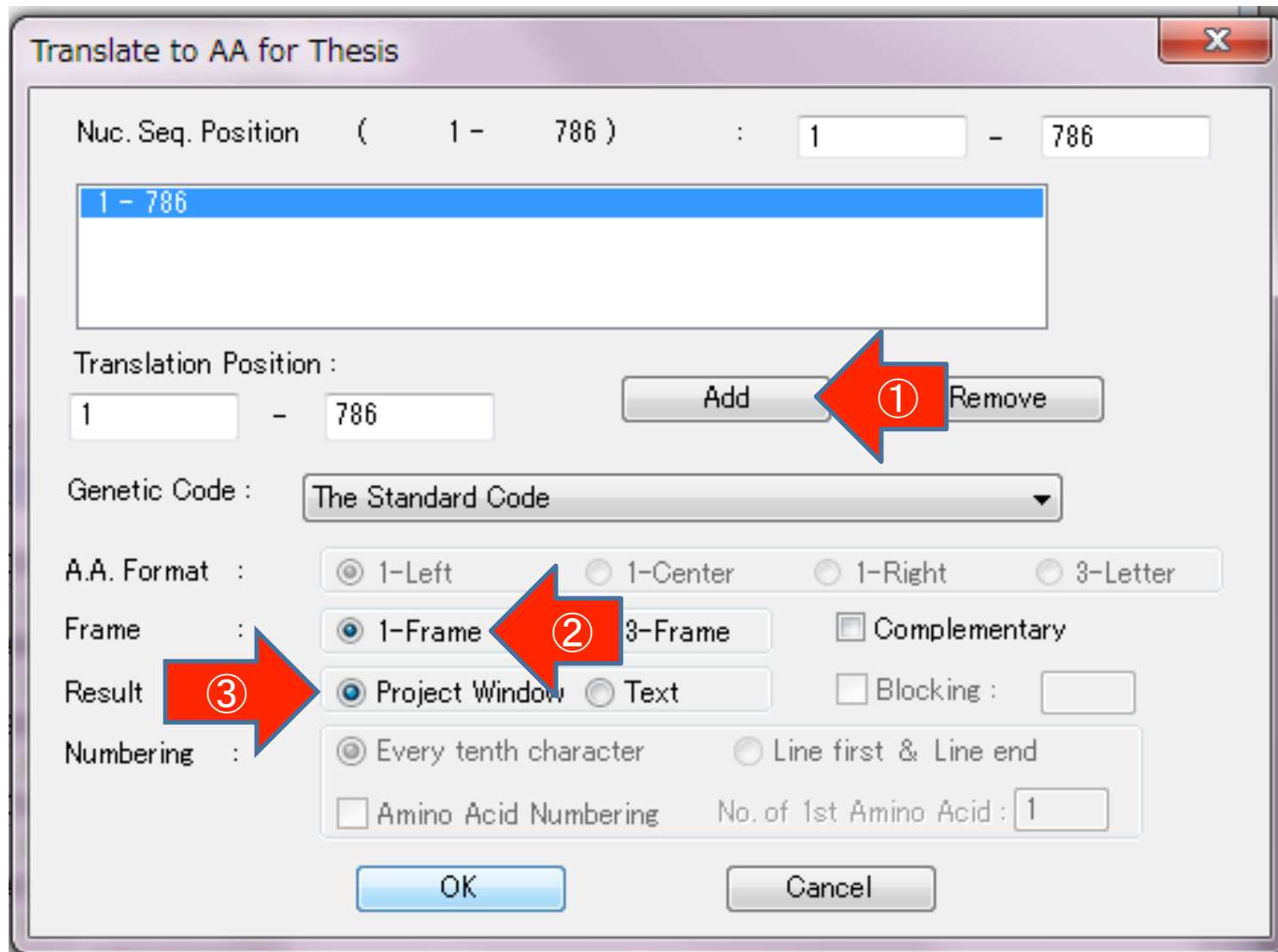
Frame :  1-Frame  3-Frame  Complementary

Result  Project Window  Text  Blocking :

Numbering :  Every tenth character  Line first & Line end

Amino Acid Numbering No. of 1st Amino Acid : 1

OK Cancel



The image shows a software dialog box titled "Translate to AA for Thesis". It contains several input fields and options for translating a nucleotide sequence. Three red arrows with circled numbers point to specific elements: arrow 1 points to the "Remove" button, arrow 2 points to the "1-Frame" radio button, and arrow 3 points to the "Project Window" radio button. The "Nuc. Seq. Position" is set to "1 - 786", and the "Translation Position" is also "1 - 786". The "Genetic Code" is "The Standard Code". The "A.A. Format" is "1-Left", "Frame" is "1-Frame", "Result" is "Project Window", and "Numbering" is "Every tenth character".

# 復習: 未知配列から、そこに含まれる遺伝子を予測する

GENETYX Ver.11 - Sequence2\_1.gpt

File Edit Find Nucleotide Peptide Homology Annotation Internet Search Application View Window Help

Genome seq.gnu

Sequence2

Sequence2\_1.gpt

Comment Map

>Sequence2\_1 Translated Sequence2 |1 frame|262 AA.

MAGSSAATSC ARFLALLATC LLWNEAASFT ASGWNKAFAT FYGGSDASGT MGGACGYGDL  
YSTGYGTINTA ALSTVLFNDG ASCGQCYRIM CDYQADRRFC ISGTSVITTA TNLCPPNYAL  
PNDAGGWCNP PRQHFDMAEP AWLKIGVYVG GIVPVMYQRV PCAKQGGVRF TINGRDYFEL  
VLVSNVGGVG SIQSVSIGKS RTGWMAMSRN WGVNWQSNAY LDGQSLSFKV TSSDGQILTF  
LDVAPAGWTF GQTFSTSQQF S\*

この配列を用いてBLAST検索

27748 MW FASTA



# これまでは。。。

Google

[ウェブ](#) [ニュース](#) [地図](#) [書籍](#) [動画](#) [もっと見る ▾](#) [検索ツール](#)

約 63,500,000 件 (0.55 秒)

**National Center for Biotechnology Information**  
[www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/) ▾ このページを訳す  
U.S. government-funded national resource for molecular biology information. Access to many public databases and other references, including the draft human genome.

**Blast**  **PubMed**  
Protein BLAST - Protein BLAST - ...  
Recent Results - Blastx - ...  
PubMed comprises more than 25 million citations for biomedical ...

**Nucleotide** **Protein**  
The Nucleotide database is a collection ... DNA & RNA ...  
Advanced search - Proteins - Batch Entrez - ...

**Gene** **NCBI**  
A portal to gene-specific content based on NCBI's RefSeq project ...  
Transfer NCBI data to your computer ... Use NCBI APIs and ...

[nih.gov からの検索結果 »](#)

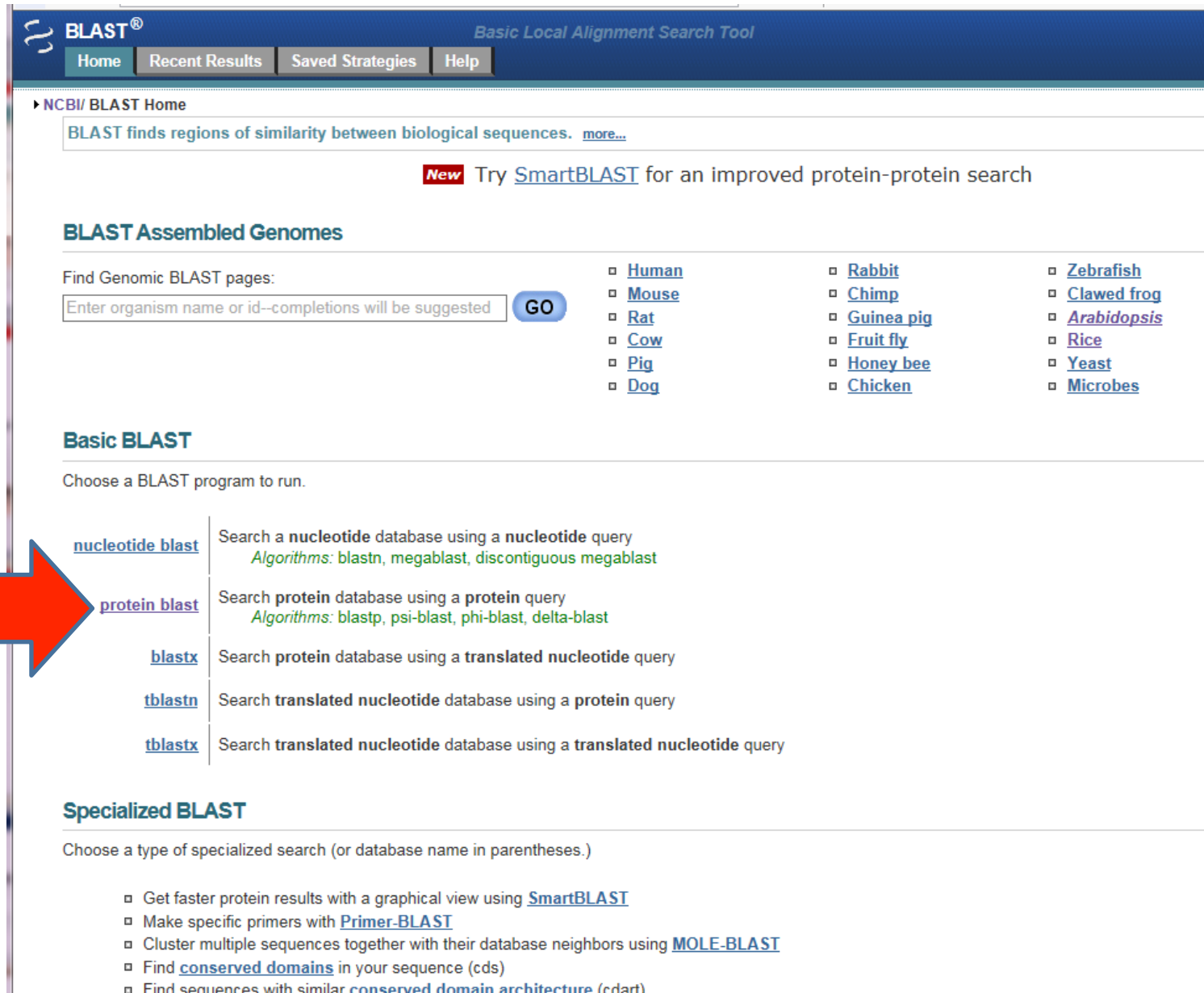
---

**国立生物工学情報センター - Wikipedia**  
<https://ja.wikipedia.org/wiki/国立生物工学情報センター> ▾  
国立生物工学情報センター(こくりつせいぶつこうがくじょうほうセンター、英: **National Center for Biotechnology Information**, **NCBI**)は、アメリカ合衆国の国立衛生研究所 (NIH) の下の国立医学図書館 (National Library of Medicine; **NLM**) の一部門として ...

**NCBIの意味 - 英和辞典 Weblio辞書**  
[eje.weblio.jp](http://eje.weblio.jp) > [英和辞典](#)・[和英辞典](#) > [ライフサイエンス辞書](#) ▾  
NCBIの意味や和訳。\*\*\* (アメリカ遺伝子情報研究機関)国立バイオテクノロジー情報センター関連語National Center for Biotechnology InformationNCBIの共起表現 - 約992万語ある英和辞典・和英辞典。発音・イディオムも分かる英語辞書。

**National Coalition Building Institute: NCBI**  
[ncbi.org/](http://www.ncbi.nlm.nih.gov/BLAST/) ▾ このページを訳す  
<http://www.ncbi.nlm.nih.gov/BLAST/> transforming communities campuses & organizations, the **National**

これまでは。。。



**BLAST®** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

► NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Try [SmartBLAST](#) for an improved protein-protein search

### BLAST Assembled Genomes

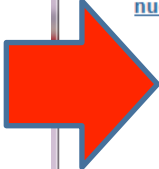
Find Genomic BLAST pages:

Enter organism name or id--completions will be suggested

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Cow](#)
- [Pig](#)
- [Dog](#)
- [Rabbit](#)
- [Chimp](#)
- [Guinea pig](#)
- [Fruit fly](#)
- [Honey bee](#)
- [Chicken](#)
- [Zebrafish](#)
- [Clawed frog](#)
- [Arabidopsis](#)
- [Rice](#)
- [Yeast](#)
- [Microbes](#)

### Basic BLAST

Choose a BLAST program to run.

<a href="#">nucleotide blast</a>	Search a <b>nucleotide</b> database using a <b>nucleotide</b> query <i>Algorithms: blastn, megablast, discontinuous megablast</i>
 <a href="#">protein blast</a>	Search <b>protein</b> database using a <b>protein</b> query <i>Algorithms: blastp, psi-blast, phi-blast, delta-blast</i>
<a href="#">blastx</a>	Search <b>protein</b> database using a <b>translated nucleotide</b> query
<a href="#">tblastn</a>	Search <b>translated nucleotide</b> database using a <b>protein</b> query
<a href="#">tblastx</a>	Search <b>translated nucleotide</b> database using a <b>translated nucleotide</b> query

### Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Get faster protein results with a graphical view using [SmartBLAST](#)
- Make specific primers with [Primer-BLAST](#)
- Cluster multiple sequences together with their database neighbors using [MOLE-BLAST](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)

これまでは。。

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blastp suite Standard Protein BL

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#)  [To](#)

```
MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYDLYSTGYGT
NTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPNYALPNDAGGWCNPPRQH
FDMAEPAWLKIGVYVGGIVPVMYQRVPCAKQGGVRFITINGRDYFELVLVSNVGGVGSIQSVSIKGR
TGWMAMSRNNGVNWQSNAYLDGQSLSFKVTSSDGGQLTFLDVAPAGWTFGQTFSTSQQFS*
```

Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Organism   Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query  [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search

Program Selection

Algorithm

blastp (protein-protein BLAST)

PSI-BLAST (Position-Specific Iterated BLAST)

PHI-BLAST (Pattern Hit Initiated BLAST)

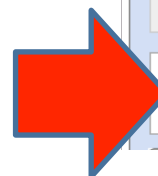
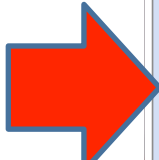
DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

Show results in a new window

配列をペースト



# これまでは。。

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastp suite/Formatting Results - 512V7MK3015

Edit and Resubmit Save Search Strategies Formatting options Download

YouTube How to read this page Blast report

### Protein Sequence (262 letters)

RID 512V7MK3015 (Expires on 11-22 22:19 pm)

Query ID Id|Query\_214894 Database Name O. sativa (japonica cultivar-group) RefSeq protein  
Description None Description Oryza sativa Japonica Group O. sativa (japonica cultivar-group) RefSeq protein  
Molecule type amino acid Program BLASTP 2.2.32+ Citation  
Query Length 262

Other reports: Search Summary Taxonomy reports Distance tree of results Multiple alignment

**New** Analyze your query with SmartBLAST

### Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. Specific hits DPBB\_1 Pollen\_allerg\_1  
Superfamilies DPBB\_1 superfamily Pollen\_allerg\_1 superfamily  
Multi-domains PLN00193

Distribution of 46 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Color key for alignment scores

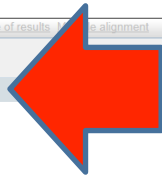
Query 1 50 100 150 200 250

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenPept Graphics Distance tree of results Alignment

	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Os04g0228400 [Oryza sativa Japonica Group]	534	534	99%	0.0	100%	NP_001052289
<input type="checkbox"/> Os06g0718100 [Oryza sativa Japonica Group]	365	365	93%	6e-127	70%	NP_001058599
<input type="checkbox"/> Os01g0823100 [Oryza sativa Japonica Group]	353	353	96%	1e-122	70%	NP_001044656
<input type="checkbox"/> Os03g0155300 [Oryza sativa Japonica Group]	347	347	84%	5e-120	72%	NP_001049005
<input type="checkbox"/> Os02g0267200 [Oryza sativa Japonica Group]	344	344	88%	9e-119	67%	NP_001046510
<input type="checkbox"/> Os01g0248900 [Oryza sativa Japonica Group]	337	337	88%	2e-116	68%	NP_001042597
<input type="checkbox"/> Os02g0268600 [Oryza sativa Japonica Group]	335	335	87%	5e-115	68%	NP_001046514
<input type="checkbox"/> Os05g0477800 [Oryza sativa Japonica Group]	333	333	92%	8e-115	67%	NP_001055842



# これまでは。。

OSU4g0228400 [Oryza sativa Japonica Group]

Sequence ID: [ref|NP\\_001](#) Length: 261

[▶ See 9 more title\(s\)](#)

See 9 more title(s)をクリック

Range 1: 1 to 261 [GenPept](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
534 bits(1376)	0.0	Compositional matrix adjust.	261/261(100%)	261/261(100%)	0/261(0%)
Query 1		MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL			60
Sbjct 1		MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL			60
Query 61		YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPPNYAL			120
Sbjct 61		YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPPNYAL			120
Query 121		PNDAGGWCNPPRQHFDMAEPAWLKIGVYVGGIVPVMYQRVPCAKQGGVRFINGRDYFEL			180
Sbjct 121		PNDAGGWCNPPRQHFDMAEPAWLKIGVYVGGIVPVMYQRVPCAKQGGVRFINGRDYFEL			180
Query 181		VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGNWQSNAYLDGQSLSFKVTSSDGQTLTF			240
Sbjct 181		VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGNWQSNAYLDGQSLSFKVTSSDGQTLTF			240
Query 241		LDVAPAGWTFGQTFSTSQQFS	261		
Sbjct 241		LDVAPAGWTFGQTFSTSQQFS	261		

[Download](#) [▼](#) [GenPept](#) [Graphics](#)

PREDICTED: expansin-A1-like [Oryza brachyantha]

Sequence ID: [ref|XP\\_006652124.1](#) Length: 262 Number of Matches: 1

Range 1: 2 to 262 [GenPept](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
497 bits(1279)	1e-175	Compositional matrix adjust.	244/261(93%)	252/261(96%)	2/261(0%)
Query 3		GSSAATSCAR--FLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL			60
Sbjct 2		GSSAATSCAR LALLATCLL++ A FTASGWN+AFATFYGGSDASGTMGGACGYGDL			61
Query 61		YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPPNYAL			120
Sbjct 62		YSTGYGTNTAALST LFNFGASCGQCYRI CDYQAD+RFCISGTSVTITATNLCPPNYAL			121
Query 121		PNDAGGWCNPPRQHFDMAEPAWLKIGVYVGGIVPVMYQRVPCAKQGGVRFINGRDYFEL			180
Sbjct 122		PNDAGGWCNPPRQHFDMAEPAWLKIGVYVGGIVPV+YQRVPCAKQGGVRF+NGRDYFEL			181
Query 181		VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGNWQSNAYLDGQSLSFKVTSSDGQTLTF			240
Sbjct 181		VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGNWQSNAYLDGQSLSF+VTSSDGQTLTF			240

# これまでは。。

Sequence ID: [ref|NP\\_001052289.1](#) Length: 261 Number of Matches: 1

▼ See 9 more title(s)

RecName: Full=Expansin-A1; AltName: Full=Alpha-expansin-1; AltName: Full=OsEXP1; AltName: Full=OsEXPA1; AltName: Full=

Sequence ID: [sp|Q7XWU8.2|EXPA1\\_ORYSJ](#)

alpha-expansin OsEXPA1 [Oryza sativa]

Sequence ID: [gb|AAL24479.1|AF394543\\_1](#)

expansin [Oryza sativa Japonica Group]

Sequence ID: [emb|CAA69105.1](#)

OSJNBa0065B15.2 [Oryza sativa Japonica Group]

Sequence ID: [emb|CAD39898.2](#)

Os04g0228400 [Oryza sativa Japonica Group]

Sequence ID: [dbj|BAF14203.1](#)

H0209A05.5 [Oryza sativa Indica Group]

Sequence ID: [emb|CAH66088.1](#)

hypothetical protein OsJ\_13482 [Oryza sativa Japonica Group]

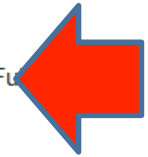
Sequence ID: [gb|EAZ29409.1](#)

unnamed protein product [Oryza sativa Japonica Group]

Sequence ID: [dbj|BAG91484.1](#)

Os04g0228400 [Oryza sativa Japonica Group]

Sequence ID: [dbj|BAS88220.1](#)



未知配列に含まれていた遺伝子は  
Expansinをコードしていることが予想された

Range 1: 1 to 261 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
534 bits(1376)	0.0	Compositional matrix adjust.	261/261(100%)	261/261(100%)	0/261(0%)
Query 1		MAGSSAATSCARFLALLLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL			60
Sbjct 1		MAGSSAATSCARFLALLLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL			60
Query 61		YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPPNYAL			120
Sbjct 61		YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVTITATNLCPPNYAL			120
Query 121		PNDAGGWCNPPRQHFDMAEPAWLKIGVYVGGIVPVMYQRVPCAQGGVRFINGRDYFEL			180
Sbjct 121		PNDAGGWCNPPRQHFDMAEPAWLKIGVYVGGIVPVMYQRVPCAQGGVRFINGRDYFEL			180
Query 181		VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGVNWQSNAYLDGQSLSFKVTSSDGQTLTF			240
Sbjct 181		VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGVNWQSNAYLDGQSLSFKVTSSDGQTLTF			240
Query 241		LDVAPAGWTFGQTFSTSQQFS	261		
Sbjct 241		LDVAPAGWTFGQTFSTSQQFS	261		

[Download](#) ▼ [GenPept](#) [Graphics](#)

# Genetyxを使うと。。。

The screenshot displays the GENETYX Ver.11 interface. The 'Internet Search' menu is open, showing options: Entrez - Nucleotide, Entrez - Protein, BLASTN, BLASTP, BLASTX, TBLASTN, TBLASTX, and BLAST Results. Two red arrows point to the menu and the 'BLASTP' option. The main window shows a protein sequence with a comment: '>Sequence2\_1 Translated Sequence2 |1 fra'. Below the sequence, a table of amino acid positions is visible.

10	20	30	40	50	60
MAGSSAATSC	ARFLALLATC	LLWNEAASFT	ASGWNKAFAT	FYGGSDASGT	MGGACGYGDL
70	80	90	100	110	120
YSTGYGNTA	ALSTVLFNDG	ASCGQCYRIM	CDYQADRRFC	ISGTSVTITA	TNLCPPNYAL
130	140	150	160	170	180
PNDAGGWCNP	PRQHFDMAEP	AWLKIGVYVG	GIVPVMYQRV	PCAKQGGVRF	TINGRDYFEL
190	200	210	220	230	240
VLVSNVGGVG	SIQSVSIKGS	RTGWMAMSRN	WGVNWQSNAY	LDGQSLSFKV	TSSDGQILTF
250	260	270	280	290	300
LDVAPAGWTF	GQTFSTSQQF	S			

# Genetyxを使うと。。。

The screenshot displays the GENETYX Ver.11 software interface. The main window is titled "BLAST Results (NCBI BLAST)" and contains a table with the following data:

Query Name	Program	Date	Status	Comment	RID
blastp_1_1_2015112...	blastp	2015/11/22 00:50:03	Finish		514RXUUN014

The interface also shows a menu bar with options like File, Edit, Homology, Internet Search, Application, View, Window, and Help. A toolbar with various icons is visible below the menu. On the left side, there are several panels, including one with a sequence logo (A, T, C, G) and another with a sequence alignment view.



# Genetyxを使うと。。。

Sequence	Expect	Score (Bits)	Definition	Identit
NP_001052289	0.0	534	Os04g0228400 [Oryza sativa Japonic...	
XP_006652124	1e-175	497	PREDICTED: expansin-A1-like [Oryza...	
XP_002446193	1e-164	469	hypothetical protein SORBIDRAFT_0...	
XP_004975119	3e-156	447	PREDICTED: expansin-A1 [Setaria it...	
AAT94291	2e-154	443	alpha-expansin EXPA1 [Triticum aest...	
NP_001105041	1e-153	441	LOC541905 precursor [Zea mays] >...	
ACG32530	1e-153	441	alpha-expansin 1 precursor [Zea ma...	
BAJ90878	8e-151	434	predicted protein [Hordeum vulgare ...	
BAJ95727	1e-150	434	predicted protein [Hordeum vulgare ...	
XP_003581525	4e-148	427	PREDICTED: expansin-A1-like [Brach...	
XP_010920093	8e-145	419	PREDICTED: expansin-A1-like [Elaeis...	
FMT11008	9e-144	416	hypothetical protein F75_25926 [Ae...	

Score = 534 bits (1376), Expect = 0.0  
Identities = 261/261 (100%), Positives = 261/261 (100%), Gaps = 0/261 (0%)

Query 1 MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL 60  
Sbjct 1 MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDL 60

Query 61 YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVITATNLCPFNAL 120  
Sbjct 61 YSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRFCISGTSVITATNLCPFNAL 120

Query 121 PNDAGGWCNPPRQHFDMAPAWLKIGVYVGGIVPVMYQRVPCAKQGGVRFINGRDYFEL 180  
Sbjct 121 PNDAGGWCNPPRQHFDMAPAWLKIGVYVGGIVPVMYQRVPCAKQGGVRFINGRDYFEL 180

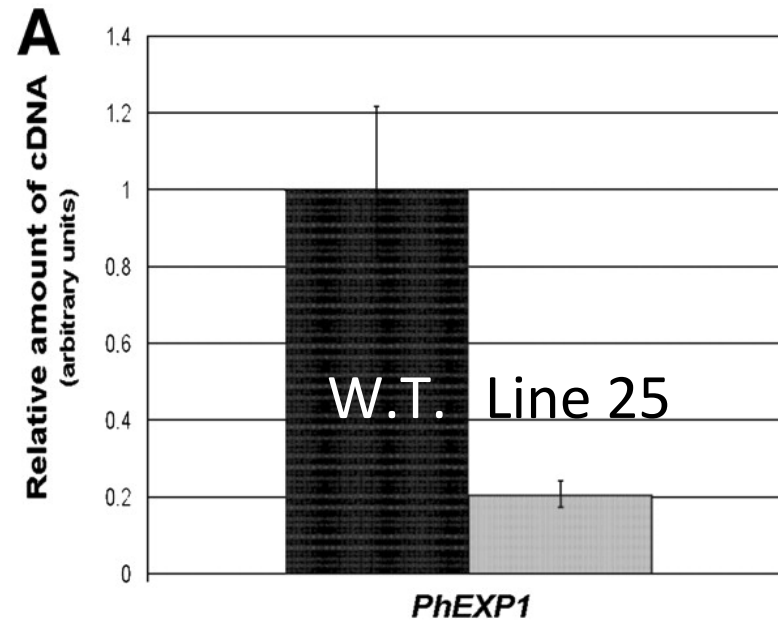
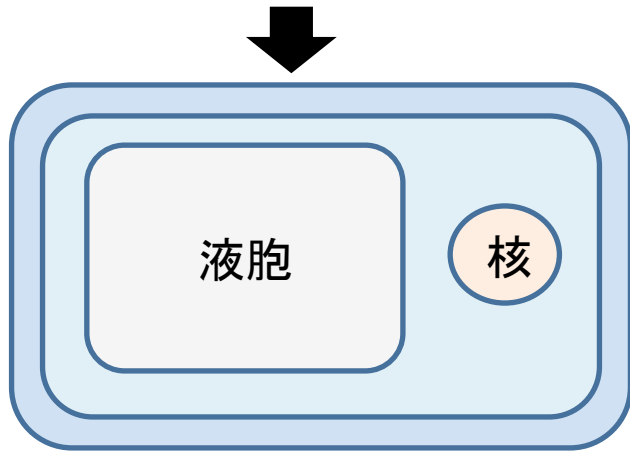
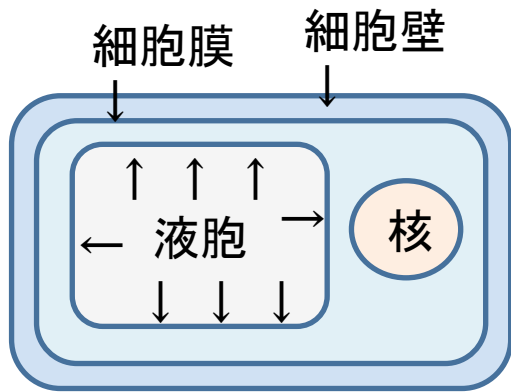
Query 181 VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGVNWQSNAYLDGQSLSFKVTSSDGQTLTF 240  
Sbjct 181 VLVSNVGGVGSIQSVSIKGSRTGWMAMSRNWGVNWQSNAYLDGQSLSFKVTSSDGQTLTF 240

Query 241 LDVAPAGWIFGQTFSTISQQFS 261  
Sbjct 241 LDVAPAGWIFGQTFSTISQQFS 261

Genetyx上で未知配列上の遺伝子がイネのEXPANSINであることが判明した

# EXPANSINの機能

細胞が大きくなるとき



# イネのゲノムに存在しているEXPANSINを系統解析する

配列を並べて(Alignment)して、配列を元に遺伝子群を分類する

FASTAフォーマットで読み込む

DNA の塩基配列やタンパク質のアミノ酸配列の配列の alignmentを行うための記述形式

(例)

>Name1

MADG....(配列).....

>Name2

MADA....(配列).....

>Name3

MADA....(配列).....

# Alignment

今回は授業用にBLAST検索の結果の中からイネのEXPANSINを集めて既にFASTAフォーマットにしてあるファイルを使用する

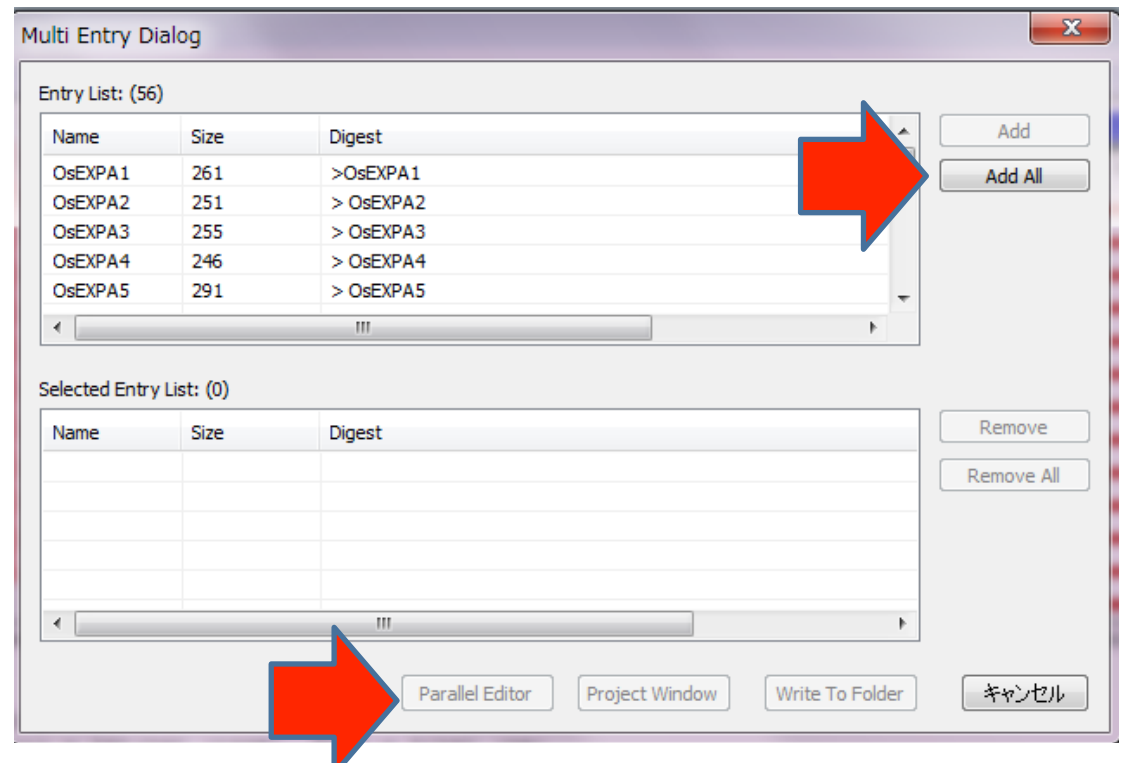
1. “OsEXPA”を開く
2. コピーしてwordに張り付ける
3. ファイル→保存 (Ctrl + S)
4. ファイルの種類”書式なし (\*.txt)”を選択
5. デスクトップに保存

```
>OsEXPA1↓
MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGT↓
MGGACGYGDLYSTGYGTNTAALSTVLFNDGASCGQCYRIMCDYQADRRC↓
ISGTSVITATNLCPNLYALPNDAGGWCNPPRQHFDMAEPAWLKIGVYVG↓
GIVPVMYQRPVCAKQGGVRFITINGRDYFELVLVSNVGGVGSIQSVIKGS↓
RTGWMAMSRNWGNVWQSNAYLDGQSLSFVKVTSDDGQTLTFLDVAAGWTF↓
GQTFSTSQQFS↓
> OsEXPA2↓
MASRSSALLLFSAFCLARRAAADYGSWQSAHATFYGGSDASGTMGGAC↓
GYGNLYSTGYGTNTAALSTVLFNDGAACGSCYELRCDNDGQWCLPGSVTV↓
TATNLCPNLYALPNDGWCNPPRPHFDMAEPAFLQIGVYRAGIVPVSYR↓
RVPCVKKGGIRFTINGHSYFNLVLTNVAGPGDVQSVIKGSSTGWQPM↓
RNWQGNWQSNYLDGQSLSFQVAVSDGRTVTNNVPAWQVQGFQTFEGGQ↓
F↓
> OsEXPA3↓
MLSGMEKQPAMLLVLTLCFAACKXVAQSAFATFYGGKDGSGTMGGAC↓
YGNLYNAGYGLYNAALSSALFNDGAMCGACYTITCDTSQTKWCKPGGNSI↓
TITATNLCPNLYALPNSGGWCNPPRQHFDMSQPAWENIAVYQAGIVPVN↓
YKRVPCQRSGGIRFAISGHDFELVTVTVNGGSGVVAQMSIKGSNTGWMA↓
MSRNWGANWQSNAYLAGQSLSFIVQLDDGRKVTAWNVAPSNWFPGATYST↓
SWVQF↓
> OsEXPA4↓
MAIAGVLFLLFLARQASAAGYGGWQSAHATFYGGSDASGTMGGACGYGNL↓
YSQGYGTNTAALSTALFNDGAACGSCYELRCDNAGSSCLPGSITVTATNF↓
CPPNYGLPSDDGWCNPPRPHFDMAEPAFLHIAQYRAGIVPVFRRVPCV↓
KKGVRFTVNGHSYFNLVLTNVAGAGDVRSVSIKGSRTGWQPMNRNWQ↓
NWQSNALDGGQSLSFQVTASDGRVTVTNNVAHPGWQFQGFEGGQF↓
> OsEXPA5↓
MSSRRDLAVLVLAALLPALSRLWLGHGLGHGHGRWRAPHVGGHGGQ↓
QGQQHAPLGGGGWSSAHATFYGGSDASGTMGGACGYGNLYSQGYGTNTA↓
ALSTALFNNGLSCGACFEVRCDAAGGGGSHSCLPGSVVVTATNFCPPNNAL↓
PSDDGWCNPPRAHFDMSQPVFQRIALFKAGIVPVSYRRVACQKKGIRF↓
TINGHSYFNLVLTNVGGAGDVHAAVKERSAAWQALSRNWQGNWQSA↓
LLDQALSFRTTGDGRSVSNNAVPRGWSFGQTFSGAQFN↓
```

# Alignment

今回は授業用にBLAST検索の結果の中からイネのEXPANSINを集めて既にFASTAフォーマットにしてあるファイルを使用する

1. “OsEXPA”を開く
2. コピーしてwordに張り付ける
3. ファイル→保存 (Ctrl + S)
4. ファイルの種類”書式なし (\*.txt)”を選択
5. デスクトップに保存
6. Genetyxにドラッグ & ドロップ
7. Add All
8. Parallel Editor



# Alignment

Multiple Alignmentをクリック

Conservation %

Alignment終了後、“Create Picture”をクリック

```
M A S S S L L L L A V L A L A A A A A A S A A A A A S A A A . G G A G G A G A G G G G G G G A . G G G G G G G G G G G G G G A G A G T . G A G A G G A  
1. OsEXPA1 M A G S S A A T S C A R F L A L L A T C L L W N E A A S F T A S G W N K A F A T F Y G G S D A S G T M G G A C G Y G D L Y S T G Y G T N T A A L S I  
2. OsEXPA2 M A S R S S A L L L L F S A F C F L A R R A A A D Y G S W Q S A H A T F Y G G G D A S G T M G G A C G Y G N L Y S T G Y G T N T A A L S T V L F N I  
3. OsEXPA3 M L S G M E K Q P A M L L V L V T L C A F A C K X S V A Q S A F A T F Y G G K D G S C T M G G A C G Y G N L Y N A G Y G L Y N A A L S S A L F N D C  
4. OsEXPA4 M A I A G V L F L L F L A R Q A S A A G Y G G W Q S A H A T F Y G G G D A S G T M G G A C G Y G N L Y S Q G Y G T N T A A L S T A L F N D G A A C C  
5. OsEXPA5 M S S R R D V L A V V L V A A L L P P A L S R G L W L G H H G L G H G H R W R A P H V G G H G Q G Q P Q Q H A P L G G G G W S S A H A T F Y G C  
6. OsEXPA6 M A P P L L L L L A S L L L V A A R R A L G L G L G Q W Q P G H A T F Y G G G D A S G T M G G A C G Y G N L Y S Q G Y G T S T A A L S T A L F N R C  
7. OsEXPA7 M S P A P R V L V L V V A T V V A L Q V S P A A G R I P G A Y G G G E W Q S A H A T F Y G G S D A S G T M G G A C G Y G N L Y S Q G Y G V N N A A I  
8. OsEXPA8 M A A A R M L V L L A S L C A L L L T A S A A K W T P A F A T F Y G G S D A S G T M G G A C G Y G D L Y G A G Y G T R T A A L S T A L F N G G A S C  
9. OsEXPA9 M E K K L L V V L F L S L C C A S R L R G E A A Q Q W T S A T A T F Y G G S D A S G T M G G S C G Y G N M Y S A G Y G T N T A L S S A L Y G D G F  
10. OsEXPA10 M A P C L L L V L F L L P A L A T G H Q H P S T L G S S A L S E W R S A K A S Y A A D P E D A I G G A C G F D L G K H G Y G M A T V G L S T A I  
11. OsEXPA11 M E L L R L L A V A A V A M A A E V A A G G D S G W S S G S A T F Y G G S D A S G T M G G A C G Y G N L Y S A G Y G T S T A A L S T A L F N N G C  
12. OsEXPA12 M A R S A F F F H C V A A V A A C I A A T A A L S G T A T F Y G G S D A S G T M G G A C G Y G N L Y S T G Y G T N T A A L S S A L F N D G A A C C  
13. OsEXPA13 M A G V A R M L A A V V C A I M P A A M A A G G V G A L E P S G W V R A H A T F Y G G A D A S G T M G G A C G Y G N L Y A Q G Y G T R T A A L S T  
14. OsEXPA14 M A S S P R A F A L V F F A I A A V G C T Q L T T A D D A A P P V W Q K A H A T F Y G G A D A S G T M G G G C G Y G D L Y S Q G Y G T R N A A L S T  
15. OsEXPA15 M A M W K K K K T P S I L P L V V V I A A A S L I A P T T A G W S S G T A T F Y G G S D A S G T M G G A C G Y G N L Y W S G Y G T N T A A L S S A I  
16. OsEXPA16 M S S V L L F L L L L L S G V S L S G C I R L G S G G Y E E W R M G S A T Y I K E S L G H P L N D G G G A C G Y G D L D I F R Y G R Y T A G V S C  
17. OsEXPA17 M A S S W N N P A I F L A A A L A V A T A A Q V V T A G F T T D L Y W Q Q Q P A P G A V T P Y K T S D W H D G S A T F Y G D P S G M G D D F G G A C  
18. OsEXPA18 M G N I V L Q L L A I L A L C I A P A R S G W L Q G T A T F Y G G A D G S G T M G G A C G Y G N L Y D Q G Y G I N N A A L S T P L F N N G A S C G C  
19. OsEXPA19 M G N I F L Q L L A V V A L C I A P A R S D W L P G T A T F Y G G A D G S G T M G G A C G Y G N L Y D Q G Y G I N N A A L S T P L F N D G A S C G C  
20. OsEXPA20 M G N I L L Q L L A V V A L C I A P A R S D W L P G T A T F Y G G A D G S G T M G G A C G Y G N L Y D Q R Y G I N N A A L S T P L F N D G A S C G C  
21. OsEXPA21 M A P P S L P I L L V L L S L S S S L S S S S A A A A G R W T D A H A T F Y G G A D A S G T M G G A C G Y G N T Y G Q G Y G T D T A A L S A V M F C  
22. OsEXPA22 M A P A R P F A L L F L A V T V G F V L L T A A D D S A N A T A T T T T A M A P S S T D D A A P P V W L K A H A T F Y G G A D A S G T M G G A C C  
23. OsEXPA23 M A P A R A F V L V L L A V A S A S T A A A N T A T T T P T N P V A A P T Q W Q K A H A T F Y G G A D A S G T M G G A C G Y G N L Y S Q G Y G T R  
24. OsEXPA24 M A D M A P A R A L A L V L L A V A V G S A L M A A A Q D A P S P P T P M A P S P S T D E T P P V W L K A H A T F Y G G A D A S G T M G G A C G Y V  
25. OsEXPA25 M E Y A I L F A T S L V I T V L A A S G F A P A H G W N K G T A T F Y G G A D A S G T M G G A C G Y G N L Y T A G Y G T N T A A L S S V L F N D G V  
26. OsEXPA26 M A P L P L L T T T S L L L F F F L A S S F A A D V V V A G G G G G G G G Y D G G G D G E G G G G G D G E G G G G G G A K M P H V N H G R Y K C C  
27. OsEXPA27 M G A M A E N L L V L C T I L A A R M A L A A A D D W I P A T A T F Y G G N D G S G T M G G A C G Y G N L Y D Q G Y G L E N A A L S T A L F N D G A  
28. OsEXPA28 M H I R F E V L A A L C I T S A A A A A G C H V S G T A T F Y G G G D A S G T M G G A C G Y G N L Y T Q G Y G V N A A L S T A L F N G C C  
29. OsEXPA29 M A P P L L L L L A S L L L V A A R R A L G L G L G Q W Q P G H A T F Y G G G D A S G T M G G A C G Y G N L Y S Q G Y G T S T A A L S T A L F N R C
```



# Alignment

Grayにすると論文のような感じになる

配列の幅

The screenshot shows a software interface for sequence alignment. At the top, there are several control panels:

- Color/Style:** Includes options for  None,  Mismatch,  Fill,  Idnt & Sim,  Gray,  Invert, and  Frame.
- Frame Settings:** Includes a dropdown for '線の幅' (line width) set to 1 and a field for '線の色' (line color).
- Match %:** A dropdown set to 50, with the label '最低一致%' (minimum identity %).
- Options:** Includes a dropdown for '行幅' (line width) set to 100, and checkboxes for  Color and  Difference, along with a 'Font...' button.

The main area displays a list of sequences, each with a label (e.g., OsEXPA1, OsEXPA2, etc.), a column number, and a sequence of characters. A vertical line is drawn through the sequences, indicating an alignment point. The sequences are: OsEXPA1 43 --GGSDASGTMGGACGY-GDLYSTGYGTNTAALS... OsEXPA2 38 --GGGDASGTMGGACGY-GNLYSTGYGTNTAALS... OsEXPA3 37 --GGKDGSTMGACGY-GNLYNAGYGLYNAALS... OsEXPA4 33 --GGGDASGTMGGACGY-GNLYSQGYGTNTAALS... OsEXPA5 73 --GGGDASGTMGGACGY-GNLYSQGYGTNTAALS... OsEXPA6 37 --GGGDASGTMGGACGY-GNLYSQGYGTNTAALS... OsEXPA7 45 --GGSDASGTMGGACGY-GNLYSQGYVNNAAAL... OsEXPA8 34 --GGSDASGTMGGACGY-GDLYGAGYGTNTAALS... OsEXPA9 36 --GGSDASGTMGGACGY-GMYSAGYGTNTAALS... OsEXPA10 42 --AADPEDAIGGACGF-GDLGKHGYGMATVGLS... OsEXPA11 36 --GGSDASGTMGGACGY-GNLYSAGYGTNTAALS... OsEXPA12 33 --GGSDASGTMGGACGY-GNLYSTGYGTNTAALS... OsEXPA13 43 --GGADASGTMGGACGY-GNLYAQGYGTNTAALS... OsEXPA14 43 --GGADASGTMGGACGY-GDLYSQGYGTNTAALS... OsEXPA15 41 --GGSDASGTMGGACGY-GNLYWSGYGTNTAALS... OsEXPA16 41 KESLGHPLNDGGGACGY-GDLDFRFGRYTAGV... OsEXPA17 61 --GDPSGMDDFGGACGYVNDIVSLYSTKTAAL... OsEXPA18 32 --GGADGSGTMGGACGY-GNLYDQGYGINNAAL... OsEXPA19 32 --GGADGSGTMGGACGY-GNLYDQGYGINNAAL... OsEXPA20 32 --GGADGSGTMGGACGY-GNLYDQRYGINNAAL... OsEXPA21 39 --GGADASGTMGGACGY-GNTYQGYGTNTAALS... OsEXPA22 61 --GGADASGTMGGACGY-GDLYSQGYGTNTAALS... OsEXPA23 48 --GGADASGTMGGACGY-GNLYSQGYGTNTAALS... OsEXPA24 59 --GGADASGTMGGACGY-VDLYSQGYGTNTAALS... OsEXPA25 36 --GGADASGTMGGACGY-GNLYTAGYGTNTAALS... OsEXPA26 56 GGGGGAKMPHVNHGKCGFPWVDGHATFYGGRD... OsEXPA27 36 --GGNDGSGTMGGACGY-GNLYDQGYGLNAAAL... OsEXPA28 35 --GGKDGSTMGACGY-GNLYTQGYGVYNAALS... OsEXPA29 43 --GGSDASGTMGGACGY-GNLYTQGYGTNTAALS... OsEXPA30 46 -GDETA-SETMGGACGY-GNLYASGYGTNTAALS... OsEXPA31 35 --GGSDGSGTMGGACGY-GNLYDQGYVDNAAALS... OsEXPA32 51 --GGRDGSGLTDGACGY-KDTSKEGYGVQTVAV... OsEXPA33 32 --GGSDGAGTMGGACGY-GNLYNAGYGLNAAALS... OsEXPB1 50 GAPKGAGPKDNGGACGFKD-VDKAPFLGMNS... OsEXPB2 40 QPNGAGPDDNGGACGFKN-TNQYPFMSMTSC... OsEXPB3 51 GAPTGAGPDDNGGACGFKN-TNQYPFMSMTSC... OsEXPB4 64 QPNGAGPDDNGGACGFKH-TNQYPFMSMTSC... OsEXPB5 54 GGPQGDGS--EGGACGYQSAVGRPFSSMTAA... OsEXPB6 53 GAPTGAGPDDNGGACGFKN-VNQYPFMSMTSC... OsEXPB7 101 GAPNGAGPDDNGGACGFKN-VNLPFFSAMTSC... OsEXPB8 57 QPNGAGAADNGGACGFKK-VNQYPFMGMTSC... OsEXPB9 52 GQPYGSGSTDNGGACGKKN-VNLPFFYNGMIS...

# Alignment

系統樹も作成できる

The screenshot shows the ParallelEditor4 interface. At the top, a conservation bar is visible with a scale from 0 to 75. Two red arrows point to the 'UPGM法' and 'neighbor-joining法' labels. Below the bar, a sequence alignment is displayed for 28 OsEXPA sequences. The sequences are color-coded by amino acid type. The first few sequences are: 1. OsEXPA1: MAGSSAATSCARFLALLATCLLWNEAASFTASGWNKAFATFYGGSDASGTMGGACGYGDLYSTGYGNTAALS...; 2. OsEXPA2: MASRSSALLLFLSAFCFLARRAAADYGSWQSAHATFYGGGSDASGTMGGACGYGNLYSTGYGNTAALSTVLFNI...; 3. OsEXPA3: MLSGMEKQPAMLLVLTLCFAFCKXSVAQSAFATFYGGKDGSC TMGGACGYGNLYNAGYGLYNAALSSALFND...; 4. OsEXPA4: MAIAGVLFLLFLARQASAAGYGGWQSAHATFYGGGSDASGTMGGACGYGNLYSQGYGNTAALSTALFNDGAAC...; 5. OsEXPA5: MSSRRDVLAVVLAALLPALSRLWLGHGHLGHGRWRAPHVGGHGQGGPQQHAPLGGGGWSSAHATFYGC...; 6. OsEXPA6: MAPPLLLLLASLLVAARRALGLGLGQWQPHATFYGGGSDASGTMGGACGYGNLYSQGYGTSTAALSTALFNRC...; 7. OsEXPA7: MSPAPRVLVVLVVALQVSPAAGRIPGAYGGGEWQSAHATFYGGSDASGTMGGACGYGNLYSQGYGVNNAAI...; 8. OsEXPA8: MAAARMLVLLASLCAALLTASAAKWTPAFATFYGGSDASGTMGGACGYGDLYGAGYGTRTAALSTALFNGGAS...; 9. OsEXPA9: MEKLLLVFLSLCCASRLRGEAAQQWTSATATFYGGSDASGTMGGSCGYGNMYSAGYGTNTTALSSALYGDG...; 10. OsEXPA10: MAPCLLLVLFLLPALATGHQHPSTLGSALSEWRSAKASYAADPEDAIGGACGFDLKGHYGMATVGLSTAI...; 11. OsEXPA11: MELLRLLAVAAVAAMAAEVAAGGDSGWSSGSATFYGGSDASGTMGGACGYGNLYSAGYGTSTAALSTALFNNG...; 12. OsEXPA12: MARSAFFHCVAAVAACIAATAAALSGTATFYGGSDASGTMGGACGYGNLYSTGYGNTAALSSALFNDGAAC...; 13. OsEXPA13: MAGVARMLAADVCAIMPAAAMAAGVGALPESGWVRAHATFYGGADASGTMGGACGYGNLYAQGYGTRTAALST...; 14. OsEXPA14: MASSPRAFALVFFAIAAVGCTQLTTADDAAPPVWQKAHATFYGGADASGTMGGGCGYGDLYSQGYGTRNAALST...; 15. OsEXPA15: MAMWKKKTPSILPLVVVIAAASLIAPT TAGWSSGTATFYGGSDASGTMGGACGYGNLYWSGYGNTAALSSAI...; 16. OsEXPA16: MSSVLLFLLLLLLSGVSLSGCIRLGSGGYEWRMG SATYIKESLGHPLNDGGGACGYGDLDIFRYGRYTAGVSC...; 17. OsEXPA17: MASSWNNPAIFLAAALAVATAAQVVTAGFTTDLYWQQQPAPGAVTPYKTS DWHG SATFYGDP SGMGDDFFGGA...; 18. OsEXPA18: MGNIVLQLLAILALCIAPARSGWLQGTATFYGGADGSGTMGGACGYGNLYDQGYGINNAALSTPLFNNGASC...; 19. OsEXPA19: MGNIFLQLLAVVALCIAPARSDWLPGTATFYGGADGSGTMGGACGYGNLYDQGYGINNAALSTPLFNDGASC...; 20. OsEXPA20: MGNILLQLLAVVALCIAPARSDWLPGTATFYGGADGSGTMGGACGYGNLYDQRYGINNAALSTPLFNDGASC...; 21. OsEXPA21: MAPPSLPILLVLLSLSSSLSSSSAAAAGRWD AHATFYGGADASGTMGGACGYGNTYGGQGYGTDTAALSAVMF...; 22. OsEXPA22: MAPARPFALLFLAVTVGFVLLTAADDSANATATTTTAMAPSSSTDDAAPPVWLKAHATFYGGADASGTMGGAC...; 23. OsEXPA23: MAPARAFVLVLLAVASASTAAANTATTTPTNPVAAPTQWQKAHATFYGGADASGTMGGACGYGNLYSQGYGTR...; 24. OsEXPA24: MADMAPARALALVLLAVAVGSALMAAAQDAPSPPTPMAPSPSTDETPPVWLKAHATFYGGADASGTMGGACGY...; 25. OsEXPA25: MEYAILFATSLVITVLAASGFAPA HGWNKGTATFYGGADASGTMGGACGYGNLYTAGYGNTAALSSVLFNDGV...; 26. OsEXPA26: MAPLPLLTTTSLLLFFFLASSFAADVVAAGGGGGGGYDGGGDGEGGGGGGDEGGGGGGGAKMPHVNHGRYKCC...; 27. OsEXPA27: MGAMAENLLVLCITILAAARMALAAADDWI PATATFYGGNDGSGTMGGACGYGNLYDQGYGLENAALSTALFNDG...; 28. OsEXPA28: ...



# Alignment

イネ以外の他の種と比較してみる

1. 授業用HPから"EXPA"を開く
2. コピーしてwordに張り付ける
3. ファイル→保存 (Ctrl + S)
4. ファイルの種類"書式なし (\*.txt)"を選択
5. デスクトップに保存
6. Genetyxにドラッグ & ドロップ
7. Add All
8. Parallel Editor
9. Multiple Alignment
10. 系統樹の作成

## タンパク質配列の情報を見る

The screenshot displays the GENETYX Ver.11 interface. The 'Peptide' menu is open, and the 'Seq Info...' option is highlighted. Two red arrows point to the menu and the selected option. The background shows a protein sequence in FASTA format.

FASTA sequence:

```
>Sequence2_1 T
MAGSSAATSC AF
YSTGYGTNTA AL
PNDAGGWCNP PR
VLVSNVGGVG SIQSVSIKGS RTGWMAMSRN WGVNWQSNAY LDGQSLSEFKV TSSDGQILTF
LDVAPAGWTF GQTFSTISQQF S*
```

Untitled1.txt

```
[GENETYX : Amino Acid Composition]
Date           : 2015.11.22
Filename       : Sequence2_1.gpt
Sequence Size  : 262
Sequence Position: 1 - 262
```

```
hydrophobic: 138( 52.67%)
neutral      : 82( 31.30%)
hydrophilic: 41( 15.65%)
other        : 1(  0.38%)
```

[ hydrophobic residues ]

```
Gly(G) 32( 12.21%) Ala(A) 27( 10.31%) Val(V) 16(  6.11%) Leu(L) 17(  6.49%)
Ile(I)  8(  3.05%) Met(M)  7(  2.67%) Phe(F) 14(  5.34%) Trp(W)  8(  3.05%)
Pro(P)  9(  3.44%)
```

[ neutral residues ]

```
Ser(S) 26(  9.92%) Thr(T) 21(  8.02%) Asn(N) 13(  4.96%) Gln(Q) 12(  4.58%)
Cys(C) 10(  3.82%)
```

[ hydrophilic residues ]

```
Asp(D) 11(  4.20%) Glu(E)  3(  1.15%) Lys(K)  5(  1.91%) His(H)  1(  0.38%)
Arg(R) 10(  3.82%) Tyr(Y) 11(  4.20%)
```

[ other residues ]

```
Asx(B)  0(  0.00%) Glx(Z)  0(  0.00%) Xaa(X)  0(  0.00%) ???(?)  1(  0.38%)
```

```
Average Molecular Weight = 27747.75
Monoisotopic Molecular Weight = 27730.0945
```

分子量約27.7kDであると予想

# 等電点の検出

The screenshot shows the GENETYX Ver.11.0 interface. The 'Peptide' menu is open, and the 'Isoelectric Point...' option is highlighted. Two red arrows point to the menu and the selected option. The background shows a protein sequence window with a multi-column alignment of amino acid sequences.

10	20	30	40	50	60
MAGSSAATSC AR				GT MGGACGYGDL	
70	80	90	100	110	120
YSTGYGTNIA AL				TA TNLCPNYAL	
130	140	150	160	170	180
PNDAGGWCNP PR				RF TINGRDYFEL	
190	200	210	220	230	240
VLVSNVGGVG SIQSVSIKGS	RTGWMAMSRN	WGVNWQSNAY	LDGQSLSFKV	TSSDGQILTF	
250	260	270	280	290	300
LDVAPAGWIF	GQTFSTISQQF	S			

# 制限酵素サイトの検出

The screenshot shows the GENEDOC software interface. The 'File' menu is open, and the 'Legacy Restriction Enzyme' option is selected. A sub-menu is displayed with 'Select Restriction Enzyme...' and 'Disp Restriction Site...' options. The main window displays a DNA sequence with a table of restriction sites.

Position	Sequence
0	50 60
C	TGGCATTGCT GCGACATGC
0	110 120
T	GGAACAAGGC GTTCGCCACC
0	170 180
G	CGTGTGGGTA CGGGGACCTG
0	230 240
A	CGGTGCTGTT CAACGACGGG
0	290 300
C	AGGCGGACAG GCGGTTCTGC
0	350 360
T	GCCCGCCGAA CTACGGGCTC
0	410 420
C	ACTTCGACAT GGCCGAACCG
0	470 480
C	CGGTGATGTA CCAGCGGGTG
0	530 540
G	GCAGGGACTA CTTCGAGCTG
0	590 600
T	CGGTGTCGAT CAAGGGGTCG

# 特許出願時のフォーム

The image shows the GENET Ver.11 software interface. The 'Nucleotide' menu is open, and the 'PTForm...' option is selected. A red arrow points to the menu, and another red arrow points to the 'PTForm...' option. The 'PTForm Dialog' is open, showing a list of positions (1-786) and a 'Translation Position' field with '1' and '786' entered. A red arrow points to the 'Add' button. The 'Genetic Code' is set to 'The Standard Code', and the 'Stop Codon' is set to 'Space'. A red arrow points to the 'OK' button.

GENET Ver.11 - 遺伝子1 CDS.gnu

File Edit Nucleotide Peptide Homology Annotation Internet Search Application View Window Help

Seq Info...  
Translate to AA...  
PTForm...  
GC Contents...  
di-Nucleotide...  
Direct Repeat...  
Inverted Repeat...  
Inverted & Complementary...  
Analyze Restriction Enzyme...  
Legacy Restriction Enzyme  
Recombinant DNA Fragment...  
Silent Mutation Restriction Site  
Search for ORF...  
Frame Analysis...  
Promoter...  
Create Probe Record...  
Search PCR Primer...  
Primer3 Search...  
Search tRNA  
Hairpin...  
Palindrome...  
Search Stacking Region...  
RNA 2nd. Structure Pred. ...  
Modify Free Energy...  
Search siRNA target...  
Coding Region  
CpG Island...  
Splice Site...  
Polymerase II Promoter...  
PolyA Signal...

PTForm Dialog

1 - 786

Translation Position :  
1 - 786 Add Remove

Genetic Code : The Standard Code

No. of 1st Amino Acid : 1  Upper Case

Stop Codon :  Space  Symbol Stop

OK Cancel