アラインメントと系統樹作成

生物情報工学II (担当　芦苅）遺伝子の配列解析

第６回目-1（11/1）遺伝子の予測と機能解析

-2（11/1）演習

第７回目-1（11/8）DNA配列から遺伝子を探す

-2（11/8）DNA、アミノ酸のアラインメント、系統樹解析

-3（11/8）演習

第８回目-1（11/15）Genetyxを用いた解析

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

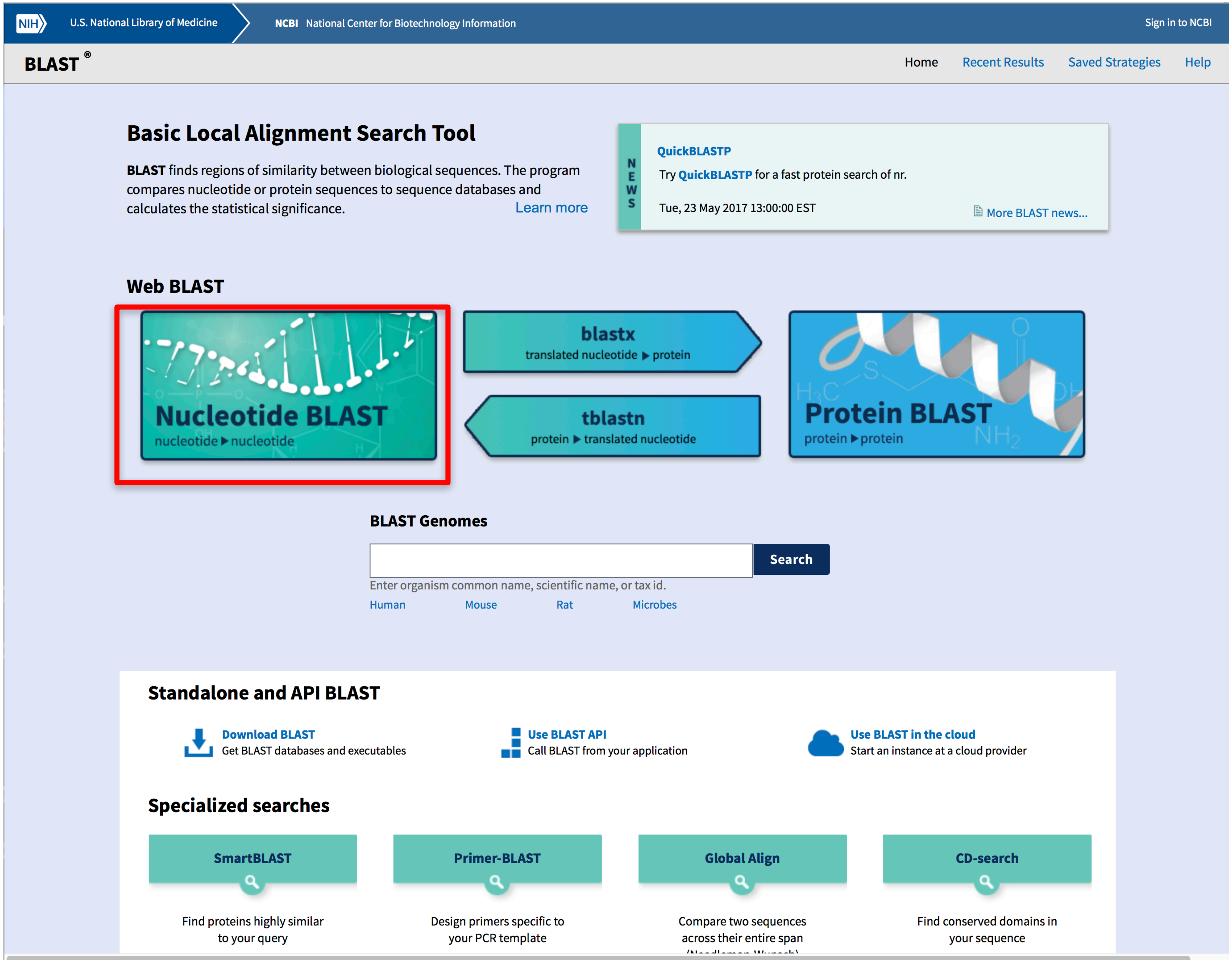
1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

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

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測



U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign in to NCBI

BLAST ®

Home Recent Results Saved Strategies Help

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

NEWS

QuickBLASTP

Try **QuickBLASTP** for a fast protein search of nr.

Tue, 23 May 2017 13:00:00 EST

More BLAST news...

Web BLAST

BLAST Genomes

☐ Search

Enter organism common name, scientific name, or tax id.

Human Mouse Rat Microbes

Standalone and API BLAST

Download BLAST
Get BLAST databases and executables

Use BLAST API
Call BLAST from your application

Use BLAST in the cloud
Start an instance at a cloud provider

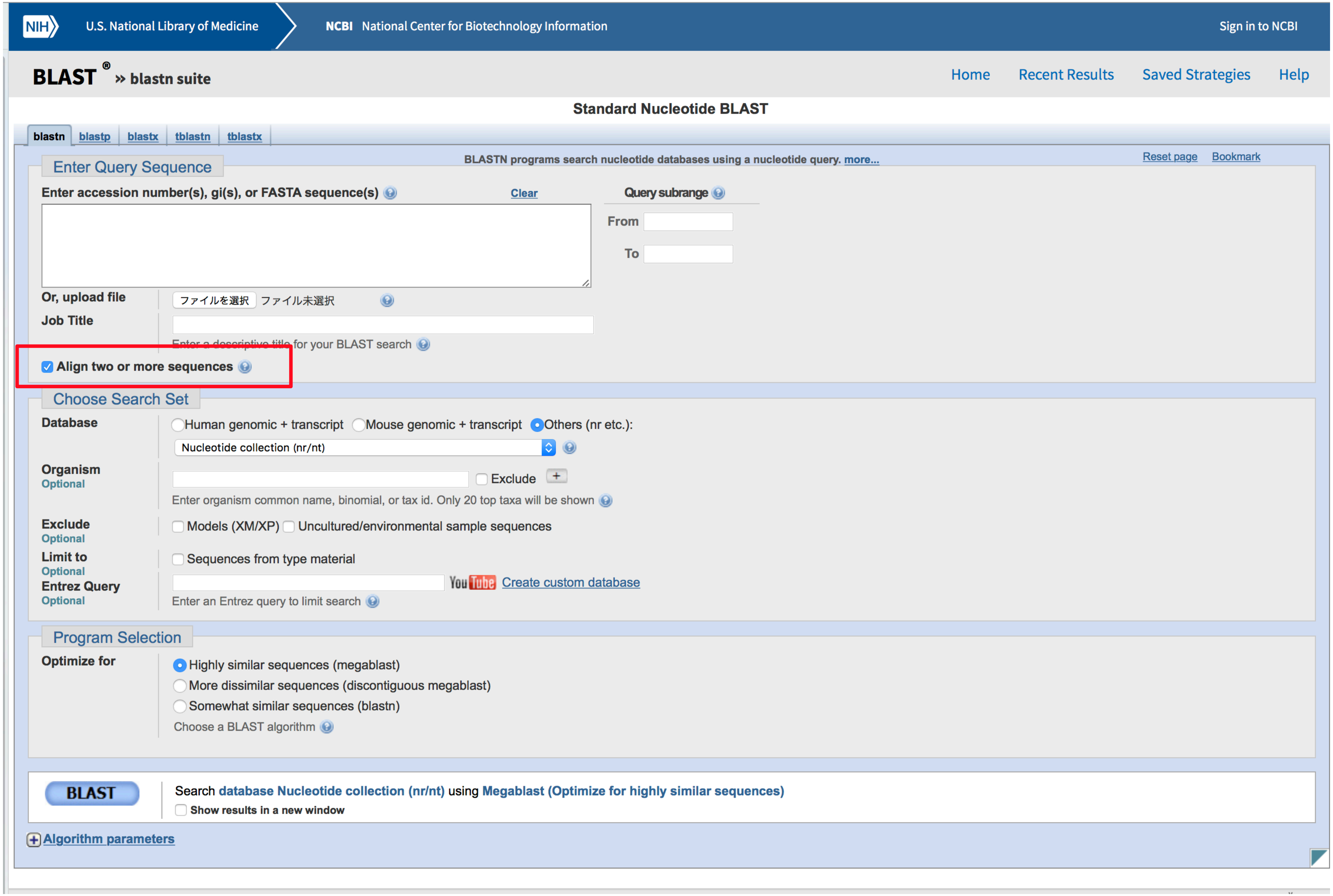
Specialized searches

SmartBLAST
Find proteins highly similar to your query

Primer-BLAST
Design primers specific to your PCR template

Global Align
Compare two sequences across their entire span

CD-search
Find conserved domains in your sequence



NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® » blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more... Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Query subrange

From

To

Or, upload file ファイルを選択 ファイル未選択

Job Title

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Choose Search Set

Database ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (nr etc.):

Nucleotide collection (nr/nt)

Organism Optional

☐ Exclude +

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

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

Limit to Optional ☐ Sequences from type material

Entrez Query Optional YouTube Create custom database

Enter an Entrez query to limit search

Program Selection

Optimize for

● Highly similar sequences (megablast)

○ More dissimilar sequences (discontiguous megablast)

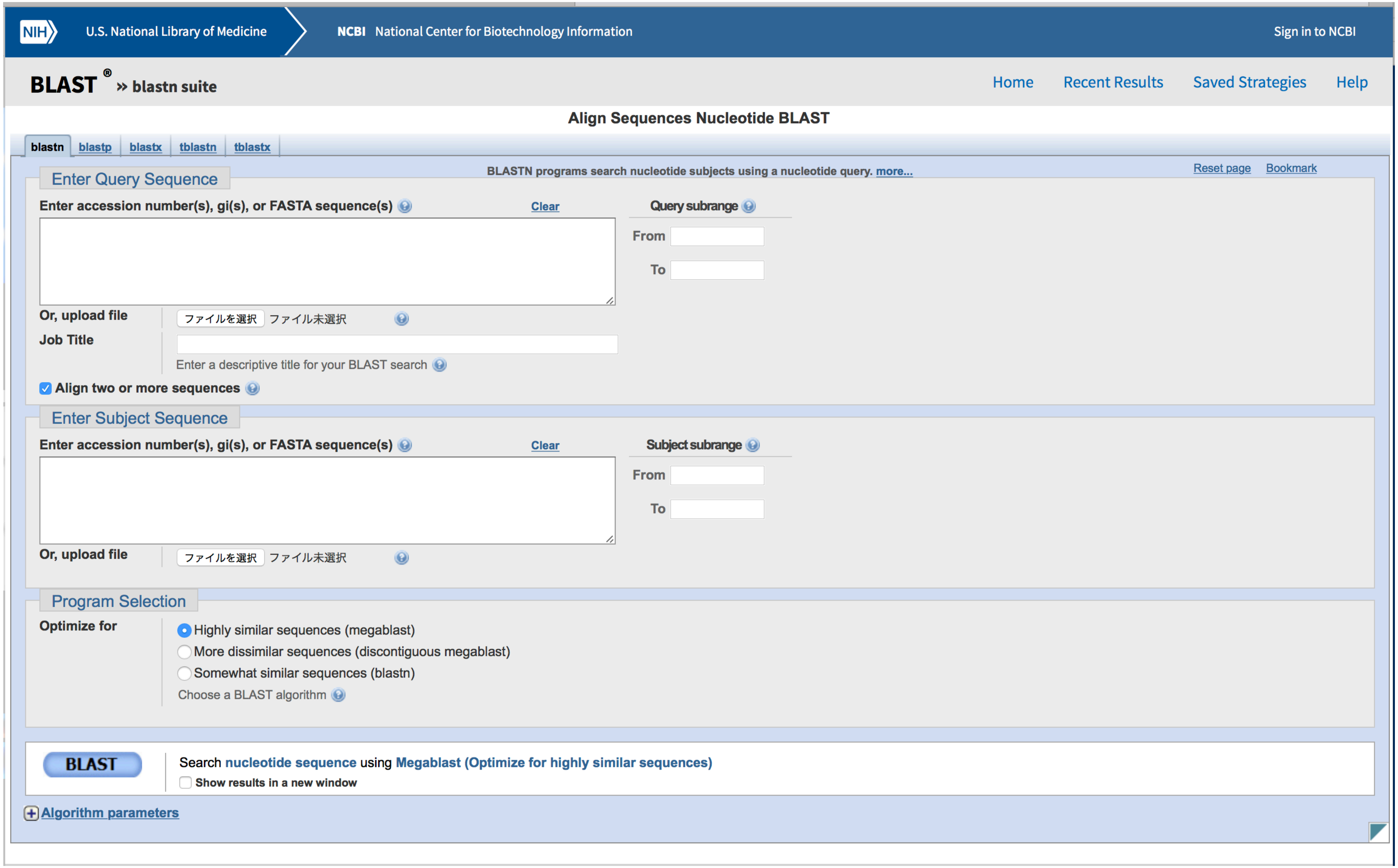
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

+ Algorithm parameters



NIH U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign in to NCBI

BLAST ® » **blastn suite**

Home Recent Results Saved Strategies Help

Align Sequences Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide subjects using a nucleotide query. more...

Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Query subrange

From

To

Or, upload file ファイルを選択 ファイル未選択

Job Title

Enter a descriptive title for your BLAST search

☑ **Align two or more sequences**

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Subject subrange

From

To

Or, upload file ファイルを選択 ファイル未選択

Program Selection

Optimize for

● Highly similar sequences (megablast)

○ More dissimilar sequences (discontiguous megablast)

○ Somewhat similar sequences (blastn)

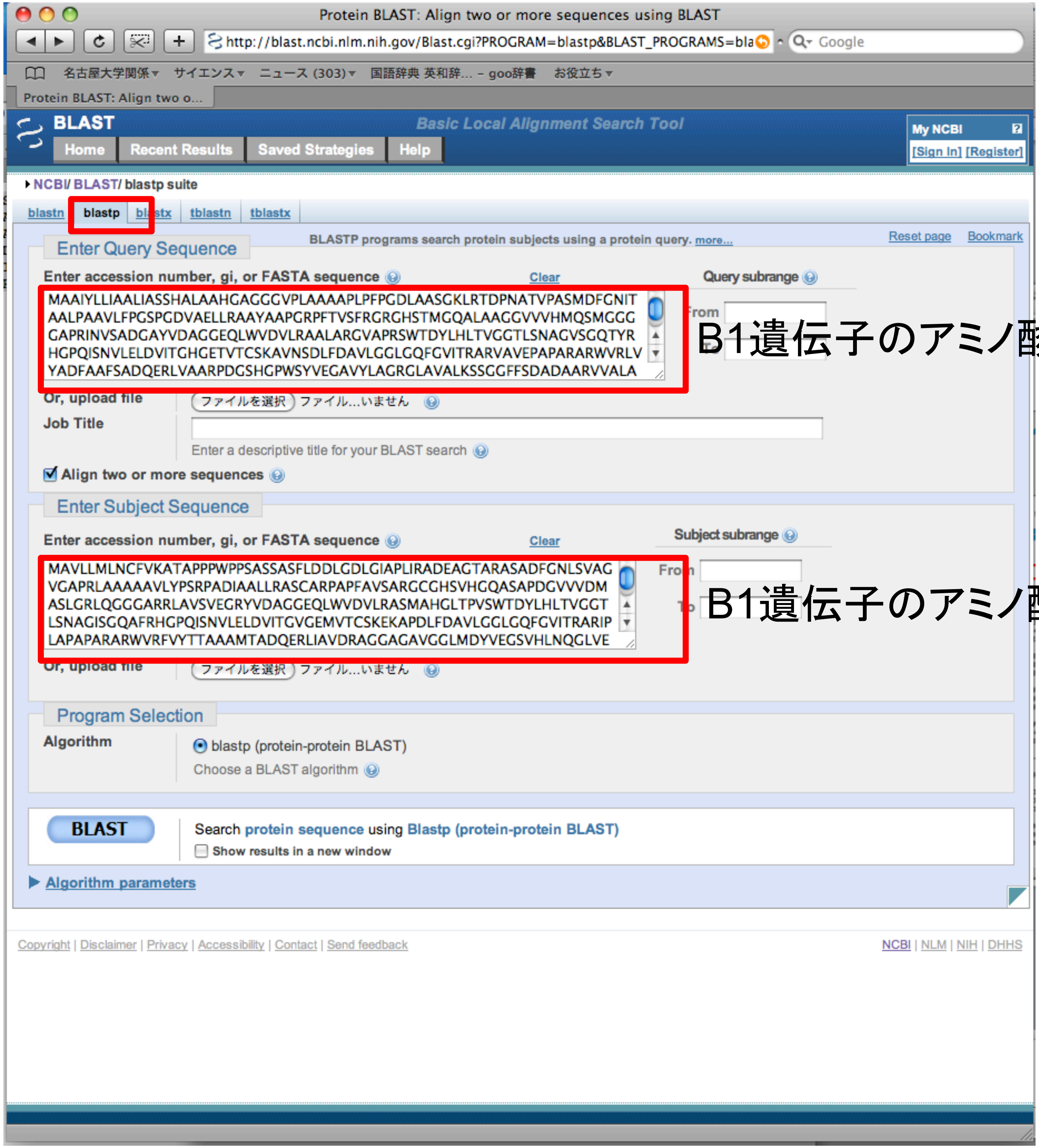
Choose a BLAST algorithm

BLAST

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ **Show results in a new window**

⊞ Algorithm parameters



BLASTP programs search protein subjects using a protein query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
MAAIYLLIAALIASSHALAAHGAGGGVPLAAAAPLPFPGDLAASGKLRTDPNATVPASMDFGNIT
AALPAAVLFPGSPGDVAELLRAAYAAPGRPFTVSFRGRGHSTMGQALAAGGVVVHMQSMGGG
GAPRINVSADGAYVDAGGEQLWVDVLRAALARGVAPRSWTDYLHLTVGGTLSNAGVSGQTYR
HGPQISNVLELDVITGHGETVTCSKAVNSDLFDAVLGGLGQFGVITRARVAVEPAPARARWVRLV
YADFAAFSADQERLVAARPDGSHGPWSYVEGAVYLAGRGLAVALKSSGGFFSDADAARVVALA
```

B1遺伝子のアミノ酸配列

Or, upload file: ファイル…いません

Job Title:

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

```
MAVLLMLNCFVKATAPPPWPPSASSASFLDDLGDLGIAPLIRADEAGTARASADFGNLSVAG
VGAPRLAAAAAVLYPSRPADIAALLRASCARPAPFAVSARGCGHSVHGQASAPDGVVVDM
ASLGRLQGGGARRLAVSVEGRYVDAGGEQLWVDVLRASMAHGLTPVSWTDYLHLTVGGT
LSNAGISGQAFRHGPQISNVLELDVITGVGEMVTCSKEKAPDLFDAVLGGLGQFGVITRARIP
LAPAPARARWVRFVYTTAAAMTADQERLIAVDRAGGAGAVGGLMDYVEGSVHLNQGLVE
```

B1遺伝子のアミノ酸配列

Or, upload file: ファイル…いません

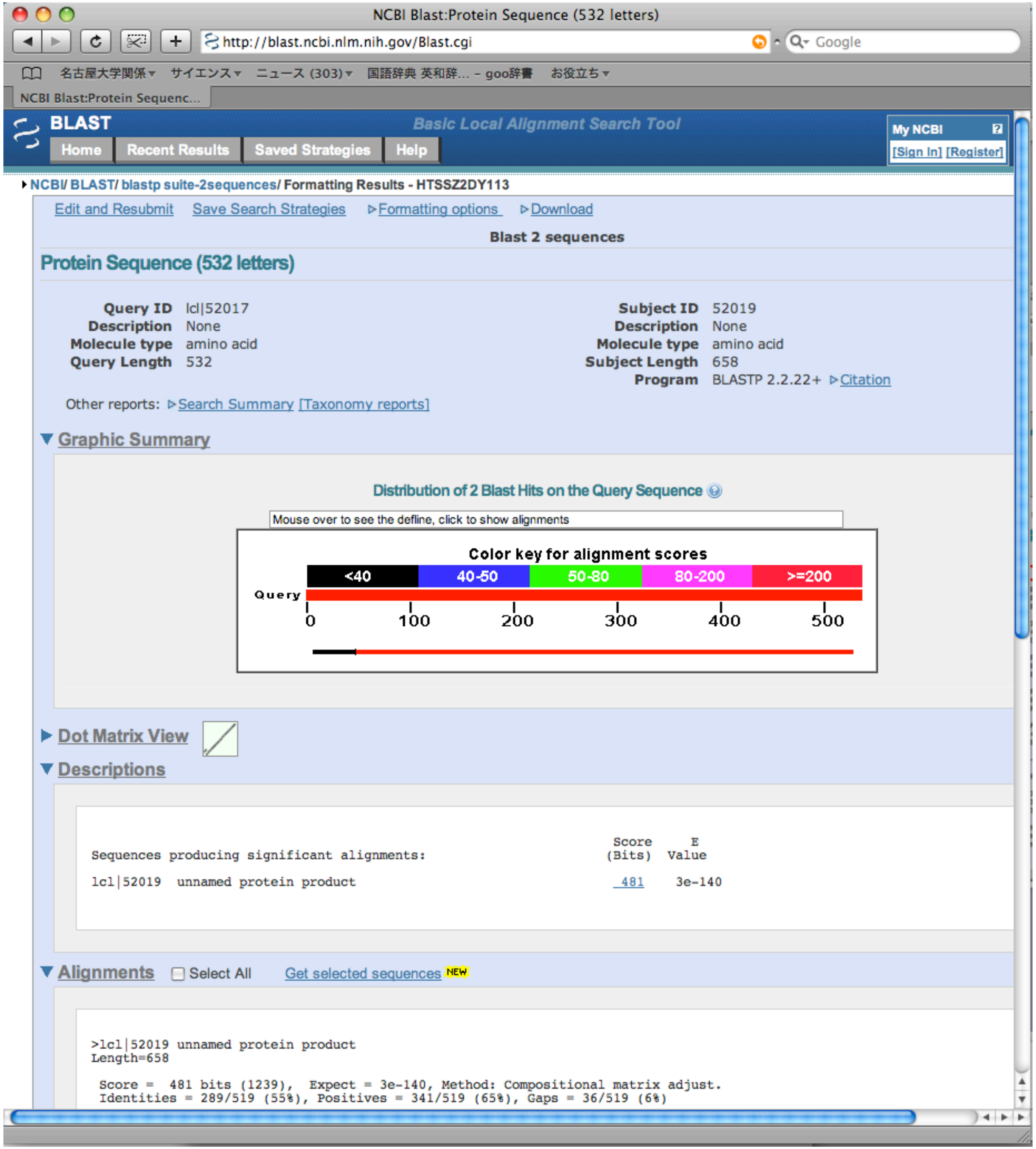
Program Selection

Algorithm: ● blastp (protein-protein BLAST)

Choose a BLAST algorithm

Search protein sequence using Blastp (protein-protein BLAST)

☐ Show results in a new window



NCBI/ BLAST/ blastp suite-2sequences/ Formatting Results - HTSSZ2DY113

Blast 2 sequences

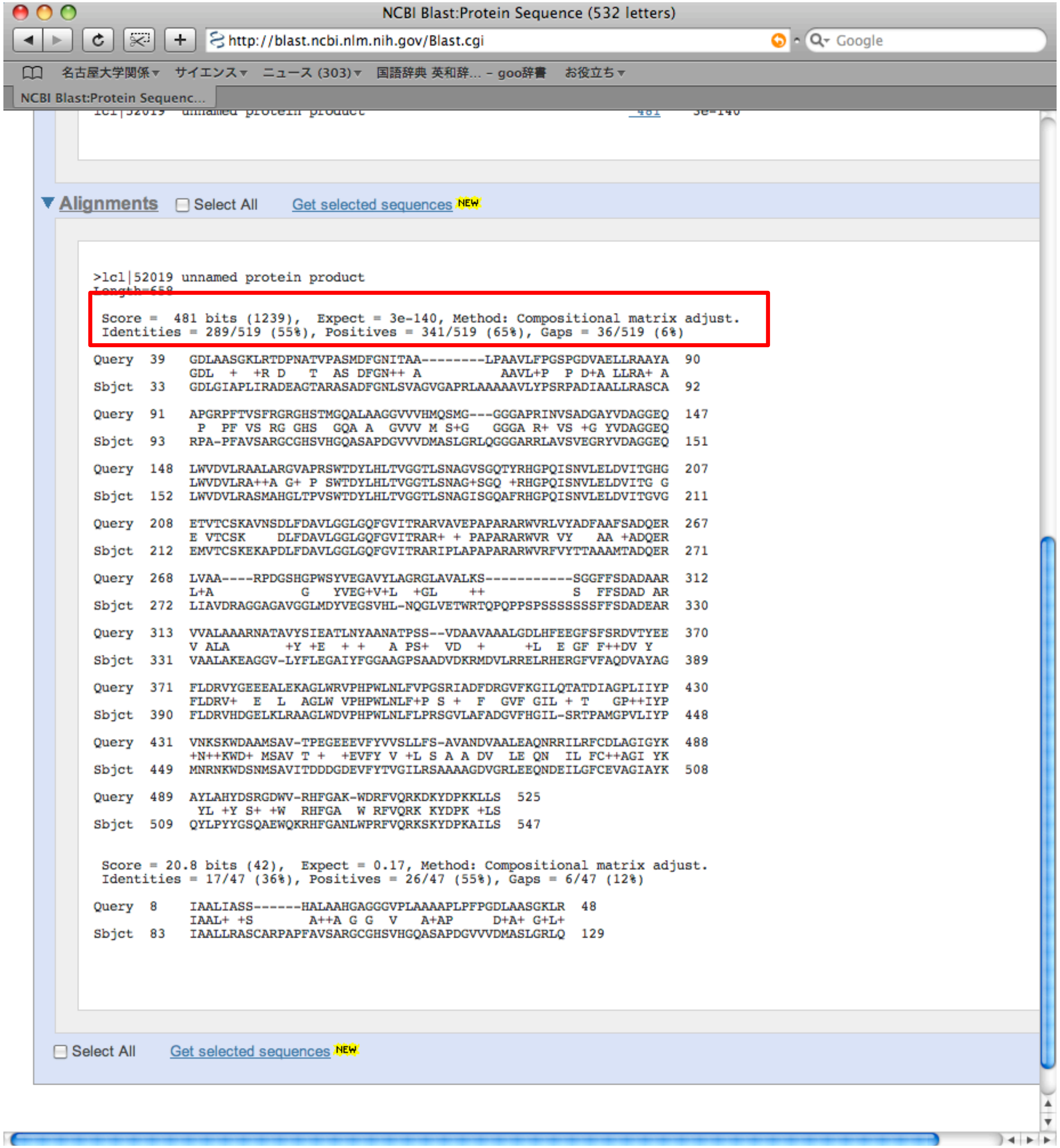
Protein Sequence (532 letters)

Query ID	lcl|52017	**Subject ID**	52019
Description	None	**Description**	None
Molecule type	amino acid	**Molecule type**	amino acid
Query Length	532	**Subject Length**	658
		Program	BLASTP 2.2.22+ ▷Citation

Other reports: ▷Search Summary [Taxonomy reports]

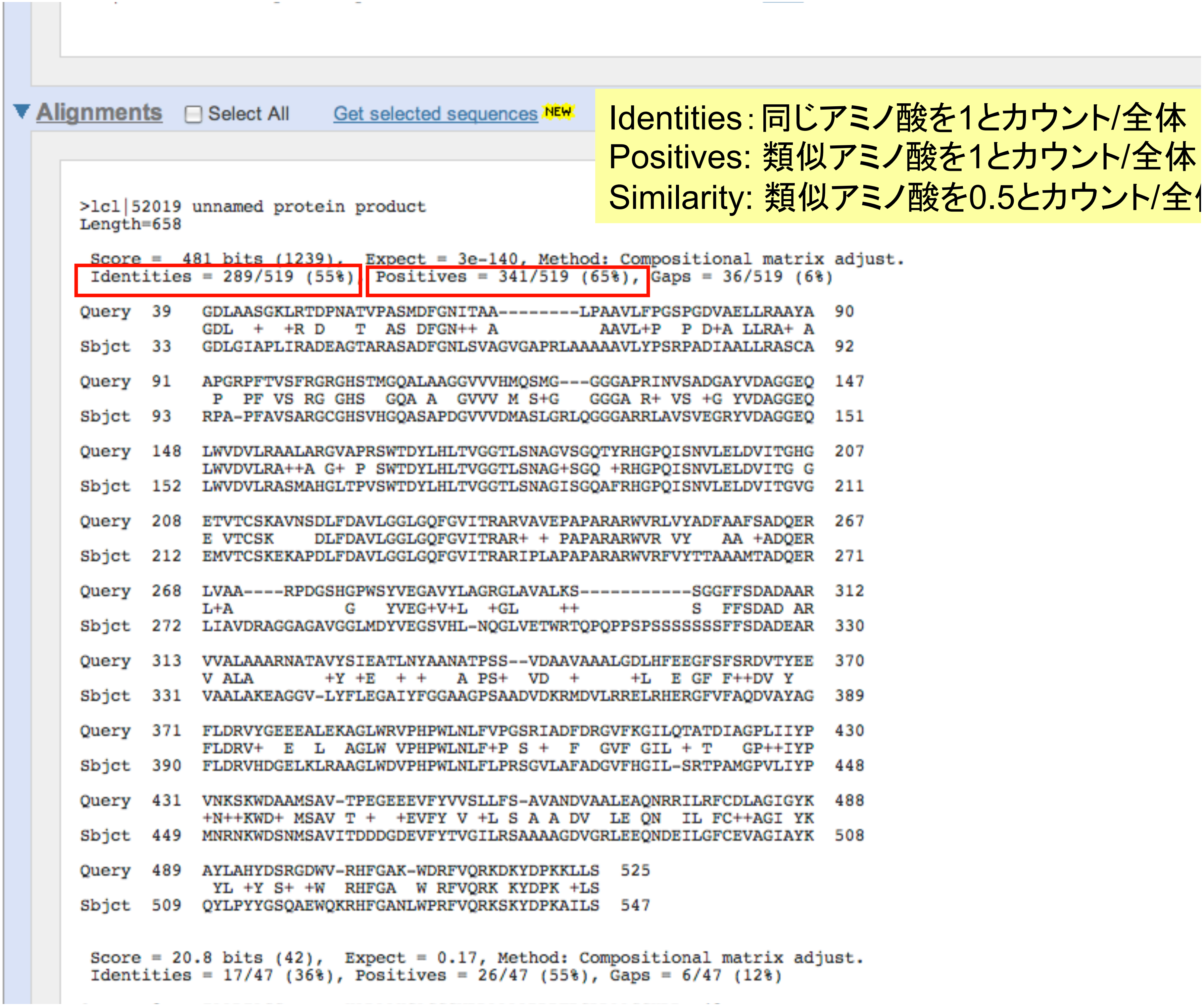
```
                                                                 Score     E
Sequences producing significant alignments:                      (Bits)  Value

lcl|52019  unnamed protein product                                 481    3e-140
```

```
>lcl|52019 unnamed protein product
Length=658

 Score =  481 bits (1239),  Expect = 3e-140, Method: Compositional matrix adjust.
 Identities = 289/519 (55%), Positives = 341/519 (65%), Gaps = 36/519 (6%)
```



```
lcl|52019  unnamed protein product                                481   3e-140

>lcl|52019 unnamed protein product
Length=658

 Score =  481 bits (1239),  Expect = 3e-140, Method: Compositional matrix adjust.
 Identities = 289/519 (55%), Positives = 341/519 (65%), Gaps = 36/519 (6%)

Query  39    GDLAASGKLRTDPNATVPASMDFGNITAA--------LPAAVLFPGSPGDVAELLRAAYA  90
             GDL  +  +R D   T  AS DFGN++ A          AAVL+P  P D+A LLRA+ A
Sbjct  33    GDLGIAPLIRADEAGTARASADFGNLSVAGVGAPRLAAAAAVLYPSRPADIAALLRASCA  92

Query  91    APGRPFTVSFRGRGHSTMGQALAAGGVVVHMQSMG---GGGAPRINVSADGAYVDAGGEQ  147
              P  PF VS RG GHS  GQA A  GVVV M S+G   GGGA R+ VS +G YVDAGGEQ
Sbjct  93    RPA-PFAVSARGCGHSVHGQASAPDGVVVDMASLGRLQGGGARRLAVSVEGRYVDAGGEQ  151

Query  148   LWVDVLRAALARGVAPRSWTDYLHLTVGGTLSNAGVSGQTYRHGPQISNVLELDVITGHG  207
             LWVDVLRA++A G+ P SWTDYLHLTVGGTLSNAG+SGQ +RHGPQISNVLELDVITG G
Sbjct  152   LWVDVLRASMAHGLTPVSWTDYLHLTVGGTLSNAGISGQAFRHGPQISNVLELDVITGVG  211

Query  208   ETVTCSKAVNSDLFDAVLGGLGQFGVITRARVAVEPAPARARWVRLVYADFAAFSADQER  267
             E VTCSK    DLFDAVLGGLGQFGVITRAR+ + PAPARARWVR VY   AA +ADQER
Sbjct  212   EMVTCSKEKAPDLFDAVLGGLGQFGVITRARIPLAPAPARARWVRFVYTTAAAMTADQER  271

Query  268   LVAA----RPDGSHGPWSYVEGAVYLAGRGLAVALKS-----------SGGFFSDADAAR  312
             L+A           G   YVEG+V+L  +GL    ++           S  FFSDAD AR
Sbjct  272   LIAVDRAGGAGAVGGLMDYVEGSVHL-NQGLVETWRTQPQPPSPSSSSSSFFSDADEAR  330

Query  313   VVALAAARNATAVYSIEATLNYAANATPSS--VDAAVAAALGDLHFEEGFSFSRDVTYEE  370
             V ALA       +Y +E  + +   A PS+  VD  +     +L  E GF F++DV Y
Sbjct  331   VAALAKEAGGV-LYFLEGAIYFGGAAGPSAADVDKRMDVLRRELRHERGFVFAQDVAYAG  389

Query  371   FLDRVYGEEEALEKAGLWRVPHPWLNLFVPGSRIADFDRGVFKGILQTATDIAGPLIIYP  430
             FLDRV+  E  L  AGLW VPHPWLNLF+P S +  F  GVF GIL + T   GP++IYP
Sbjct  390   FLDRVHDGELKLRAAGLWDVPHPWLNLFLPRSGVLAFADGVFHGIL-SRTPAMGPVLIYP  448

Query  431   VNKSKWDAAMSAV-TPEGEEEVFYVVSLLFS-AVANDVAALEAQNRRILRFCDLAGIGYK  488
             +N++KWD+ MSAV T +  +EVFY V +L S A A DV  LE QN  IL FC++AGI YK
Sbjct  449   MNRNKWDSNMSAVITDDDGDEVFYTVGILRSAAAAGDVGRLEEQNDEILGFCEVAGIAYK  508

Query  489   AYLAHYDSRGDWV-RHFGAK-WDRFVQRKDKYDPKKLLS  525
              YL +Y S+ +W  RHFGA  W RFVQRK KYDPK +LS
Sbjct  509   QYLPYYGSQAEWQKRHFGANLWPRFVQRKSKYDPKAILS  547


 Score = 20.8 bits (42),  Expect = 0.17, Method: Compositional matrix adjust.
 Identities = 17/47 (36%), Positives = 26/47 (55%), Gaps = 6/47 (12%)

Query  8     IAALIASS------HALAAHGAGGGVPLAAAAPLPFPGDLAASGKLR  48
             IAAL+ +S       A++A G G  V   A+AP     D+A+ G+L+
Sbjct  83    IAALLRASCARPAPFAVSARGCGHSVHGQASAPDGVVVDMASLGRLQ  129
```

Identities：同じアミノ酸を1とカウント/全体
Positives: 類似アミノ酸を1とカウント/全体
Similarity: 類似アミノ酸を0.5とカウント/全



```
>lcl|52019 unnamed protein product
Length=658

 Score =  481 bits (1239),  Expect = 3e-140, Method: Compositional matrix adjust.
 Identities = 289/519 (55%), Positives = 341/519 (65%), Gaps = 36/519 (6%)

Query  39   GDLAASGKLRTDPNATVPASMDFGNITAA--------LPAAVLFPGSPGDVAELLRAAYA  90
            GDL  +  +R D   T  AS DFGN++ A           AAVL+P  P D+A LLRA+ A
Sbjct  33   GDLGIAPLIRADEAGTARASADFGNLSVAGVGAPRLAAAAAVLYPSRPADIAALLRASCA  92

Query  91   APGRPFTVSFRGRGHSTMGQALAAGGVVVHMQSMG---GGGAPRINVSADGAYVDAGGEQ  147
             P  PF VS RG GHS  GQA A  GVVV M S+G   GGGA R+ VS +G YVDAGGEQ
Sbjct  93   RPA-PFAVSARGCGHSVHGQASAPDGVVVDMASLGRLQGGGARRLAVSVEGRYVDAGGEQ  151

Query  148  LWVDVLRAALARGVAPRSWTDYLHLTVGGTLSNAGVSGQTYRHGPQISNVLELDVITGHG  207
            LWVDVLRA++A G+ P SWTDYLHLTVGGTLSNAG+SGQ +RHGPQISNVLELDVITG G
Sbjct  152  LWVDVLRASMAHGLTPVSWTDYLHLTVGGTLSNAGISGQAFRHGPQISNVLELDVITGVG  211

Query  208  ETVTCSKAVNSDLFDAVLGGLGQFGVITRARVAVEPAPARARWVRLVYADFAAFSADQER  267
            E VTCSK    DLFDAVLGGLGQFGVITRAR+ + PAPARARWVR VY   AA +ADQER
Sbjct  212  EMVTCSKEKAPDLFDAVLGGLGQFGVITRARIPLAPAPARARWVRFVYTTAAAMTADQER  271

Query  268  LVAA----RPDGSHGPWSYVEGAVYLAGRGLAVALKS-----------SGGFFSDADAAR  312
            L+A           G   YVEG+V+L  +GL    ++           S  FFSDAD AR
Sbjct  272  LIAVDRAGGAGAVGGLMDYVEGSVHL-NQGLVETWRTQPQPPSPSSSSSSFFSDADEAR  330

Query  313  VVALAAARNATAVYSIEATLNYAANATPSS--VDAAVAAALGDLHFEEGFSFSRDVTYEE  370
            V ALA       +Y +E  + +   A PS+  VD  +     +L  E GF F++DV Y
Sbjct  331  VAALAKEAGGV-LYFLEGAIYFGGAAGPSAADVDKRMDVLRRELRHERGFVFAQDVAYAG  389

Query  371  FLDRVYGEEEALEKAGLWRVPHPWLNLFVPGSRIADFDRGVFKGILQTATDIAGPLIIYP  430
            FLDRV+  E  L  AGLW VPHPWLNLF+P S +  F  GVF GIL + T   GP++IYP
Sbjct  390  FLDRVHDGELKLRAAGLWDVPHPWLNLFLPRSGVLAFADGVFHGIL-SRTPAMGPVLIYP  448

Query  431  VNKSKWDAAMSAV-TPEGEEEVFYVVSLLFS-AVANDVAALEAQNRRILRFCDLAGIGYK  488
            +N++KWD+ MSAV T +  +EVFY V +L S A A DV  LE QN  IL FC++AGI YK
Sbjct  449  MNRNKWDSNMSAVITDDDGDEVFYTVGILRSAAAAGDVGRLEEQNDEILGFCEVAGIAYK  508

Query  489  AYLAHYDSRGDWV-RHFGAK-WDRFVQRKDKYDPKKLLS  525
             YL +Y S+ +W  RHFGA  W RFVQRK KYDPK +LS
Sbjct  509  QYLPYYGSQAEWQKRHFGANLWPRFVQRKSKYDPKAILS  547


 Score = 20.8 bits (42),  Expect = 0.17, Method: Compositional matrix adjust.
 Identities = 17/47 (36%), Positives = 26/47 (55%), Gaps = 6/47 (12%)
```

アミノ酸２０種類

- 水になじむもの
(親水性)
(極性を持つもの)
  - 電荷を持つもの
    - 酸性
      - アスパラギン酸(D), グルタミン酸(E)
    - 塩基性
      - アルギニン(R), リシン(K), ヒスチジン(H)
  - 電荷を持たないもの
    - グリシン(G), セリン(S), トレオニン(T), システイン(C)
アスパラギン(N), グルタミン(Q), チロシン(Y)
- 水になじまないもの
(疎水性)
(非極性のもの)
  - アラニン(A), バリン(V), ロイシン(L), イソロイシン(I)
メチオニン(M), トリプトファン(W), フェニルアラニン(F)
プロリン(P)

生物情報工学II (担当　芦苅）遺伝子の配列解析

第６回目-1（11/1）遺伝子の予測と機能解析

-2（11/1）演習

第７回目-1（11/8）DNA配列から遺伝子を探す

-2（11/8）DNA、アミノ酸のアラインメント、系統樹解析

-3（11/8）演習

第８回目-1（11/15）Genetyxを用いた解析

-2（11/15）演習

第９回目-1（11/22）KEGGを用いた解析

連鎖解析

出席と宿題の提出　提出方法

リンク集

データベース検索：

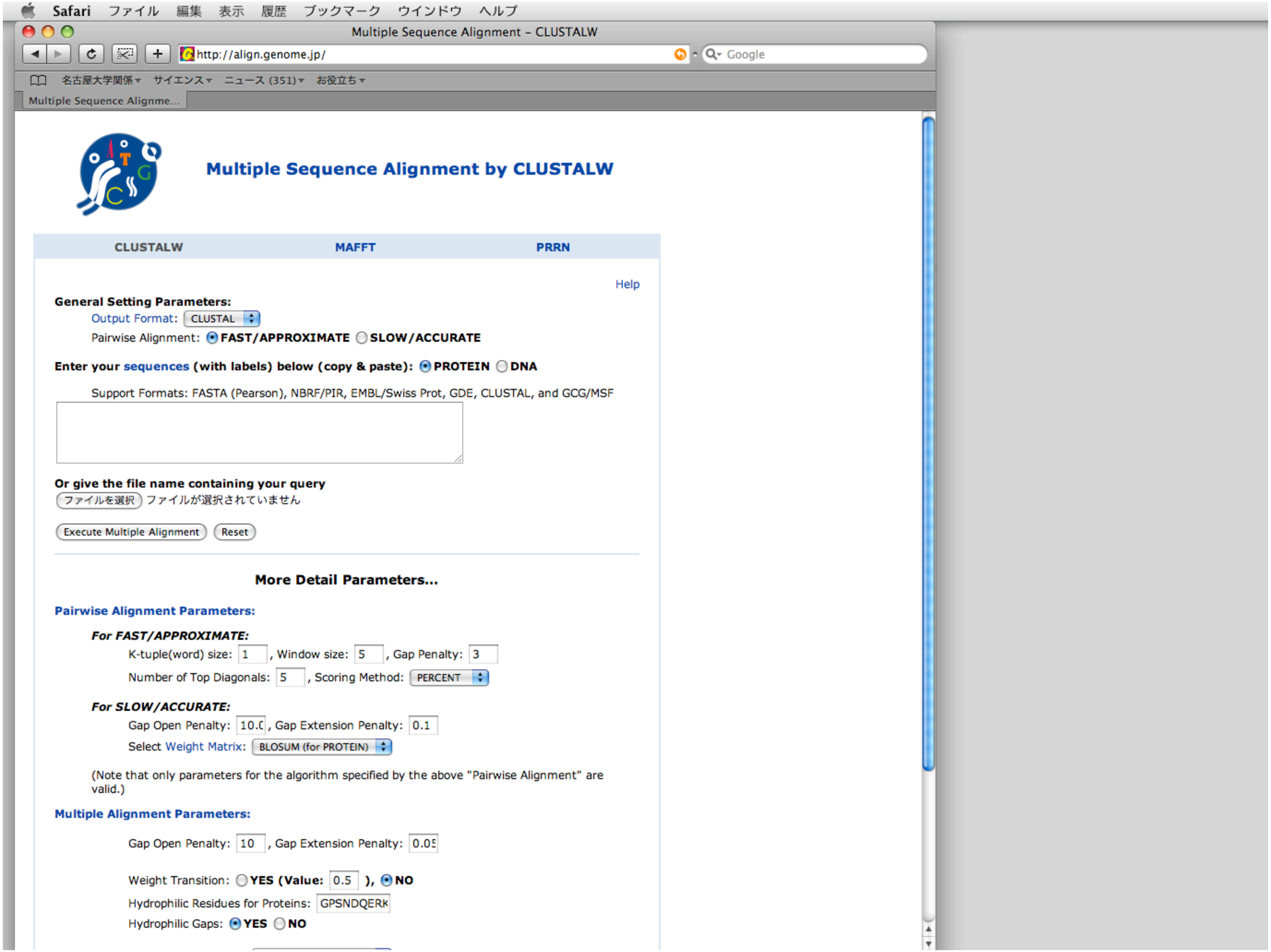
1. PubMed: 論文検索
2. Entrez：総合データベース
3. Google：何でも検索
4. Google Scholar：科学に関する検索
5. 特許データベース

ホモロジー検索：

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

配列解析：

1. Splign ：DNAとcDNAのマッチング
2. Blast align ：DNAとcDNAのマッチング
3. EMBOSS program" ：DNAとcDNAのマッチング
4. EMBOSS program" ：DNA配列をアミノ酸配列に変換
5. BLAST(NCBI) ：遺伝子のホモロジー検索
6. BLAST(DDBJ) ：遺伝子のホモロジー検索
7. Softberry/FGENESH ：遺伝子の予測
8. GeneMark ：遺伝子の予測
9. GENESCAN ：遺伝子の予測
10. clustalW： アライメント
11. InterProScan： ドメイン解析
12. PSORT： 細胞内局在性予測
13. GEO：遺伝子発現解析
14. Gene Ontology：用語を用いた検索
15. SSPN：スプライシングの予測



Multiple Sequence Alignment by CLUSTALW

CLUSTALW MAFFT PRRN

Help

General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment: ● **FAST/APPROXIMATE** ○ **SLOW/ACCURATE**

Enter your sequences (with labels) below (copy & paste): ● **PROTEIN** ○ **DNA**

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

Or give the file name containing your query

ファイルが選択されていません

More Detail Parameters...

Pairwise Alignment Parameters:

For FAST/APPROXIMATE:

K-tuple(word) size: 1 , Window size: 5 , Gap Penalty: 3

Number of Top Diagonals: 5 , Scoring Method: PERCENT

For SLOW/ACCURATE:

Gap Open Penalty: 10.0 , Gap Extension Penalty: 0.1

Select Weight Matrix: BLOSUM (for PROTEIN)

(Note that only parameters for the algorithm specified by the above "Pairwise Alignment" are valid.)

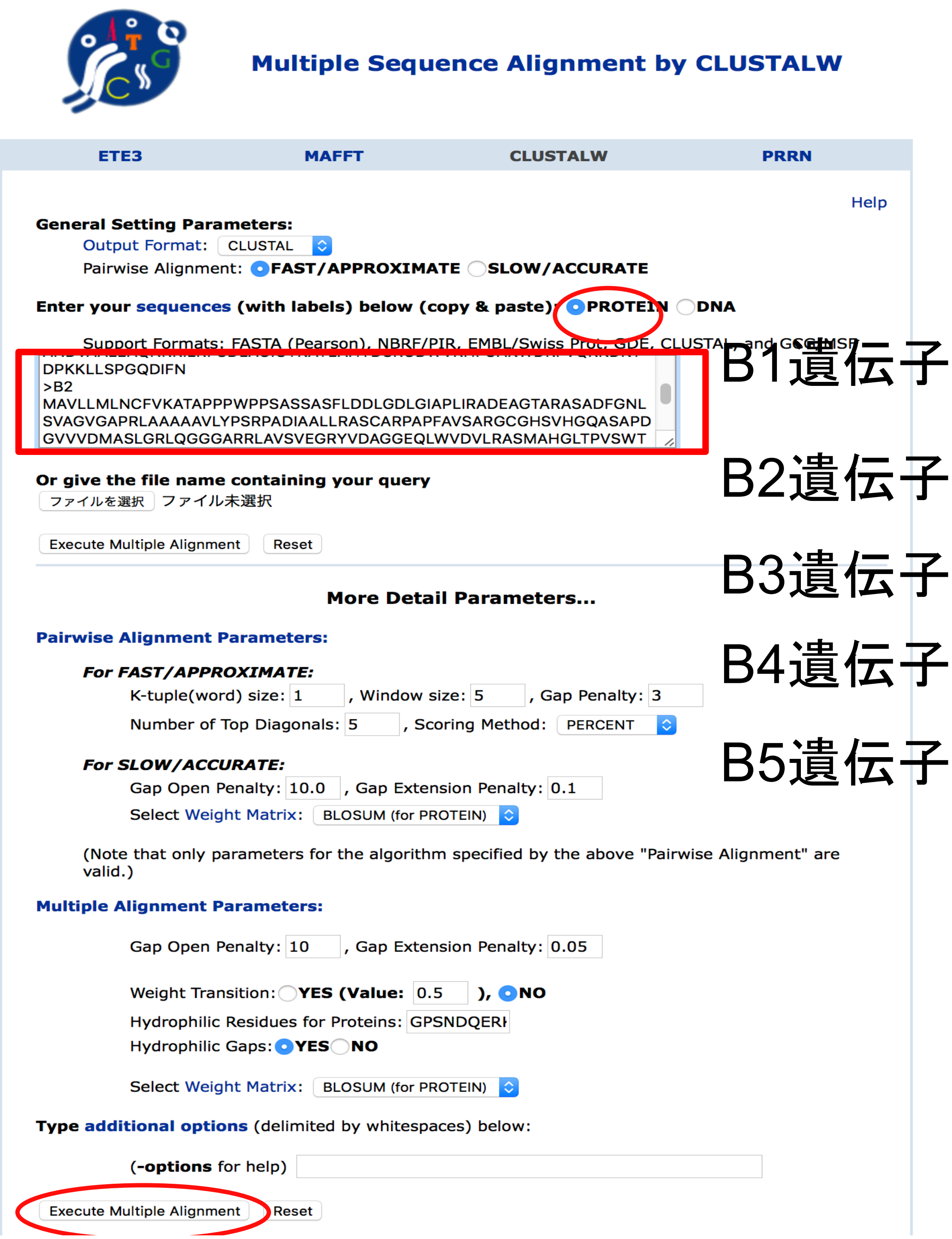
Multiple Alignment Parameters:

Gap Open Penalty: 10 , Gap Extension Penalty: 0.05

Weight Transition: ○ **YES (Value:** 0.5 **),** ● **NO**

Hydrophilic Residues for Proteins: GPSNDQERK

Hydrophilic Gaps: ● **YES** ○ **NO**



Multiple Sequence Alignment by CLUSTALW

General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment: ● **FAST/APPROXIMATE** ○ **SLOW/ACCURATE**

Enter your sequences (with labels) below (copy & paste): ● **PROTEIN** ○ **DNA**

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

```
DPKKLLSPGQDIFN
>B2
MAVLLMLNCFVKATAPPPWPPSASSASFLDDLGDLGIAPLIRADEAGTARASADFGNL
SVAGVGAPRLAAAAAVLYPSRPADIAALLRASCARPAPFAVSARGCGHSVHGQASAPD
GVVVDMASLGRLQGGGARRLAVSVEGRYVDAGGEQLWVDVLRASMAHGLTPVSWT
```

Or give the file name containing your query

ファイル未選択

More Detail Parameters...

Pairwise Alignment Parameters:

For FAST/APPROXIMATE:

K-tuple(word) size: 1 , Window size: 5 , Gap Penalty: 3

Number of Top Diagonals: 5 , Scoring Method: PERCENT

For SLOW/ACCURATE:

Gap Open Penalty: 10.0 , Gap Extension Penalty: 0.1

Select Weight Matrix: BLOSUM (for PROTEIN)

(Note that only parameters for the algorithm specified by the above "Pairwise Alignment" are valid.)

Multiple Alignment Parameters:

Gap Open Penalty: 10 , Gap Extension Penalty: 0.05

Weight Transition: ○ **YES (Value:** 0.5 **),** ● **NO**

Hydrophilic Residues for Proteins: GPSNDQER

Hydrophilic Gaps: ● **YES** ○ **NO**

Select Weight Matrix: BLOSUM (for PROTEIN)

Type additional options (delimited by whitespaces) below:

(**-options** for help)

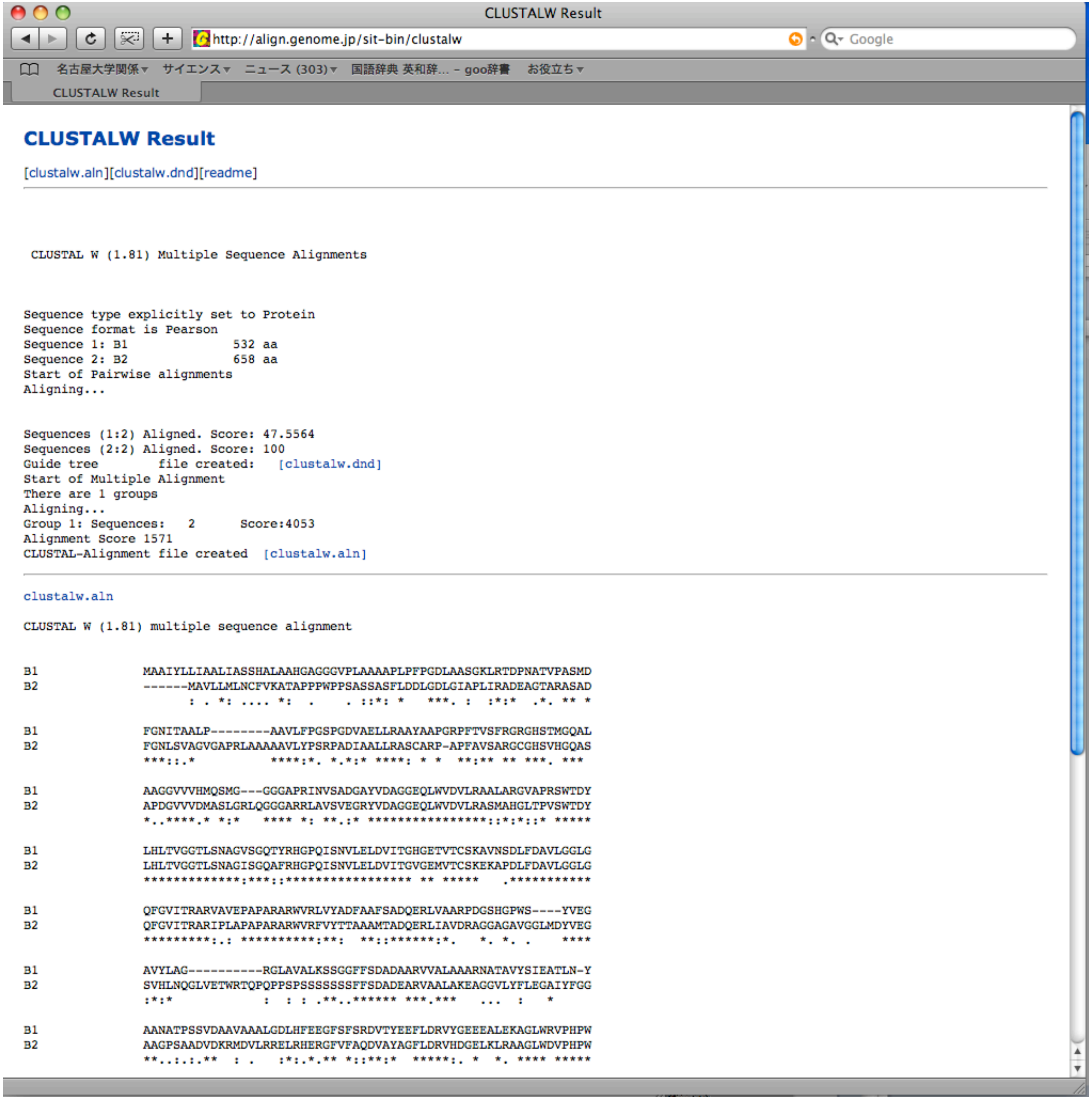
B1遺伝子のアミノ酸配列

B2遺伝子のアミノ酸配列

B3遺伝子のアミノ酸配列

B4遺伝子のアミノ酸配列

B5遺伝子のアミノ酸配列



CLUSTALW Result

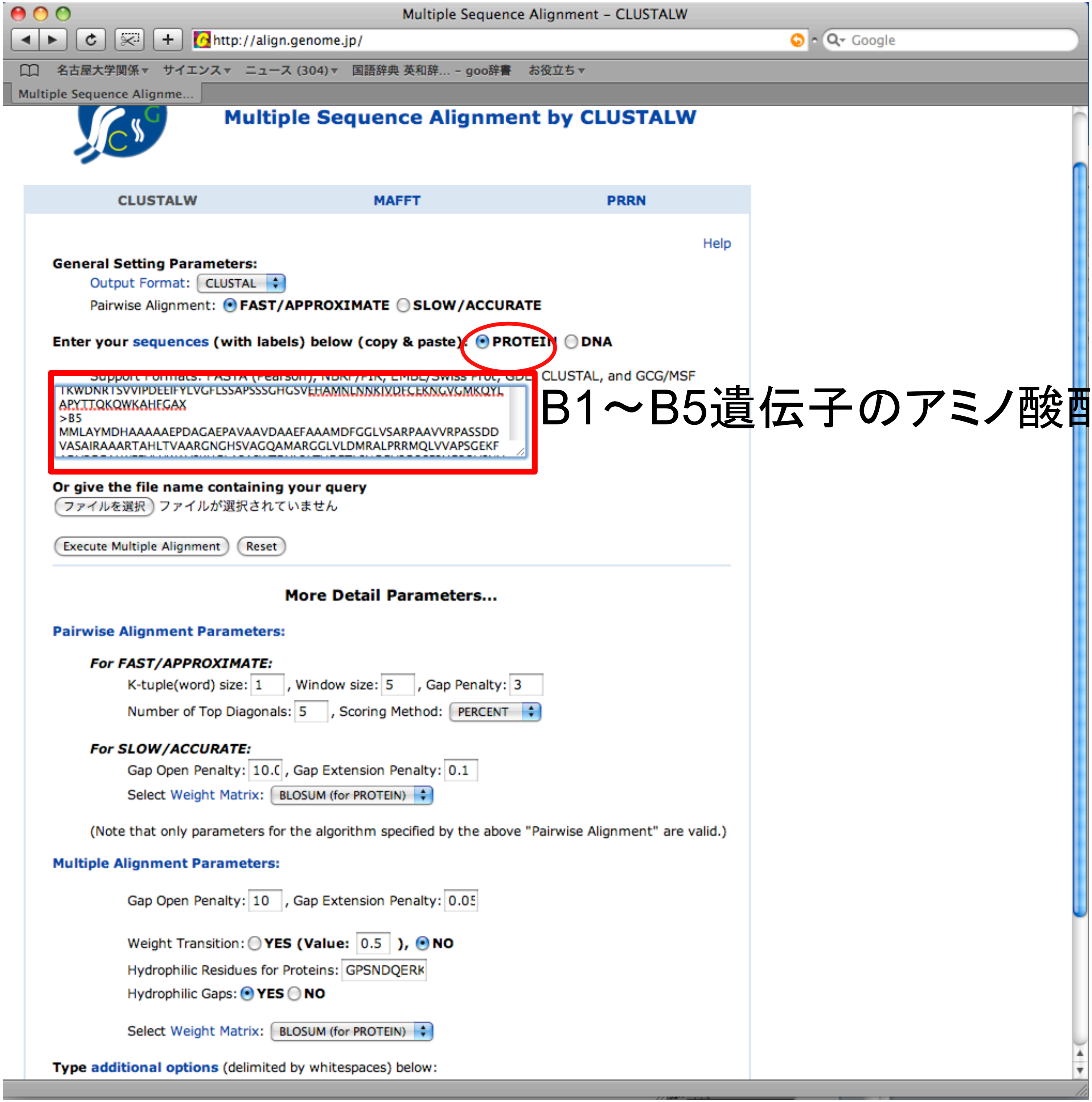
[clustalw.aln][clustalw.dnd][readme]

```
 CLUSTAL W (1.81) Multiple Sequence Alignments



Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: B1           532 aa
Sequence 2: B2           658 aa
Start of Pairwise alignments
Aligning...


Sequences (1:2) Aligned. Score: 47.5564
Sequences (2:2) Aligned. Score: 100
Guide tree        file created:   [clustalw.dnd]
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:4053
Alignment Score 1571
CLUSTAL-Alignment file created  [clustalw.aln]
```

---

clustalw.aln

```
CLUSTAL W (1.81) multiple sequence alignment


B1              MAAIYLLIAALIASSHALAAHGAGGGVPLAAAAPLPFPGDLAASGKLRTDPNATVPASMD
B2              ------MAVLLMLNCFVKATAPPPWPPSASSASFLDDLGDLGIAPLIRADEAGTARASAD
                      : . *: .... *:  .    . ::*: *   ***. :  :*:*  .*. ** *

B1              FGNITAALP--------AAVLFPGSPGDVAELLRAAYAAPGRPFTVSFRGRGHSTMGQAL
B2              FGNLSVAGVGAPRLAAAAAVLYPSRPADIAALLRASCARP-APFAVSARGCGHSVHGQAS
                ***::.*          ****:*. *.*:* ****: * *  **:** ** ***. ***

B1              AAGGVVVHMQSMG---GGGAPRINVSADGAYVDAGGEQLWVDVLRAALARGVAPRSWTDY
B2              APDGVVVDMASLGRLQGGGARRLAVSVEGRYVDAGGEQLWVDVLRASMAHGLTPVSWTDY
                *..****.* *:*   **** *: **.:* ****************::*:*::* *****

B1              LHLTVGGTLSNAGVSGQTYRHGPQISNVLELDVITGHGETVTCSKAVNSDLFDAVLGGLG
B2              LHLTVGGTLSNAGISGQAFRHGPQISNVLELDVITGVGEMVTCSKEKAPDLFDAVLGGLG
                *************:***::***************** ** *****   .***********

B1              QFGVITRARVAVEPAPARARWVRLVYADFAAFSADQERLVAARPDGSHGPWS----YVEG
B2              QFGVITRARIPLAPAPARARWVRFVYTTAAAMTADQERLIAVDRAGGAGAVGGLMDYVEG
                *********:.: **********:**:  **::******:*.   *. *. .    ****

B1              AVYLAG----------RGLAVALKSSGGFFSDADAARVVALAAARNATAVYSIEATLN-Y
B2              SVHLNQGLVETWRTQPQPPSPSSSSSSSFFSDADEARVAALAKEAGGVLYFLEGAIYFGG
                :*:*            :  : : .**..****** ***.***   ...  :   *

B1              AANATPSSVDAAVAAALGDLHFEEGFSFSRDVTYEEFLDRVYGEEEALEKAGLWRVPHPW
B2              AAGPSAADVDKRMDVLRRELRHERGFVFAQDVAYAGFLDRVHDGELKLRAAGLWDVPHPW
                **..:.:.**  : .   :*:.*.** *::**:*  *****:. *  *. **** *****
```



Multiple Sequence Alignment by CLUSTALW

General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment: ● **FAST/APPROXIMATE** ○ **SLOW/ACCURATE**

Enter your sequences (with labels) below (copy & paste): ● **PROTEIN** ○ **DNA**

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

```
TKWDNRTSVVIPDEEIFYLVGFLSSAPSSSGHGSVEHAMNLNNKIVDFCEKNGVGMKQYL
APYTTQKQWKAHFGAX
>B5
MMLAYMDHAAAAAEPDAGAEPAVAAVDAAEFAAAMDFGGLVSARPAAVVRPASSDD
VASAIRAAARTAHLTVAARGNGHSVAGQAMARGGLVLDMRALPRRMQLVVAPSGEKF
```

B1～B5遺伝子のアミノ酸配列

Or give the file name containing your query

ファイルが選択されていません

More Detail Parameters...

Pairwise Alignment Parameters:

For FAST/APPROXIMATE:

K-tuple(word) size: 1 , Window size: 5 , Gap Penalty: 3

Number of Top Diagonals: 5 , Scoring Method: PERCENT

For SLOW/ACCURATE:

Gap Open Penalty: 10.(, Gap Extension Penalty: 0.1

Select Weight Matrix: BLOSUM (for PROTEIN)

(Note that only parameters for the algorithm specified by the above "Pairwise Alignment" are valid.)

Multiple Alignment Parameters:

Gap Open Penalty: 10 , Gap Extension Penalty: 0.0

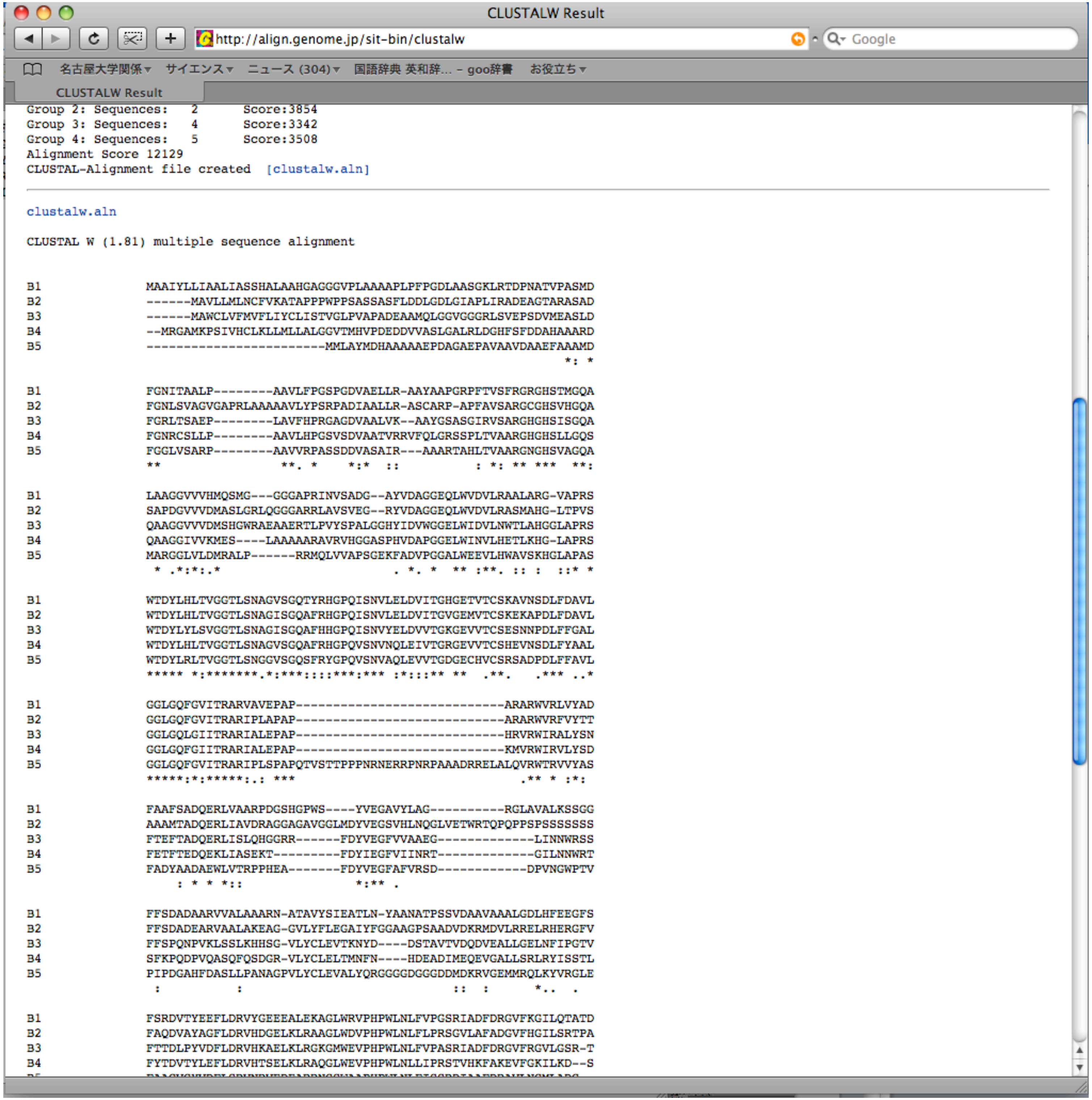
Weight Transition: ○ **YES (Value:** 0.5 **),** ● **NO**

Hydrophilic Residues for Proteins: GPSNDQERK

Hydrophilic Gaps: ● **YES** ○ **NO**

Select Weight Matrix: BLOSUM (for PROTEIN)

Type additional options (delimited by whitespaces) below:

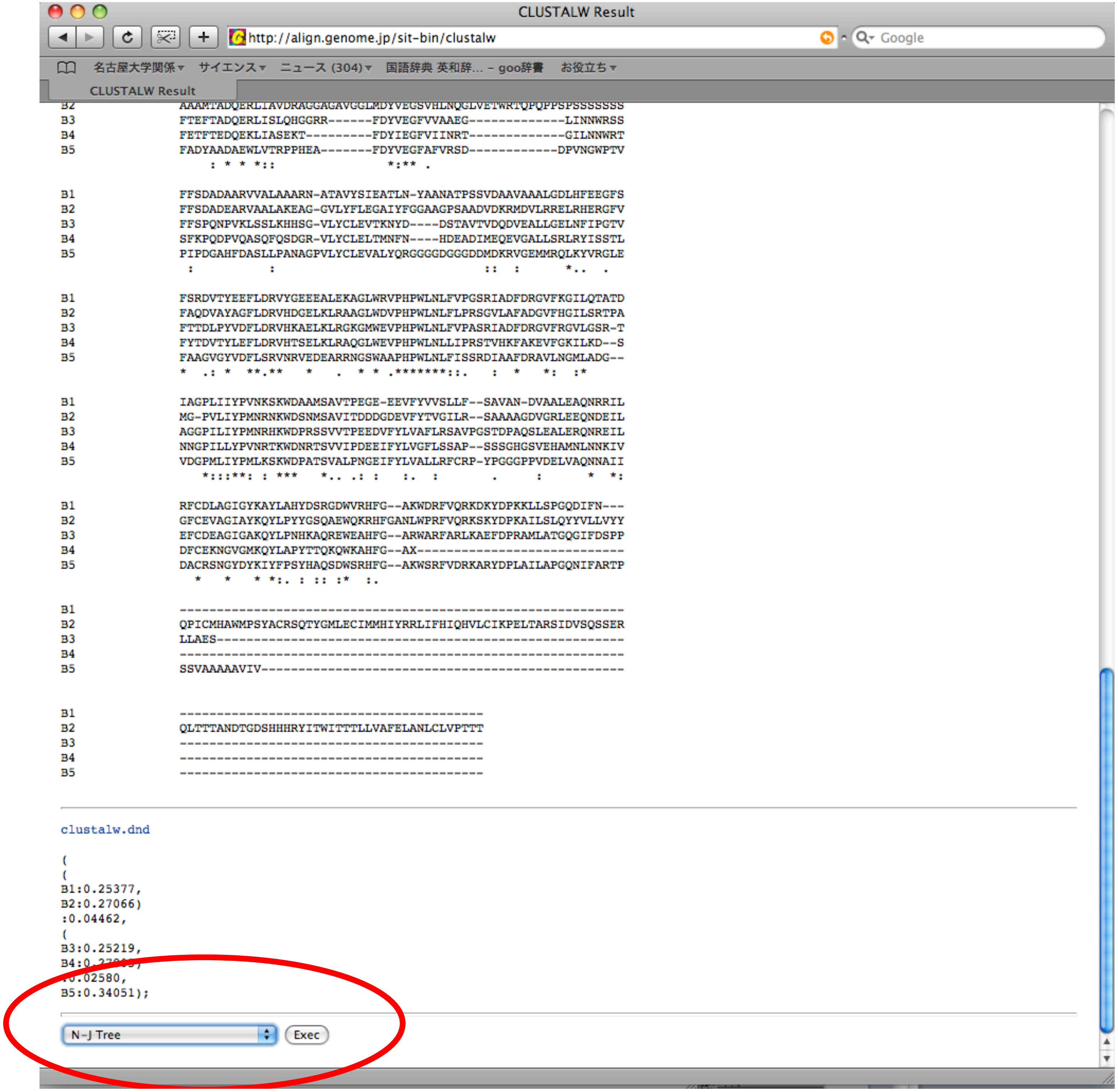


```
Group 2: Sequences:   2      Score:3854
Group 3: Sequences:   4      Score:3342
Group 4: Sequences:   5      Score:3508
Alignment Score 12129
CLUSTAL-Alignment file created  [clustalw.aln]
```

---

clustalw.aln

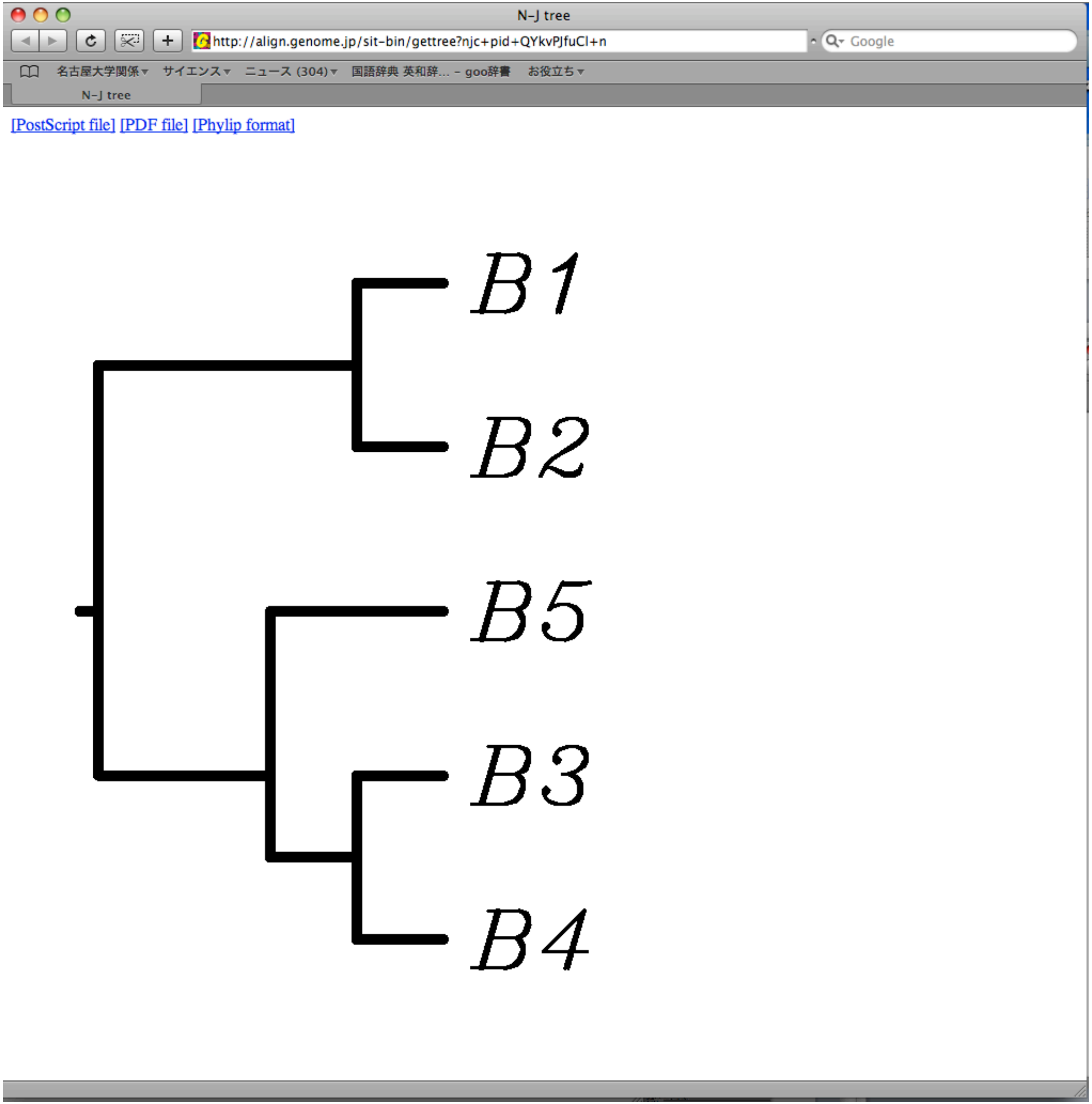
```
CLUSTAL W (1.81) multiple sequence alignment


B1              MAAIYLLIAALIASSHALAAHGAGGGVPLAAAAPLPFPGDLAASGKLRTDPNATVPASMD
B2              ------MAVLLMLNCFVKATAPPPWPPSASSASFLDDLGDLGIAPLIRADEAGTARASAD
B3              ------MAWCLVFMVFLIYCLISTVGLPVAPADEAAMQLGGVGGGRLSVEPSDVMEASLD
B4              --MRGAMKPSIVHCLKLLMLLALGGVTMHVPDEDDVVASLGALRLDGHFSFDDAHAAARD
B5              ------------------------MMLAYMDHAAAAAEPDAGAEPAVAAVDAAEFAAAMD
                                                                        *: *

B1              FGNITAALP--------AAVLFPGSPGDVAELLR-AAYAAPGRPFTVSFRGRGHSTMGQA
B2              FGNLSVAGVGAPRLAAAAAVLYPSRPADIAALLR-ASCARP-APFAVSARGCGHSVHGQA
B3              FGRLTSAEP--------LAVFHPRGAGDVAALVK--AAYGSASGIRVSARGHGHSISGQA
B4              FGNRCSLLP--------AAVLHPGSVSDVAATVRRVFQLGRSSPLTVAARGHGHSLLGQS
B5              FGGLVSARP--------AAVVRPASSDDVASAIR---AAARTAHLTVAARGNGHSVAGQA
                **                **. *    *:*  ::          : *: ** ***  **:

B1              LAAGGVVVHMQSMG---GGGAPRINVSADG--AYVDAGGEQLWVDVLRAALARG-VAPRS
B2              SAPDGVVVDMASLGRLQGGGARRLAVSVEG--RYVDAGGEQLWVDVLRASMAHG-LTPVS
B3              QAAGGVVVDMSHGWRAEAAERTLPVYSPALGGHYIDVWGGELWIDVLNWTLAHGGLAPRS
B4              QAAGGIVVKMES----LAAAAARAVRVHGGASPHVDAPGGELWINVLHETLKHG-LAPRS
B5              MARGGLVLDMRALP------RRMQLVVAPSGEKFADVPGGALWEEVLHWAVSKHGLAPAS
                 * .*:*:.*                       . *. *  ** :**. :: :  ::* *

B1              WTDYLHLTVGGTLSNAGVSGQTYRHGPQISNVLELDVITGHGETVTCSKAVNSDLFDAVL
B2              WTDYLHLTVGGTLSNAGISGQAFRHGPQISNVLELDVITGVGEMVTCSKEKAPDLFDAVL
B3              WTDYLYLSVGGTLSNAGISGQAFHHGPQISNVYELDVVTGKGEVVTCSESNNPDLFFGAL
B4              WTDYLHLTVGGTLSNAGVSGQAFRHGPQVSNVNQLEIVTGRGEVVTCSHEVNSDLFYAAL
B5              WTDYLRLTVGGTLSNGGVSGQSFRYGPQVSNVAQLEVVTGDGECHVCSRSADPDLFFAVL
                ***** *:*******.*:***::::***:*** :*:::** **  .**.   .*** ..*

B1              GGLGQFGVITRARVAVEPAP----------------------------ARARWVRLVYAD
B2              GGLGQFGVITRARIPLAPAP----------------------------ARARWVRFVYTT
B3              GGLGQLGIITRARIALEPAP----------------------------HRVRWIRALYSN
B4              GGLGQFGIITRARIALEPAP----------------------------KMVRWIRVLYSD
B5              GGLGQFGVITRARIPLSPAPQTVSTTPPPNRNERRPNRPAAADRRELALQVRWTRVVYAS
                *****:*:*****:.: ***                              .** * :*:

B1              FAAFSADQERLVAARPDGSHGPWS----YVEGAVYLAG----------RGLAVALKSSGG
B2              AAAMTADQERLIAVDRAGGAGAVGGLMDYVEGSVHLNQGLVETWRTQPQPPSPSSSSSSS
B3              FTEFTADQERLISLQHGGRR------FDYVEGFVVAAEG------------LINNWRSS
B4              FETFTEDQEKLIASEKT---------FDYIEGFVIINRT------------GILNNWRT
B5              FADYAADAEWLVTRPPHEA-------FDYVEGFAFVRSD------------DPVNGWPTV
                    : * * *::               *:** .

B1              FFSDADAARVVALAAARN-ATAVYSIEATLN-YAANATPSSVDAAVAAALGDLHFEEGFS
B2              FFSDADEARVAALAKEAG-GVLYFLEGAIYFGGAAGPSAADVDKRMDVLRRELRHERGFV
B3              FFSPQNPVKLSSLKHHSG-VLYCLEVTKNYD----DSTAVTVDQDVEALLGELNFIPGTV
B4              SFKPQDPVQASQFQSDGR-VLYCLELTMNFN----HDEADIMEQEVGALLSRLRYISSTL
B5              PIPDGAHFDASLLPANAGPVLYCLEVALYQRGGGGDGGGDDMDKRVGEMMRQLKYVRGLE
                 :          :                            ::  :      *..  .

B1              FSRDVTYEEFLDRVYGEEEALEKAGLWRVPHPWLNLFVPGSRIADFDRGVFKGILQTATD
B2              FAQDVAYAGFLDRVHDGELKLRAAGLWDVPHPWLNLFLPRSGVLAFADGVFHGILSRTPA
B3              FTTDLPYVDFLDRVHKAELKLRGKGMWEVPHPWLNLFVPASRIADFDRGVFRGVLGSR-T
B4              FYTDVTYLEFLDRVHTSELKLRAQGLWEVPHPWLNLLIPRSTVHKFAKEVFGKILKD--S
```



```
B2              AAAMTADQERLIAVDRAGGAGAVGGLMDYVEGSVHLNQGLVETWRTQPQPPSPSSSSSSS
B3              FTEFTADQERLISLQHGGRR------FDYVEGFVVAAEG-------------LINNWRSS
B4              FETFTEDQEKLIASEKT---------FDYIEGFVIINRT-------------GILNNWRT
B5              FADYAADAEWLVTRPPHEA-------FDYVEGFAFVRSD------------DPVNGWPTV
                    : * * *::               *:** .

B1              FFSDADAARVVALAAARN-ATAVYSIEATLN-YAANATPSSVDAAVAAALGDLHFEEGFS
B2              FFSDADEARVAALAKEAG-GVLYFLEGAIYFGGAAGPSAADVDKRMDVLRRELRHERGFV
B3              FFSPQNPVKLSSLKHHSG-VLYCLEVTKNYD----DSTAVTVDQDVEALLGELNFIPGTV
B4              SFKPQDPVQASQFQSDGR-VLYCLELTMNFN----HDEADIMEQEVGALLSRLRYISSTL
B5              PIPDGAHFDASLLPANAGPVLYCLEVALYQRGGGGDGGGDDMDKRVGEMMRQLKYVRGLE
                 :          :                            ::  :      *..  .

B1              FSRDVTYEEFLDRVYGEEEALEKAGLWRVPHPWLNLFVPGSRIADFDRGVFKGILQTATD
B2              FAQDVAYAGFLDRVHDGELKLRAAGLWDVPHPWLNLFLPRSGVLAFADGVFHGILSRTPA
B3              FTTDLPYVDFLDRVHKAELKLRGKGMWEVPHPWLNLFVPASRIADFDRGVFRGVLGSR-T
B4              FYTDVTYLEFLDRVHTSELKLRAQGLWEVPHPWLNLLIPRSTVHKFAKEVFGKILKD--S
B5              FAAGVGYVDFLSRVNRVEDEARRNGSWAAPHPWLNLFISSRDIAAFDRAVLNGMLADG--
                *  .: *  **.**   *   .  * * .*******::.   :  *   *:  :*

B1              IAGPLIIYPVNKSKWDAAMSAVTPEGE-EEVFYVVSLLF--SAVAN-DVAALEAQNRRIL
B2              MG-PVLIYPMNRNKWDSNMSAVITDDDGDEVFYTVGILR--SAAAAGDVGRLEEQNDEIL
B3              AGGPILIYPMNRHKWDPRSSVVTPEEDVFYLVAFLRSAVPGSTDPAQSLEALERQNREIL
B4              NNGPILLYPVNRTKWDNRTSVVIPDEEIFYLVGFLSSAP--SSSGHGSVEHAMNLNNKIV
B5              VDGPMLIYPMLKSKWDPATSVALPNGEIFYLVALLRFCRP-YPGGGPPVDELVAQNNAII
                   *:::**: : ***   *.. .: :   :.  :       .     :      *  *:

B1              RFCDLAGIGYKAYLAHYDSRGDWVRHFG--AKWDRFVQRKDKYDPKKLLSPGQDIFN---
B2              GFCEVAGIAYKQYLPYYGSQAEWQKRHFGANLWPRFVQRKSKYDPKAILSLQYYVLLVYY
B3              EFCDEAGIGAKQYLPNHKAQREWEAHFG--ARWARFARLKAEFDPRAMLATGQGIFDSPP
B4              DFCEKNGVGMKQYLAPYTTQKQWKAHFG--AX----------------------------
B5              DACRSNGYDYKIYFPSYHAQSDWSRHFG--AKWSRFVDRKARYDPLAILAPGQNIFARTP
                  *   *   * *:. : :: :*  :.

B1              ------------------------------------------------------------
B2              QPICMHAWMPSYACRSQTYGMLECIMMHIYRRLIFHIQHVLCIKPELTARSIDVSQSSER
B3              LLAES-------------------------------------------------------
B4              ------------------------------------------------------------
B5              SSVAAAAAVIV-------------------------------------------------


B1              -----------------------------------------
B2              QLTTTANDTGDSHHHRYITWITTTLLVAFELANLCLVPTTT
B3              -----------------------------------------
B4              -----------------------------------------
B5              -----------------------------------------
```

clustalw.dnd

```
(
(
B1:0.25377,
B2:0.27066)
:0.04462,
(
B3:0.25219,
B4:0.27[illegible])
:0.02580,
B5:0.34051);
```


N–J tree
http://align.genome.jp/sit-bin/gettree?njc+pid+QYkvPJfuCl+n
Google
名古屋大学関係
サイエンス
ニュース (304)
国語辞典 英和辞… – goo辞書
お役立ち
N–J tree
[PostScript file] [PDF file] [Phylip format]
B1
B2
B5
B3
B4