

BLASTクイックスタート

このミニコースでは、配列相同性検索プログラムであるBLASTファミリーについて実用的な紹介をしていきます。その課題は単純な探索から、ある特別な目的の探索をBLASTの創造的な使い方を実現するといった幅の広いものになっています。

課題.1 blastnの利用

■ 課題. 1-1 プライマーでの増幅領域の特定

下に示したプライマーを用いることで増幅できる、GenBankに登録されているヒトゲノムの配列を BLASTを用いて探さない。

Primer1: GTACCTTGATTTCGTATTC

Primer2: GACTCTACTACCTTTACCC

■ ヒント

forwardとreverse両方のプライマーをBLASTのinput boxに貼り付けてみましょう。この時に最初のプライマー配列の後ろに30文字ほどのNを入れてみてください。

探索範囲をヒトゲノムのみに絞り込んでください。

結果が得られたら、それらが同じ配列に当たっていることを確認してください。またこれらのヒットは DNAの相補鎖にあたっていないなければいけません

いくつのヒットが得られましたか？

■ 課題. 1-2 一塩基多型 (SNPs) の特定

Hermankova, Mらは、子供と大人が併用療法で受けているHIV-1薬剤の耐性についての研究をしています。これらの患者から単離したHIV-1配列上のSNPsを特定するために、下のテキストボックスに示したある患者から得られた配列を用いてblsat探索をしてください。結果は'Flat Query with Identities'フォーマットで出力してください。アライメント上で6番目のポジションに認められるSNPを確認してください。

BLAST 入力配列

```
ATGACCTCAAATCACTCTTTGGCAACGACCCCTCGTCACAATAAGATAGGGGGGCAAC
T
AAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGGAGTTGC
C
AGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGT
```

課題.2 blastpの利用

■ 課題. 2-1 タンパク質断片配列の素性を知る

下に示されたアミノ酸配列の素性を調べてください。この配列はショウジョウバエの5エクソンを翻訳して作成されたものです。このタンパク質の素性を知るために、swissprotデータベースに対して blastpによる検索をかけてください。

BLAST 入力配列

```
MSQICKRGLLISNRLAPAALRCKSTWFSEVQMGPDAIL
GVTEAFKKDTNPKKINLGAGAYR
DDNTQPFVLPVREAEKRVVSRSLDKEYATIIGIPEFYN
KAIELALGKGSKRLLAAKHNVTAQ
SIGTGAIRIGAAFLAKFWQGNREIYIPSPSWGHNHVAIF
```

■ 課題. 2-2 タンパク質断片配列の素性を知る

下に示されたアミノ酸配列の素性を調べてください。この配列はショウジョウバエの4エクソンを翻訳して作成されたものです。このタンパク質の素性を知るために、swissprotデータベースに対して blastpによる検索をかけてください。

```
MSLTVEIVATKPYEGQKPGTSGLRKKVKVFTQPNYTENF  
VQAILEANGAALAGSTLVVGGDGRF  
YCKEAAELIVRLSAANGVSKLLVGQNGILSTPAVSSLIR  
HNKALGGIVLTASHNPGGPENDFGI  
KFNCENGGPAPDAFTNHIYKITTEIKEYKLVRNLQIDIS
```

Step. 3 blastxの利用

■ 課題. 3-1 フレームシフト位置の特定

下に示された核酸配列のどこにフレームシフトが入っていますか？

BLAST 入力配列

```
AGAAGAAGACATAGTAATTAGATCTGAAAATTTACGAACAATGCTAAAACCATAATAG  
TACAGCTGAAGGAATCTATAAAAA  
TTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATACCTATAGCAACGGGGGGA  
GCAATTTATGCAACAGGAGACATA  
ATAGGAGATATAAGACAAGCACATTGTAACTTAGTAGAGACCAATGGGATAACACTTT
```

■ 課題. 3-2 フレームシフト位置の特定

下に示された核酸配列のどこにフレームシフトが入っていますか？

BLAST 入力配列

```
ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGCACCATGCTCCTTGGGT  
GTTGATGATCCGTAGTGC  
TGCAGACCAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGAAGCAACCA  
CCACTCCATTTTGTGCAT  
CAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACACGCCTGTGTA
```

課題. 4 短いほとんど完全にマッチする配列の探索

■ 課題. 4-1 EinsteinはGenBankにいるの？

いったいいくつの'einstein'がGenBankのアミノ酸配列には含まれているのでしょうか？ “短いほとんど完全にマッチする配列の探索”用に用意された特別なページに初期設定されているパラメータに注目してください。特に wordsize, expect, filter オプション, 比較マトリクスに注目してください。

(以前のblastには“Search for short, nearly exact matches”という特別に用意されたサービスがあったのですが、現在はありません。)

■ 課題. 4-2 ValentineはGenBankにあるの？

いったいいくつの'valentein'がGenBankのアミノ酸配列には含まれているのでしょうか？ “短いほとんど完全にマッチする配列の探索”用に用意された特別なページに初期設定されているパラメータに注目してください。特に wordsize, expect, filter オプション, 比較マトリクスに注目してください。

(以前のblastには“Search for short, nearly exact matches”という特別に用意されたサービスがあったのですが、現在はありません。)

課題. 5 mouseに特化したホモロジー検索

■ 課題. 5-1 Hoxbホモログの探索

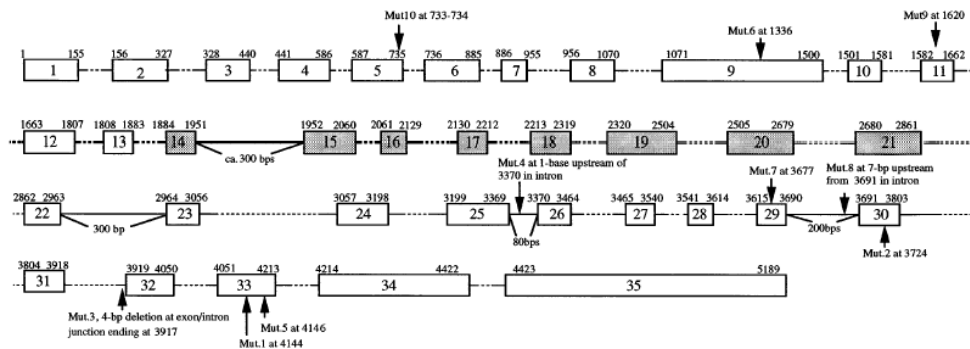
マウスHoxbに関するmRNA(NM_008268), アミノ酸Refseq配列(NP_032294)を用いて reference genome に対する blast探索で何個のホモログがマウスで見つかるかを調べてみてください。

■ 課題. 5-2 プロトカドヘリンホモログの探索

マウスプロトカドヘリンに関するmRNA(AY013770), アミノ酸配列(AAK26059)を用いて reference genome に対する blast探索で何個のホモログがマウスで見つかるかを調べてみてください。

課題. 6 blastを用いた二配列のアライメント

■ 課題. 6-1 Exon位置の探索



下に示した配列はGenBankのHTGに登録されたWRN Werenerシンドローム遺伝子の一部を含んだものです。この遺伝子は35個のExonからなっており、上に示したようなcDNA構造をとっています。BLAST2seqを用いて WRN遺伝子のcDNA配列とこのHTGに登録された配列とを比べることで、このHTGの配列がどのExonのものかを調べてください。Subject配列の向きに注意してください。

BLAST 入力配列

```
>HTG sequence
GAATTCATTTAAGGAAAGAAAATGAAAATTTGATCCCTAATATTATTT
AATGAAGTGGCTAAATGAATATCTCTGCTTTGTGGTTTGAAAATTAAT
ATTGATTTTTTTTCCCTAGAGGAAGAAATCCTACTTAGCGACATGA
ACAAACAGTTGACTTCAATCTCTGAGGAAGTGATGGATCTGGCTAAGC

>cDNA of WRN gene
TGTGCGCCGGGAGGCGCGGCTTGTACTCGGCAGCGCGGAATAAAG
TTTGTGATTTGGTGTCTAGCCTGGATGCCTGGGTTGCAGCCCTGCTT
GTGGTGGCGCTCCACAGTCATCCGGCTGAAGAAGACCTGTTGGACTGG
ATCTTCTCGGTTTTCTTTCAGATATTGTTTTGTATTTACCCATGAAG
```

■ 課題. 6-2 欠落した配列は？

下に示された配列をSequence1, Sequence2の両方に入れてBLAST2seqを実行してみてください。どうして配列が二つに分断されるのでしょうか？欠落した配列は何でしょう？

BLAST 入力配列

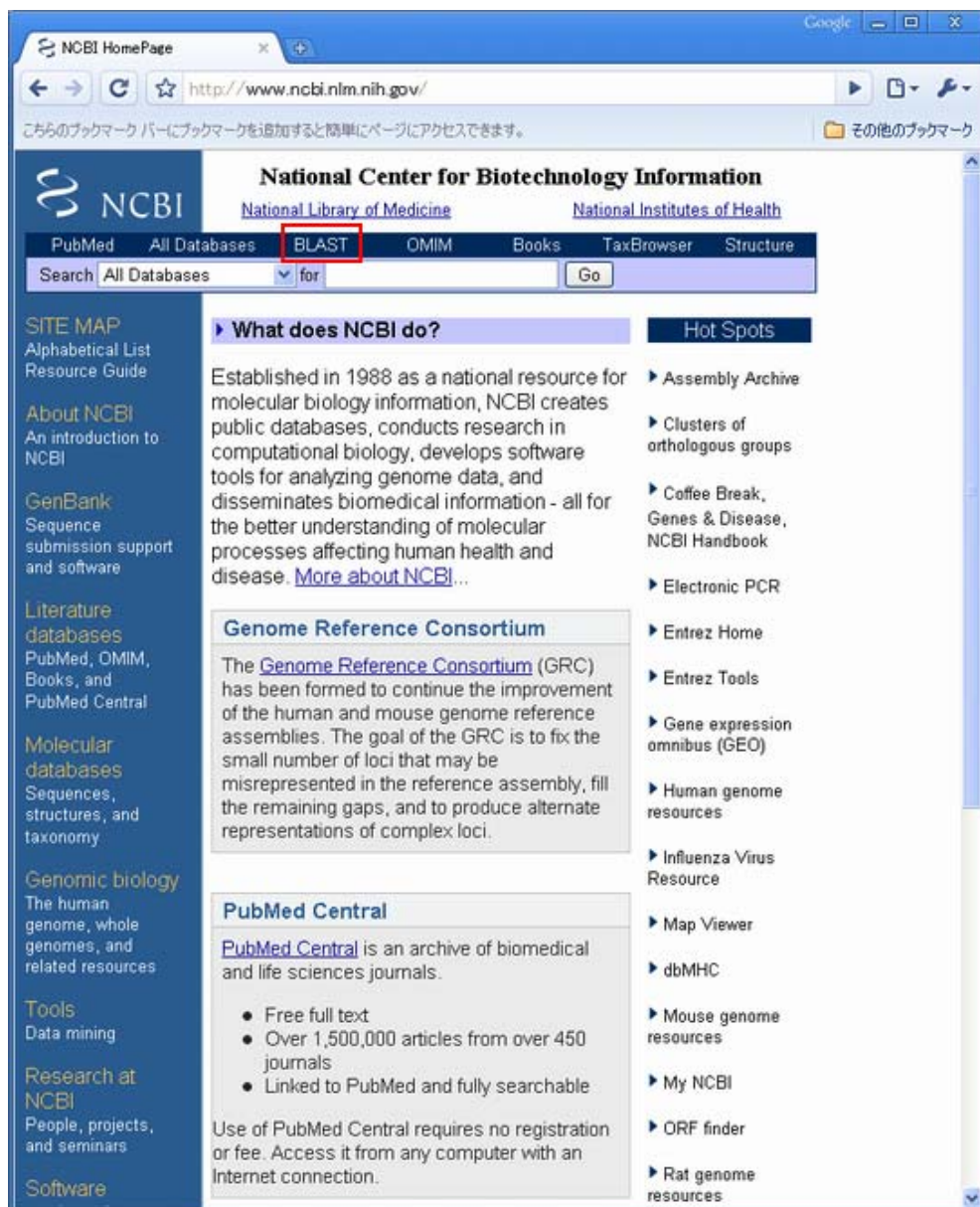
```
AGCCCCCTCACCTCACTCGCAGGCATACAGCCCCAGAGGCTCCCGATGGCGAGATTAT
GGTGC
CTTGGCTATCATCATGGCAGGAATTGCATTTGGCTTTCACCAACTCTACAAGAGGTACC
TGCTG
CCCCCTCATCCTGGGAGGCCGAGAGGACAGAAAGCAGCTGGAGAGGATGGCAGCGAGTCT
```

解答・解説

■ 解答1-1

[blastnを用いたプライマー探索](#)

NCBIのトップページ上部からBLASTを選択し、BLASTメニューのページに移動します。



そこからまずblastn(核酸問い合わせ配列 vs 核酸データベース) (赤四角で囲まれた箇所)を選択します。

BLAST: Basic Local Align... x

http://blast.ncbi.nlm.nih.gov/Blast.cgi

こちらのブックマーク バーにブックマークを追加すると簡単にページにアクセスできます。 その他のブックマーク

BLAST

Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Re...]

► NCBI/BLAST Home

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

New Designing or Testing PCR Primers? Try your search in **Primer-BLAST**. [Go](#)

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

☐ Human	☐ Oryza sativa	☐ Gallus gallus
☐ Mouse	☐ Bos taurus	☐ Pan troglodytes
☐ Rat	☐ Danio rerio	☐ Microbes
☐ Arabidopsis thaliana	☐ Drosophila melanogaster	☐ Apis mellifera

Basic BLAST

Choose a BLAST program to run.

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

Specialized BLAST

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

- ☐ Make specific primers with [Primer-BLAST](#)
- ☐ Search [trace archives](#)
- ☐ Find [conserved domains](#) in your sequence (cds)
- ☐ Find sequences with similar [conserved domain architecture](#) (cdart)
- ☐ Search sequences that have gene expression profiles (GEP)

News

[Align Sequences with BLAST](#)

A new BL2seq functionality has been added to the standard BLAST pages that allows you to align a query against a set of subject sequences.

2008-09-04 12:58:57

[More BLAST news...](#)

Tip of the Day

How to Search Custom Database in Web-Blast Using Entrez Queries.

A powerful feature of the BLAST Web interface is the ability to limit BLAST searches to a subset of any database using a standard Entrez query. Skillful use of Entrez queries allows the equivalent of on-the-fly construction of databases of exact composition.

[More tips...](#)

Nucleotide BLAST: Search... x

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&MEGABLAST=

こちらのブックマークバーにブックマークを追加すると簡単にページにアクセスできます。

その他のブックマーク

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI/BLAST/blastn suite

blastn blastp blastx tblastn tblastx

BLASTH programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Book](#)

Enter Query Sequence

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

Or, upload file [ファイルを選択](#) 選択されていません [?](#)

Job Title [Enter a descriptive title for your BLAST search](#) [?](#)

☐ Blast 2 sequences

Choose Search Set

Database ☒ Human genomic + transcript ☐ Mouse genomic + transcript ☐ Others (nr etc.):
Human genomic plus transcript (Human G+T) [?](#)

Entrez Query Optional [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 Human G+T using Megablast (Optimize for highly similar sequences) ☐ Show results in a new window

Algorithm parameters

[Copyright](#) | [Disclaimer](#) | [Privacy](#) | [Accessibility](#) | [Contact](#) | [Send feedback](#) [NCBI](#) | [BLM](#) | [NBI](#) | [CHHS](#)

上部のテキストボックスに探索したい配列(二本のプライマーの間にNを30文字入れたもの)を入力し①、生物種を Homo Sapiensに限定し②、検索エンジンをblastにした後③、 BLASTボタンをクリックして④探索を開始します。

Nucleotide BLAST: Search... x

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&MEGABLAST=

こちらのブックマーク バーにブックマークを追加すると簡単にページにアクセスできます。

その他のブックマーク

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI/BLAST/blastn suite

blastn blasto blastx tblastn tblastx

BLASTn programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Book](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

GTACCTTGATTTCGTATTC
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GACTCTACTACCTTACCC

Clear Query subrange

From To

Or, upload file

ファイルを選択 選択されていません

Job Title

Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database

☒ Human genomic + transcript ☐ genomic + transcript ☐ Others (nr etc.):

Human genomic plus transcript (Human G+T)

Entrez Query

Optional

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 Human G+T using Blastn (Optimize for somewhat similar sequences)

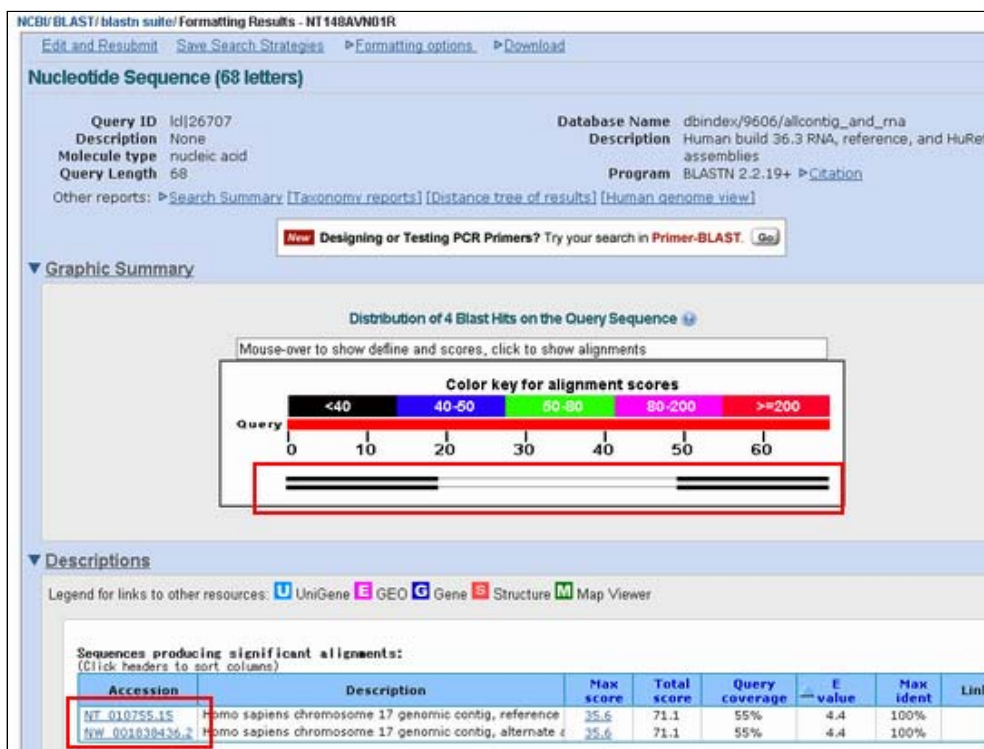
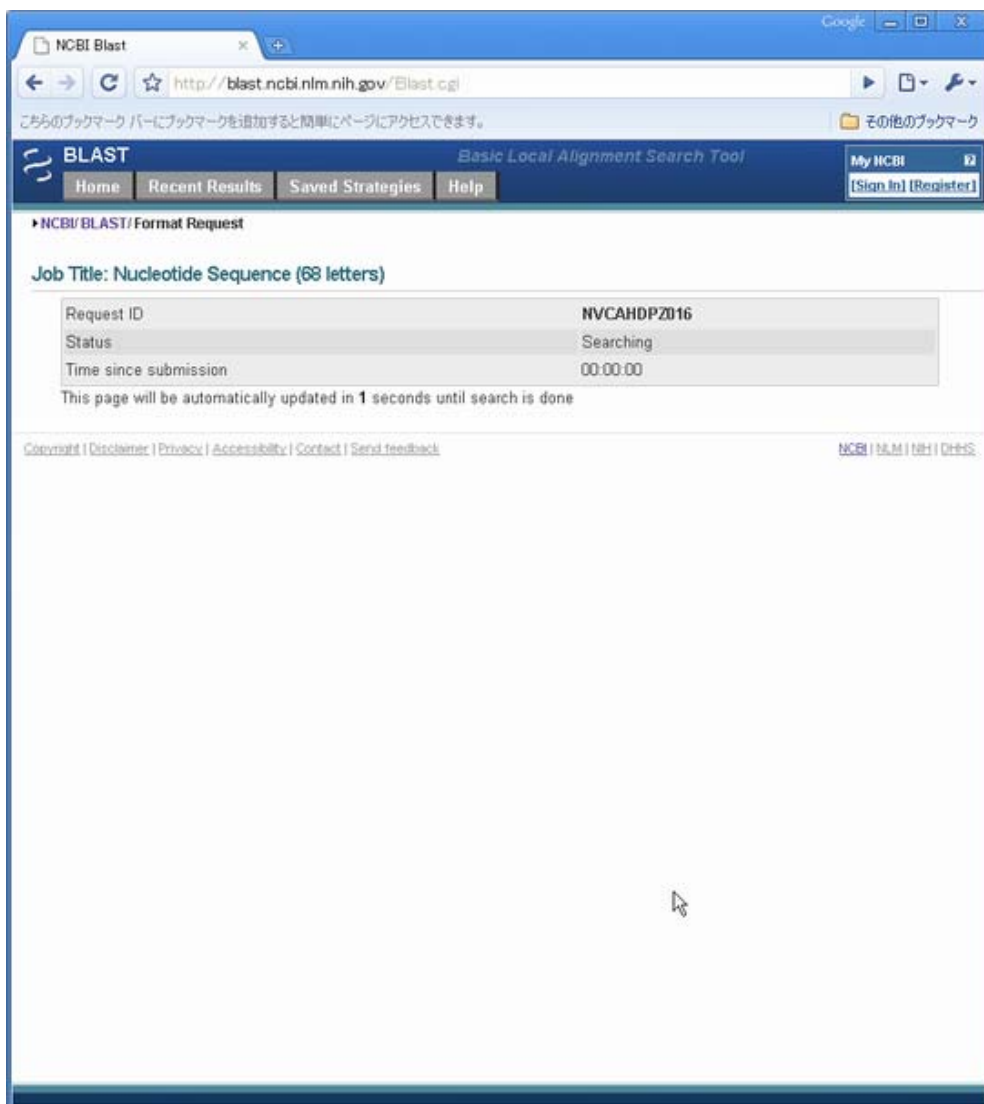
☐ Show results in a new window

Algorithm parameters

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback

NCBI | NLM | NIH | DITIS

すると下図のように進捗表示画面になりしばらく待つと結果画面へと自動的に移行します。



図から2本の配列に対してヒットがあることが確認できます(赤四角)。これらのヒットはいずれも17番染色体のものですが、下のNWで始まるエントリは国際コンソーシアムによるreferenceアセンブルとは異なった個人のゲノムを再読したものであり、実質上は17番染色体の一箇所にヒットしていることになります。

結果画面を下にスクロールしてそれぞれの結果のアライメントを見ると、プライマーのforward/reverse双方の部分がPlusのstrandとMinusのstrandにヒットしていることが確認できます。(forwardのマッピング情報部分が青四角で囲まれた箇所、reverseのマッピング情報が赤四角で囲まれた箇所) これにより、forward/reverseのプライマーが逆向きに正しくマッピングされていることが確認できます。このプライマーにより増幅される領域は、17番染色体の5,001,573bp~5,001,796bpになります。

```
>[ ]ref|NT_010755.15|Hs17_10912 [D] Homo sapiens chromosome 17 genomic cont
Length=5072491

Sort alignments f
E value Score
Query start pos

Features flanking this part of subject sequence:
1164 bp at 5' side: breast cancer 1, early onset isoform BRCA1-delta11
44590 bp at 3' side: similar to 60S acidic ribosomal protein P1

Score = 35.6 bits (38), Expect = 4.4
Identities = 19/19 (100%), Gaps = 0/19 (0%)
Strand=Plus/Plus
Query 50      GACTCTACTACCTTTACCC 68
          |||||
Sbjct 5001573 GACTCTACTACCTTTACCC 5001591

Features flanking this part of subject sequence:
1369 bp at 5' side: breast cancer 1, early onset isoform BRCA1-delta11
44385 bp at 3' side: similar to 60S acidic ribosomal protein P1

Score = 35.6 bits (38), Expect = 4.4
Identities = 19/19 (100%), Gaps = 0/19 (0%)
Strand=Plus/Minus
Query 1      GTACCTTGATTTTCGTATTC 19
          |||||
Sbjct 5001796 GTACCTTGATTTTCGTATTC 5001778
```

■ 解答1-2

blastnを用いた相同性検索

同様にblastnの配列投入画面において課題に示された配列を入力します。ただし、今回はアライメント結果を見た
いので、出力のAlignment Viewオプションの箇所を flat-query anchored with identitiesに変更してBLASTを実
行します。(以前のバージョンでは blast実行時にフォーマットの設定ができましたが、現在のバージョンでは実
行後に設定します。)

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastn suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter accession number, gi, or FASTA sequence Clear

CCAAAGTAGTATGACAAAATTTTAGAGCCTTTAGAGACAAAATCCAGAAATAGTTAT
 CTATCAATACATGGATGATTTGTATGAGGATCTGACTTAGAACTAGGGCAGCATAGAGC
 AAAAATAGAGGAACCTGAGACGACATCTGTTGAGGTGGGGATTTACACACAGACAAAAA
 ACATCAGAAAGAGCCTCCATTCCTTTGGA

Query subrange
 From
 To

Or, upload file 選択されていません

Job Title
 Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database
☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.)
 Nucleotide collection (nr/nt)

Organism
 Optional

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

Entrez Query
 Optional

 Enter an Entrez query to limit search

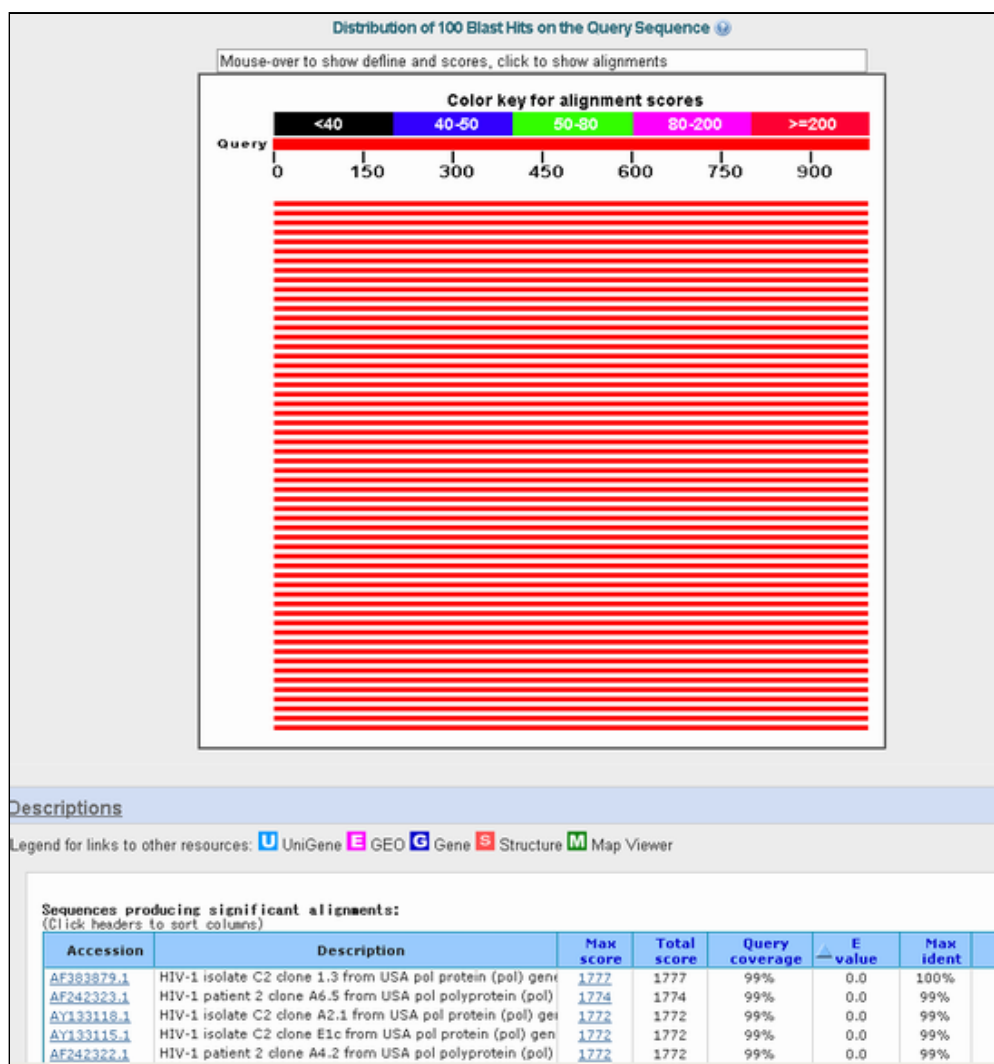
Program Selection

Optimize for
☐ Highly similar sequences (megablast)
☐ More dissimilar sequences (discontiguous megablast)
☒ Somewhat similar sequences (blastn)

Search database nr using Blastn (Optimize for somewhat similar sequences)
☐ Show results in a new window

上部のテキストボックスに探索したい配列を入力し①、データベースをnr/ntにし②、検索エンジンをblastにした後③、BLASTボタンをクリックして④探索を開始します。

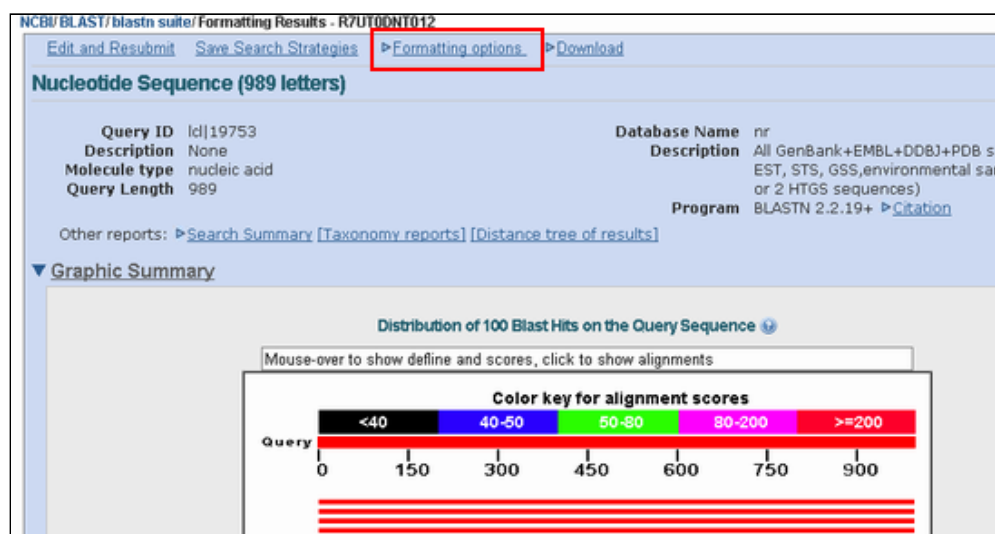
すると、図に示されたようにHIV-1への多くのヒットが認められます。



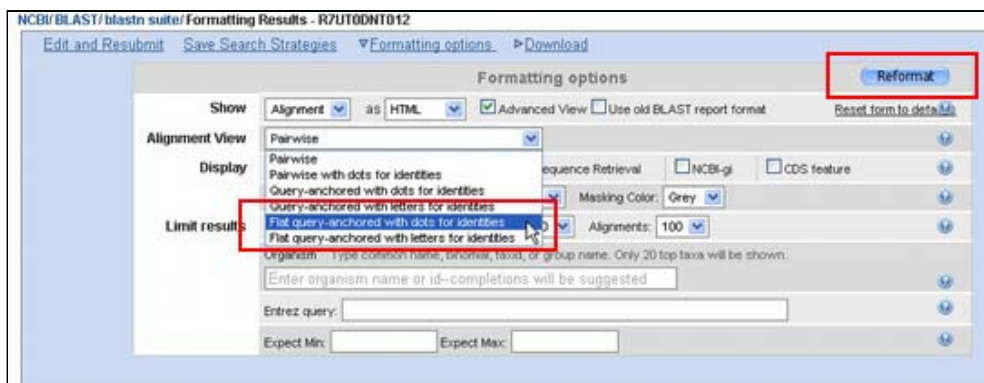
blast検索結果の表示フォーマット変更

続いて、出力フォーマットを変更します。

結果画面の上部の“Formatting options”をクリックしてください（図中赤四角）。



すると以下のように出力フォーマットを指定する画面が表示されますので、Displayプルダウンメニューから“Flat query-anchored with dots for identities”を選択し、右上の“Reformat”ボタンをクリックして再描画してください。



画面を下にスクロールと、blastの結果がマルチプルアライメント様に表示されています。 Query配列と同じ配列は"."で、違うところだけがACGTで示されています。

☐ Query	5	CCTCAAA	CACCTCTTTGGCAACGACCCCTCGTCACAATAAGATAGGGGGGCAACTAAAG	64
☐ AF383879	1			60
☐ AF242323	1			60
☐ AY133118	1			60
☐ AY133115	1			60
☐ AF242322	1			60
☐ AY133113	1			60
☐ AF383881	1			60
☐ AY133117	1			60
☐ AY133114	1			60
☐ AF242321	1			60
☐ AF383880	1	G.		60
☐ AY133116	1		G.....	60
☐ AF251949	72	G.		131
☐ AY835770	2253	G.		2312
☐ AF072959	19	G.		77
☐ AY835779	2253			2312
☐ AY835771	2253	G.		2312
☐ AF072937	19	G.		78
☐ AY032316	1	G.		60
☐ AY835749	2253	G.		2312
☐ AF075719	2276		A.T.....	2335
☐ AF004394	2253		A.....	2312
☐ K02013	1835	G.		1894
☐ M17449	2268		A.T.....	2327
☐ EF363124	2253	G.		2312
☐ DQ487190	1795	G.	A.....A.....	1854
☐ AY471887	1	G.		60
☐ AJ971093	1	G.		60
☐ AF042100	2252	G.		2311
☐ AY964424	1	G.		60
☐ DQ009844	1	G.		60

図から6塩基目にはA/Gの多型があることがわかります。

解答・解説2

■ 解答2-1

blastpを用いた相同性検索

まず、BLASTのトップページから"protein blast"を選択します。

BLAST: Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI/BLAST Home

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

New Designing or Testing PCR Primers? Try your search in **Primer-BLAST**. [Go](#)

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

☐ Human	☐ Oryza sativa	☐ Gallus gallus
☐ Mouse	☐ Bos taurus	☐ Pan troglodytes
☐ Rat	☐ Danio rerio	☐ Microbes
☐ Arabidopsis thaliana	☐ Drosophila melanogaster	☐ Apis mellifera

Basic BLAST

Choose a BLAST program to run:

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

Specialized BLAST

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

- ☐ Make specific primers with [Primer-BLAST](#)
- ☐ Search [trace archives](#)
- ☐ Find [conserved domains](#) in your sequence (cds)
- ☐ Find sequences with similar [conserved domain architecture](#) (cdart)
- ☐ Search sequences that have [gene expression profiles](#) (GEO)
- ☐ Search [immunoglobulins](#) (IgBLAST)
- ☐ Search for [SNPs](#) (snp)
- ☐ Screen sequence for [vector contamination](#) (vecscreen)
- ☐ [Align two sequences using BLAST \(h2seq\)](#)

Your Recent Results

News

[Align Sequences with BLAST](#)

A new BL2seq functionality has been added to the standard BLAST pages that allows you to align a query against a set of subject sequences.
2008-09-04 12:56:52

[More BLAST news...](#)

Tip of the Day

How to Search Custom Databases in Web-Blast Using Entrez Queries.

A powerful feature of the BLAST Web interface is the ability to limit BLAST searches to a subset of any database using a standard Entrez query. Skillful use of Entrez queries allows the equivalent of on-the-fly construction of databases of exact composition.

[More tips...](#)

問い合わせ配列を入力①、データベースからswissprotを選択し②、検索エンジンとしてblastpを選択後③、blastボタンを押して④検索を実行します。

NCBI/BLAST/blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTP programs search protein databases using a protein query. [more...](#)

Enter accession number, gi, or FASTA sequence Clear Query subrange

From

To

Or, upload file 選択されていません

Job Title

Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database ②

Organism Optional

Enter organism name or id-completions will be suggested

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

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Algorithm ☒ blastp (protein-protein BLAST) ③

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

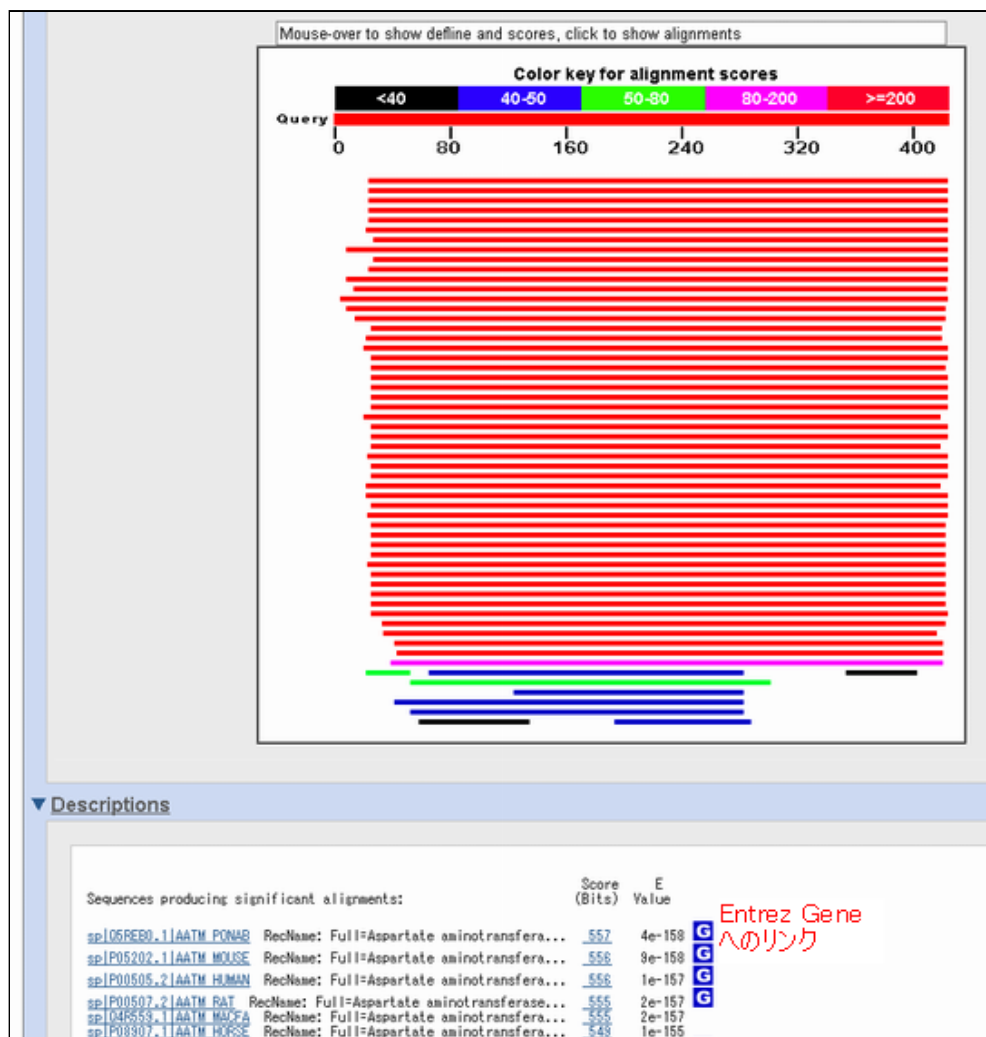
④

Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

[Algorithm parameters](#) Note: Parameter values that differ from

しばらくして得られる結果を見ると、その上位はいずれもAspartate aminotransferaseで、Pongo, Mouse, Rat, Pigなどさまざまな生物のタンパク質にヒットしていることがわかります。



ヒットした配列の詳細な情報は結果の右横に“G”マークのついている箇所をクリックすることでEntrez Geneのページへ移動することで確認できます。

NCBI Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals

Search Gene for 75042478[PUI] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Your browser version may not work well with NCBI's Web applications. More information [here...](#)

Display Full Report Show 20 Sort by Relevance Send to

All: 1 Current Only: 1 Genes Genomes: 0 SNP GeneView: 0

☐ 1: GOT2 glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2) [*Pongo abelii*]

GeneID: 100171753 updated 16-Dec-2008

Summary

Gene description glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)

Gene type protein coding

RefSeq status PROVISIONAL

Organism [Pongo abelii](#)

Lineage Eukaryota; Metazoa; Chordata; Oraniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Pongo

Also known as GOT2; DKFZp459L1636;

Bibliography

GeneRIFs: Gene References Into Function [What's a GeneRIF?](#)

Submit: [New GeneRIF](#)

General protein information

Preferred Names
glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)

Names
glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)
Aspartate aminotransferase, mitochondrial

NP_001124888.1
EC [2.6.1.1](#)

これらの情報から問い合わせ配列は ショウジョウバエのAspartate aminotransferaseであることが類推されます。

■ 解答2-2

[blastpを用いた相同性検索2](#)

同様にblastpのページに課題の問い合わせ配列を入力①、データベースからswissprotを選択し②、検索エンジンとしてblastpを選択後③、blastボタンを押して④検索を実行します。

NCBI/BLAST/blastp suite

blastn **blastp** blastx tblastn tblastx

Enter Query Sequence BLASTP programs search protein databases using a protein query. [more...](#)

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

PYFQKNGVGGFARSMPTASAVDLVGRKLGKEVFEVPTGMYFGNLMADAGRLCLCGEESFGTGSN
 HIREKGGI WAVLAWI SVMRHTGKGI EDILKQHSVYGRNYFTRYDYEECASDPCNEMVATMEKT
 ITAPEFVGKSYSSGGKTYKVKEDNFSTDPYDKSVATKQGLRI VFEDGSRI VVRLSGTGSSGA
 TVRLYIDSYEKENVLGQASVMLKPLIDIALEISQLPKFTGRNAPTIVT

From

To

Or, upload file 選択されていません

Job Title

Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database

Organism

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

Entrez Query

Optional

Program Selection

Algorithm ☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

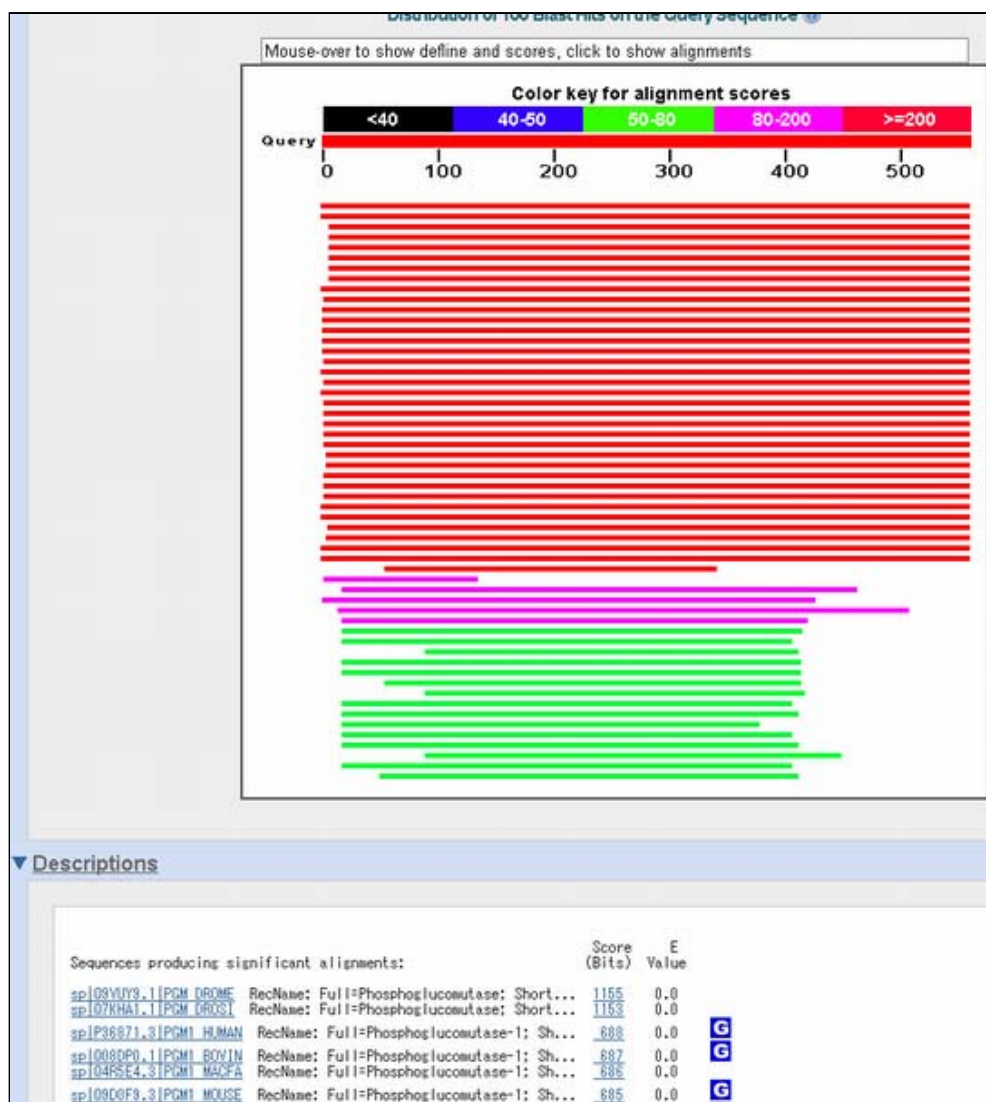
☐ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

しばらくして得られる結果を見ると、その上位はいずれもPhosphoglucomutase で、ショウジョウバエ、Humanなどさまざまな生物のタンパク質にヒットしています。これらの情報から問い合わせ配列は ショウジョウバエのPhosphoglucomutaseであることが類推されます。



解答・解説3

■ 解答3-1 _____

blastxを用いた相同性検索

与えられた核酸配列のどこにフレームシフトが入っているかを調べるためにblastxを用います。BLASTメニューのページからblastx(核酸問い合わせ配列 vs アミノ酸データベース)を選択します。


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
 Designing or Testing PCR Primers? Try your search in **Primer-BLAST**.
 [Go](#)

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

☐ [Human](#)
☐ [Mouse](#)
☐ [Rat](#)
☐ [Arabidopsis thaliana](#)

☐ [Oryza sativa](#)
☐ [Bos taurus](#)
☐ [Danio rerio](#)
☐ [Drosophila melanogaster](#)

☐ [Gallus gallus](#)
☐ [Pan troglodytes](#)
☐ [Microbes](#)
☐ [Apis mellifera](#)

Basic BLAST

Choose a BLAST program to run.

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

Specialized BLAST

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

- ☐ Make specific primers with [Primer-BLAST](#)
- ☐ Search [trace archives](#)
- ☐ Find [conserved domains](#) in your sequence (cds)
- ☐ Find sequences with similar [conserved domain architecture](#) (cdart)
- ☐ Search sequences that have [gene expression profiles](#) (GEO)
- ☐ Search [immunoglobulins](#) (IgBLAST)
- ☐ Search for [SNPs](#) (snp)
- ☐ Screen sequence for [vector contamination](#) (vecscreen)
- ☐ [Align](#) two sequences using BLAST (bl2seq)
- ☐ Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay

問い合わせ配列として与えられた配列を入力し①、そのほかはデフォルトのままでBLASTボタンをクリックする②ことでホモロジー検索を実施します。

結果を見ると問い合わせ配列はenvelope glycoproteinであることが推察され、下にスクロールしてそのアライメントを見るとアライメントが二つに分断されフレームがずれていることがわかります。二つ目のアライメントが+2のフレームに対するもので、問い合わせ配列の268bpで終わっており（赤く囲まれた箇所を参照）、一つ目のアライメントが+1のフレームに対するもので、問い合わせ配列の268bpから始まっている（青く囲まれた箇所を参照）ことから、その分断箇所は問い合わせ配列の268塩基目付近であることが確認できます。

```
>gb|AAL71628.1| envelope glycoprotein [Human immunodeficiency virus 1]
gb|AAL71647.1| envelope glycoprotein [Human immunodeficiency virus 1]
Length=201

Score = 232 bits (591), Expect(2) = 2e-107
Identities = 110/112 (98%), Positives = 110/112 (98%), Gaps = 1/112 (0%)
Frame = +1
Query 268 TIAFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTWPTNK-KSTNKTGTITLPCRIKQ 444
Sbjct 90 TIAFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTWPTN KSTNKTGTITLPCRIKQ 149

Query 445 IINRWQEVGKAMYAPPIKGQIRCSSNITGIFLTRDGGNASDETETFRPGGKN 600
Sbjct 150 IINRWQEVGKAMYAPPIKGQIRCSSNITGIFLTRDGGNASDETETFRPGGKN 201

Score = 181 bits (460), Expect(2) = 2e-107
Identities = 89/89 (100%), Positives = 89/89 (100%), Gaps = 0/89 (0%)
Frame = +2
Query 2 EEDIVIRSENFNNAKTIIIVQLKESIKINCTRPNNNTRKSIPIATGGAIYATGDIIGDIR 181
Sbjct 1 EEDIVIRSENFNNAKTIIIVQLKESIKINCTRPNNNTRKSIPIATGGAIYATGDIIGDIR 60

Query 182 QAHCNLSRDQWDNTLSQLVTKLREQFGNK 268
Sbjct 61 QAHCNLSRDQWDNTLSQLVTKLREQFGNK 89
```

■ 解答3-2

[blastxを用いた相同性検索2](#)

blastx(核酸問い合わせ配列 vs アミノ酸データベース)のページから、問い合わせ配列として与えられた配列を入

カシ①、そのほかはデフォルトのままでBLATボタンをクリックする②ことでホモロジー検索を実施します。

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastx

blastn blastp **blastx** tblastn tblastx

Enter Query Sequence BLASTX search protein databases using a translated nucleotide query. [more...](#)

Enter accession number, gi, or FASTA sequence [Clear](#)

GGGCT
AATTTTAACAAGAGATGGTGGGAACGAACTGATGGGAACACACTGATGGGAATGAGACCTTCAGACCTGG
AGGAG
GGAAATGAGGGGACAAATTGGA

Query subrange

From

To

①

Or, upload file 選択されていません

Genetic code

Job Title

Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database

Organism

Enter organism name or id--completions will be suggested

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

Entrez Query

Enter an Entrez query to limit search

BLAST

② Search database nr using Blastx (search protein databases using a translated nucleotide query)

☐ Show results in a new window

結果を見ると問い合わせ配列はenvelope glycoproteinであることが推察され、下にスクロールしてそのアライメントを見るとアライメントが二つに分断されフレームがずれていることがわかります。二つ目のアライメントが+1のフレームに対するもので、問い合わせ配列の564bpで終わっており（赤く囲まれた箇所を参照）、一つ目のアライメントが+21のフレームに対するもので、問い合わせ配列の566bpから始まっている（青く囲まれた箇所を参照）ことから、その分断箇所は問い合わせ配列の565塩基目付近であることが確認できます。

```

>gb|AAL31985.1| envelope glycoprotein [Human immunodeficiency virus type 1]
Length=514

Score = 559 bits (1440), Expect(2) = 0.0
Identities = 301/306 (98%), Positives = 304/306 (99%), Gaps = 0/306 (0%)
Frame = +2
Query 566 DIVPIDNDRNDSTSYRLSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNKTFSGT 745
Sbjct 185 DIVPIDNDRNDSTSYRLSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNKTFSGT 244

Query 746 GPCTNVSTVQCTHGIRPIVSTQLLNGSLAEEGIVIRYENITDNAKSIIQLNETVQINC 925
Sbjct 245 GPCTNVSTVQCTHGIRPIVSTQLLNGSLAEEGIVIRYENITDNAKSIIQLNETVQINC 304

Query 926 TRPNNTRKSPIGPGRAFYATGDIIGDIRKAYCNISGAKWNNLTKRIAYKLKEQFPNKT 1105
Sbjct 305 TRPNNTRKSPIGPGRAFYATGDIIGDIRKAYCNISGAKWNNLTKRIAYKLKEQFPNKT 364

Query 1106 IVFKPSSGGDPEIYMHSFNCRGEFFYCNTTKLFDSSWDntnlkntwnntwnknsIILPC 1285
Sbjct 365 IVFKPSSGGDPEIYMHSFNCRGEFFYCNTTKLFDSSWDNTNLNKTWNNTWNKNSIILPC 424

Query 1286 RIKQIINMWQEVGKAMYAPPIEGPLYCLSNITGLILTRdggnetdgnntdgnnetFRPGGG 1465
Sbjct 425 RIKQIINMWQEVGKAMYAPPIEGPLYCLSNITGLILTRDGGNETDGNNT G ETRPGGG 484

Query 1466 NMRDNW 1483
Sbjct 485 DMRDNW 490

Score = 310 bits (793), Expect(2) = 0.0
Identities = 165/188 (87%), Positives = 172/188 (91%), Gaps = 4/188 (2%)
Frame = +1
Query 1 MRVKEKYQHLWRWGTMLLGLLMIRSAADQLWVTYYYGVVWKEATTTFCASDAKAYDTE 180
Sbjct 1 MRVKEKYQHLWRWGTMLLGLLMI SAADQLWVTYYYGVVWKEATTT FCASDAKAYDTE 60

Query 181 VHNWVWATHACVPTDNPQEVVLANVAENFNMWDMNMVEQMHEIISLWDQSLKPCVKLTP 360
Sbjct 61 VHNWVWATHACVPTDNPQEVVLANV ENFNMWDMNMVEQMHEIISLWDQSLKPCVKLTP 120

Query 361 LCVTLNCTDK InnktipnntstipptivtpisnssmtGEGEIKNCSFNITTAIRDKVQR 540
Sbjct 121 LCVTLNCTD++NNTKTTPNNT TT T +S+ GEIKNCSFNITTAIRDKVQ+ 176

Query 541 GYALFSKL 564
Sbjct 177 GYALFNKL 184

```

解答・解説4

■ 解答4-1

blastpを用いた短い配列に対する同源性検索

以前のblastには“Search for short, nearly exact matches”という特別に用意されたサービスがあったのですが、現在はありません。従って、通常のblast探索のパラメータを変更することで、同様の結果が得られるように対応した回答を紹介します。

BLASTメニューのページからblastpを選択し、問い合わせ配列として“einstein”と入力し①、データベースとして“nr”を選択してください②。

NCBI/BLAST/blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTP programs search protein databases using a protein query. [more...](#)

Enter accession number, gi, or FASTA sequence [Clear](#)

einstein

Query subrange

From

To

Or, upload file ファイルを選択 選択されていません

Job Title

Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional

Enter organism name or id; completions will be shown

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

Entrez Query Optional

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)

Choose a BLAST algorithm

BLAST Search database nr using Blastp (protein-protein BLAST)

☐ Show results in a new window

Algorithm parameters

Note: Parameter values that di

次に、画面下部の青四角で囲まれた“Algorithm parameters”をクリックして、詳細なパラメータ設定画面を表示させてください。

BLAST Search database nr using Blastp (protein-protein BLAST)
☐ Show results in a new window

▼ Algorithm parameters Note: Parameter highlighted in yellow

General Parameters

Max target sequences: 1000 ①
Select the maximum number of aligned sequences to display

Short queries: ☒ Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 2 ②

Scoring Parameters

Matrix: PAM30 ③

Gap Costs: Existence: 9 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: ☐ Low complexity regions

Mask: ☐ Mask for lookup table only
☐ Mask lower case letters

データベースへのヒットが多いことを想定して、“Max target sequences”（何本までのヒットを表示するか）を1000に①、“Word size”をより短い2に②、“Matrix”を PAM30という短い問い合わせ配列向きのものに③変更します。その後、blastボタンを押してホモロジー検索を実行します。

しばらく待つと結果が表示されます。極めて多くのヒットが認められます。トップヒットのアライメントを見てみましょう。

▼ Descriptions

Sequences producing significant alignments:

		Score (Bits)	E Value
ref YP_849739.1 	acetylglutamate kinase [Listeria welshimeri ...	28.5	221
ref XP_001802616.1 	hypothetical protein SNOG_12333 [Phaeosph...	28.3	221
ref YP_001250878.1 	hypothetical protein LPC_1378 [Legionella...	28.1	237
ref XP_446631.1 	unnamed protein product [Candida glabrata] >...	28.1	237
ref XP_727987.1 	hypothetical protein [Plasmodium yoelii voel...	28.1	237

Score部分に書かれた数字をクリックしてみてください。該当するアライメントにジャンプします。

```

>[ref|YP_849799.1|] G acetylglutamate kinase [Listeria welshimeri serovar 6b str. SLCC5334]
sp|A0AJ38.1|ARGB_LISW6 G RecName: Full=Acetylglutamate kinase; AltName: Full=NAG kinase;
Short=AGK; AltName: Full=N-acetyl-L-glutamate 5-phosphotransferase
emb|CAK21020.1| G argB [Listeria welshimeri serovar 6b str. SLCC5334]
Length=250

GENE ID: 4464962 argB | argB [Listeria welshimeri serovar 6b str. SLCC5334]
(10 or fewer PubMed links)

Score = 26.5 bits (55), Expect = 221
Identities = 7/8 (87%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query 1  EINSTein 8
Sbjct 193 EINTTEIN 200

>[ref|XP_001802616.1|] G hypothetical protein SNOG_12393 [Phaeosphaeria nodorum SN15]
gb|EAT80206.1| G hypothetical protein SNOG_12393 [Phaeosphaeria nodorum SN15]
Length=536

GENE ID: 5979525 SNOG_12393 | hypothetical protein [Phaeosphaeria nodorum SN15]

Score = 26.5 bits (55), Expect = 221
Identities = 7/8 (87%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query 1  EINSTein 8
Sbjct 377 EINSTein+ 884

```

アライメントより明らかなように、完全一致するものは1本も得られなかったことがわかります。

ちなみに現在のblastp検索では、結果の最上部を見ると分かるように、問い合わせ配列が短い場合には自動的にそれに合わせたパラメータが設定されてホモロジー検索が実行されています。

NCBI/BLAST/blastp suite/ Formatting Results - R88N1CYW01R

ⓘ Your search parameters were adjusted to search for a short input sequence.

[Edit and Resubmit](#) [Save Search Strategies](#) [▶Formatting options](#) [▶Download](#)

Protein Sequence (8 letters)

■ 解答4-2

blastpを用いた短い配列に対する相同性検索2

以前のblastには“Search for short, nearly exact matches”という特別に用意されたサービスがあったのですが、現在はありません。従って、通常のblast探索のパラメータを変更することで、同様の結果が得られるように対応した回答を紹介します。

4-1と同様の操作を‘valentine’配列に変えて実行してみてください。BLASTメニューのページからblastpを選択し、問い合わせ配列として‘valentine’と入力し①、データベースとして‘nr’を選択してください②。

NCBI/BLAST/blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTP programs search protein databases using a protein query. [more...](#)

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

valentein

Or, upload file ファイルを選択 選択されていません

Job Title Enter a descriptive title for your BLAST search

☐ Blast 2 sequences

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional Enter organism name or id--completions will be suggested Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Entrez Query Optional 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) Choose a BLAST algorithm

BLAST Search database nr using Blastp (protein-protein BLAST) ☐ Show results in a new window

Algorithm parameters **Note: Parameter values that differ from defaults are shown in yellow.**

次に、画面下部の青四角で囲まれた“Algorithm parameters”をクリックして、詳細なパラメータ設定画面を表示させてください。


```

>[kb|AA053202.1] similar to Plasmodium falciparum. Hypothetical protein [Dictyostelium
discoideum]
Length=1893

Score = 26.9 bits (56), Expect = 186
Identities = 9/11 (81%), Positives = 9/11 (81%), Gaps = 2/11 (18%)

Query 1  VALENTE--IN 9
        VALENTE  IN
Sbjct 809  VALENTEODIN 819

>[ref|NP_757534.1] [G] 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Mycoplasma penetrans
HF-2]
sp|Q8EWQ6.1|DXR MYCPE RecName: Full=1-deoxy-D-xylulose 5-phosphate reductoisomerase;
Short=DXP reductoisomerase; AltName: Full=1-deoxyxylulose-5-phosphate
reductoisomerase; AltName: Full=2-C-methyl-D-erythritol
4-phosphate synthase
dbj|BAC43938.1| [G] 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Mycoplasma penetrans
HF-2]
Length=377

GENE ID: 1040787 MYPE1470 | 1-deoxy-D-xylulose 5-phosphate reductoisomerase
[Mycoplasma penetrans HF-2] (10 or fewer PubMed links)

Score = 26.9 bits (56), Expect = 186
Identities = 7/9 (77%), Positives = 9/9 (100%), Gaps = 0/9 (0%)

Query 1  VALENTEIN 9
        +ALENTE+N
Sbjct 343  IALENTEVN 351

```

アライメントより明らかなように、完全一致するものは1本も得られなかったことがわかります。

解答・解説5

■ 解答5-1

mRNAを用いたホモログ探索

まずは、mRNA配列を用いてホモログ候補をblastnの利用によりマウスゲノム配列中から探してみましょう。

►NCBI/ BLAST Home

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

New Designing or Testing PCR Primers? Try your search in **Primer-BLAST**.

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

☐ Human	☐ <i>Oryza sativa</i>	☐ <i>Gallus gallus</i>
☒ Mouse	☐ <i>Bos taurus</i>	☐ <i>Pan troglodytes</i>
☐ Rat	☐ <i>Danio rerio</i>	☐ <i>Microbes</i>
☐ <i>Arabidopsis thaliana</i>	☐ <i>Drosophila melanogaster</i>	☐ <i>Apis mellifera</i>

Basic BLAST

Choose a BLAST program to run.

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

BLASTメニューのページから“BLAST Assembled Genomes”⇒“Mouse”を選択します。

BLAST
Overview
FAQs
News
Manual
References
Retrieve results
Genome Project

BLAST Mouse Sequences.

Enter an accession, gi, or a sequence in FASTA format:
NM_008268 ①

Or, choose a file to upload
 選択されていません

Set subsequence: (optional)
From: To:

Database: genome (reference only) ② 316 sequences

Program: BLASTN: Compare nucleotide sequences ③

Optional parameters

Expect	Filter	Descriptions	Alignments
0.01	default	100	100

Advanced options:

④

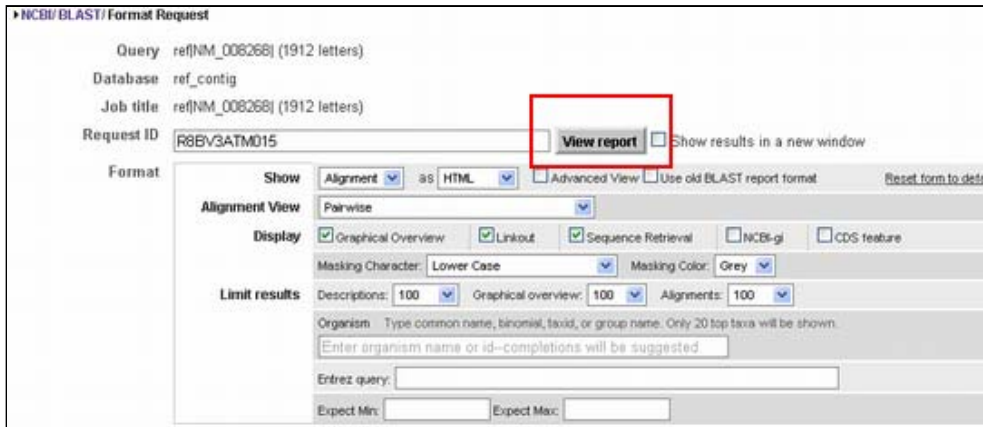
queryのテキストボックスに“NM_008268”と入力し①、Databaseは“genome (reference only)”を②、Programは“blastn”を選択して③ “Begin Search”ボタンを押すことで④、ホモロジー検索を実施します。

このようにNCBIで提供しているblast検索では問い合わせ配列として実際の配列以外にアクセッション番号を入力することも可能です。また、データベースとしてgenomeに二種類(all genomes, reference only)が存在します

が、前者では公的機関が公開したゲノムアセンブリの他にcelera社が公開したゲノムデータなどが含まれており、紛らわしくなるため後者を選択しました。

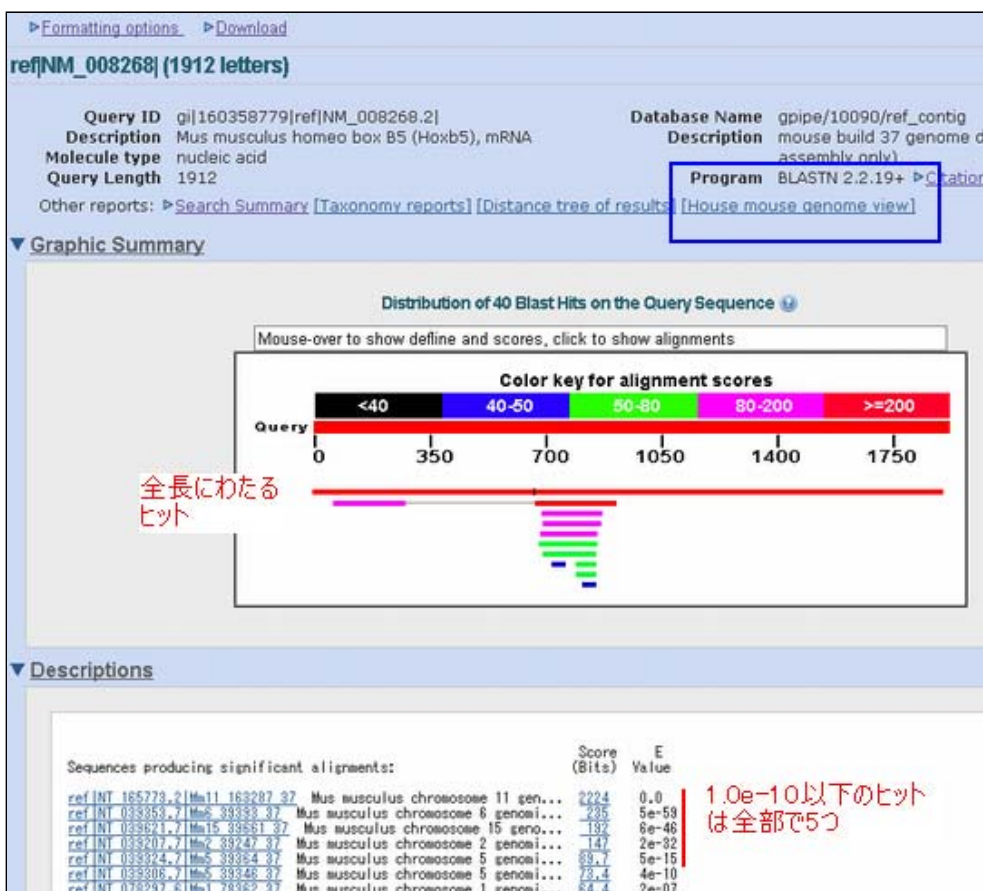
核酸 vs 核酸の検索が実行できるものとしてはmegablastもありますが、megablastでは高速に探索が実行できる反面、問い合わせ配列と類似度の高い配列しか探索することができません。そのため、この課題ではblastnを選択しています。

実行すると、以下のような表示方法を指定する画面へと遷移します。ここでは、デフォルトのまま“View report”ボタンをクリックします。

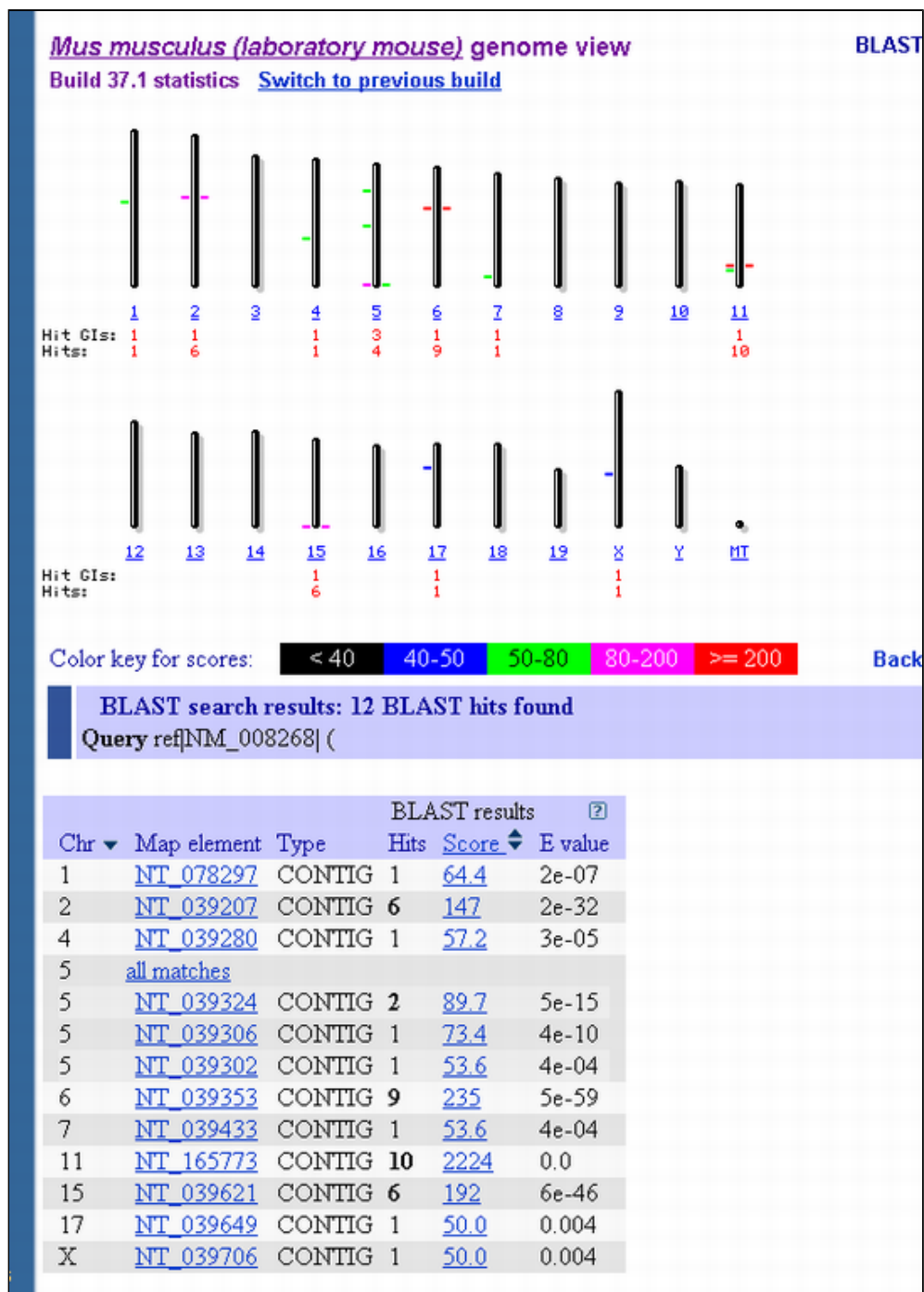


The image shows the NCBI/BLAST/Format Request form. The 'Query' field contains 'refNM_008268| (1912 letters)'. The 'Database' is 'ref_contig'. The 'Job title' is 'refNM_008268| (1912 letters)'. The 'Request ID' is 'R8BV3ATMD15'. The 'View report' button is highlighted with a red box. The 'Format' section shows 'Show Alignment' as 'HTML' and 'Advanced View' as 'Use old BLAST report format'. The 'Alignment View' section shows 'Pairwise' and 'Display' options like 'Graphical Overview', 'Linkout', 'Sequence Retrieval', 'NCBI-gi', and 'CDS feature'. The 'Limit results' section shows 'Descriptions: 100', 'Graphical overview: 100', and 'Alignments: 100'.

結果としては、下図に示したようにNM_008268が実際にゲノム上に存在する箇所(11番染色体)に一番高いスコアで全長にわたってヒットし、それ以外に全部で11箇所、E-value:1.0e-10以下に絞ると問い合わせ配列の一部が4箇所にヒットしていることが確認できます。



また、画面上部の“House mouse genome view”（青四角）リンクをクリックするとイデオグラム上でこれらのヒット位置の概要を見ることができます。



アミノ酸配列を用いたホモログ探索

次に、アミノ酸配列を用いたホモログ探索を行ってみましょう。

マウスに特化したblast検索画面から、問い合わせ配列としてアミノ酸配列のアクセッション番号である“NP_032294”を入力し①、Databaseは“genome (reference only)”を②、Programは“tblastn”を選択して③ “Begin Search”ボタンを押すことで④、ホモロジー検索を実施します。

tblastnではアミノ酸配列を入力とし、核酸データベースに対して検索が可能です。この際データベース側は6フレームすべてをアミノ酸に変換し、アミノ酸 vs アミノ酸の検索が実行されます。

BLAST Mouse Sequences.

Enter an accession, gi, or a sequence in FASTA format:

NP_032294

Or, choose a file to upload

ファイルを選択 選択されていません

Set subsequence: (optional)

From:

To:

Database:

genome (reference only)

316 sequences

Program:

TBLASTN: Compare a protein sequence against a nucleotide database

Optional parameters

Expect

Filter

Descriptions

Alignments

0.01

default

100

100

Advanced options:

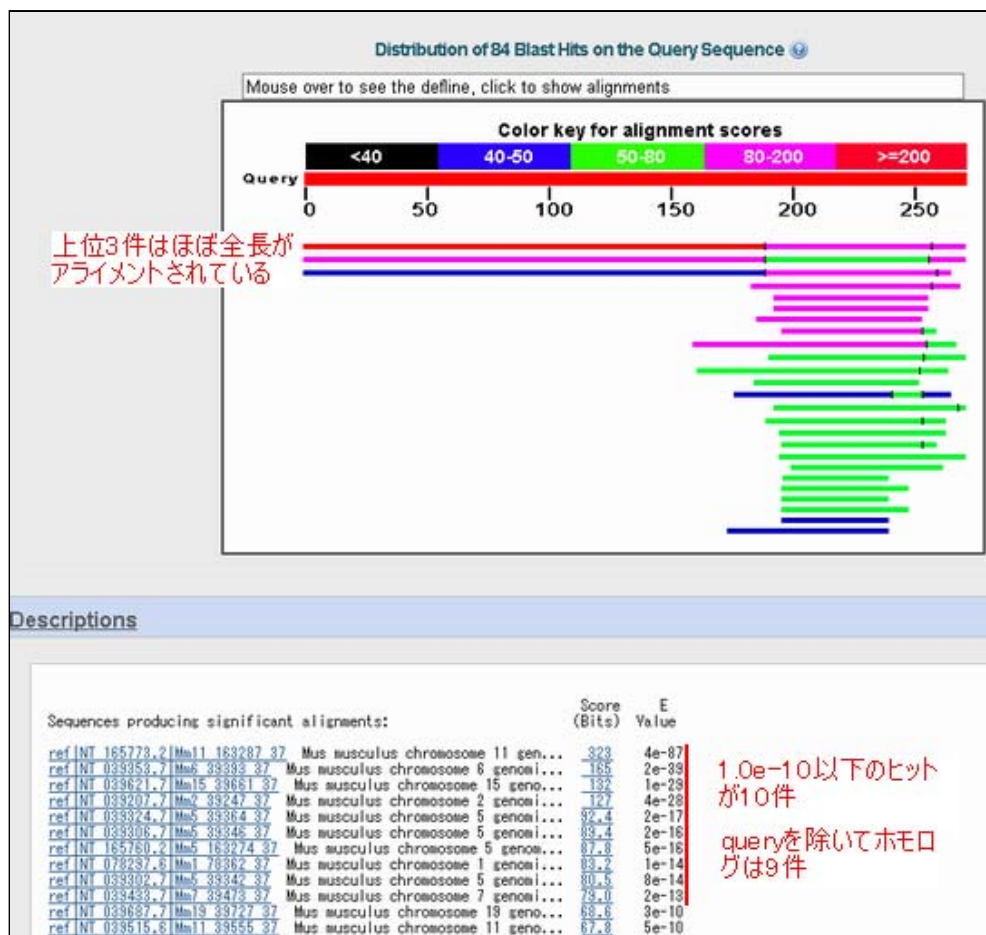
Begin Search

Clear Input

Get the URL with preset values ?

Get URL

今回は以下に示すように非常に多くのヒットが見られます。核酸同士の比較よりもそれをアミノ酸に変換し比較するほうが感度が高いことがわかります。mRNAとの比較では一部しかヒットが認められなかった2件（6, 15番染色体へのヒット）もほぼ全長にわたって類似していることが確認できます。課題にあったホモログの数ですが、ホモログの定義をすることは難しいですが、例えば閾値をE-value : $1.0e-10$ とすると 9個となります ($2e-39 \sim 2e-13$)。



■ 解答5-2

mRNAを用いたホモログ探索2

全く同様にBLASTメニューのページからmouseに特化したblastを選択し、“AY013770”を入力とし①、Databaseは“genome (reference only)”を②、Programは“blastn”を③選択して “Begin Search”ボタンを押すことで④、ホモロジー検索を実施します。

BLAST Mouse Sequences.

Enter an accession, gi, or a sequence in FASTA format:

AY013770 ①

Or, choose a file to upload

ファイルを選択 選択されていません

Set subsequence: (optional)

From: To:

Database: genome (reference only) ② 316 sequences ③

Program: BLASTN: Compare nucleotide sequences ④

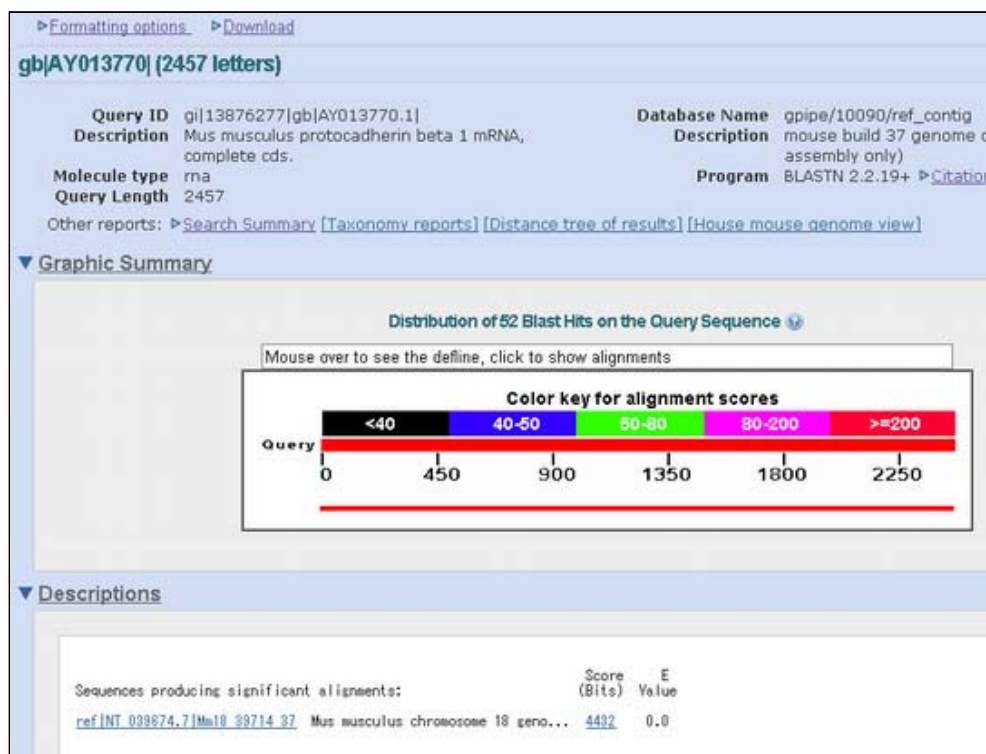
Optional parameters

Expect: 0.01 Filter: default Descriptions: 100 Alignments: 100

Advanced options:

Begin Search Clear Input

AY013770を入力とした例では、この配列が存在する18番染色体への完全なヒットが1件見られるのみでホモログは検出されませんでした。



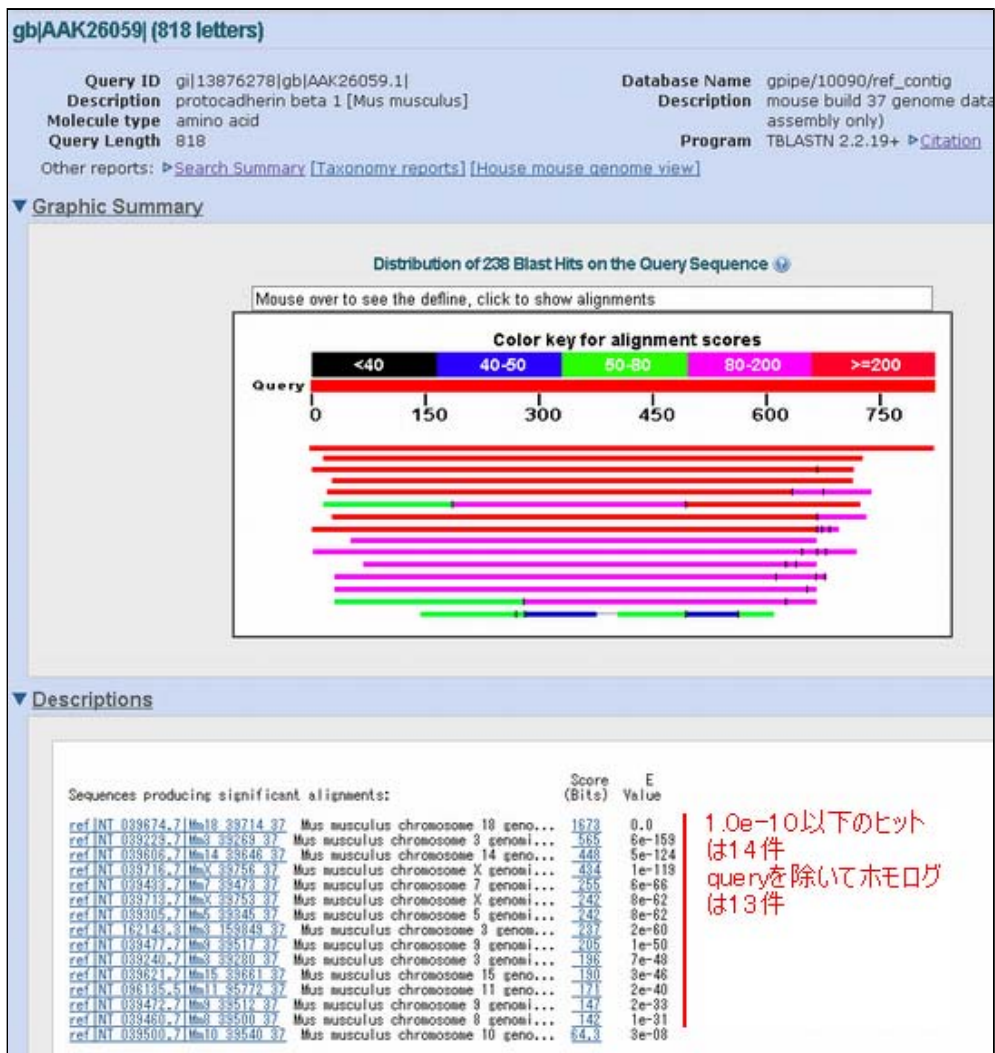
アミノ酸配列を用いたホモログ探索2

引き続いてアミノ酸配列のアクセッション番号である“AAK26059”を入力し①、 Databaseは“genome (reference

only)”を②、Programは”tblastn”を選択して③ “Begin Search”ボタンを押すことで④、ホモロジー検索を実施してみましょう。

The screenshot shows the BLAST Mouse Sequences web interface. The title is "BLAST Mouse Sequences." Below the title, there are two main input sections. The first section is for entering a sequence, with a radio button selected for "Enter an accession, gi, or a sequence in FASTA format:". A text box contains the accession "AAK26059", which is highlighted with a red box and a circled 1. Below this is a radio button for "Or, choose a file to upload" with a "ファイルを選択" button and the text "選択されていません". The second section is for setting the database and program. The "Database:" dropdown is set to "genome (reference only)", highlighted with a red box and a circled 2. To its right, it says "316 sequences" with a circled 3. The "Program:" dropdown is set to "TBLASTN: Compare a protein sequence against a nucleotide database", also highlighted with a red box. Below these are "Optional parameters" including "Expect" (0.01), "Filter" (default), "Descriptions" (100), and "Alignments" (100), each in a dropdown menu. There is also an "Advanced options:" text box. At the bottom, there are two buttons: "Begin Search" (highlighted with a red box and a circled 4) and "Clear Input".

核酸同士の比較では見つけれなかったホモログが数多く認められます。例えば閾値をE-value : 1.0e-10とすると13個となります(6e-159～1e-31)。



解答・解説6

■ 解答6-1

BLASTメニューのページからAlign two sequences using BLAST(b12seq)を選択します（赤四角）。

► NCBI/BLAST Home

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

New Designing or Testing PCR Primers? Try your search in **Primer-BLAST**. [Go](#)

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#)

- ▣ [Human](#)
- ▣ [Mouse](#)
- ▣ [Rat](#)
- ▣ [Arabidopsis thaliana](#)
- ▣ [Oryza sativa](#)
- ▣ [Bos taurus](#)
- ▣ [Danio rerio](#)
- ▣ [Drosophila melanogaster](#)
- ▣ [Gallus gallus](#)
- ▣ [Pan troglodytes](#)
- ▣ [Microbes](#)
- ▣ [Apis mellifera](#)

Basic BLAST

Choose a BLAST program to run.

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

Specialized BLAST

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

- ▣ Make specific primers with [Primer-BLAST](#)
- ▣ Search [trace archives](#)
- ▣ Find [conserved domains](#) in your sequence (cds)
- ▣ Find sequences with similar [conserved domain architecture](#) (cdart)
- ▣ Search sequences that have [gene expression profiles](#) (GEO)
- ▣ Search [immunoglobulins](#) (IgBLAST)
- ▣ Search for [SNPs](#) (snp)
- ▣ Screen sequence for [vector contamination](#) (vecscreen)
- ▣ [Align](#) two sequences using BLAST (b2seq)
- ▣ Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay

続いて配列投入画面でSequence1, 2にそれぞれ課題で示された配列を入力し、そのほかはデフォルトのパラメータのままでホモロジー検索を実施します。BLAST 2 sequencesは、blastを用いて2本の配列のどの部分類似しているかを高速に検索する目的に特化したものです。

Google

http://blast.ncbi.nlm.nih.gov/bl2seq/wblast2.cgi

こちらのブックマーク バーにブックマークを追加すると簡単にページにアクセスできます。

その他のブックマーク

NCBI Entrez BLAST 2 sequences BLAST Example Help

BLAST 2 SEQUENCES

This functionality is now available in the main BLAST pages and this link will be removed in the near future. Step-by-step [instructions](#) for use are available.

This tool produces the alignment of two given sequences using BLAST engine for local alignment. The stand-alone executable for blasting two sequences (bl2seq) can be retrieved from [NCBI ftp site](#)

Reference: Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250

Program: **blastn** Matrix: **Not Applicable**

Parameters used in BLASTN program only:
 Reward for a match: **1** Penalty for a mismatch: **-2**

☐ Use [Mega BLAST](#) Strand option: **Both strands** View option: **Standard**
 Masking character option: **X for protein, n for nucleotide** Masking color option: **Black**
☐ Show CDS translation

Open gap: **5** and extension gap: **2** penalties
 gap x dropoff: **50** **expect** **10.0** word size: **11** **Filter** ☒ **Align**

Sequence 1
 Enter accession, GI or sequence in FASTA format from: **0** to: **0**
 CTGCTCAGGCTCCGAGTAGCTGGGACTACGGCGCGCCGCCA000CA
 OCTGGCTAATTTTTTTTATTTTATAGTAGAGACGGGGTTTCA00GTG
 TTAGCCAGGATGCTCGATCTCGGACCTCGTGAATCTG000G00CTCA
 G0CTCCAAAGTGCTGGGATTACGGGCTAGGCCACGGG00CA00CT
 G

or upload FASTA file **ファイルを選択** 選択されていません

Sequence 2
 Enter accession, GI or sequence in FASTA format from: **0** to: **0**
 AATACGTTGTTAAATGTTATTAGTTGACCAAGGCGAGTAAATGAAAC
 OGCATTTTGGTGCCATTAAATAGGGAACCAATGTAATGTAATGTA
 AATGGAGCAATTCGACTAGGCAAGTGTATTTTGTATTTTATATA
 CAATTTCTATTATTTTCAAGTAATAAACAATGTTTTTCATACTGAA
 TATTA

or upload FASTA file **ファイルを選択** 選択されていません

Align **Clear Input**

Comments and suggestions to blast-help@ncbi.nlm.nih.gov

結果の画面ではまず上部にこの二本の配列のどの領域が類似したかが模式図およびドットマトリクス（ハープロット）に似た形で示されます(図中赤で囲った部分)。その下にアライメントの詳細が示されています。今回用いた配列では一本目の配列の116bpから233bp目が二本目の配列の954bp目から1071bp目にヒットしていることがわかります。

BLAST 2 Sequences

http://blast.ncbi.nlm.nih.gov/bt2seq/wblast2.cgi

NCBI Entrez BLAST 2 sequences BLAST Example Help

BLAST 2 SEQUENCES

This functionality is now available in the main BLAST pages and this link will be removed in the near future. Step-by-step [instructions](#) for use are available.

This tool produces the alignment of two given sequences using BLAST engine for local alignment. The stand-alone executable for blasting two sequences (bt2seq) can be retrieved from [NCBI ftp site](#).
 Reference: Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250

Program: **blastn** Matrix: Not Applicable

Parameters used in BLASTN program only:
 Reward for a match: 1 Penalty for a mismatch: -2

☐ Use Mega BLAST Strand option: Both strands View option: Standard
 Masking character option: X for protein, n for nucleotide Masking color option: Black
☐ Show CDS translation

Open gap: 5 and extension gap: 2 penalties
 gap x dropoff: 50 expect: 10.0 word size: 11 Filter: ☒ Align

Sequence 1
 Enter accession, GI or sequence in FASTA format from: 0 to: 0
 GATGAGGATGATGAGGACGATGACGTCAGCCATGTGGATGAGGAGGACGTCTGGGGT
 GCAGA
 GGGAGGACCGACGGGGCGGCGATGGGCAGATCAATGAGCAGGTAGAGAAGCTGCGGGG
 CCGGA
 AGGTGCCAGCAATGAGACTGAGGAGACTAG

or upload FASTA file: ファイルを選択 選択されていません

Sequence 2
 Enter accession, GI or sequence in FASTA format from: 0 to: 0
 GATGAGGATGATGAGGACGATGACGTCAGCCATGTGGATGAGGAGGACGTCTGGGGT
 GCAGA
 GGGAGGACCGACGGGGCGGCGATGGGCAGATCAATGAGCAGGTAGAGAAGCTGCGGGG
 CCGGA
 AGGTGCCAGCAATGAGACTGAGGAGACTAG

or upload FASTA file: ファイルを選択 選択されていません

Align Clear Input

Comments and suggestions to blast-help@ncbi.nlm.nih.gov

すると、以下のように二つに分断された形で結果が得られます。両者は同じ配列なので対角線上に1本の線状な結果が期待されるのですが、赤く○で印をつけた領域が欠けています。

Blast Result

http://blast.ncbi.nlm.nih.gov/Blast2.cgi?0

こちらのブックマーク バーにブックマークを追加すると簡単にページにアクセスできます。

NCBI

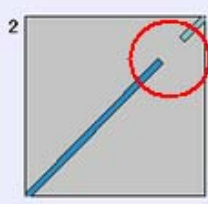
PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.18 [Mar-02-2008]

Match: 1 Mismatch: -2 gap open: 5 gap extension: 2
 xdropoff: 0 expect: 10.00 wordsize: 11 Filter ☒ View option: Standard
 Masking character option: X for protein, n for nucleotide Masking color option: Black
☐ Show CDS translation

Sequence 1: |c||1
 Length = 863 (1 .. 863)

Sequence 2: |c||1
 Length = 863 (1 .. 863)



NOTE: Bitscore and expect value are calculated based on the size of the nr database.
 NOTE: If protein translation is reversed, please repeat the search with reverse strand of the query sequence.

Score = 1246 bits (648), Expect = 0.0
 Identities = 648/648 (100%), Gaps = 0/648 (0%)
 Strand=Plus/Plus

```

Query 1  AGCCCCCTCACTCACTCGCAGCCATACAGCCCCAGAGGCTCCGATGGGAGATTATG  60
Sbjct 1  AGCCCCCTCACTCACTCGCAGCCATACAGCCCCAGAGGCTCCGATGGGAGATTATG  60

Query 61  GTGCTTGGCTATCATCATGGCAGGAATTGCATTGGCTTCACTCACTCACTCACTCACT  120
Sbjct 61  GTGCTTGGCTATCATCATGGCAGGAATTGCATTGGCTTCACTCACTCACTCACTCACT  120

Query 121  ACCTGCTGCCCCATCTCTGGAGGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  180
Sbjct 121  ACCTGCTGCCCCATCTCTGGAGGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  180

Query 181  CGAGTCTCTGGAAGTGAAGTGGCAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG  240
Sbjct 181  CGAGTCTCTGGAAGTGAAGTGGCAAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG  240
  
```

下にスクロールしてアライメントを見ると649bp目から752bp目からが欠けているようです。

```

Query 541  AGTCTACGTCATCTCTCTCTGGGAAGGACAGCCACAGCCCGGAGGGCTCCACGGCCACAT  600
Sbjct 541  AGTCTACGTCATCTCTCTCTGGGAAGGACAGCCACAGCCCGGAGGGCTCCACGGCCACAT  600

Query 601  ACCACCTGCTGGGTCCCCAGGAAGAGGGTGAAGGGGTGCTAGATGTCA 648
Sbjct 601  ACCACCTGCTGGGTCCCCAGGAAGAGGGTGAAGGGGTGCTAGATGTCA 648

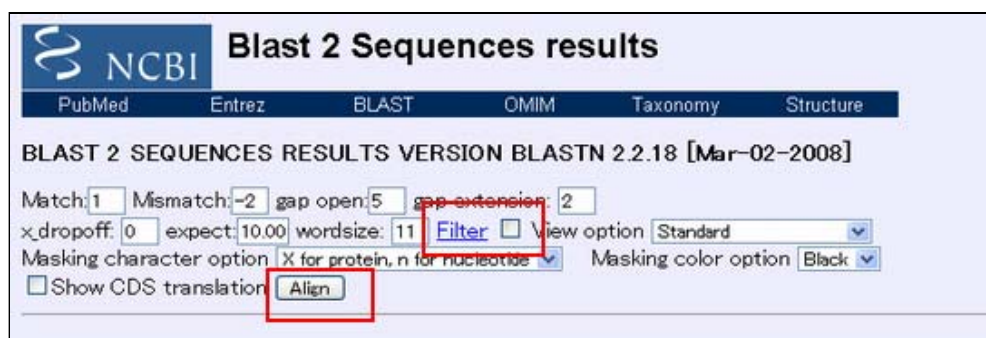
Query 753  GTCCTTGGGGTGCAGAGGGAGGACCGAGGGGCGGCGATGGGCAGATCAATGAGCAGGTA  812
Sbjct 753  GTCCTTGGGGTGCAGAGGGAGGACCGAGGGGCGGCGATGGGCAGATCAATGAGCAGGTA  812

Query 813  GAGAAGCTGCGGCGGCGGAGAGGTGCCAGCAATGAGACTGAGCGAGACTAG  863
Sbjct 813  GAGAAGCTGCGGCGGCGGAGAGGTGCCAGCAATGAGACTGAGCGAGACTAG  863
  
```

Score = 214 bits (111), Expect = 8e-52
 Identities = 111/111 (100%), Gaps = 0/111 (0%)
 Strand=Plus/Plus

これはデフォルトでフィルタオプションがonになっており、この領域がLow Complexityと判断されたことに起因す

と思われます。そこで、画面上部のオプションでFilterにつけられたチェックを外して再度Alignボタンを押して検索を実行してみましょう。



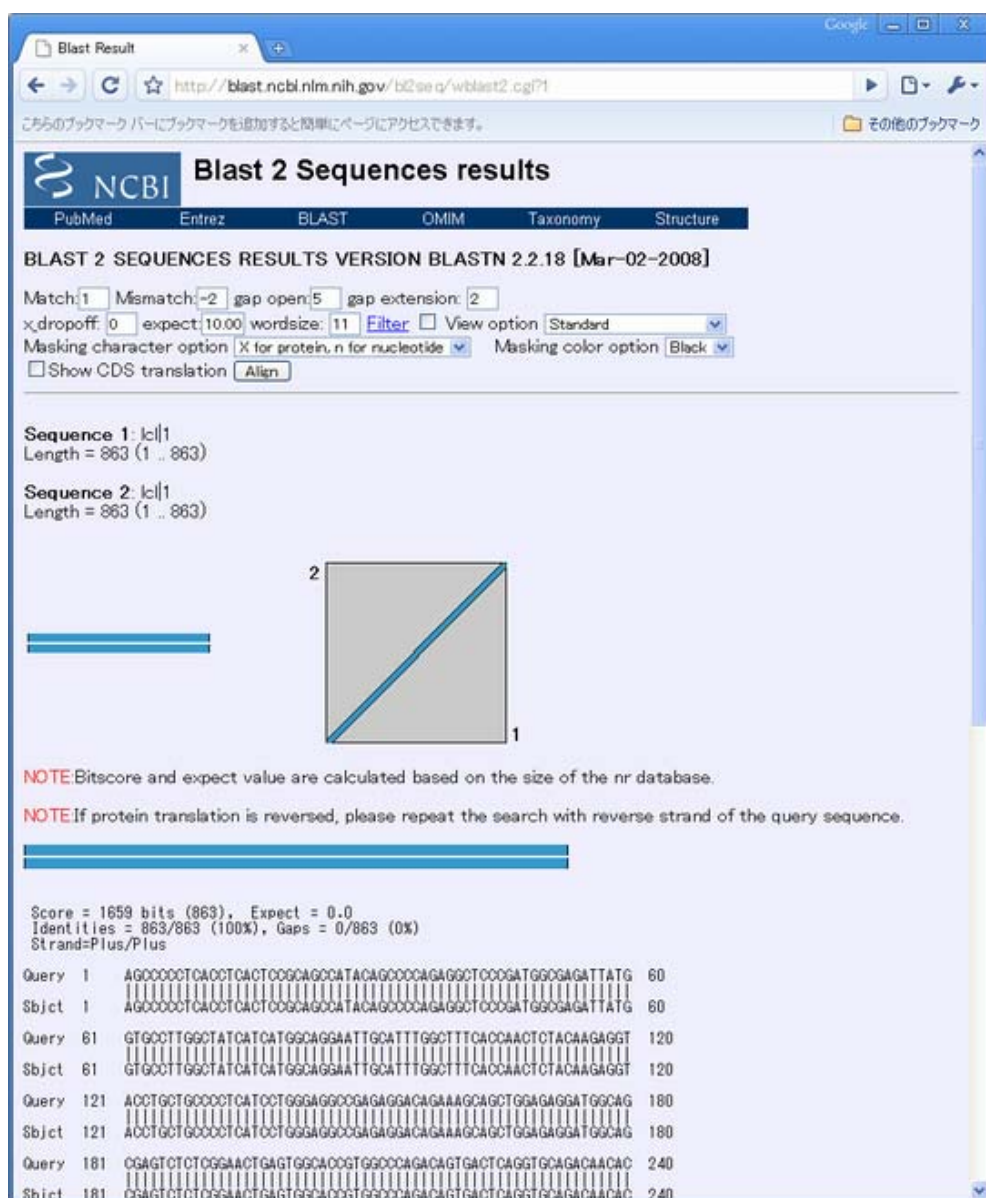
NCBI Blast 2 Sequences results

PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.18 [Mar-02-2008]

Match: 1 Mismatch: -2 gap open: 5 gap extension: 2
x_dropoff: 0 expect: 10.00 wordsize: 11 Filter View option: Standard
Masking character option: X for protein, n for nucleotide Masking color option: Black
☐ Show CDS translation Align

今度は分断されることなく全長がアライメントできたことが確認できます。



Blast Result

http://blast.ncbi.nlm.nih.gov/b2seq/wblast2.cgi?1

NCBI Blast 2 Sequences results

PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.18 [Mar-02-2008]

Match: 1 Mismatch: -2 gap open: 5 gap extension: 2
x_dropoff: 0 expect: 10.00 wordsize: 11 Filter View option: Standard
Masking character option: X for protein, n for nucleotide Masking color option: Black
☐ Show CDS translation Align

Sequence 1: |c|1
Length = 863 (1 .. 863)

Sequence 2: |c|1
Length = 863 (1 .. 863)

NOTE: Bitscore and expect value are calculated based on the size of the nr database.

NOTE: If protein translation is reversed, please repeat the search with reverse strand of the query sequence.

Score = 1659 bits (863), Expect = 0.0
Identities = 863/863 (100%), Gaps = 0/863 (0%)
Strand=Plus/Plus

Query 1 AGCCCCCTCACTCACTCCGAGCCATACAGCCCCAGAGGCTCCGATGGGAGATTATG 60
Sbjct 1 AGCCCCCTCACTCACTCCGAGCCATACAGCCCCAGAGGCTCCGATGGGAGATTATG 60

Query 61 GTGCCCTGGCTATCATCATGGCAGGAATTGCATTGGCTTTCAACCACTCTACAAGAGGT 120
Sbjct 61 GTGCCCTGGCTATCATCATGGCAGGAATTGCATTGGCTTTCAACCACTCTACAAGAGGT 120

Query 121 ACCTGCTGCCCTCATCCTGGAGGCCGAGAGGACAGAAAGCAGCTGGAGAGGATGGCAG 180
Sbjct 121 ACCTGCTGCCCTCATCCTGGAGGCCGAGAGGACAGAAAGCAGCTGGAGAGGATGGCAG 180

Query 181 CGAGTCTCTCGGAAGTGAAGTGGCAAGCTGGCCAGACAGTGAAGTGAAGTGAAGTGAAGT 240
Sbjct 181 CGAGTCTCTCGGAAGTGAAGTGGCAAGCTGGCCAGACAGTGAAGTGAAGTGAAGTGAAGT 240

原文更新日: 2005年6月7日 日本語版更新日: 2009年1月19日

All Rights Reserved, Copyright(C) 1997-2009 Japan Science and Technology Agency(JST)

