

NCBI BLASTチュートリアル

このチュートリアルでは、NCBIサイトでのBLASTによる相同性検索の方法について、一般的な使い方を紹介しています。

はじめに。

■BLASTとは

まずはじめに、簡単にBLASTについて紹介することにしましょう。

BLASTは、**B**asic **L**ocal **A**lignment **S**earch **T**ool の略で、ペアワイズの局所的なアライメント／相同性検索（ホモロジー検索）を比較的高速に行うプログラムのことを指します。厳密な解を提供する Smith-Watermanアルゴリズムを少しヒューリスティックにすることで、完全な厳密解は与えないものの実用的には十分な精度を持ちつつ、Smith-Watermanよりはるかに高速に検索を実現します。

また、BLASTではペアワイズの相同性検索の結果に対して、その類似度からスコアや偶然そのような配列の一致が起こる期待値e-valueを出力し、閾値以上でデータベースとヒットした場合、その結果を出力します。

このBLASTプログラムをWWWを介して使用可能にしたサービスの一つがここで紹介するNCBI BLASTサービスになります。具体的にはユーザが手元に持っている塩基配列、あるいはアミノ酸配列を問い合わせ配列（query配列）として、GenBank/EMBL/DDBJなどの公共データベースに登録されている塩基配列あるいはアミノ酸配列に対して相同性検索を実施し、その結果をWWWを介してユーザに提供するサービスです。

BLASTの基本的なサービスは、問い合わせ配列とデータベース配列の組み合わせから、次の5種類が用意されています。

blastn

塩基配列を問い合わせ配列として、塩基配列データベースから相同性を持つ配列（エントリ）を検索します。

blastp

アミノ酸配列を問い合わせ配列として、アミノ酸配列データベースから相同性を持つ配列（エントリ）を検索します。

blastx

塩基配列を問い合わせ配列として、それを6フレーム分アミノ酸配列に翻訳し、アミノ酸配列データベースから相同性を持つ配列（エントリ）を検索します。この検索はESTの断片などを配列決定した場合に、既知のアミノ酸配列情報との相同性を調べる場合などに有効です。

tblastn

アミノ酸配列を問い合わせ配列として、塩基配列データベースを6フレームでアミノ酸に翻訳しながら相同性を持つ配列（エントリ）を検索します。

tblastx

塩基配列を問い合わせ配列として、これを6フレームアミノ酸に翻訳しつつ塩基配列データベースも6フレームアミノ酸に翻訳して、アミノ酸配列同士を比較することにより相同性を持つ配列（エントリ）を検索します。弱い相同性しかない場合でも検出できることが期待されます。

以下では、これらの中から最も代表的なblastn, blastpについて、実際のデータを用いたBLAST検索を実行してみましょう。

1. blastnの使い方

■1-1 blastnを用いたホモロジー検索を行ってみる

[blastnサービスのページへ移動する](#)

まず、BLAST検索のトップページに移ってみましょう。BLAST検索のページへはNCBIトップページの上部にあるとかかれたリンクをクリックすることで移動できます。

NCBI National Center for Biotechnology Information
 National Library of Medicine National Institutes of Health

PubMed All Databases **BLAST** OMIM Books TaxBrowser Structure

Search All Databases for **BLAST 検索ページへ** Go

SITE MAP
 Alphabetical List
 Resource Guide

About NCBI
 An introduction to
 NCBI

GenBank
 Sequence
 submission support
 and software

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More about NCBI...](#)

Hot Spots

- ▶ Assembly Archive
- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR

このページからは、様々なBLAST検索ができるようになっています。大きく分けて上部の赤四角で囲んだ部分では、ゲノム配列決定がされた生物種に関する各種配列データへのホモロジー検索が可能です。

次に中央部の青四角で囲んだ部分が、上のはじめにで紹介した、基本的な5つのBLASTサービスになっています。一番上の“nucleotide blast”と書かれたリンクが blastnによるサービスで、次の“protein blast”と書かれたリンクが blastpによるサービスです。

一番下の緑四角で囲んだ領域では、blastを用いた特別な検索サービスが並んでいます。primerの作成やtrace archiveと呼ばれる配列群への検索やベクターの検索などです。これらについては、別のチュートリアルで紹介する予定にしています。


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

ゲノム解読された生物種のデータに対する検索

Choose a species genome to search, or [list all genomes available](#)

| | | |
|---|--|--|
| <input type="checkbox"/> Human | <input type="checkbox"/> Oryza sativa | <input type="checkbox"/> Gallus gallus |
| <input type="checkbox"/> Mouse | <input type="checkbox"/> Bos taurus | <input type="checkbox"/> Pan troglodytes |
| <input type="checkbox"/> Rat | <input type="checkbox"/> Danio rerio | <input type="checkbox"/> Microbes |
| <input type="checkbox"/> Arabidopsis thaliana | <input type="checkbox"/> Drosophila melanogaster | <input type="checkbox"/> Apis mellifera |

Basic BLAST

Choose a BLAST program to run. 基本的なblast検索

| | |
|----------------------------------|---|
| nucleotide blast | Search a nucleotide database using a nucleotide query |
| blastn | <i>Algorithms:</i> blastn, megablast, discontinuous megablast |
| protein blast | Search protein database using a protein query |
| blastp | <i>Algorithms:</i> blastp, psi-blast, phi-blast |
| tblastx | 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.) 特定の用途向けの検索

- ☐ 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

では、青四角で囲まれた“Basic BLAST”から一番上の“nucleotide blast”をクリックしてみてください。
 下のような設定画面に移動します。

NCBI/BLAST/blastn suite: BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

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](#) ファイルが...いてません [?](#)

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](#) **パラメータの設定**

この画面では、上から順に赤四角で囲まれた領域で問い合わせ配列の設定を、次の青四角の領域でデータベースの設定を、緑の領域で検索プログラムの設定を行い、“BLAST”ボタンをクリックすることでホモロジー検索が実行されます。

さらに下の“Algorithm Parameters”をクリックすると、検索時のパラメータ設定が可能になります。

[問い合わせ配列を設定する \(blastn\)](#)

まずは、問い合わせ配列を設定してみましょう。問い合わせ配列は基本的に一本のFASTA形式からなっている必要があります。（ヘッダがなくて配列のみでも検索はできます）

BLAST利用時の問い合わせ配列の設定には、大きく分けると以下に挙げた3通りの方法があります。

- テキストボックスへの問い合わせ配列の入力（カットアンドペースト）
- 問い合わせ配列のアップロード
- アクセッション番号での指定

順に見ていきましょう。

例として、

```
>test
GTTTCGAGAGGTGTGCTCTGAACAAGCCGAGACGGGGCCGTGCCGAGCAATGATCTCCGCTGGTACTTTG
ATGTGACTGAAGGGAAGTGTGCCCCATTCTTTACGGCGGATGTGGCGGCAACCGGAACAACCTTTGACAC
AGAAGAGTACTGCATGGCCGTGTGTGGCAGCGCCATGTCCCAAAGTTTACTCAAGACTACCCAGGAACCT
CTTGCCCGAGATCCTGTAACTTCCTACAACA
```

この配列を問い合わせ配列として入力してみましょう。ちなみにこの配列はデータベースに登録されており、アクセッション番号X06981です。

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](#) ファイルが...ていません [?](#)

Job Title

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

☐ Blast 2 sequences

一番オーソドックスな方法は、図中赤四角で囲まれた領域に直接上の配列を打ち込むか、あるいは、ユーザ側のコンピュータの機能を使ってカットアンドペーストするかです。入力すると下図のようになります。

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](#) ファイルが...ていません [?](#)

Job Title

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

☐ Blast 2 sequences

次の問い合わせ配列のアップロードですが、これはユーザ側のコンピュータに、問い合わせ配列がテキスト形式として保存されている場合に使える方法です。

配列のアップロードには、上で示した問い合わせ配列入力画面で青四角で囲まれた“ファイルを選択”（あるいは“参照”など）と書かれている場合もあります。ユーザが使用しているコンピュータのOSやブラウザに依存します）をクリックしてください。

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](#) ファイルが...ていません [?](#)

Job Title

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

☐ Blast 2 sequences

Choose Sequence Database

Organism Optional

Entrez Query Optional

Program S Optimize for

マイドキュメント

最近使ったファイル

デスクトップ

マイドキュメント

マイコンピュータ

マイネットワーク

ファイル名(N):

ファイルの種類(R):

☐ 読み取り専用ファイルとして開く(R)

開く(O)

キャンセル

ファイル選択のダイアログが開きます。ここで適切なファイルを選択し、ファイルをアップロードしてください。

最後に、アクセッション番号での指定方法についてです。これは、問い合わせに使いたい配列のアクセッション番号がわかっている場合にしか使えませんが、簡単で便利な方法です。

ここではアクセッション番号がX06981とわかっています。これを先ほど問い合わせ配列を入力した箇所と同じ赤四角で囲まれた領域に入力してください。

これだけで、後は自動的に該当する配列が問い合わせ配列となります。

問い合わせ配列のレンジを指定する (blastn)

問い合わせ配列に関する箇所でのその他の設定を見てみましょう。まずは、問い合わせ配列のレンジ指定です。

これは、上図の赤四角の部分で行うのですが、問い合わせ配列のある一部分のみを問い合わせとして使う場合に指定します。例えば問い合わせ配列全体が600bpあったとしてその中の先頭300bpのみを問い合わせ配列として使いたい場合には、“From”の欄に1を“To”の欄に300を入れて検索を実行します。

デフォルト（空欄）の場合には入力された配列の全長を問い合わせ配列として使います。

データベースを指定する (blastn)

続いてデータベースを設定しましょう。

下図の領域で、データベースは設定します。

blastn検索時のデフォルトは、“Human genomic + transcript”（ヒトゲノムと転写産物）となっています。これを変更したい場合には、その右横の“Mouse genomic + transcript”（マウスゲノムと転写産物）を選択するか、あるいは下のプルダウンメニューから検索を希望するデータベースを選択します。

プルダウンメニューから選択できるデータベースの中で代表的なものには、以下のようなものがあります。

Nucleotide Collection(nr/nt)

GenBank, EMBL, DDBJ, PDBに含まれている塩基配列の全エントリ (HTGS0, 1, 2, EST, GSS, STS, PAT, WGSをのぞく) およびRefseq配列から重複を除いたもの

Reference mRNA sequences(refseq_rna)

NCBI Refseqプロジェクトで作成しているmRNA配列

Reference genomic sequences(refseq_genomic)

NCBI Refseqプロジェクトで作成しているゲノム配列

NCBI Genomes (chromosome)

NCBI Genomesに含まれているゲノム配列

Expressed sequence tags (est)

GenBank, EMBL, DDBJに含まれるEST配列

Non-human, non-mouse ESTs (est_others)

ヒト、マウス以外のEST配列

[検索エンジンを選択する \(blastn\)](#)

続いて塩基同士の配列比較時の検索エンジンを設定しましょう。最初には説明しませんでした。塩基の問い合わせ配列を塩基のデータベースに対して検索する場合にはblastnが用いられますが、データベースサイズの肥大化に伴い、検索時間がblastといえども長くかかるようになってきました。

このため、より高速に検索することの可能な検索エンジンがblastnから派生してできてきました。その検索エンジンを設定する箇所が下図の領域です。

上から順に、

megablast

問い合わせ配列と非常に近い塩基配列のみを高速に検索する。約95%程度の塩基一致率以上のものを検出する

discontiguous megablast

megablastよりも多少のミスマッチを許したシードから検索を開始する。生物種間を越えた程度の検索が可能

blastn

7塩基以上一致するシードからの検索。網羅性は高い

となっています。目的に合わせて選択してください。

この例では、例として示したアクセッション番号X06981の塩基配列をnr/ntデータベースに対して blastnを用いて検索することになります。

Enter Query Sequence

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

X06981 From To

Or, upload file [ファイルを選択](#) ファイルが...ていません

Job Title X06981:Human mRNA fragment for amyloid beta-protein...
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 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)
Choose a BLAST algorithm

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

今までの解説に従って必要な情報を入力し、“BLAST”ボタンをクリックして検索を実行してみましょう。

すると、下図のようにジョブを受け付けた画面が表示されます。この画面は自動的に更新され、検索が終了すると検索結果画面が表示されます。

► NCBI/ BLAST/ blastn suite/ Formatting Results - GGF7HXX601R [\[Formatting options\]](#)

Job Title: X06981:Human mRNA fragment for amyloid beta-protein...

| | |
|-----------------------|--------------------------|
| Request ID | GGF7HXX601R |
| Status | Searching |
| Submitted at | Wed Oct 29 06:37:37 2008 |
| Current time | Wed Oct 29 06:37:39 2008 |
| Time since submission | |

This page will be automatically updated in 16 seconds

[Copyright](#) | [Disclaimer](#) | [Privacy](#) | [Accessibility](#) | [Contact](#) | [Send feedback](#)

■ 1-2 blastn検索結果の見方

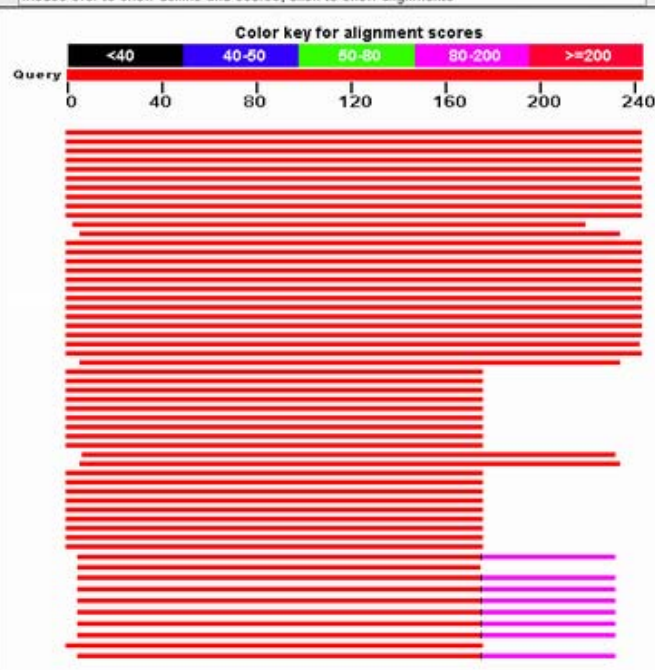
[結果の概要\(blastn\)](#)

では、続いて検索結果を見ていきましょう。検索結果は非常に長いページになっています。これをスクロールして見ていくと、上からGraphic Summary, Descriptions, Alignmentsの3領域からなっていることがわかります。いずれの領域もデータベースに対してヒットした各エントリに対し、スコアの高い順にソートされて表示されています。

Graphic Summary

Distribution of 109 Blast Hits on the Query Sequence

Mouse-over to show details and scores, click to show alignments



Descriptions

Legend for links to other resources: [U](#) UniGene [G](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

| Accession | Description | Max score | Total score | Query coverage | E value | Max ident | Links |
|--------------------------------|--|-----------|-------------|----------------|---------|-----------|---|
| NM_001136139.1 | Homo sapiens amyloid beta (A4) precursor protein (APP), | 439 | 439 | 100% | 3e-120 | 100% | G |
| AB394791.1 | Synthetic construct DNA, clone: pF1KB3334, Homo sapien | 439 | 439 | 100% | 3e-120 | 100% | |
| NM_000484.2 | Homo sapiens amyloid beta (A4) precursor protein (APP), | 439 | 439 | 100% | 3e-120 | 100% | U G G |
| BC018937.2 | Homo sapiens amyloid beta (A4) precursor protein (peptid | 439 | 439 | 100% | 3e-120 | 100% | |
| X04981.1 | Human mRNA fragment for amyloid beta-protein (AP) inse | 439 | 439 | 100% | 3e-120 | 100% | U G G |
| AK294534.1 | Homo sapiens cDNA FLJ50491 complete cds, highly similar | 437 | 437 | 99% | 9e-120 | 100% | U G G |
| AK312326.1 | Homo sapiens cDNA, FLJ92638, Homo sapiens amyloid bet | 434 | 434 | 100% | 1e-118 | 99% | U G G |
| NM_001013018.1 | Pan troglodytes amyloid beta (A4) precursor protein (APP) | 430 | 430 | 100% | 1e-117 | 99% | G |
| M58726.1 | M. fascicularis myloid b-protein precursor Kunitz protease i | 416 | 416 | 100% | 3e-113 | 97% | U |
| NM_214372.1 | Sus scrofa amyloid beta (A4) precursor protein (APP), mR | 394 | 394 | 100% | 1e-106 | 95% | U G |
| X52438.1 | Artificial gene for APP1-72 | 390 | 390 | 88% | 1e-105 | 100% | |
| X15885.1 | Macaca mulatta APP-70 mRNA fragment for amyloid precu | 389 | 389 | 93% | 4e-105 | 97% | U G G |
| AY265533.1 | Stenella coeruleoalba beta-amyloid precursor protein 749 | 389 | 389 | 100% | 4e-105 | 95% | |
| NM_001006601.1 | Canis lupus familiaris amyloid beta (A4) precursor protein | 376 | 376 | 100% | 3e-101 | 94% | U G G |
| AY26579.1 | Canis familiaris beta-amyloid precursor protein 770 mRNA | 376 | 376 | 100% | 3e-101 | 94% | U G G |
| AF197164.1 | Cavia sp. putative amyloid precursor protein mRNA, partic | 343 | 343 | 100% | 2e-91 | 91% | |
| BC005499.1 | Mus musculus amyloid beta (A4) precursor protein, mRNA | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| AK159245.1 | Mus musculus osteoclast-like cell cDNA, RIKEN full-length | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| AK148479.1 | Mus musculus B16 F10Y cells cDNA, RIKEN full-length enri | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| AK159999.1 | Mus musculus osteoclast-like cell cDNA, RIKEN full-length | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| AK159718.1 | Mus musculus osteoclast-like cell cDNA, RIKEN full-length | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| AY267348.1 | Mus musculus amyloid-beta precursor protein-like protein | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| M24397.1 | Mouse amyloid beta protein 69S (mBPP69S) protease inhibi | 336 | 336 | 100% | 2e-89 | 90% | U G G |
| NM_019288.1 | Rattus norvegicus amyloid beta (A4) precursor protein (A4) | 329 | 329 | 99% | 3e-87 | 90% | U G G |
| AK159352.1 | Mus musculus osteoclast-like cell cDNA, RIKEN full-length | 324 | 324 | 100% | 1e-85 | 89% | U G G |
| X15210.1 | Murine mRNA fragment for A4 amyloid precursor protein | 324 | 324 | 93% | 1e-85 | 91% | U G G |
| NM_001136016.2 | Homo sapiens amyloid beta (A4) precursor protein (APP), | 318 | 318 | 72% | 6e-84 | 100% | G |
| AK295621.1 | Homo sapiens cDNA FLJ50531 complete cds, highly similar | 318 | 318 | 72% | 6e-84 | 100% | U G G |
| EU032797.1 | Synthetic construct Homo sapiens clone HA1B:100067826; | 318 | 318 | 72% | 6e-84 | 100% | G |
| NM_201413.1 | Homo sapiens amyloid beta (A4) precursor protein (APP), | 318 | 318 | 72% | 6e-84 | 100% | U G G |
| BC065578.1 | Homo sapiens amyloid beta (A4) precursor protein, mRNA | 318 | 318 | 72% | 6e-84 | 100% | U G G |

▼ Alignments ☐ Select All [Get selected sequences](#) [Distance tree of results](#)

> [ref|NM_001136130.1](#) **G** Homo sapiens amyloid beta (A4) precursor protein (APP), transcript variant 6, mRNA
Length=3473

GENE ID: 351 APP | amyloid beta (A4) precursor protein [Homo sapiens]
([Over 100 PubMed links](#))

Score = 439 bits (486), Expect = 3e-120
Identities = 243/243 (100%), Gaps = 0/243 (0%)
Strand=Plus/Plus

```

Query 1      GTTCGAGAGGTGTGCTCTGACAAAGCCGAGACGGGGGCGTGGCGAGCAATGATCTCCGGC 60
                |||
Sbjct 885     GTTCGAGAGGTGTGCTCTGACAAAGCCGAGACGGGGGCGTGGCGAGCAATGATCTCCGGC 944

Query 61      TGGTACTTTGATGTGACTGAAAGGGAAGTGTGCCCATCTCTTTACGGGGATGTGGGGC 120
                |||
Sbjct 945     TGGTACTTTGATGTGACTGAAAGGGAAGTGTGCCCATCTCTTTACGGGGATGTGGGGC 1004

Query 121     AACCGGAACAACCTTTGACACAGAAAGTACTGATGGCGGTGTGTGGAGGGCCATGTCC 180
                |||
Sbjct 1005    AACCGGAACAACCTTTGACACAGAAAGTACTGATGGCGGTGTGTGGAGGGCCATGTCC 1064

Query 181     CAAAGTTTACTCAAGACTACCCAGGAACCTCTTGCCCGAGATCCTGTAAACTTCCAC 240
                |||
Sbjct 1065    CAAAGTTTACTCAAGACTACCCAGGAACCTCTTGCCCGAGATCCTGTAAACTTCCAC 1124

Query 241     ACA 243
                |||
Sbjct 1125    ACA 1127

```

> [gb|AF534731.1](#) Synthetic construct DNA, clone: sF1K8334, Homo sapiens APP gene for amyloid beta A4 protein precursor, complete cds, without stop codon, in Flexi system
Length=2327

Score = 439 bits (486), Expect = 3e-120
Identities = 243/243 (100%), Gaps = 0/243 (0%)
Strand=Plus/Plus

```

Query 1      GTTCGAGAGGTGTGCTCTGACAAAGCCGAGACGGGGGCGTGGCGAGCAATGATCTCCGGC 60
                |||
Sbjct 888     GTTCGAGAGGTGTGCTCTGACAAAGCCGAGACGGGGGCGTGGCGAGCAATGATCTCCGGC 927

Query 61      TGGTACTTTGATGTGACTGAAAGGGAAGTGTGCCCATCTCTTTACGGGGATGTGGGGC 120
                |||
Sbjct 928     TGGTACTTTGATGTGACTGAAAGGGAAGTGTGCCCATCTCTTTACGGGGATGTGGGGC 987

Query 121     AACCGGAACAACCTTTGACACAGAAAGTACTGATGGCGGTGTGTGGAGGGCCATGTCC 180
                |||
Sbjct 989     AACCGGAACAACCTTTGACACAGAAAGTACTGATGGCGGTGTGTGGAGGGCCATGTCC 1047

Query 181     CAAAGTTTACTCAAGACTACCCAGGAACCTCTTGCCCGAGATCCTGTAAACTTCCAC 240
                |||
Sbjct 1048    CAAAGTTTACTCAAGACTACCCAGGAACCTCTTGCCCGAGATCCTGTAAACTTCCAC 1107

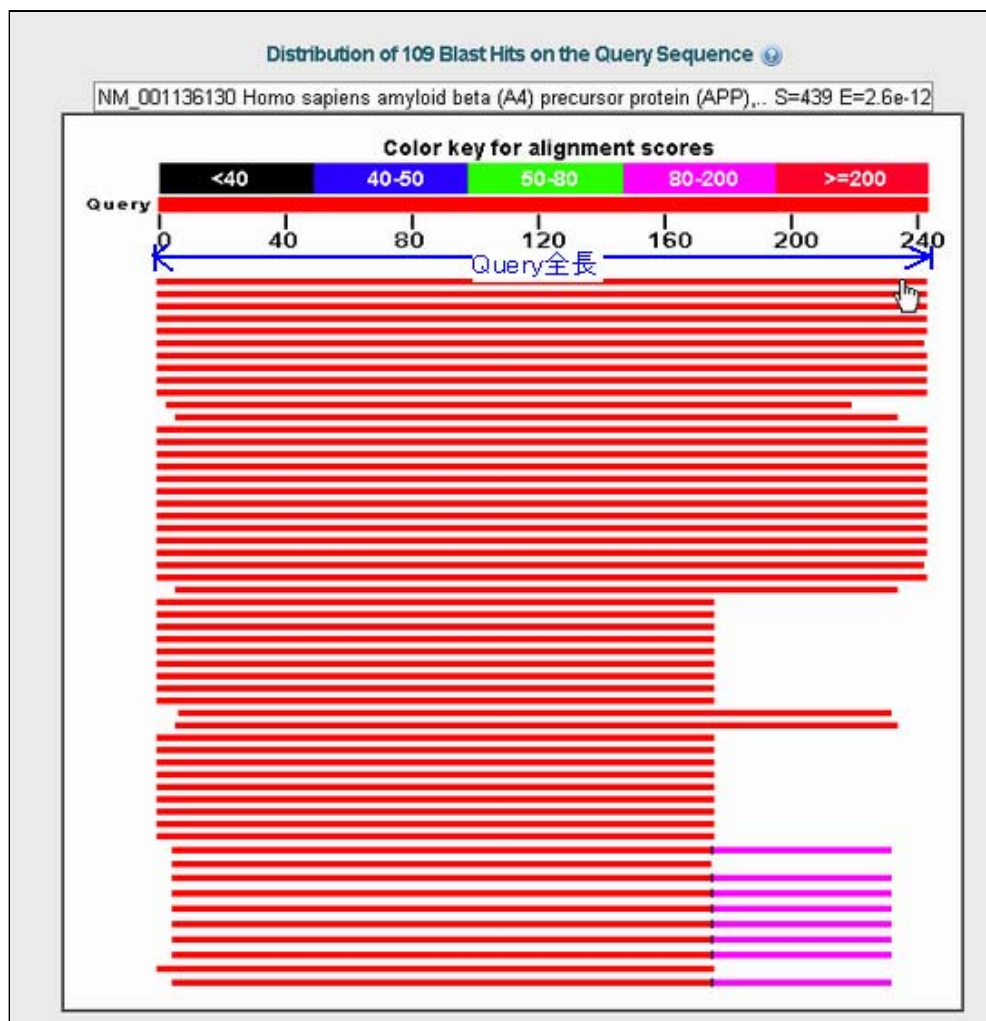
Query 241     ACA 243
                |||
Sbjct 1108    ACA 1110

```

順に見ていきましょう。

[Graphic Summaryの見方\(blastn\)](#)

まずは、Graphic Summaryです。文字通りこの部分ではblast検索結果の概要が示されています。



Query全体に対して、どこからどこあたりにヒットが見られたのかが表示されています。バーで表示されている各行がデータベース内の1エントリに相当し、Query上でヒットのあった領域に色がつけられています。色は上に示されているようにスコアで区別されています。

この例では、Queryの全長に渡ってヒットするエントリが上の方に並んでおり、真ん中ほどから下にはQueryの1-170bpほどが当たっているエントリが並んでいることが分かります。

各行にマウスを重ねると、上部のテキストボックスにそのエントリに関する説明とスコア、e-valueなどが表示されます。

またクリックすると、そのエントリに対するアライメント結果（後述）へと移動します。

Descriptionsの見方(blastn)

次に“Descriptions”の領域です。この部分ではヒットしたエントリについて、その概要が表形式で示されています。

Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structures [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

| Accession | Description | Max score | Total score | Query coverage | E-value | Max ident | Links |
|----------------|--|-----------|-------------|----------------|---------|-----------|---------------------|
| NM_001136130.1 | Homo sapiens amyloid beta (A4) precursor protein (APP), | 439 | 439 | 100% | 3e-120 | 100% | G |
| AB394791.1 | Synthetic construct DNA, clone: pF1KB3334, Homo sapien | 439 | 439 | 100% | 3e-120 | 100% | |
| NM_000484.2 | Homo sapiens amyloid beta (A4) precursor protein (APP), | 439 | 439 | 100% | 3e-120 | 100% | UEG |
| BC018937.2 | Homo sapiens amyloid beta (A4) precursor protein (peptid | 439 | 439 | 100% | 3e-120 | 100% | |

左から順に、Accession番号、エントリの説明、スコア（そのエントリと問い合わせ配列とが複数個所でヒットした場合最大のスコア）、合計スコア（そのエントリと問い合わせ配列とが複数個所でヒットした場合、そのスコアの合計）、カバー率、（問い合わせ配列に対して、ヒットした領域がどれだけカバーするか）、E-value（エントリと問い合わせ配列とのヒットが偶然起こる期待値。小さいほどそのヒットが偶然出ないことを表す。）、一致率（エントリと問い合わせ配列がヒットしている領域での（塩基）一致率）、リンクを表しています。

スコアやE-value、一致率などと書かれたヘッダをクリックすることで、それぞれの値でソートできるようになっている。

また、アクセッション番号をクリックするとそのエントリへ、それ以外の値の箇所をクリックするとアライメントへと移動します。

リンクに関しては、そのエントリのデータがUnigeneやGE0などに含まれている場合、クリックすることでそのデータへと移動します。

[Alignmentsの見方\(blastn\)](#)

最後にアライメントの見方です。

```
>[ref|NM_001136130.1] G Homo sapiens amyloid beta (A4) precursor protein (APP), transcript
variant 8, mRNA
Length=3473

GENE_ID: 351 APP | amyloid beta (A4) precursor protein [Homo sapiens]
(Over 100 PubMed links)

Score = 439 bits (486), Expect = 3e-120
Identities = 243/243 (100%), Gaps = 0/243 (0%)
Strand=Plus/Plus

Query 1      GTTCGAGAGGTGCTCTGAACAAGCCGAGACGGGCCGTGCCGAGCAATGATCTCCCGC 60
Sbjct 885    GTTCGAGAGGTGCTCTGAACAAGCCGAGACGGGCCGTGCCGAGCAATGATCTCCCGC 944

Query 81     TGGTACTTTGATGTGACTGAAGGGAAGTGTGCCCATTCCTTTACGGCGGATGTGGCGGC 120
Sbjct 945    TGGTACTTTGATGTGACTGAAGGGAAGTGTGCCCATTCCTTTACGGCGGATGTGGCGGC 1004

Query 121    AACCGGAACAACCTTTGACACAGAAGAGTACTGCATGGCCGTGTGTGGCAGCGCCATGTCC 180
Sbjct 1005   AACCGGAACAACCTTTGACACAGAAGAGTACTGCATGGCCGTGTGTGGCAGCGCCATGTCC 1064

Query 181    CAAAGTTTACTCAAGACTACCCAGGAACCTCTTGCCCGAGATCCTGTTAACTTCCTACA 240
Sbjct 1065   CAAAGTTTACTCAAGACTACCCAGGAACCTCTTGCCCGAGATCCTGTTAACTTCCTACA 1124

Query 241    ACA 243
Sbjct 1125   ACA 1127
```

この領域では、問い合わせ配列とヒットしたデータベースのエントリ（Sbjctと書かれている）との間の配列アライメントを詳細に見ることができます。

2. blastpの使い方

■2-1 blastpを用いたホモロジー検索を行ってみる

[blastpサービスのページへ移動する](#)

続いては、blastpを用いたアミノ酸配列同士のホモロジー検索について、blastnによる場合と異なる箇所を中心にみていくことにしましょう。

まずは、blastp検索のページへと移動します。

BLAST検索のトップページから、Basic BLAST内の“protein blast”をクリックします。

► 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](#).

| | | |
|---|--|--|
| <input type="checkbox"/> Human | <input type="checkbox"/> Oryza sativa | <input type="checkbox"/> Gallus gallus |
| <input type="checkbox"/> Mouse | <input type="checkbox"/> Bos taurus | <input type="checkbox"/> Pan troglodytes |
| <input type="checkbox"/> Rat | <input type="checkbox"/> Danio rerio | <input type="checkbox"/> Microbes |
| <input type="checkbox"/> Arabidopsis thaliana | <input type="checkbox"/> Drosophila melanogaster | <input type="checkbox"/> 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

すると、以下のようなblastp検索の設定画面へ移動します。

NCBI/BLAST/blastp suite: BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

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

From

To

Or, upload file ファイルを選択 ファイルが...ていません [v](#)

Job Title

Enter a descriptive title for your BLAST search [v](#)

☐ Blast 2 sequences

Choose Search Set

Database Non-redundant protein sequences (nr) [v](#) [v](#)

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. [v](#)

Entrez Query Optional

Enter an Entrez query to limit search [v](#)

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm [v](#)

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

☐ Show results in a new window

[Algorithm parameters](#)

[問い合わせ配列を設定する \(blastp\)](#)

問い合わせ配列の設定はblastnと同じです。

Enter Query Sequence

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

From

To

Or, upload file ファイルを選択 ファイルが...ていません [v](#)

Job Title

Enter a descriptive title for your BLAST search [v](#)

☐ Blast 2 sequences

blastnの時と同様に問い合わせ配列の設定をしましょう。

[データベースを設定する \(blastp\)](#)

データベースの設定も基本的にはblastnと同じです。

Choose Search Set

Database Non-redundant protein sequences (nr) [v](#) [v](#)

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. [v](#)

Entrez Query Optional

Enter an Entrez query to limit search [v](#)

Non-redundant protein sequences (nr)

Reference proteins (refseq_protein)

Swissprot protein sequences (swissprot)

Patented protein sequences (pat)

Protein Data Bank proteins (pdb)

Environmental samples (env_nr)

但し、選択できるデータベースはblastnとは異なります。代表的なものには以下のようなものがあります。

Non-redundant protein sequences (nr)

GenBank含まれている全エントリに書かれたCDS情報を翻訳したもの、RefSeqアミノ酸配列、PDB、SwissProt PIR、PRFから冗長度を除いたもの

Reference proteins (refseq_protein)

NCBI Refseqプロジェクトで作成しているアミノ酸配列

Swissprot protein sequences (swissprot)

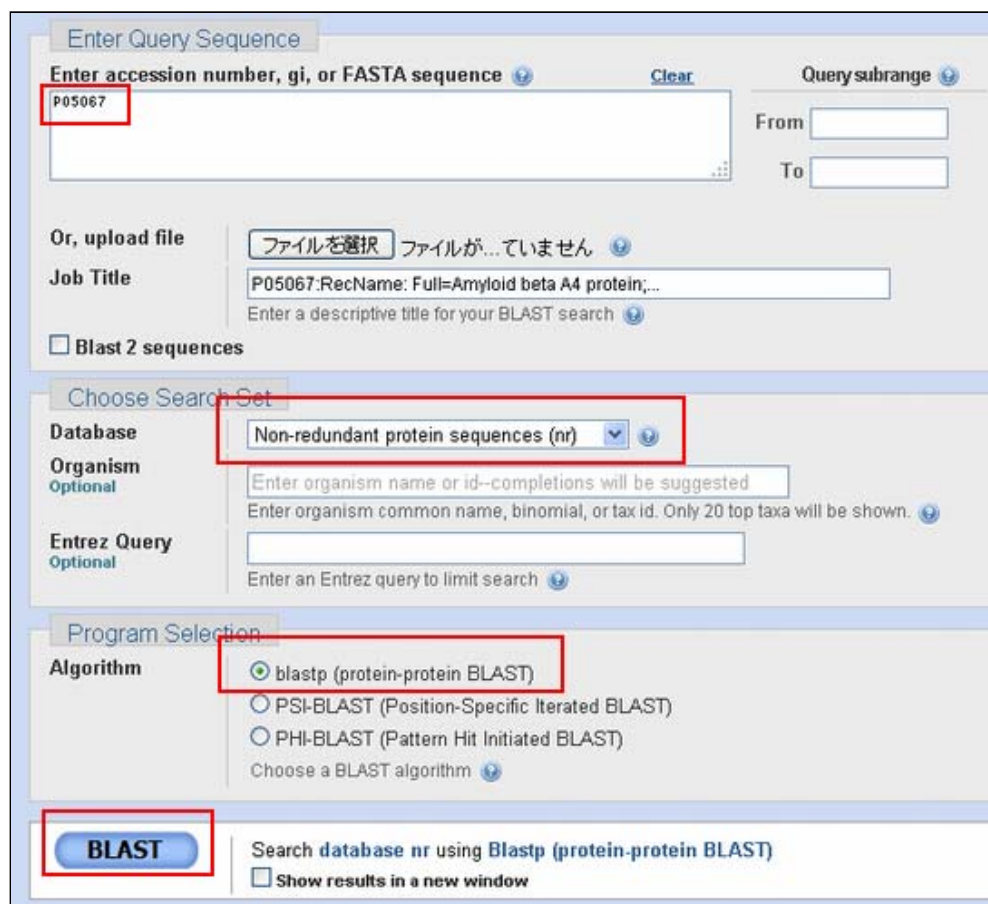
SWISSPROTデータベースに含まれるアミノ酸配列

検索エンジンを選択する (blastp)

続いてアミノ酸配列同士の配列比較時の検索エンジンを設定しましょう。アミノ酸配列同士の比較では通常blastpが用いられますが、よい弱いホモロジーを検索するためにPSI-BLAST, PHI-BLASTといった派生プログラムが作られました。（別のチュートリアルで詳細に紹介します）



以上の設定条件を参考にして、ここでは例としてアクセッション番号P05067を問い合わせ配列とし、nrに対するblastp検索を実行してみましょう。



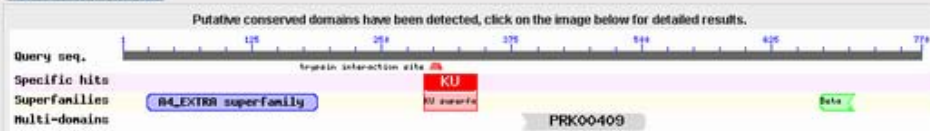
■2-2 blastp検索結果の見方

検索結果の見方 (blastp)

では、続いて検索結果を見ていきましょう。blastnの場合と同様に上から“Graphic Summary”, “Descriptions”, “Alignments”から構成されています。表示される情報もblastnの場合と同じです。

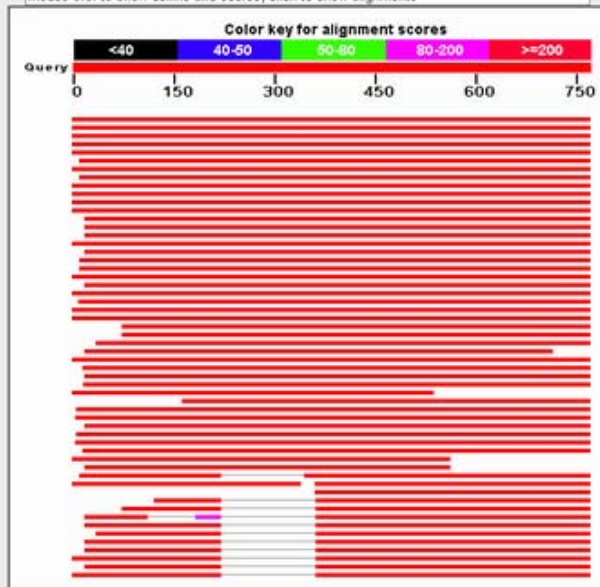
Graphic Summary

Show Conserved Domains



Distribution of 129 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



Descriptions

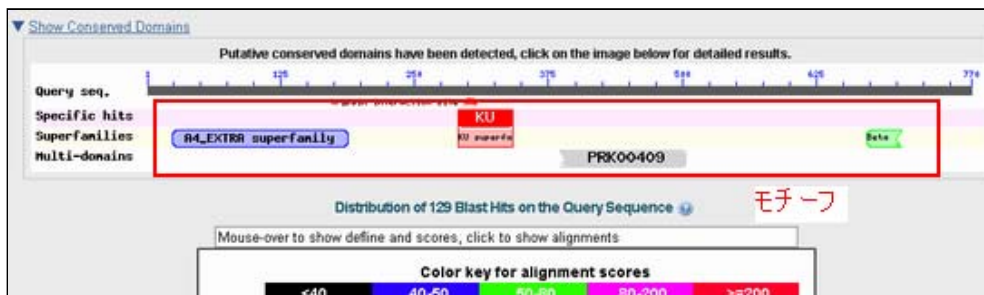
Sequences producing significant alignments:

| | Score (bits) | E Value |
|---|-----------------|------------|
| ref NP_058475.1 sayloid beta A4 protein isoform a precursor ... | 1618 | 0.0 |
| ref NP_031010.1 sayloid beta A4 protein [Pan troglodytes]... | 1608 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1607 | 0.0 |
| ref NP_058537.1 sayloid beta A4 protein [Sus scrofa] >sp P79... | 1424 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1374 | 0.0 |
| ref NP_001066801.1 sayloid beta A4 protein [Canis lupus fami... | 1393 | 0.0 |
| ref NP_062181.1 sayloid beta (A4) precursor protein [Rattus ... | 1393 | 0.0 |
| gb AA01308.1 beta-sayloid precursor protein 770 [Canis fami... | 1393 | 0.0 |
| gb AA02183.1 sayloid-beta precursor protein-like protein lo... | 1387 | 0.0 |
| sp P12023.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1384 | 0.0 |
| ref XP_001493900.1 PREDICTED: similar to sayloid precursor p... | 1384 | 0.0 |
| gb AA02183.1 sayloid precursor protein variant 2 [Sus scrofa] | 1381 | 0.0 |
| ref NP_058816.1 sayloid beta A4 protein isoform b precursor ... | 1365 | 0.0 |
| ref NP_001124408.1 sayloid beta A4 protein isoform d [Homo p... | 1365 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1364 | 0.0 |
| gb AA01310.1 beta-sayloid precursor protein 749 [Stenella c... | 1363 | 0.0 |
| sp E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1362 | 0.0 |
| gb AA01310.1 beta-sayloid precursor protein isoform APP751 ... | 1343 | 0.0 |
| gb AA01310.1 beta-sayloid precursor protein 751 [Canis fami... | 1342 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1332 | 0.0 |
| ref XP_001104639.1 PREDICTED: similar to sayloid beta A4 pro... | 1312 | 0.0 |
| ref NP_058639.1 sayloid beta A4 protein [Gallus gallus] >sb ... | 1315 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1313 | 0.0 |
| gb AA01310.1 beta-sayloid precursor protein 749 [Stenella c... | 1309 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1292 | 0.0 |
| ref NP_001124408.1 sayloid beta A4 protein isoform f precurs... | 1293 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1292 | 0.0 |
| ref XP_001104639.1 PREDICTED: similar to sayloid beta A4 pro... | 1272 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1245 | 0.0 |
| ref NP_001082039.1 beta-sayloid precursor protein A [Xenopus... | 1233 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1222 | 0.0 |
| ref NP_001066801.1 sayloid beta (A4) precursor protein [Xeno... | 1214 | 0.0 |
| ref XP_001124408.1 sayloid beta A4 protein isoform d [Homo p... | 1193 | 0.0 |
| gb AA01310.1 beta-sayloid precursor protein 749 [Stenella c... | 1112 | 0.0 |
| gb AA01310.1 beta-sayloid precursor protein 749 [Stenella c... | 1104 | 0.0 |
| gb E05301.1 A4_MOUSE RefName: FullA4sayloid beta A4 protein; A... | 1093 | 0.0 |

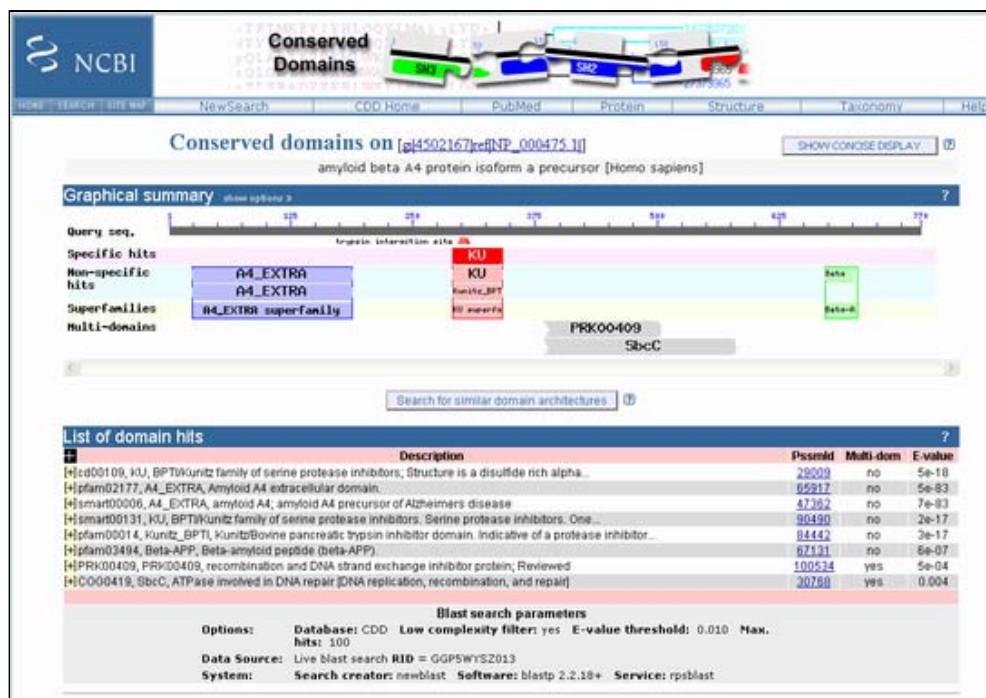
▼ Alignments ☐ Select All [Get selected sequences](#) [Distance tree of results](#)

>ref|NP_004751.1| **UG** amyloid beta A4 protein isoform a precursor [Homo sapiens]
 sp|P55357.3|A4_HUMAN **G** RecName: Full=amyloid beta A4 protein; AltName: Full=Alzheimer disease amyloid protein; AltName: Full=ABPP; AltName: Full=APP3;
 Short=APP; AltName: Full=ProA4; AltName: Full=Cerebral vascular amyloid peptide; Short=CYP; AltName: Full=Protease nexin-II; Short=PN-II; Contains: RecName: Full=Soluble APP-alpha; Short=S-APP-alpha; Contains: RecName: Full=Soluble APP-beta; Short=S-APP-beta; Contains: RecName: Full=C89; Contains: RecName: Full=Beta-amyloid protein 42; AltName: Full=Beta-APP42; Contains: RecName: Full=Beta-amyloid protein 40; AltName: Full=Beta-APP40; Contains: RecName: Full=C83; Contains: RecName: Full=P9(42); Contains: RecName: Full=P9(40); Contains: RecName: Full=Gamma-secretase C-terminal fragment 53; AltName: Full=Gamma-CTF(53); AltName: Full=amyloid intracellular domain 59; Short=AICD-59; Short=AID(59); Contains: RecName: Full=Gamma-secretase C-terminal fragment 57; AltName: Full=Gamma-CTF(57); AltName: Full=amyloid intracellular domain 57; Short=AICD-57; Short=AID(57); Contains: RecName: Full=Gamma-secretase C-terminal fragment 50; AltName: Full=Gamma-CTF(50); AltName: Full=amyloid intracellular domain 50; Short=AICD-50; Short=AID(50); Contains: RecName: Full=C81; Flags: Precursor
 sh|AA059507.1| **G** amyloid-beta protein [Homo sapiens]
 P-1_smc sequence 111111
 Length=770
 GENE ID: 351 APP amyloid beta (A4) precursor protein [Homo sapiens]
 (Over 100 PubMed links)
 Score = 1810 bits (4180), Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 770/770 (100%), Positives = 770/770 (100%), Gaps = 0/770 (0%)
 Query 1 MLPGALLLLAANTARALEVPTDGNAGLLAEPTAMFCGRLNHHYKNGKWCSPGSK 60
 Sbjct 1 MLPGALLLLAANTARALEVPTDGNAGLLAEPTAMFCGRLNHHYKNGKWCSPGSK 60
 Query 61 TCIDTKGILQYDQVYPELQITNVEANGPYTIQNGKRGKCKTHPHFYIPYRLVG 120
 Sbjct 61 TCIDTKGILQYDQVYPELQITNVEANGPYTIQNGKRGKCKTHPHFYIPYRLVG 120
 Query 121 EFYSQALLYPKQKFLHGRMVCETHLHHYVAKETCEKSTNLHYQMLPGIDKFR 180
 Sbjct 121 EFYSQALLYPKQKFLHGRMVCETHLHHYVAKETCEKSTNLHYQMLPGIDKFR 180
 Query 181 QVEFYDCLAEEDQVYDGADEEDQGVVWGGADTVADGSEKYYVEAEAEAEAEVEE 240
 Sbjct 181 QVEFYDCLAEEDQVYDGADEEDQGVVWGGADTVADGSEKYYVEAEAEAEAEVEE 240
 Query 241 EADGDEDEDGVEEAEAEPEYTERITSIATTTTTTSEVEYVREVCSEDAETGPC 300
 Sbjct 241 EADGDEDEDGVEEAEAEPEYTERITSIATTTTTTSEVEYVREVCSEDAETGPC 300
 Query 301 RAMISRYVFDYEDKCAPFFYGGCGGNRNFTTEYQNAVCGSAMGSLKLTDEPLAD 360
 Sbjct 301 RAMISRYVFDYEDKCAPFFYGGCGGNRNFTTEYQNAVCGSAMGSLKLTDEPLAD 360
 Query 361 PVKLPTTAASTPQVQKYLETPQGENHAFQAKERLEAKHREKSYNREVEAEERDA 420
 Sbjct 361 PVKLPTTAASTPQVQKYLETPQGENHAFQAKERLEAKHREKSYNREVEAEERDA 420
 Query 421 KNLPAKAKAVIQHFQEKVESLEGEAANERQOLYETHMARYEAMLRPRRLAENYITAL 480
 Sbjct 421 KNLPAKAKAVIQHFQEKVESLEGEAANERQOLYETHMARYEAMLRPRRLAENYITAL 480
 Query 481 QAVPPFRPHVFNMLKYYRAEDKDRHTLKHFEHYRMDPKKAQIRSDVMTLSVYIER 540
 Sbjct 481 QAVPPFRPHVFNMLKYYRAEDKDRHTLKHFEHYRMDPKKAQIRSDVMTLSVYIER 540

ただし“Graphic Summary”に違いがあります。



“Graphic Summary”の上部にQuery配列が持つconserved domain（モチーフ）が示されており、これらをクリックすると、“Conserved Domain”データベースの該当するエントリへと移動することです。



それ以外に主だった違いは見当たりません。

作成日：2008年10月29日

All Rights Reserved, Copyright(C) 1997-2008 Japan Science and Technology Agency(JST)