

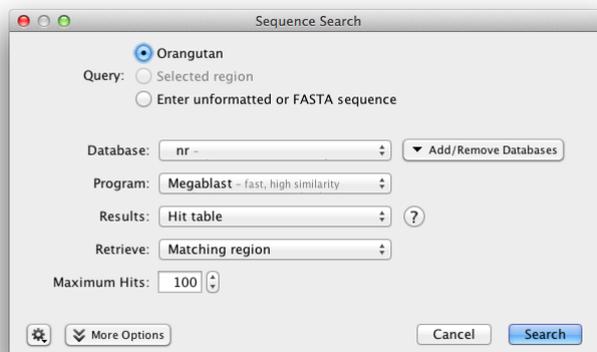
Geneious Prime でシーケンス解析



第 2 回 BLAST 検索

BLAST (Basic Local Alignment Search Tool)は、シーケンス間で部分的に類似する領域を見つけるツールです。ヌクレオチドやタンパク質のシーケンスをデータベースと比較して、一致した部分の統計的有意性を計算します。新規/未知シーケンスの同定、機能・進化的関係性の推測、遺伝子ファミリーのメンバーの同定などに使用することができます。

Geneious から BLAST 検索をするには、クエリーシーケンス(複数選択可)を選択し、ツールバーの **BLAST** ボタンをクリックするか、**右クリック** → **BLAST** を選択します。



Database で使用したいデータベースを、**Program** でどのような BLAST 検索を行いたいかを選択します。デフォルトでは選択したクエリーシーケンスとデータベースのタイプに最も適した BLAST アルゴリズムが自動的に選択されています(付録参照)。

Search ボタンで検索が始まり、クエリーシーケンスと同じフォルダ内にサブフォルダが作成されて結果が保存されます。

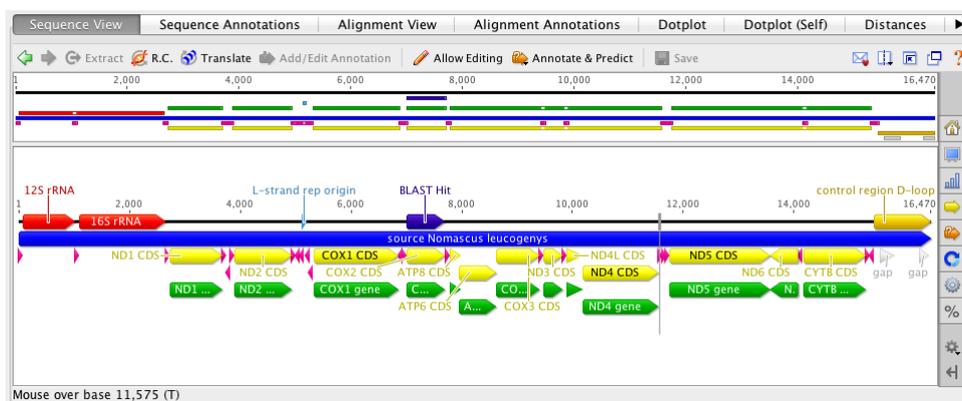
デフォルトでは、クエリー毎にヒットしたデータベースのシーケンスがアラインメントされた結果が、表(Hit table)として表示されます。

| Bit-Score | E Value | % Pairwise Id... | Grade | Name | Description | Sequence Le... | Hit start | Hit end |
|-----------|-----------|------------------|--------|--------------|--|----------------|-----------|---------|
| 311.612 | 1.18e-107 | 100.0% | 100.0% | NP_001014408 | ribonuclease pancreatic precursor [Bos t... | 150 | 1 | 150 |
| 313.153 | 2.63e-107 | 100.0% | 100.0% | CDG32088 | TPA: ribonuclease A C2 [Bos taurus] | 150 | 66 | 215 |
| 310.071 | 3.83e-107 | 99.3% | 99.7% | XP_005901936 | PREDICTED: ribonuclease pancreatic [Bo... | 150 | 1 | 150 |
| 270.396 | 9.61e-92 | 100.0% | 92.7% | 1C08_A | Chain A, Bovine Pancreatic Ribonuclease ... | 128 | 1 | 128 |
| 269.24 | 7.28e-91 | 86.1% | 91.0% | AAA72757 | RNase A [synthetic construct] | 151 | 6 | 156 |
| 266.929 | 4.03e-90 | 91.3% | 95.7% | XP_005960934 | PREDICTED: ribonuclease pancreatic-like... | 150 | 1 | 146 |
| 261.922 | 2.22e-88 | 100.0% | 91.3% | P61824 | RecName: Full=Ribonuclease pancreatic;... | 124 | 1 | 124 |
| 261.922 | 2.40e-88 | 100.0% | 91.3% | CA837066 | artificial [synthetic construct] | 124 | 5 | 128 |
| 261.536 | 2.65e-88 | 100.0% | 91.3% | CAA33801 | unnamed protein product [Bos taurus] | 124 | 2 | 125 |
| 260.381 | 8.76e-88 | 99.2% | 90.9% | 1E1E_A | Chain A, Crystal Structure Of F120w Muta... | 124 | 1 | 124 |
| 260.381 | 9.25e-88 | 99.2% | 90.9% | 3DH6_A | Chain A, Crystal Structure Of Bovine Panc... | 124 | 1 | 124 |
| 260.381 | 9.25e-88 | 99.2% | 90.9% | 3D17_A | Chain A, Crystal Structure Of Bovine Panc... | 124 | 1 | 124 |
| 260.381 | 9.25e-88 | 99.2% | 90.9% | 3D18_A | Chain A, Crystal Structure Of Bovine Panc... | 124 | 1 | 124 |
| 260.381 | 9.25e-88 | 99.2% | 90.9% | 3D1C_A | Chain A, Crystal Structure Of Bovine Panc... | 124 | 1 | 124 |
| 260.381 | 9.89e-88 | 99.2% | 90.9% | 4WYN_A | Chain A, The Crystal Structure Of The A1... | 124 | 2 | 125 |
| 265.003 | 1.11e-87 | 100.0% | 91.7% | 3MMR_A | Chain A, Crystal Structure Of Ribonuclease | 125 | 120 | 254 |

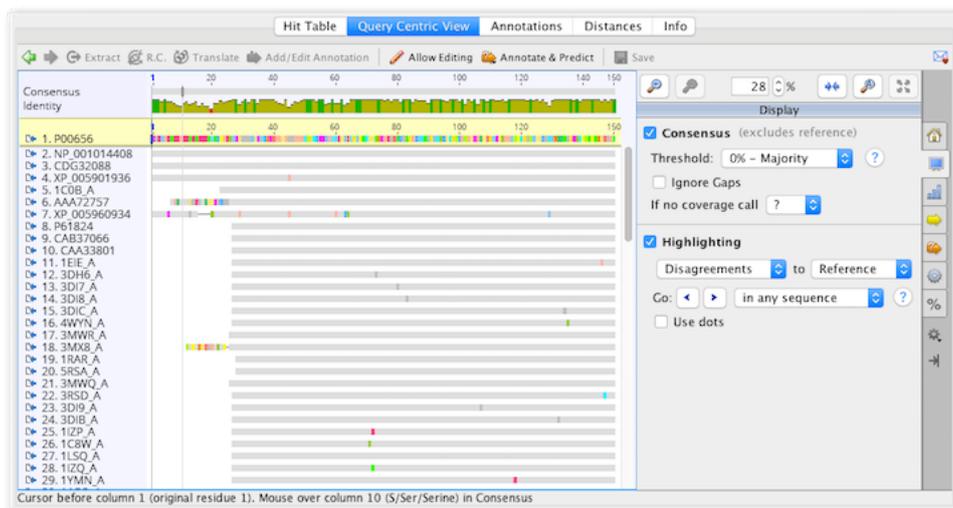
結果の並び順は、カラムのヘッダーをクリックすることで変更することができます。

E-value は、結果が偶然である予想頻度を示す値で、低いほど結果が偶然ではないことを示しています。**% Pairwise Identity** は、クエリーとヒットしたシーケンスのアラインメント領域で、どれくらいシーケンスが一致するかを示しています。シーケンス長の違いによる差異は考慮されないことに注意が必要です。**Grade** は、Query Coverage、E-value、Identity から計算された Geneious 独自の値で、これでソートすると、一番長く、一番相同性が高いヒットを一番上に表示することができます。

アラインメントビューでははじめ、クエリーとヒットシーケンス間のアラインメント領域のみが表示されています。完全なシーケンスとアノテーションは、**File** → **Download Document** をクリックするか、アラインメントビューの上にある **Download** ボタンで取得することができます。



Query-centric view ではクエリーに対するすべてのヒットシーケンスのアラインメントを1つにして視覚化することができます。クエリーシーケンスは一番上に黄色の背景で表示されます。



複数のクエリーシーケンスをバッチで BLAST 検索する場合、**Result** の設定で **bin into "has hit" vs "no hit" in database** を選択すると、クエリーをデータベースで「ヒットした」シーケンスリストと「ヒットしなかった」シーケンスリストに分けることができます。

付録

ヌクレオチドのデータベース

| Database | Nucleotide searches |
|---|--|
| Nucleotide collection (nr) | All non-redundant GenBank+EMBL+DDBJ+PDB sequences (no EST, STS, GSS or HTGS sequences) |
| 16S ribosomal RNA | 16S rRNA sequences from bacteria and archaea |
| 18S ribosomal RNA | 18S rRNA sequences (Fungal) |
| 28S ribosomal RNA | 28S rRNA sequences (Fungal) |
| Environmental samples (env_nt) | Nucleotide sequences from large environmental sequence projects |
| Expressed sequence tags (est) | Database of GenBank + EMBL + DDBJ sequences from EST Divisions |
| EST human | Human subset of est |
| EST mouse | Mouse subset of est |
| EST others | Non-Human, non-mouse subset of est |
| Genomic Survey Sequences (gss) | Genome Survey Sequence, includes single-pass genomic data, exon-trapped sequences, and Alu PCR sequences |
| High Throughput Genomic Sequences (htgs) | Unfinished HTGS: phases 0, 1 and 2 (finished, phase 3 HTG sequences are in nr) |
| Human ALU repeat elements (alu_repeats) | A small database of Human ALU repeat elements |
| Human RefSeqGene (RefSeq_Gene) | NCBI transcript reference sequences from human |
| Internal transcribed spacer region (ITS) | ITS region from fungal type and reference material |
| NCBI Genomes (chromosome) | Complete genomes and chromosomes from the NCBI Reference Sequence project. |
| NCBI Reference Genomic Sequences (refseq_genomic) | Genomic Reference sequences |
| Patented Protein Sequences (pat) | Nucleotide sequences derived from the Patent division of GenBank |
| Protein Data Bank (PDB) | Sequences derived from the 3D-structures of proteins from PDB |
| Reference RNA (refseq_rna) | NCBI Transcript Reference Sequences |
| RefSeq Representative genomes | Best quality and minimum redundancy genomes from NCBI Refseq Genomes |
| Sequence Tagged Sites (dbsts) | Database of GenBank+EMBL+DDBJ sequences from STS Divisions |
| WGS Human | Whole-genome shotgun contigs for Homo sapiens |

タンパク質のデータベース

| Database | Protein searches |
|-------------------------------------|--|
| Nucleotide collection (nr) | All non-redundant GenBank coding region (CDS) translations+PDB+SwissProt+PIR+PRF |
| Metagenomic proteins (env_nr) | Translations of sequences in env_nt |
| Patented Protein Sequences (pat) | Protein sequences derived from the Patent division of GenBank |
| Protein Data Bank (PDB) | Sequences derived from 3D structure Brookhaven PDB |
| Reference Proteins (refseq_protein) | NCBI protein reference sequences |
| UniProtKB/SwissProt | Non-redundant protein sequences information from EMBL |

BLAST アルゴリズム

blastn: ヌクレオチドのデータベースに対してヌクレオチドのクエリー。遠縁でも検出だが低速。

Discontiguous Megablast: 異なる種間でも高感度に検出。中間。

Megablast: 相同性の高いものに限定して検出。高速。

blastp: タンパク質のデータベースに対してアミノ酸のクエリー

blastx: タンパク質のデータベースに対して 6 フレームで翻訳したヌクレオチドのクエリー

tblastn: 6 フレームで翻訳したヌクレオチドのデータベースに対してアミノ酸のクエリー

tblastx: 6 フレームで翻訳したヌクレオチドのデータベースに対して 6 フレームで翻訳したヌクレオチドのクエリー

深みへ

More Options ボタンから、さらに詳細な設定を行うことができます。

例えば **Max E-value** を 1e-3 などとすると偶然のヒットではなく統計的に有意と考えられる結果のみを表示することができます。

Entrez query では、”**sequence from type[filter]**”と指定すると Type material のみを、”**all[filter] NOT uncultured[filter] NOT “environmental”[filter] NOT unclassified[filter]**”とすると、未培養、環境サンプル、未分類の結果を除外できます。

もっと深みへは[コチラ\(BLAST Help\)](#)