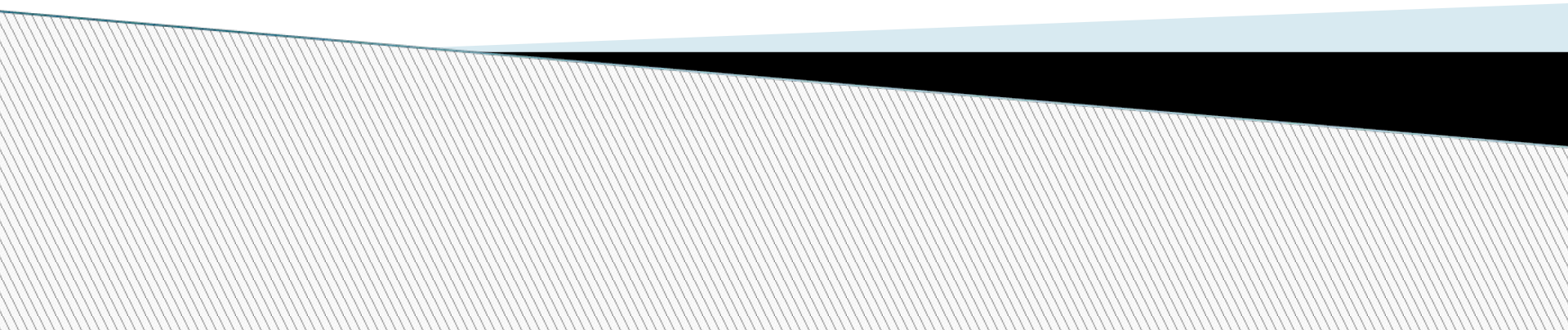


BLAST:

Basic Local Alignment Search Tool



BLAST – алгоритм для нахождения участков локального сходства между последовательностями.

Алгоритм сравнивает входную последовательность с последовательностями в базе данных, ищет сходные последовательности в базе данных и оценивает статистическую значимость находок.

Protein BLAST: поиск аминокислотной последовательности в базе данных белков

Алгоритмы

-blastp

-psi-blast

-phi-blast

Здесь описан интерфейс, установленный на «родине» BLAST: National Center for Biotechnology Information (NCBI) в США,
<http://blast.ncbi.nlm.nih.gov/>

<http://blast.ncbi.nlm.nih.gov/> → protein blast

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

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(s), gi(s), or FASTA sequence(s)

Clear Query subrange

From

To

Or, upload file

Выберите файл Файл не выбран

Job Title

Enter a descriptive title for your BLAST search

Align two or more 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.

Exclude

Optional

Models (XM/XP) Uncultured/environmental sample sequences

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

Show results in a new window

Algorithm parameters

ВВОДИМ ПОСЛЕДОВАТЕЛЬНОСТЬ

база данных

организм (если надо ограничить)

дополнительные параметры

Параметры сервиса

максимальный
размер выдачи

Algorithm parameters

General Parameters

Max target sequences: 100
Select the maximum number of aligned sequences to display

Short queries: Automatically adjust parameters for short input sequences

Expect threshold: 10
порог на E-value

Word size: 3

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62
параметры выравнивания

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment
борьба с «участками малой сложности»

Filters and Masking

Filter: Low complexity regions

Mask: Mask for lookup table only
 Mask lower case letters

Участок малой сложности

Ищем: белок P02929

если отключить “Compositional adjustments” и фильтр, то одной из находок (18-ой от начала) будет следующее:

Query: P02929 TONB_ECOLI; Subject: Q95P09 TSEP_GLOMM

Score = 63.5 bits (153), Expect = 1e-09

Identities = 32/76 (42%), Positives = 47/76 (62%), Gaps = 6/76 (8%)

```
Query  56  EPPQAVQPPPEPVVEPEPEPEPEPIPEP-PKEAPVVIEKPKPKPKPKPKPVKKVQEQPKRDV  114
      EP  +P PEP  EPEPEPEP PEP P+  P  +P+P+P+P+P+P  + + +P+ +
Sbjct  243  EPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEP  302

Query  115  KP-----VESRPASPF  125
      +P      ES+P S F
Sbjct  303  EPEPQPEPESKPNSLF  318
```

в исходном белке имеется участок, содержащий очень много пролина и глутаминовой кислоты

Данное выравнивание **не свидетельствует о гомологии**, несмотря на хорошее значение E-value (10^{-9})

Переход к текстовому виду

Чтобы увидеть выдачу самой программы (а не его обработку интерфейсом), можно поступить так:

выбираем formatting options

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

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

Formatting options

[Reformat](#)

Show as Advanced View Use old BLAST report format [Reset form to defaults](#)

Alignment View

Display Graphical Overview Linkout Sequence Retrieval NCBI-gi

Masking Character: Color:

Limit results Descriptions: Graphical overview: Alignments:

Organism Type common name, binomial, taxid, or group name. Only 20 top taxa will be shown.
 Exclude [+](#)

Entrez query:

Expect Min: Expect Max:

Percent Identity Min: Percent Identity Max:

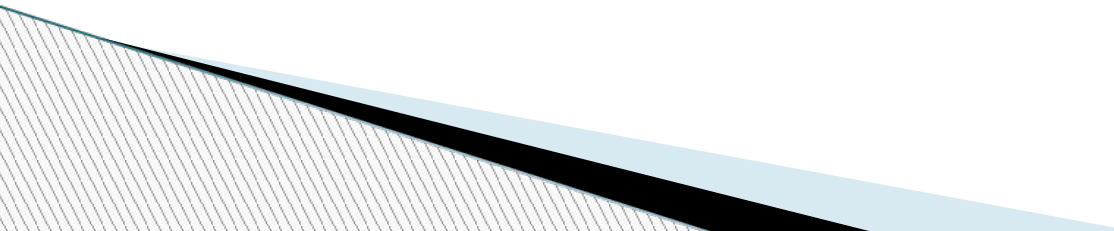
Format for PSI-BLAST with inclusion threshold:

подтверждаем выбор

Что выдает BLAST?

Набор последовательностей, сходных с входной последовательностью

для каждой находки приведены

- E-value (“Expect”), Bit Score и Score
 - процент идентичности, сходства (Positives) и пробелов (Gaps) в выравнивании
 - информация о найденной последовательности
- 

Выравнивание, выданное BLAST

Length=129 Длина найденного белка

Score = 78.6 bits (192), Expect = 9e-15, Method: Compositional matrix adjust.
Identities = 34/73 (47%), Positives = 50/73 (68%), Gaps = 0/73 (0%)

Query 17 YRLEE VQKHNSQSTWII VHHRIYDITKFLDEHPGGEEVLREQAGGDATENFEDVGHSTD 76
 Y EEV +H W+I++ ++Y+I+ ++DEHPGGEEV+ + AG DATE F+D+GHS +
Sbjct 11 YTHEEVAQHTTHDDLWVILNGKVYNISNYIDEHPGGEEVILDCAGTDATEAFDDIGHSDE 70

Query 77 ARALSETFIIGEL 89
 A + E IG L
Sbjct 71 AHEILEKLYIGNL 83

Вес в битах Вес E-value

Число совпадений Длина выравнивания

E-value – ожидаемое количество **случайных** находок с таким же и лучшим Score (в той же базе данных, с теми же параметрами):

$$E\text{-value} = Kmn \cdot e^{-\lambda S}$$

S – Score (вес)

m – длина исходной последовательности

n – размер базы данных (суммарная длина всех последовательностей)

K и λ – параметры

$$E\text{-value} = \frac{mn}{2^B} \text{ Bit score (вес в битах)}$$

$$E\text{-value} = \frac{mn}{2^B} \quad \text{Выражение E-value через биты}$$

Чем **меньше** E-value, тем **больше** значимость находки.