

Семинар 4. Программа BLAST. Поиск схожих последовательностей. Определение последовательностей из метагеномного исследования с помощью BLAST.

Изучить web-interface программы BLAST

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Показать существования различных программ семейства BLAST, но мы будем использовать только protein и nucleotide.

При запуске BLAST выбрать кнопку “Show results in a new window”

Расширить “[Algorithm parameters](#)” и изучить параметры. Дать интуитивное (строгое будет на лекции) объяснение матрицы счета BLOSUM.

Объяснить

- Query/sequence start and query/sequence end (могут быть полные выравнивания, могут быть частичные)
- Percent Identity
- Score
- E-value

1. Protein BLAST

> Unknown Sequence

```
MLSVRVAAAVVRALPRRAAALVSRNALGSSFIAARNFHASNTHLQKTGTAEMSSILEERILGADTSVDLEE
TGRVLSIGDGIARVHGLRNVQAEEMVEFSSGLKGMSLNLEPDNVGVVVFNGNDALIKEGDIVKRTGAIVDG
PIGSKTRRRVGLKAPGIIPRISVREPMQTGIKAVDSLVPIGRGQRELIIGDRQTGKTSIAIDTIINQKRF
NDGSDEKKKLYCIYVAIMAMAQKRSTVAQLVKRLTDADAMKYTIVVSATASDAAPLQYLAPYSGCSMGGEYFRD
NGKHALIYDDLSKQAVAYRQMSLLRRPPGREAYPGDVFYLSRLLERA AKMNDAFGGGSLTALPVIET
QAGDVSAIYPTNVISITDGGQIFLETIFYKGIRPAINVGLSVSRVGS PAINVGLSVSRVGSAAQTRAMKQVAGTMKLELAQYRE
VAAFAQFGSQVALDAATQQLSRGVRLTELLKQGQYSPMAIEEQ
```

Изучить выравнивание: увидеть, где замены и вставки; оно полное или оборванное.

2. Nucleotide BLAST

Изучается на примере метагеномного исследования. Данные берутся из статьи

Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics

<https://www.nature.com/articles/s41598-018-30515-5>

На гитхабе есть искусственно сгенерированный набор ридов для тестирования

<https://github.com/Ales-ibt/Metagenomic-benchmark>

Скачать первый файл с полутора миллионами ридов (20Mb)

Metagenomic-benchmark/**datasets_16SrRNA**/v3v4_lib1.1.fasta.gz

```
$ less v3v4_lib1.1.fasta | wc -l
```

525000

Поизучать вручную через Nucleotide BLAST, из каких видов последовательности, и полное или частичное выравнивание с геном 16S RNA (вспомнить, из чего состоит рибосома). Можно прогнать штук 5 последовательностей.

Рассказать, что есть command-line BLAST (прокрутить вниз страницу <https://blast.ncbi.nlm.nih.gov/Blast.cgi> до “**Standalone and API BLAST**” и нажать на Download BLAST.

Посмотреть доступность различных программ и баз данных для скачивания (НИЧЕГО НЕ СКАЧИВАЕМ НА СЕМИНАРЕ).

Extra Reading on Metagenomic Studies:

From Genomics to Metagenomics: First Steps

<https://www.nap.edu/read/11902/chapter/5>

Human Microbiome Dataset

<https://ubiome.com/blog/post/ubiome-largest-human-microbiome-dataset/>