



BLAST

14 апреля 2026

Спирин Сергей Александрович
Алешина Юлия Александровна

Повторение: сходство и гомология

- ❖ Гомология — общность происхождения
 - ✓ У гомологичных нуклеотидных последовательностей/белков можно говорить о парах гомологичных нуклеотидов/ амк остатков
 - ✓ В эволюционно правильном выравнивании все остатки в одной колонке гомологичны друг другу

Предок	TATGCGAAT-G-CCCTGAA
Потомок1	TATGCATA-CG-C---GAA
Потомок2	TAT--CAAT-GCCCTGGTA
Потомок3	TAT--CAAG-GCCATGGGA

Повторение: сходство и гомология

- ❖ Признак гомологии — сходство последовательностей
 - ✓ Для выявления сходства последовательности надо выровнять
 - ✓ Подбирают оптимальное выравнивание, то есть имеющее наибольший вес
 - ✓ Оптимальное выравнивание существует для любых последовательностей, в том числе негомологических
 - ✓ Для двух последовательностей можно рассматривать или **глобальное**, или **локальное** выравнивание

Повторение: Алгоритмы и программы парного выравнивания

❖ Оптимальное **глобальное** выравнивание: алгоритм Нидлмана – Вунша

Needleman & Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins". *Journal of Molecular Biology*. **48** (3): 443–53

В оригинальной работе предлагалось оценивать выравнивание с линейными штрафами за гэпы (одинаковый штраф за каждый гэп) и описывался алгоритм нахождения оптимального выравнивания с таким весом. Позднее был предложен вес с аффинными штрафами и алгоритм модифицирован для этой ситуации, ещё позднее введены матрицы замен.

✓ Программы в EMBOSS: `needle` и `stretcher`

❖ Оптимальное **локальное** выравнивание: алгоритм Смита – Уотермена (=Смита – Ватермана)

Smith & Waterman (1981). "Identification of Common Molecular Subsequences". *Journal of Molecular Biology*. **147** (1): 195–197

✓ Программа в EMBOSS: `water`

❖ Заданное число лучших выравниваний: алгоритм Уотермена – Эггерта

(Waterman & Eggert (1987). *Journal of Molecular Biology*. **197** (4): 723-728)

✓ Программа в EMBOSS: `matcher`

Повторение: Алгоритмы и программы парного выравнивания

Параметры этих программ = параметры вычисления веса:

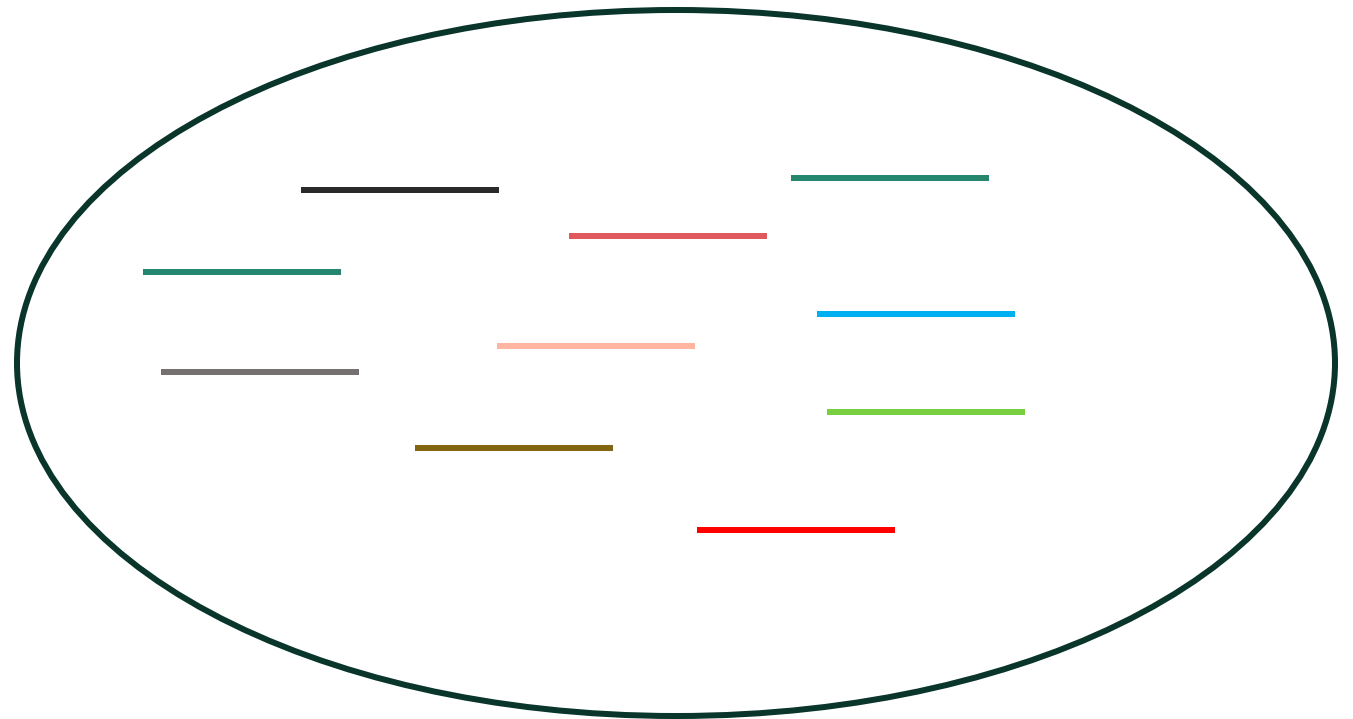
- ❖ матрица замен (для белков) или веса за совпадение и несовпадение (для ДНК/РНК)
- ❖ штраф за первый гэп инделя (gap opening penalty)
- ❖ штраф за следующие гэпы инделя (gap extension penalty)

Что это за последовательность?

MVHLTPEEKSAVTALWGKVNDEVGGGEALGRLLVVYPWTQRFF
ESFGDLSTPDAVMGNPVKAHGKKVLGAFSDGLAHL DNLKGTFA
TLSELHCDKLHVDPENFRLLGNVLCVLAH HFGKEFTPPVQAAY
QKVVAGVANALAHKYH

База данных последовательностей

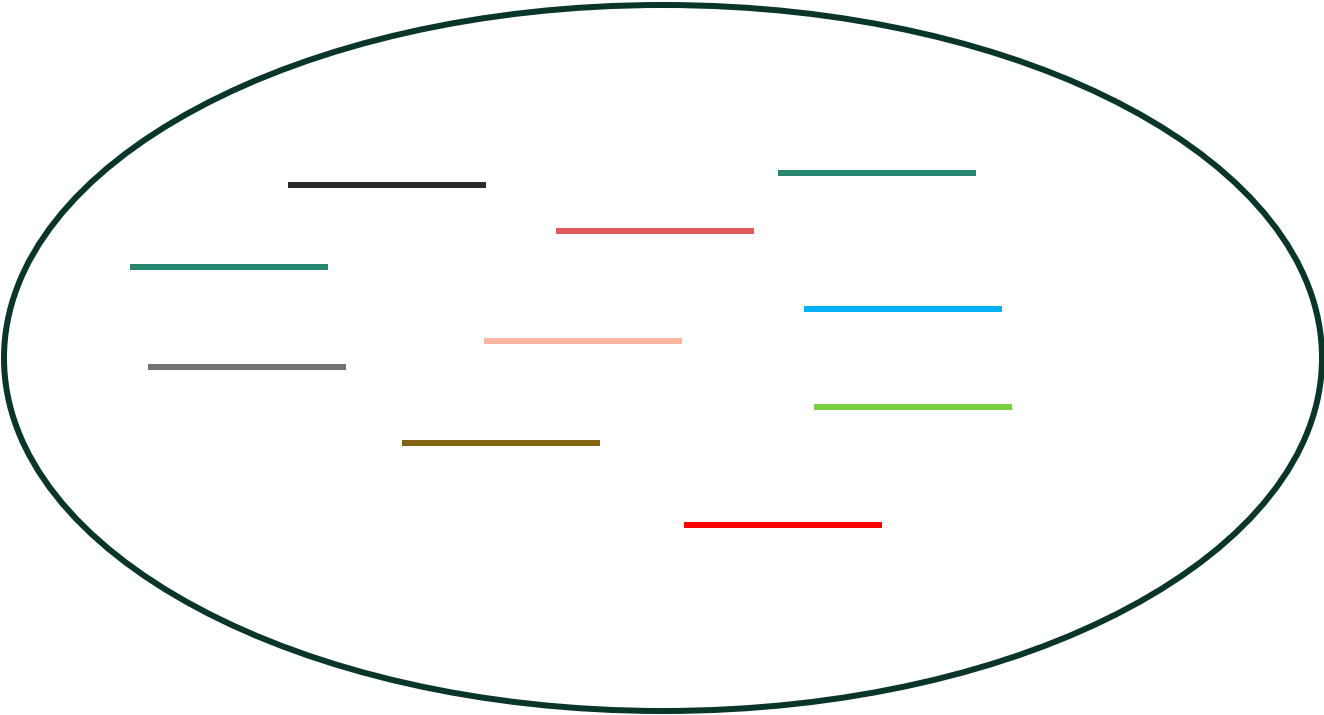
MVHLTPEEKSAVTALWGKLVNDEVGGEALGRLLVVYP
WTQRRFFESFGDLSTPDAVMGNPVKAHGKKVLGAFSDG
LAHLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLC
VLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH



Какие Вы знаете?

База данных последовательностей

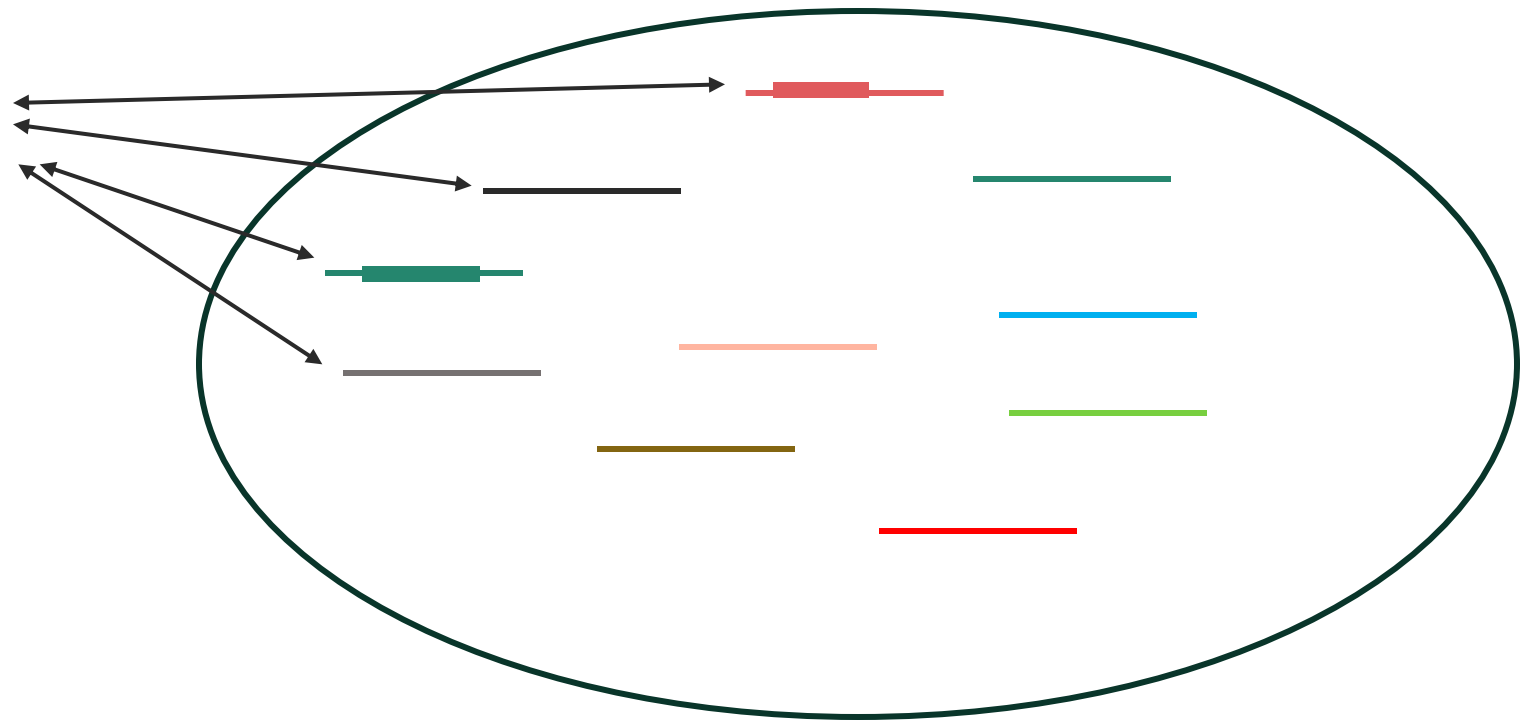
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVWYP
WTQRRFFESFGDLSTPDAVMGNPVKAHGKKVLGAFSDG
LAHLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLC
VLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH



База данных последовательностей

Как будем сравнивать с последовательностями из базы?

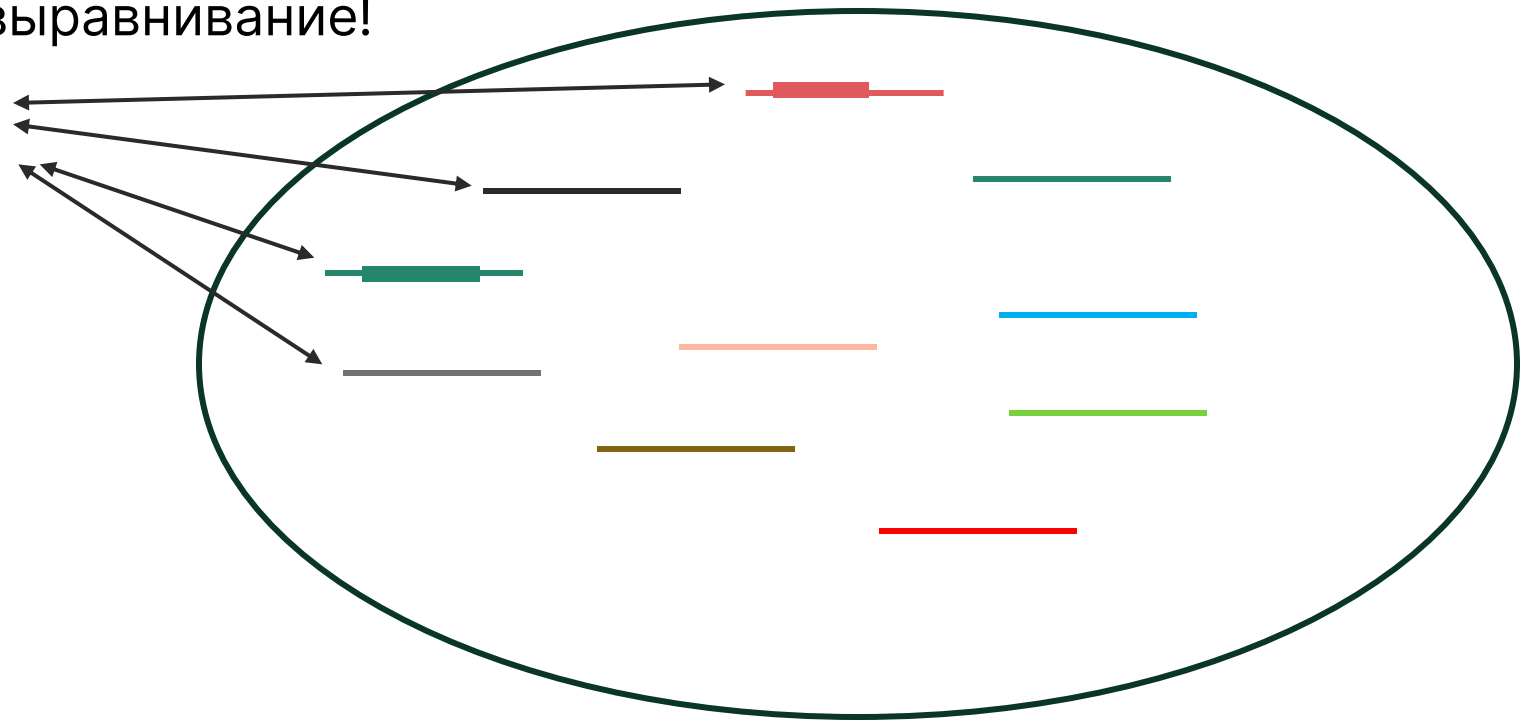
MVHLTPEEKSAVTALWGKLVNDEVGGEALGRLLVWYP
WTQRRFFESFGDLSTPDAVMGNPVKAHGKKVLGAFSDG
LAHLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLVC
VLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH



База данных последовательностей

Локальное
выравнивание!

MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYP
WTQRRFFESFGDLSTPDAVMGNPVKAHGKKVLGAFSDG
LAHLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLVC
VLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH



Идея поиска гомологов в банке последовательностей

- ❖ На входе — последовательность, для которой хочется найти гомологичные («запрос»), и банк последовательностей
- ❖ Выравниваем запрос с каждой последовательностью банка, посчитаем веса этих парных выравниваний
- ❖ Отберём те последовательности банка («находки»), для которых вес **существенно выше, чем мог бы быть по случайным причинам.**

Basic Local Alignment Search Tool

- ❖ BLAST – это 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**

BLAST – это семейство программ

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

ElasticBLAST 1.0.0 is Now available!

ElasticBLAST version 1.0.0 has support for faster cheaper disks at AWS and better supports Kubernetes on GCP!

Mon, 09 Jan 2023

[More BLAST news...](#)

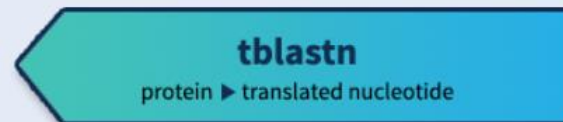
Web BLAST



Nucleotide BLAST
nucleotide ► nucleotide



blastx
translated nucleotide ► protein



tblastn
protein ► translated nucleotide



Protein BLAST
protein ► protein

BLAST – почему локальное выравнивание?

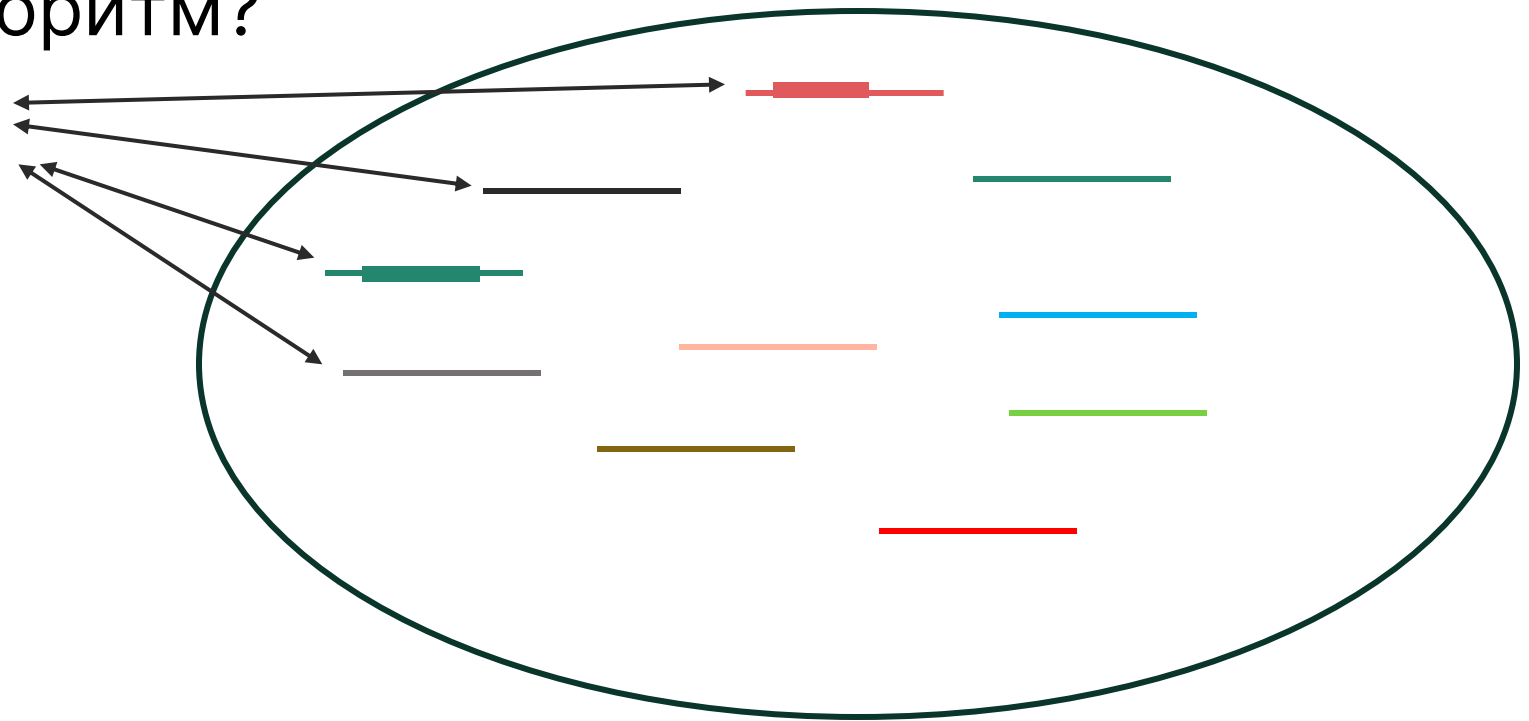
BLAST – почему локальное выравнивание?

- ❖ Глобальное выравнивание следует применять только в случае заранее известной гомологии последовательностей по всей длине.
- ❖ Часто у последовательностей гомологичны только отдельные части (примеры: гомеобелки, полипротеины, ...)
- ❖ Если про последовательности заранее ничего не известно, то более информативным будет локальное выравнивание.

База данных последовательностей

Локальное выравнивание!
Алгоритм?

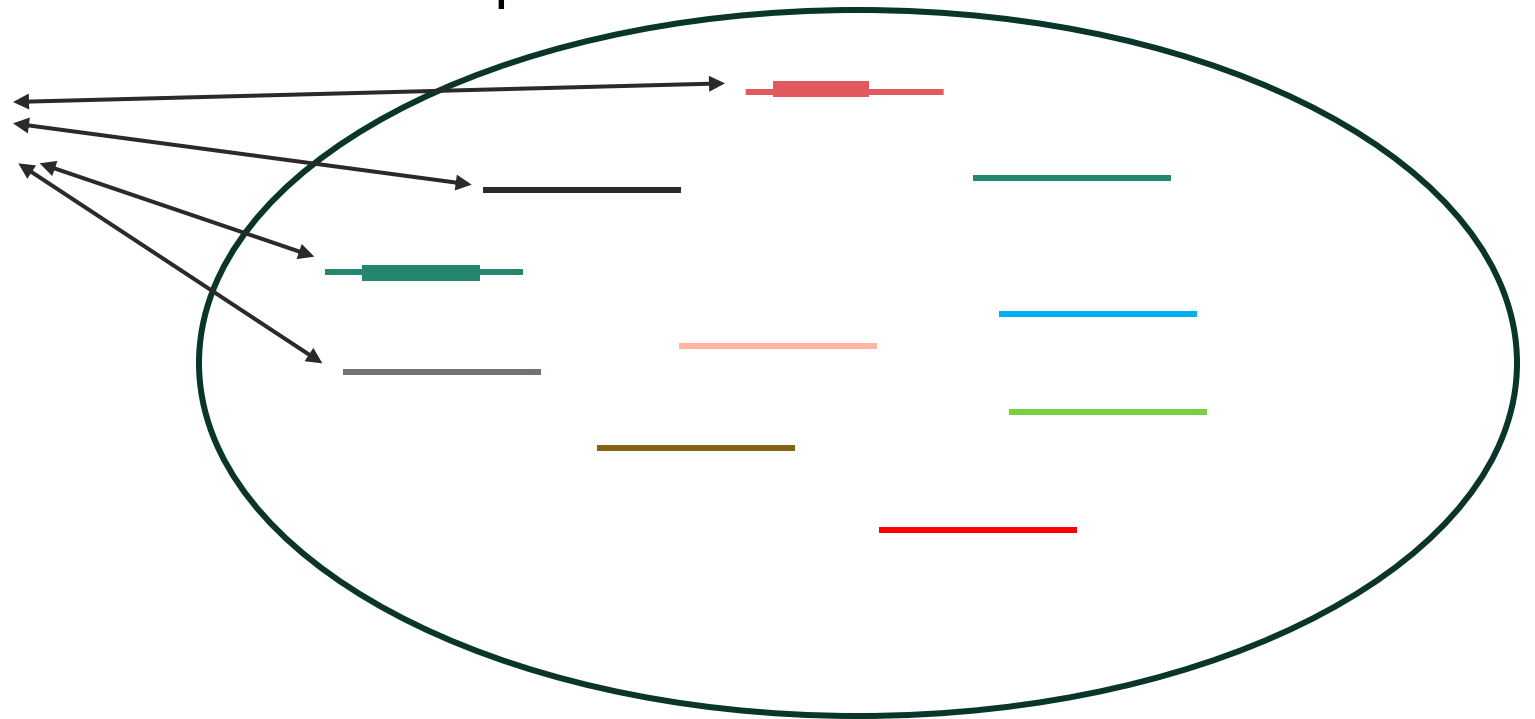
MVHLTPEEKSAVTALWGKVNVDVEVGGEALGRLLVWYP
WTQRRFFESFGDLSTPDAVMGNPVKAHGKKVLGAFSDG
LAHLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLC
VLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH



База данных последовательностей

Алгоритм Смита-Ватермана

MVHLTPEEKSAVTALWGKLVNDEVGGEALGRLLVVYP
WTQRRFFESFGDLSTPDAVMGNPVKAHGKKVLGAFSDG
LAHLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLVC
VLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH



Только в SwissProt ~250 млн последовательностей....

BLAST – это эвристический алгоритм

Алгоритмы биоинформатики можно разделить на точные и эвристические.

- ❖ Точные алгоритмы решают какую-либо точно сформулированную формализованную задачу. Пример: алгоритм Нидлмана – Вунша, который для данных последовательностей находит выравнивание с максимальным весом.
- ❖ Эвристический алгоритм — алгоритм решения задачи, включающий практический метод, не являющийся гарантированно точным или оптимальным, но достаточный для решения поставленной задачи

BLAST не гарантирует нахождение оптимального локального выравнивания. За счёт этого достигается высокая скорость работы.

Но теоретически возможно, что BLAST не найдёт в банке имеющийся там вполне достоверный (судя по выравниванию) гомолог.

Идея алгоритма BLAST

- ❖ Задача – найти в банке последовательности, хорошо (то есть с большим весом) выравнивающиеся с последовательностью запроса.
- ❖ Можно последовательно выравнивать каждую банковскую последовательность с запросом алгоритмом Смита – Ватермана
- ❖ При нынешних объемах банков это очень медленно

Проиндексируем базу

АЛФАВИТНЫЙ УКАЗАТЕЛЬ

(цифры обозначают номера экспериментов или параграфов)

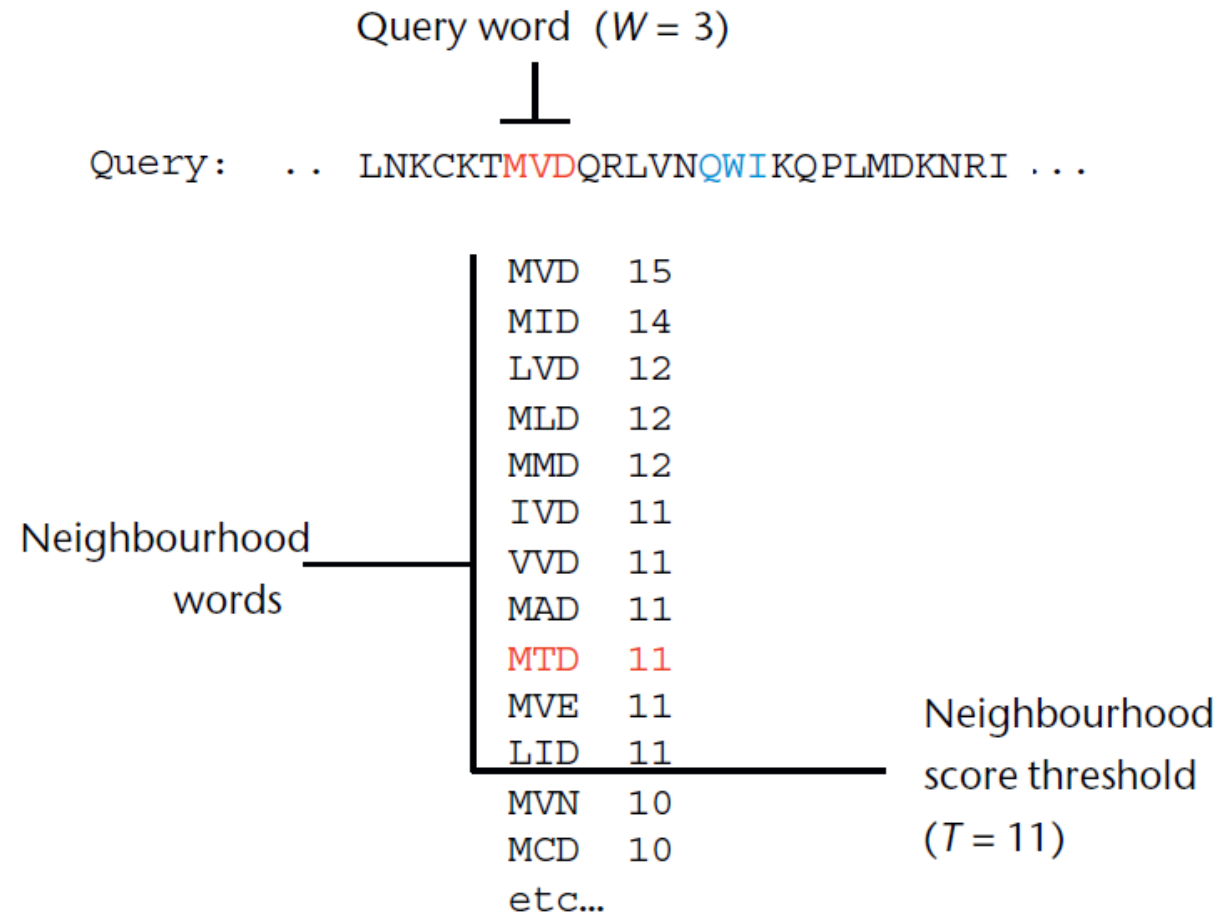
- | | |
|---|---|
| Агрегатное состояние 18, 19. | Время, деление на равные промежутки 15, 16. |
| Акустический указатель 169. | Время, измерение 13—15, 113, § 3. |
| Акция 128. | Время падения 120. |
| Амплитуда колебания 162, 191, 196, 197, 211, 217. | Высота падения 118, 120. |
| Аперiodические колебания 205. | Вытесняемость жидкости 8, 9, 21, 22. |
| Балансирование 65, 66, 70. | Вытесняемость твердых тел 20. |
| Барометр чашечный § 1. | Гармоническое колебание 191, 196, § 28. |
| Батавские слезки 61. | Градуирование шкалы динамометра 55. |
| Биение 217. | Грамм § 7. |
| Бифилярный подвес 150, 156, 162, 197, 207. | Графики 55, 147, 183, 193, 194, 199. |
| Блок 84—86, § 2 — 1, 3, 4. | Грузики с крючками § 2—10. |
| Блок ступенчатый § 2—5. | Давления, сила 53, 135. |
| Болонская колбочка 61. | Дальность полета 118, 122, 157. |
| | Движение волновое 201. |

Алгоритм BLAST: индексация

- ❖ В случае BLAST индексами служат **слова** заданной длины из букв, встречающихся в наших последовательностях
- ❖ **Длина слова** (word_size) – это параметр алгоритма
- ❖ Например, для белков и при длине слова 3 это AAA, AAC, AAD, ..., YYY, всего $20^3 = 8000$ слов.
- ❖ Индекс можно представить себе как таблицу, в которой для каждого слова указано, в какой последовательности банка и в каком месте это слово встретилось.

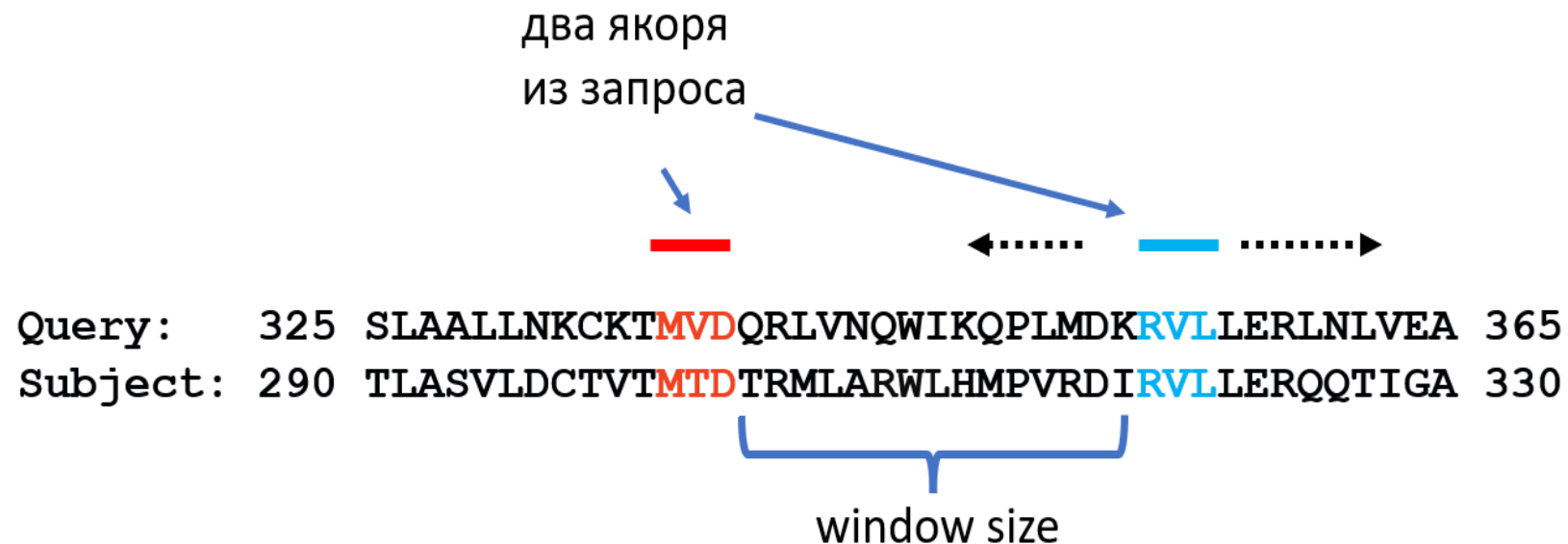
Алгоритм BLAST: разбиение последовательности-запроса

- ❖ Два параметра
 - ✓ длина слова (word_size, ≥ 2 , в standalone по умолчанию 3)
 - ✓ порог на сходство слов (threshold, ≥ 0 , по умолчанию 11)
- ❖ Берутся все слова из запроса (query) длины word_size
- ❖ В индексах ищутся слова, имеющие сходство со словами из запроса на уровне не менее threshold



Алгоритм BLAST: от якоря к выравниванию

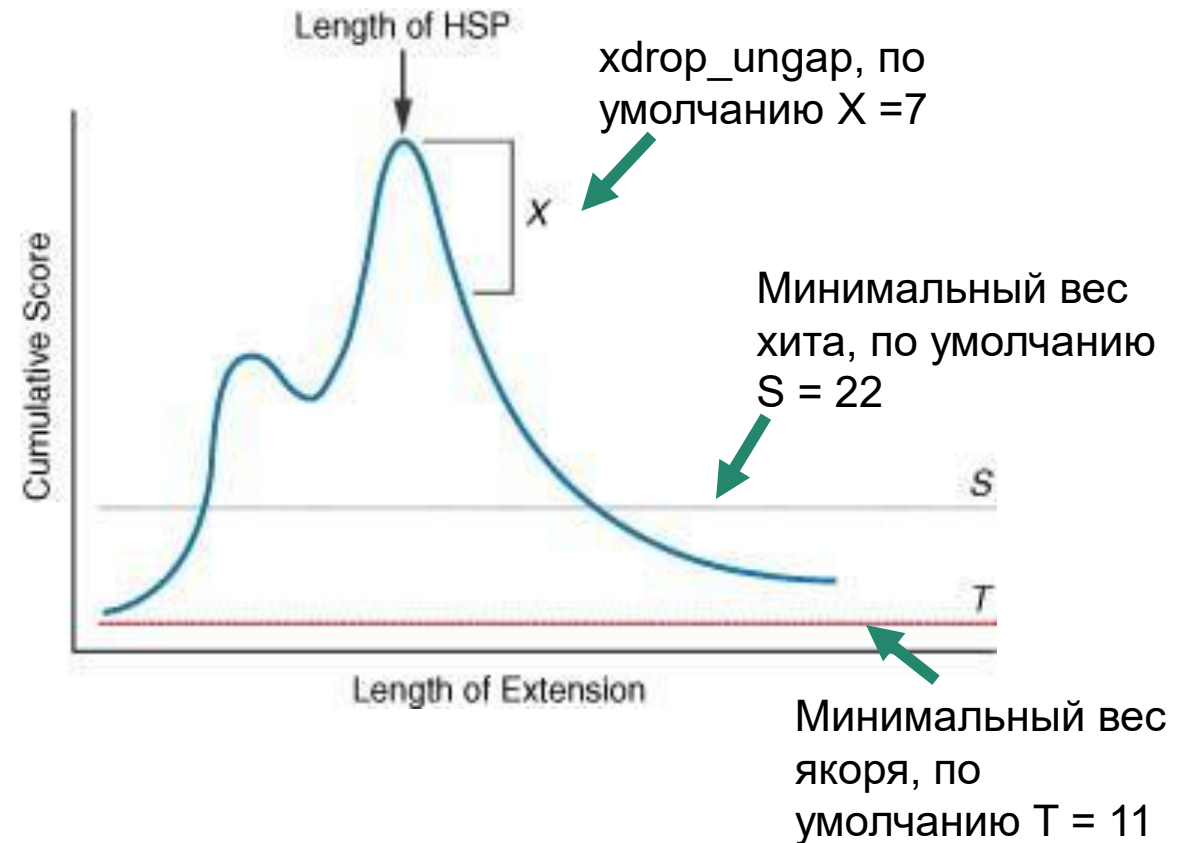
- ❖ Выравнивание начинает строиться, если в запросе есть пара слов на расстоянии, меньшем параметра `window_size` (по умолчанию 40), для которых нашлась пара сходных слов в одной банковской последовательности на том же расстоянии. В результате получаем два якоря — выравнивания длины `word_size`.



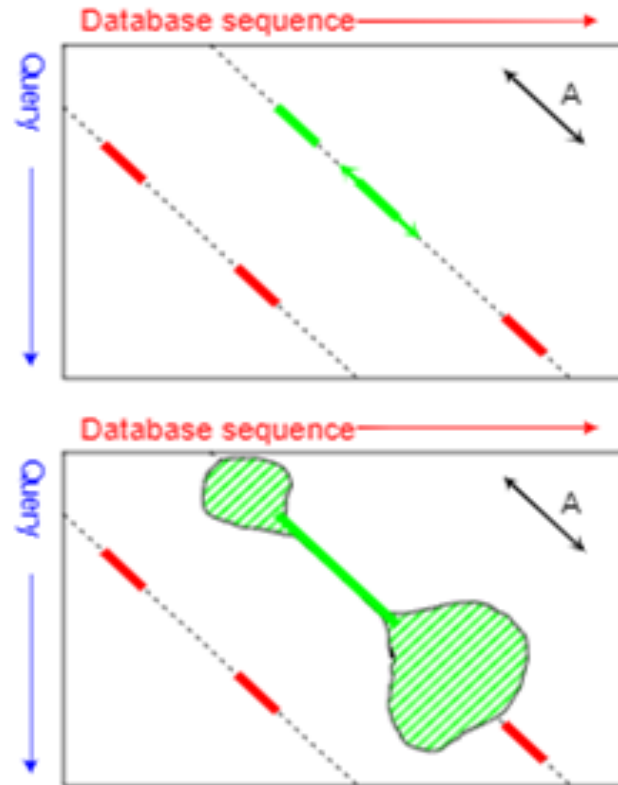
Алгоритм BLAST: от якоря к выравниванию

- ❖ Второй якорь расширяется без гэпов в обе стороны, пока вес не упадёт на заданную величину от максимально достигнутого (по умолчанию этот параметр `xdrop_ungap` = 7 бит)
- ❖ Если максимально достигнутый вес больше 22 бит, то соответствующее выравнивание расширяется уже с гэпами (аналогично алгоритму Нидлмана – Вунша). Расширение продолжается, пока вес не упадёт ниже максимально достигнутого на величину, большую `xdrop_gap`, по умолчанию 15 бит

Схема расширения в одну сторону



Алгоритм BLAST: от якоря к выравниванию



<https://docplayer.net/15013198-Databases-indexation.html>

Автор: Laurent Falquet, SIB

Локальное выравнивание как диагональ в матрице

Subject

Query

	G	T	A	T	A	G	T	C
G	G							
T		T						
T				T				
A					A			
G						G		
T							T	
A								

Subject: GTATAGTC

|| |||

Query: GT-TAGTA

Локальное выравнивание как диагональ в матрице

Subject

	G	T	A	T	A	G	T	C
G	G							
T		T						
T				T				
A					A			
G						G		
T							T	
A								

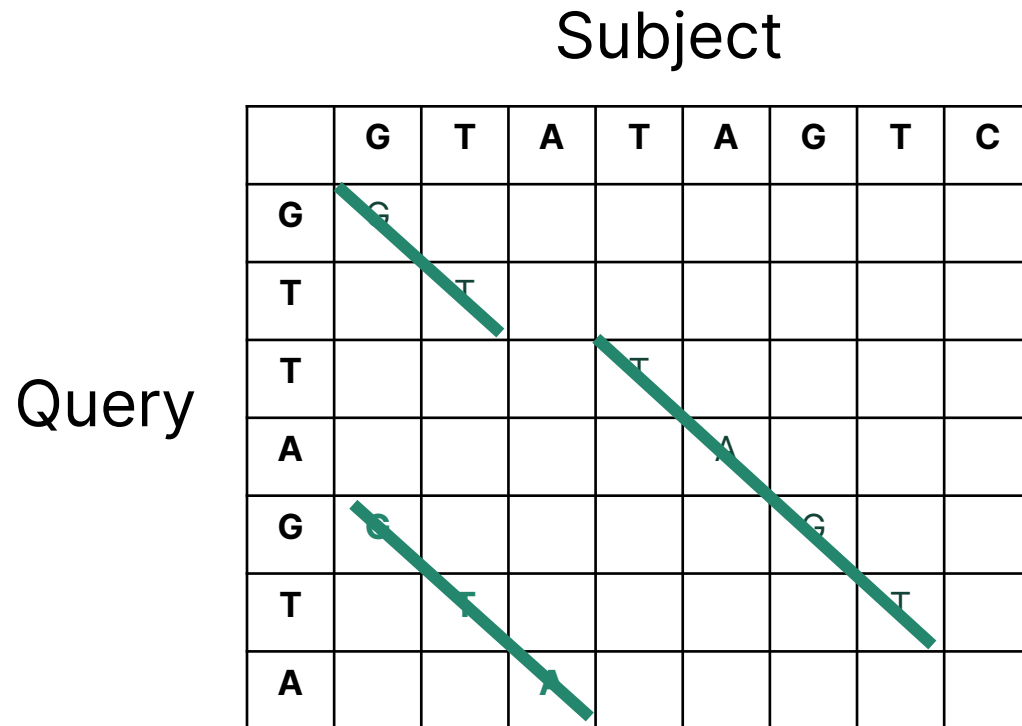
Query

Subject: GTATAGTC

|| ||||

Query: GT-TAGTA

Локальное выравнивание как диагональ в матрице



Subject: GTATAGTC
 | | | | |
Query: GT-TAGTA

Subject: ----GTATAGTC
 | | |
Query: GTTAGTA-----

Роль длины слова (эксперимент)

- ❖ Вход: последовательность из 466 остатков
- ❖ NCBI BLAST (<https://blast.ncbi.nlm.nih.gov/>)
- ❖ Область поиска: Swiss-Prot, белки из бактерий
- ❖ Порог E-value = 10, остальные параметры, кроме "Word Size", по умолчанию
- ❖ $W = 6$
 - ✓ Найдено 16 последовательностей, в них 18 находок
 - ✓ 8 находок с $E < 0,001$
 - ✓ Время работы сервиса NCBI – менее одной минуты
- ❖ $W = 2$
 - ✓ Найдено 69 последовательностей, в них 75 находок
 - ✓ 12 находок с $E < 0,001$
 - ✓ Время работы сервиса NCBI – около 35 мин

Роль длины слова

- ❖ Чем больше длина слова, тем быстрее работает BLAST, но тем меньше его **чувствительность**. Это означает, что вероятность пропустить гомологи возрастает.
- ❖ В веб-версии blastp на сайте NCBI значение длины слова по умолчанию равно 5, доступны значения 2, 3 и 6.

Вопросы и ответы про BLAST

За счёт чего BLAST работает быстро?

За счёт просмотра не всех возможных выравниваний, а только полученных расширением "затравок". Каждая "затравка" получается из слова длины k ($k = 2, 3, \dots, 6$), встреченного в запросе, и очень сходного слова из какой-либо банковской последовательности.

"Затравки" находятся очень быстро благодаря предварительной индексации всех слов в банке. В результате индексации для каждого слова указано, в каких местах каких банковских последовательностей это слово встречается.

Что может поменяться при изменении параметра "Word size"?

Чем длиннее слово, тем меньше машинного времени займёт поиск.

Чем короче слово, тем чувствительнее поиск (меньше опасность пропустить хорошее выравнивание).

Protein BLAST – поиск гомологов данного белка в банке аминокислотных последовательностей

Алгоритмы:

- ❖ BLASTP
- ❖ Quick BLASTP
- ❖ PSI-BLAST
- ❖ PHI-BLAST
- ❖ DELTA-BLAST

Можно использовать:

- из командной строки (standalone BLAST)
- через веб-интерфейс

Что подается на вход программе BLAST?

- ✓ Последовательность-запрос (query)
- ✓ Банк последовательностей
- ✓ Параметры:
 - Выравнивания (матрица амк замен, штрафы за гэпы)
 - Поиска (длина слова и др)
 - Выдачи (максимальное число находок, пороги на качество выравнивания, форма выдачи и др)

Что выдает программа BLAST?

На выходе:

- ✓ Заголовок с описанием программы, банка, запроса (query)
- ✓ Список находок
- ✓ Выравнивание запроса с находками

Веб-интерфейсы тем или иным способом перерабатывают выдачу программы. Часто вставляется графическое изображение находок.

Выравнивание, выданное BLAST

ID находки - subject

Длина найденного белка

Sequence ID: Q51368.2 Length: 342

Range 1: 234 to 338

Участок найденного белка,
попавший в выравнивание

Score:80.9 bits(198), Expect:1e-16,

Method: Compositional matrix adjust.,

Identities:46/115(40%), Positives:63/115(54%), Gaps:15/115(13%)

```
Query 123  SPFENTAPARLTSSTATAATSKPVTSVASGPRALSRNQPPQYPARAQALRIEGQVKVKFDV 182
          +P + PA L S + + KP L + P YP AQA IEG+VKV F +
Sbjct 234  APSGSQGPAGLPSGSLNDSDIKP-----LRMDPPVYPRMAQARGIEGRVKVLFTI 283

Query 183  TPDGRVDNVQILSAKPANMFEREVKNAMRRWRYEPGKPGSGIVVN-----ILFKI 232
          T DGR+D++Q+L + P+ MF+REV+ AM +WR+EP G IV FKI
Sbjct 284  TSDGRIDDIQVLESVPSRMFDREVRQAMAKWRFEPRVSGGKIVARQATKMFFFKI 338
```

Выравнивание, выданное BLAST

Вес в битах Вес

Sequence ID: Q51368.2 Length: 342
Range 1: 234 to 338

E-value

Score: 80.9 bits (198), Expect: 1e-16,
Method: Compositional matrix adjust.,
Identities: 46/115 (40%), Positives: 63/115 (54%), Gaps: 15/115 (13%)

```
Query 123 SPFENTAPARLTSSTATAATSKPVTSVASGPRALSRNQPYPARAQALRIEGQVKVKFDV 182
      +P + PA L S + + KP L + P YP AQA IEG+VKV F +
Sbjct 234 APSGSQGPAGLPSGSLNDSDIKP-----LRMDPPVYPRMAQARGIEGRVKVLFTI 283

Query 183 TPDGRVDNVQILSAKPANMFEREVKNAMRRWRYEPGKPGSGIVVN-----ILFKI 232
      T DGR+D++Q+L + P+ MF+REV+ AM +WR+EP G IV FKI
Sbjct 284 TSDGRIDDIQVLESVPSRMFDREVRQAMAKWRFEPRVSGGKIVARQATKMFFFKI 338
```

Число совпадений Длина выравнивания!
(не находки)

Число сходных букв Число символов гэпа

Словарик BLAST

- ❖ Identities – совпадения (кол-во + % от длины выравнивания)
- ❖ Positives – сходные буквы = значение весовой матрицы положительно (кол-во + % от длины выравнивания)
- ❖ Gaps – знаки гэпов (кол-во + % от длины выравнивания)
- ❖ Score – вес выравнивания (в битах и обычный = сумма значений матрицы по сопоставлениям минус штраф за гэпы)
- ❖ Evalue – e-value, ожидаемое число выравниваний с тем же или большим весом. Запись вида $9e-15$ означает $9 \cdot 10^{-15}$.

E-value

E-value – ожидаемое количество случайных находок с таким же и лучшим весом выравнивания, при поиске в той же базе данных, со случайным запросом той же длины и состава, с теми же параметрами на вычисление веса выравнивания.

❖ Зависит от:

- ✓ Веса выравнивания
- ✓ Размера банка
- ✓ Длины запроса
- ✓ Параметров

Чем ниже e-value,
тем выше значимость находки

E-value

- ❖ E-value – **ожидаемое** количество **случайных** находок с таким же и лучшим весом выравнивания, при поиске в той же базе данных, со случайным запросом той же длины и состава, с теми же параметрами на вычисление веса выравнивания.

Что означает слово «ожидаемое»?

Формально это то, что называется «математическое ожидание случайной величины». Случайной величиной в данном случае является **число находок** (*NB! Просьба запомнить!*)

На практике ожидаемое вычисляется как среднее по достаточно большому количеству испытаний.

Другое ключевое слово — «случайных». Нам нужно понять, сколько можно ожидать именно случайных, то есть бессмысленных, негомологичных находок, чтобы оценить, насколько надёжно утверждение, что данная находка — действительно гомолог

Как посчитать E-value

- ❖ Прямой способ — вычислительный эксперимент: перемешать буквы в запросе очень много раз, каждый раз запуская BLAST, и посмотреть, сколько в среднем при одном запуске бывает находок с весом выше данного.
- ❖ Такой способ, естественно, не применяется :)
- ❖ *Стоит подумать: от чего и как может зависеть число случайных находок*

E-value

$$E\text{-value} = Kmn \cdot e^{-\lambda S} \quad (\text{Karlin \& Altschul, 1990})$$

- ❖ S — Score (вес)
- ❖ m — длина исходной последовательности
- ❖ n — размер базы данных (суммарная длина всех последовательностей)
- ❖ K и λ — две константы, зависящие только от параметров вычисления веса

BLAST хранит значения K и λ для нескольких наборов параметров вычисления веса, их раз и навсегда нашли посредством вычислительного эксперимента.

Вес в битах

- ❖ Вес в битах B зависит от обычного веса S и параметров вычисления веса.

$$B = (\lambda S - \ln K) / \ln 2$$

- ❖ Зависимость подогнана так, чтобы

$$E\text{-value} = mn \cdot 2^{-B}$$

- ✓ m — длина исходной последовательности
- ✓ n — размер базы данных
(констант K и λ теперь нет, они “загнаны внутрь B ”)

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

Алгоритм blastp

BLAST® » blastp suite Home Recent Results Saved Strategies Help

blastn **blastp** blastx tblastn tblastx Standard Protein BLAST

BLASTP programs search protein databases using a protein query. more... Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange From To

Or, upload file Файл не выбран

Job Title

Align two or more sequences

Choose Search Set

Databases Standard databases (nr etc.): New Experimental databases < Try experimental clustered nr databases For more info see What is clustered nr?

Compare Select to compare standard and experimental database

Standard

Database

Organism Optional exclude

Exclude Optional Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm Quick BLASTP (Accelerated protein-protein BLAST) **blastp (protein-protein BLAST)** PSI-BLAST (Position-Specific Iterated BLAST) PHI-BLAST (Pattern Hit Initiated BLAST) DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

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

+ Algorithm parameters

Вводим последовательность

Банк

- Non-redundant protein sequences (nr)
- RefSeq Select proteins (refseq_select)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot (swissprot)**
- Patented protein sequences (pataa)
- Protein Data Bank proteins (pdb)
- Metagenomic proteins (env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

Таксон организмов (если надо ограничить)

Дополнительные параметры

Дополнительные параметры

Algorithm parameters

Restore default search parameters

General Parameters

Max target sequences [?](#)
Select the maximum number of aligned sequences to display [?](#)

Short queries Automatically adjust parameters for short input sequences [?](#)

Expect threshold [?](#)

Word size [?](#)

Max matches in a query range [?](#)

Максимальный размер выдачи

Порог на E-value

Длина слова

Scoring Parameters

Matrix [?](#)

Gap Costs [?](#)

Compositional adjustments [?](#)

Параметры выравнивания

Filters and Masking

Filter Low complexity regions [?](#)

Mask Mask for lookup table only [?](#)
 Mask lower case letters [?](#)

Борьба с «участками малой сложности»

BLAST

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

Участок малой сложности

- ❖ Определяется как участок со смещенным составом (biased composition)
 - ✓ Гомополимерные участки
 - ✓ Короткие повторы
 - ✓ Перепредставленность отдельных остатков
- ❖ Вычисление E-value (параметры K и λ) опирается на средние по всем белкам частоты аминокислотных остатков, поэтому на участках малой сложности оно становится некорректным -> ложное предсказание гомологии

Выдача BLAST в интерфейсе NCBI

Job Title	Q51368:RecName: Full=Protein TonB
RID	JR0G8S5T013 Search expires on 11-07 21:49 pm Download All ▾
Program	BLASTP ? Citation ▾
Database	swissprot See details ▾
Query ID	P0C6X7.1
Description	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; A ...
Molecule type	amino acid
Query Length	7073
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Общий вес

Максимальный вес

Покрытие запроса

- Descriptions**
- Graphic Summary
- Alignments
- Taxonomy

Sequences producing significant alignments

[Download](#) ▾ [Select columns](#) ▾ Show [?](#)

select all 82 sequences selected

[GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	14775	14775	100%	0.0	100.00%	7073	P0C6X7.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat SARS CoV R...	14454	14454	100%	0.0	97.50%	7071	P0C6W6.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat SARS coron...	14296	14296	100%	0.0	95.86%	7067	P0C6W2.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat CoV 279/2005	14230	14230	100%	0.0	95.51%	7079	P0C6V9.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	12939	12939	100%	0.0	86.19%	7096	P0DTD1.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecName: Fu...	Severe acute res...	9103	9103	61%	0.0	100.00%	4382	P0C6U8.1

Чтобы скачать выдачу самой программы (а не её обработку интерфейсом), можно поступить так:

Job Title	Q51368:RecName: Full=Protein TonB
RID	JR0G8S5T013 <small>Search expires on 11-07 21:49 pm</small> Download All ▾
Program	BLASTP ? Citation ▾
Database	swissprot See details ▾
Query ID	P0C6X7.1
Description	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; A ...
Molecule type	amino acid
Query Length	7073
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

- Descriptions**
- Graphic Summary
- Alignments
- Taxonomy

Sequences producing significant alignments

Download ▾ Select columns ▾ Show 100 ▾ [?](#)

select all 82 sequences selected [GenPept](#) [Graph](#) [Multiple alignment](#) [MSA Viewer](#)

	Description	Score	Query cover	E value	Per. Ident	Acc. Len	Accession		
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Sev...	100%	0.0	100.00%	7073	P0C6X7.1		
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat...	100%	0.0	97.50%	7071	P0C6W6.1		
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat...	100%	0.0	95.86%	7067	P0C6W2.1		
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat...	100%	0.0	95.51%	7079	P0C6V9.1		
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	12939	12939	100%	0.0	86.19%	7096	P0DTD1.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecName: Fu...	Severe acute res...	9103	9103	61%	0.0	100.00%	4382	P0C6U8.1

1 RID: ZTMWDYK013
 2 Job Title: sp|P0C6X7|R1AB_SARS Replicase polyprotein...
 3 Program: BLASTP
 4 Query: sp|P0C6X7|R1AB_SARS Replicase polyprotein lab OS=Severe acute respiratory syndrome coronavirus OX=694009 GN=rep PE=1 SV=1 ID: lcl|Query_9160139(amino acid) Length: 7073
 5 Database: swissprot Non-redundant UniProtKB/SwissProt sequences

7 Sequences producing significant alignments:

Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query cover	E Value	Per. Ident	Acc. Len	Accession
10 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Severe acute... NA			694009	14775	14775	100%	0.0	100.00	7073	P0C6X7.1
11 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bat SARS CoV... NA			349344	14454	14454	100%	0.0	97.50	7071	P0C6W6.1
12 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bat SARS cor... NA			442736	14296	14296	100%	0.0	95.86	7067	P0C6W2.1
13 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bat CoV 279/... NA			389167	14230	14230	100%	0.0	95.51	7079	P0C6V9.1
14 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Severe acute... NA			2697049	12939	12939	100%	0.0	86.19	7096	PODTD1.1
15 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Severe acute... NA			694009	9103	9103	62%	0.0	100.00	4382	P0C6U8.1
16 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bat SARS CoV... NA			349344	8822	8822	62%	0.0	96.55	4380	P0C6T7.1
17 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bat SARS cor... NA			442736	8547	8547	62%	0.0	93.93	4376	P0C6F8.1
18 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bat CoV 279/... NA			389167	8495	8495	62%	0.0	93.26	4388	P0C6F5.1
19 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Severe acute... NA			2697049	7475	7475	62%	0.0	80.42	4405	PODTC1.1
20 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Pipistrellus... NA			694008	6150	6150	100%	0.0	45.84	7182	P0C6W4.1
21 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Rousettus ba... NA			694006	6106	6395	97%	0.0	52.41	6930	P0C6W5.1
22 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Betacoronavi... NA			1263720	5980	6287	99%	0.0	49.62	7078	K9N7C7.1
23 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bovine respi... NA			233264	5451	5525	82%	0.0	49.55	7094	P0C6W8.1
24 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bovine enter... NA			233262	5449	5522	82%	0.0	49.55	7094	P0C6W7.1
25 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Human corona... NA			31631	5437	5589	84%	0.0	49.56	7095	P0C6X6.1
26 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Murine hepat... NA			76344	5429	5507	85%	0.0	49.18	7124	P0C6X8.1
27 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bovine coron... NA			11132	5418	5488	81%	0.0	49.28	7094	P0C6W9.1
28 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Murine hepat... NA			11142	5408	5487	84%	0.0	49.28	7176	P0C6X9.1
29 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Murine hepat... NA			11144	5356	5432	84%	0.0	48.72	7180	P0C6Y0.1
30 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Human corona... NA			443240	5319	5400	83%	0.0	48.57	7152	P0C6X3.1
31 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Human corona... NA			443241	5317	5396	83%	0.0	48.48	7132	P0C6X4.1
32 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Human corona... NA			443239	5312	5391	83%	0.0	48.49	7182	P0C6X2.1
33 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Bovine coron... NA			11133	5287	5355	80%	0.0	48.91	7059	P0C6X0.1
34 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Scotophilus ... NA			693999	4328	4398	71%	0.0	46.40	6793	P0C6W0.1
35 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Porcine epid... NA			229032	4295	4383	72%	0.0	46.48	6781	P0C6Y4.1
36 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Human corona... NA			277944	4286	4370	73%	0.0	45.79	6729	P0C6X5.1
37 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Human corona... NA			11137	4258	4345	72%	0.0	45.56	6758	P0C6X1.1
38 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Porcine tran... NA			11151	4194	4261	72%	0.0	45.60	6684	P0C6Y5.1
39 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Feline infec... NA			33734	4173	4233	70%	0.0	45.61	6709	Q98VG9.2
40 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Avian infect... NA			11122	4133	4183	71%	0.0	45.35	6629	P0C6Y1.1
41 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Avian infect... NA			11127	4133	4190	73%	0.0	45.31	6631	P0C6Y3.1
42 RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Avian infect... NA			160235	4131	4181	71%	0.0	45.35	6629	P0C6Y2.1
43 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Pipistrellus... NA			694008	2335	2335	62%	0.0	33.82	4481	P0C6T5.1
44 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Rousettus ba... NA			694006	2288	2577	59%	0.0	40.41	4248	P0C6T6.1
45 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Betacoronavi... NA			1263720	2184	2489	60%	0.0	37.74	4391	K9N638.1
46 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bovine respi... NA			233264	1887	1961	44%	0.0	38.08	4383	P0C6T9.1
47 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bovine enter... NA			233262	1885	1958	44%	0.0	38.08	4383	P0C6T8.1
48 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Human corona... NA			31631	1879	2031	46%	0.0	38.07	4383	P0C6U7.1
49 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bovine coron... NA			11133	1875	1942	43%	0.0	37.72	4383	P0C6U1.1
50 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Bovine coron... NA			11132	1873	1943	43%	0.0	37.82	4383	P0C6U0.1
51 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Murine hepat... NA			76344	1860	1937	46%	0.0	37.67	4416	P0C6U9.1
52 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Murine hepat... NA			11142	1850	1929	46%	0.0	37.80	4468	P0C6V0.1
53 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Murine hepat... NA			11144	1843	1918	45%	0.0	37.32	4474	P0C6V1.1
54 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Human corona... NA			443240	1810	1890	45%	0.0	36.95	4441	P0C6U4.1
55 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Human corona... NA			443241	1808	1887	45%	0.0	37.03	4421	P0C6U5.1
56 RecName: Full=Replicase polyprotein la; Short=ppla; AltName:... Human corona... NA			443239	1803	1882	45%	0.0	37.02	4471	P0C6U3.1

89	RecName: Full=Replicase polyprotein lab; Short=pplab; AltName:... Alphamesoniv... NA	1552985	52.0	52.0	7%	3e-04	22.56	5088	F8RL29.1
90	RecName: Full=Probable RNA helicase SDE3; AltName:... Arabidopsis ... thale cress	3702	49.3	49.3	4%	0.002	25.52	1002	Q8GYD9.1
91	RecName: Full=ADP-ribose glycohydrolase AF_1521; AltName:... Archaeoglobu... NA	224325	43.1	43.1	2%	0.032	29.93	192	O28751.2

92
93 Alignments:

94
95 >RecName: Full=Replicase polyprotein lab; Short=pplab; AltName: Full=ORFlab polyprotein; Contains: RecName: Full=Host translation inhibitor nsp1; AltName: Full=Leader
Contains: RecName: Full=Non-structural protein 2; Short=nsp2; AltName: Full=p65 homolog; Contains: RecName: Full=Papain-like protease nsp3; Short=PL-PRO; AltName: Ful
Contains: RecName: Full=Non-structural protein 4; Short=nsp4; Contains: RecName: Full=3C-like proteinase nsp5; Short=3CL-PRO; Short=3CLp; AltName: Full=Main protease;
Short=nsp5; AltName: Full=SARS coronavirus main proteinase; Contains: RecName: Full=Non-structural protein 6; Short=nsp6; Contains: RecName: Full=Non-structural prote
protein 8; Short=nsp8; Contains: RecName: Full=Viral protein genome-linked nsp9; AltName: Full=Non-structural protein 9; Short=nsp9; AltName: Full=RNA-capping enzyme
10; Short=nsp10; AltName: Full=Growth factor-like peptide; Short=GFL; Contains: RecName: Full=RNA-directed RNA polymerase nsp12; Short=Pol; Short=RdRp; AltName: Full=
Full=Helicase nsp13; Short=Hel; AltName: Full=Non-structural protein 13; Short=nsp13; Contains: RecName: Full=Guanine-N7 methyltransferase nsp14; AltName: Full=Non-st
exoribonuclease nsp14; Short=ExoN; Contains: RecName: Full=Uridylate-specific endoribonuclease nsp15; AltName: Full=NendoU; AltName: Full=Non-structural protein 15; S
nsp16; AltName: Full=Non-structural protein 16; Short=nsp16 [Severe acute respiratory syndrome-related coronavirus]

96 Sequence ID: P0C6X7.1 Length: 7073

97 Range 1: 1 to 7073

98
99 Score:14775 bits(38346), Expect:0.0,

100 Method:Compositional matrix adjust.,

101 Identities:7073/7073(100%), Positives:7073/7073(100%), Gaps:0/7073(0%)

102
103 Query 1 MESLVLGVNEKTHVQLSLPVLQVRDVLVRGFGDSVEEALSEAREHLKNGTCGLVELEKGV 60
104 | MESLVLGVNEKTHVQLSLPVLQVRDVLVRGFGDSVEEALSEAREHLKNGTCGLVELEKGV
105 Sbjct 1 MESLVLGVNEKTHVQLSLPVLQVRDVLVRGFGDSVEEALSEAREHLKNGTCGLVELEKGV 60
106
107 Query 61 LPQLEQPYVFIKRSDALSTNHGHKVVVELVAEMDGIQYGRSGITLGVLPVPHVGETPIAYRN 120
108 | LPQLEQPYVFIKRSDALSTNHGHKVVVELVAEMDGIQYGRSGITLGVLPVPHVGETPIAYRN
109 Sbjct 61 LPQLEQPYVFIKRSDALSTNHGHKVVVELVAEMDGIQYGRSGITLGVLPVPHVGETPIAYRN 120
110
111 Query 121 VLLRKNNGKAGGHSYGIDLKSYDLGDELGTDPIEDYEQNWNNTKHGSGALRELTRELNNG 180
112 | VLLRKNNGKAGGHSYGIDLKSYDLGDELGTDPIEDYEQNWNNTKHGSGALRELTRELNNG
113 Sbjct 121 VLLRKNNGKAGGHSYGIDLKSYDLGDELGTDPIEDYEQNWNNTKHGSGALRELTRELNNG 180
114
115 Query 181 AVTRYVDNDFCGPDGYPLDCIKDFLARAGKSMCTLSEQLDYIESKRGVYCCRDHEHEIAW 240
116 | AVTRYVDNDFCGPDGYPLDCIKDFLARAGKSMCTLSEQLDYIESKRGVYCCRDHEHEIAW
117 Sbjct 181 AVTRYVDNDFCGPDGYPLDCIKDFLARAGKSMCTLSEQLDYIESKRGVYCCRDHEHEIAW 240
118
119 Query 241 FTERSDKSYEHQTPFEIKSAKKFDTFKGECPKFVFPVPLNSKVKVIQPRVEKKKTEGFMGRI 300
120 | FTERSDKSYEHQTPFEIKSAKKFDTFKGECPKFVFPVPLNSKVKVIQPRVEKKKTEGFMGRI
121 Sbjct 241 FTERSDKSYEHQTPFEIKSAKKFDTFKGECPKFVFPVPLNSKVKVIQPRVEKKKTEGFMGRI 300
122
123 Query 301 RSVYPVASPQECNNMHLSTLMKCNHCDEVSWQTCDFLKATCEHCVTENLVIEGPTTCGYL 360
124 | RSVYPVASPQECNNMHLSTLMKCNHCDEVSWQTCDFLKATCEHCVTENLVIEGPTTCGYL
125 Sbjct 301 RSVYPVASPQECNNMHLSTLMKCNHCDEVSWQTCDFLKATCEHCVTENLVIEGPTTCGYL 360
126
127 Query 361 PTNAVVKMPCPACQDPEIGPEHSVADYHNHNSNIETRLRKGGRTRCFGGCVFAYVGCYNKR 420
128 | PTNAVVKMPCPACQDPEIGPEHSVADYHNHNSNIETRLRKGGRTRCFGGCVFAYVGCYNKR
129 Sbjct 361 PTNAVVKMPCPACQDPEIGPEHSVADYHNHNSNIETRLRKGGRTRCFGGCVFAYVGCYNKR 420
130

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	14775	14775	100%	0.0	100.00%	7073	P0C6X7.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat SARS CoV R...	14454	14454	100%	0.0	97.50%	7071	P0C6W6.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat SARS coron...	14296	14296	100%	0.0	95.86%	7067	P0C6W2.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat CoV 279/2005	14230	14230	100%	0.0	95.51%	7079	P0C6V9.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	12939	12939	100%	0.0	86.19%	7096	P0DTD1.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Rousettus bat co...	6106	6395	97%	0.0	52.41%	6930	P0C6W5.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Betacoronavirus ...	5980	6287	98%	0.0	49.62%	7078	K9N7C7.1
✓	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bovine respirator...	5451	5525	81%	0.0	49.55%	7094	P0C6W8.1

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	14775	14775	100%	0.0	100.00%	7073	P0C6X7.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat SARS CoV R...	14454	14454	100%	0.0	97.50%	7071	P0C6W6.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat SARS coron...	14296	14296	100%	0.0	95.86%	7067	P0C6W2.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bat CoV 279/2005	14230	14230	100%	0.0	95.51%	7079	P0C6V9.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Severe acute res...	12939	12939	100%	0.0	86.19%	7096	P0DTD1.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Rousettus bat co...	6106	6395	97%	0.0	52.41%	6930	P0C6W5.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Betacoronavirus ...	5980	6287	98%	0.0	49.62%	7078	K9N7C7.1
<input checked="" type="checkbox"/>	RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: RecName...	Bovine respirator...	5451	5525	81%	0.0	49.55%	7094	P0C6W8.1

BLAST может найти несколько локальных выравниваний!

Таблица находок BLAST

Max Score: самый большой из весов (в битах) выравниваний запроса с данной находкой

Total Score: суммарный вес (в битах) всех выравниваний запроса с данной находкой

Query cover: процент длины запроса, покрытый выравниваниями

E Value: в таблице находок это E-value, подсчитанное по особой формуле на основе **всех** выравниваний запроса с данной находкой

Per. Ident: процент идентичных букв в лучшем (по весу) из выравниваний запроса с данной находкой

Standalone BLAST

BLAST можно установить на своём компьютере
(а на kodo10 он уже установлен)

Предположим, вам нужно найти гомологи белка, чья последовательность — в файле `myprot.fasta`, в протеоме, содержащемся в файле `proteom.fasta` (всё в `fasta`-формате, BLAST других не понимает).

Придётся сначала проиндексировать ваш банк программой `makeblastdb`, подав ей на вход протеом (читайте `makeblastdb -help`)

Эта программа создаст несколько файлов, необходимых для поиска, в том числе тот самый индекс якорей (сразу для всех допустимых длин слов)

После этого можно искать программой `blastp`, указав ей имя файла с запросом и название проиндексированного банка
(читайте `blastp -help`, нужные опции: `-query`, `-db`, `-out`)

Standalone BLAST

Впрочем, можно использовать BLAST и для обычного локального выравнивания двух последовательностей, безо всякой индексации:

❖ `blastp -query seq1.fasta -subject seq2.fasta -out result.blastp`

Но имейте в виду, что BLAST и в таком варианте не гарантирует оптимального выравнивания (это эвристический алгоритм)! Зато можно быстро выровнять очень длинные последовательности (команде `water` может не хватить памяти) и получить не одно, а много локальных выравниваний.

(На самом деле в этом варианте BLAST «на ходу» индексирует вторую последовательность)

BLAST: варианты формата выходного файла

```
-outfmt <String>
  alignment view options:
    0 = Pairwise,
    1 = Query-anchored showing identities,
    2 = Query-anchored no identities,
    3 = Flat query-anchored showing identities,
    4 = Flat query-anchored no identities,
    5 = BLAST XML,
    6 = Tabular,
    7 = Tabular with comment lines,
    8 = Seqalign (Text ASN.1),
    9 = Seqalign (Binary ASN.1),
    10 = Comma-separated values,
    11 = BLAST archive (ASN.1),
    12 = Seqalign (JSON),
    13 = Multiple-file BLAST JSON,
    14 = Multiple-file BLAST XML2,
    15 = Single-file BLAST JSON,
    16 = Single-file BLAST XML2,
    18 = Organism Report
```

0–4 — чтобы смотреть глазами

5–12 — чтобы парсить программами.

6, 7 и 10 можно импортировать в электронные таблицы



Спасибо за внимание!