



Гомология и множественное выравнивание Карты локального сходства

22 апреля 2025

Алексеевский Андрей Владимирович
Алешина Юлия Александровна

Выравнивание как матрица

Seq 2

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

Seq 1

- Любое выравнивание можно визуализировать в виде матрицы, где по одной оси (горизонтальной) будут буквы первой последовательности, а по другой (вертикальной) – буквы второй последовательности.

```

Seq 1:  GTATAGTC
         ||  |||
Seq 2:  GT-TAGTA
    
```

Выравнивание как матрица

Seq 2

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

Seq 1

Seq 1: GTATAGTC
 || |||
 Seq 2: GT-TAGTA

Seq 2

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

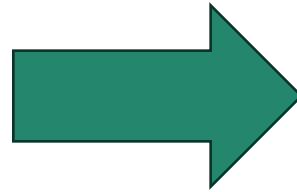
Seq 1

Seq 1: GT-A-TAGTC
 || | ||
 Seq 2: GTTAGTA---

Seq 2

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

Seq 1

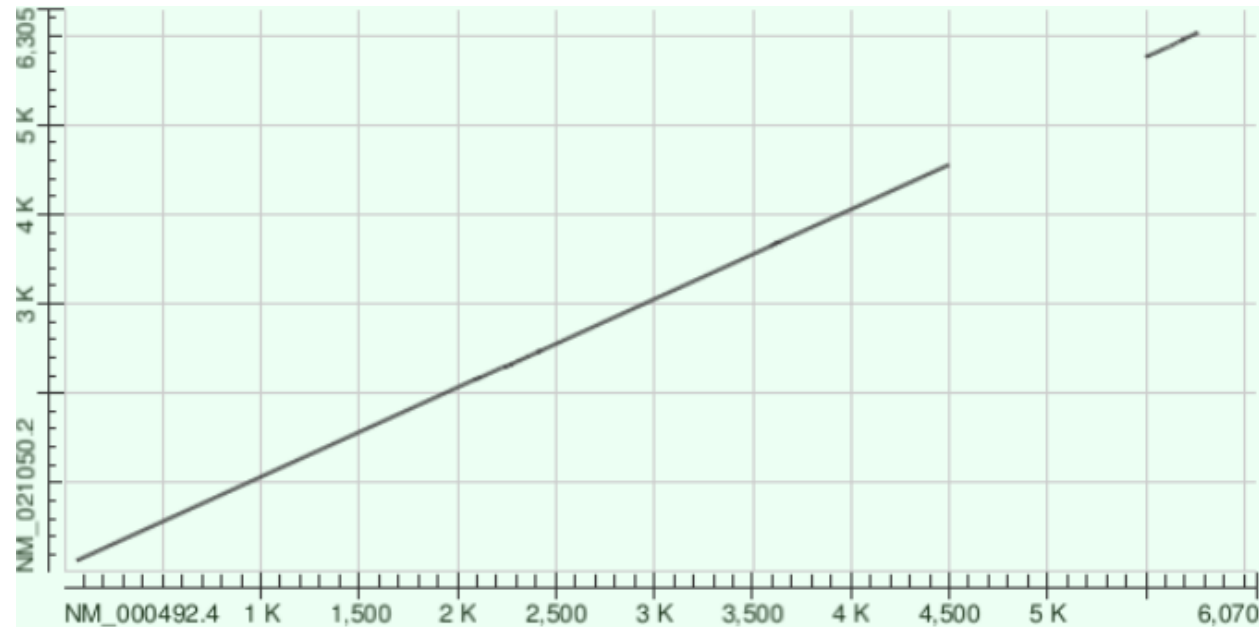


A								
T								
G								
A								
T								
T								
G								
	G	T	A	T	A	G	T	C

Seq 1: GTATAGTC
 || |||
 Seq 2: GT-TAGTA

Карта локального сходства

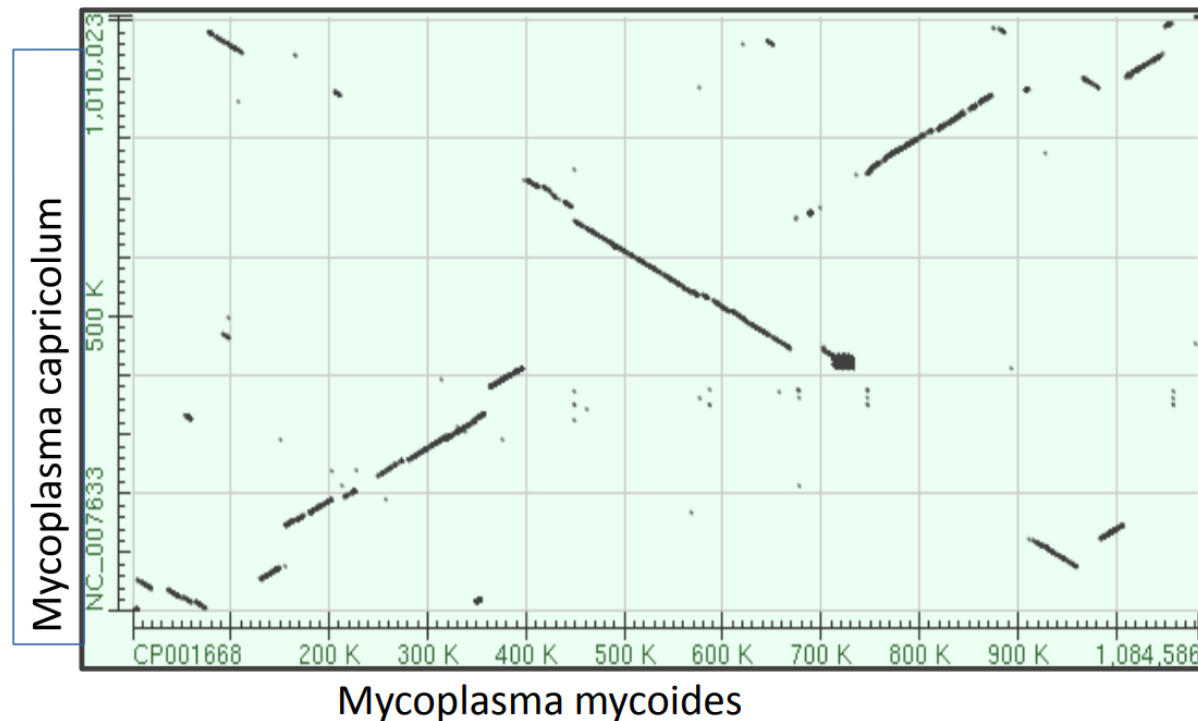
- ❖ Визуальное представление парного выравнивания
- ❖ Обычно для длинных последовательностей (геномы)
- ❖ Отрезки отображают наличие локального выравнивания



Карта локального сходства

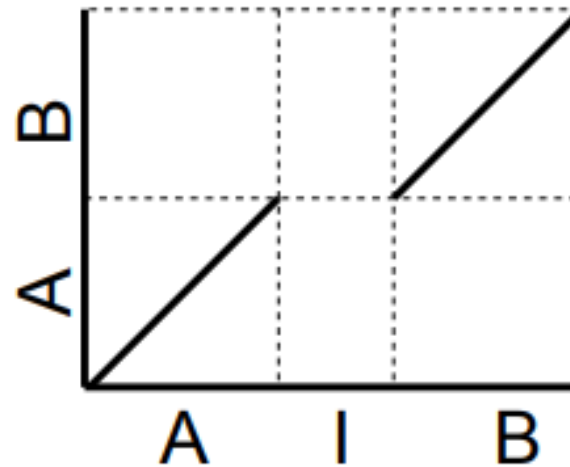
- ❖ Карту локального сходства можно строить для изучения глобальных перестроек геномов и белков

Карта локального сходства геномов *M.capricolum* и *M.mycoides*



Делеция или вставка

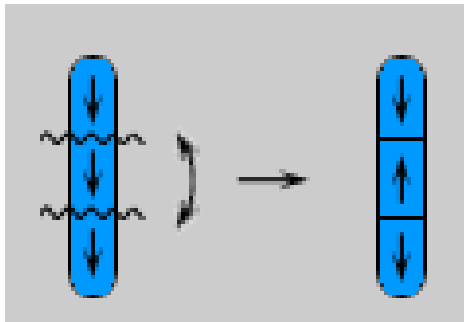
R: AIB
Q: AB



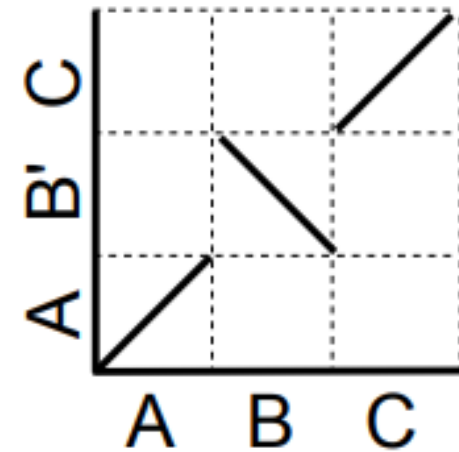
❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

Инверсия

Инверсия – кусочек вырезан и вставлен, но прямая и обратная цепочка перепутаны

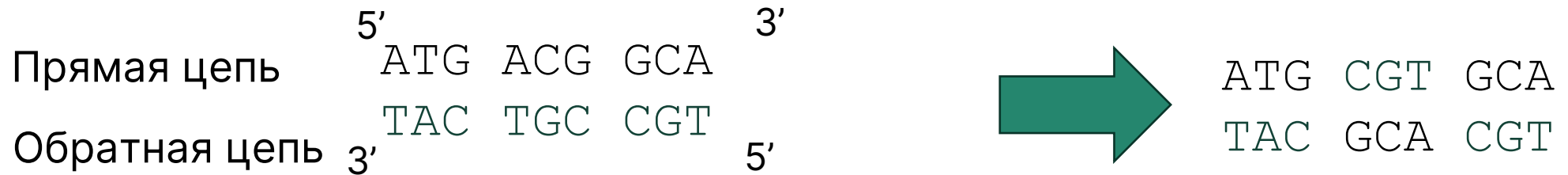


R: ABC
Q: AB'C



❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

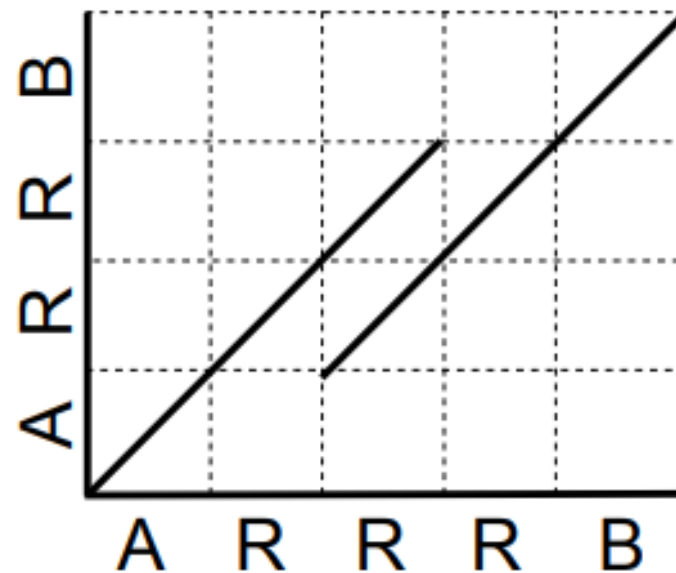
Инверсия



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

Дупликация

R: ARRRB
Q: ARRB



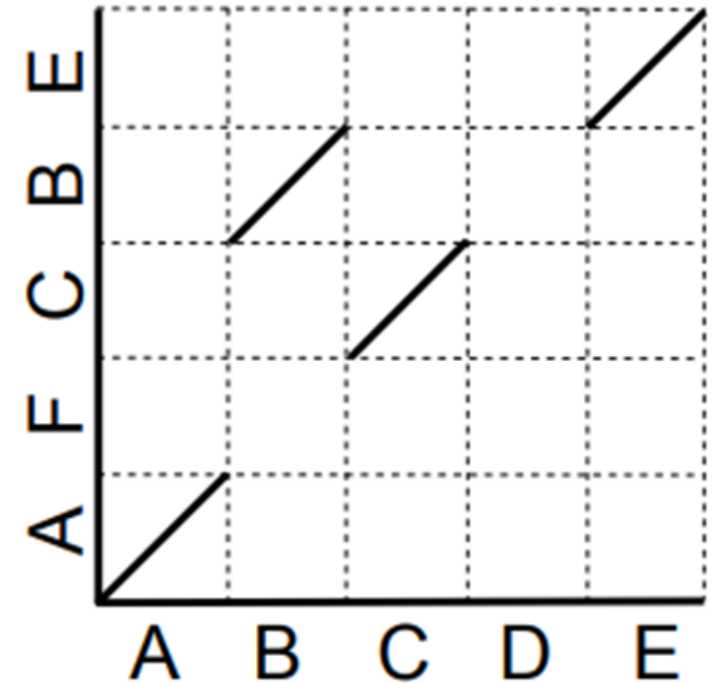
❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

Транслокация

Транслокация – перенесение фрагмента в другое место.

С сохранением ориентации или нет.

R: ABCDE
Q: AFCBE

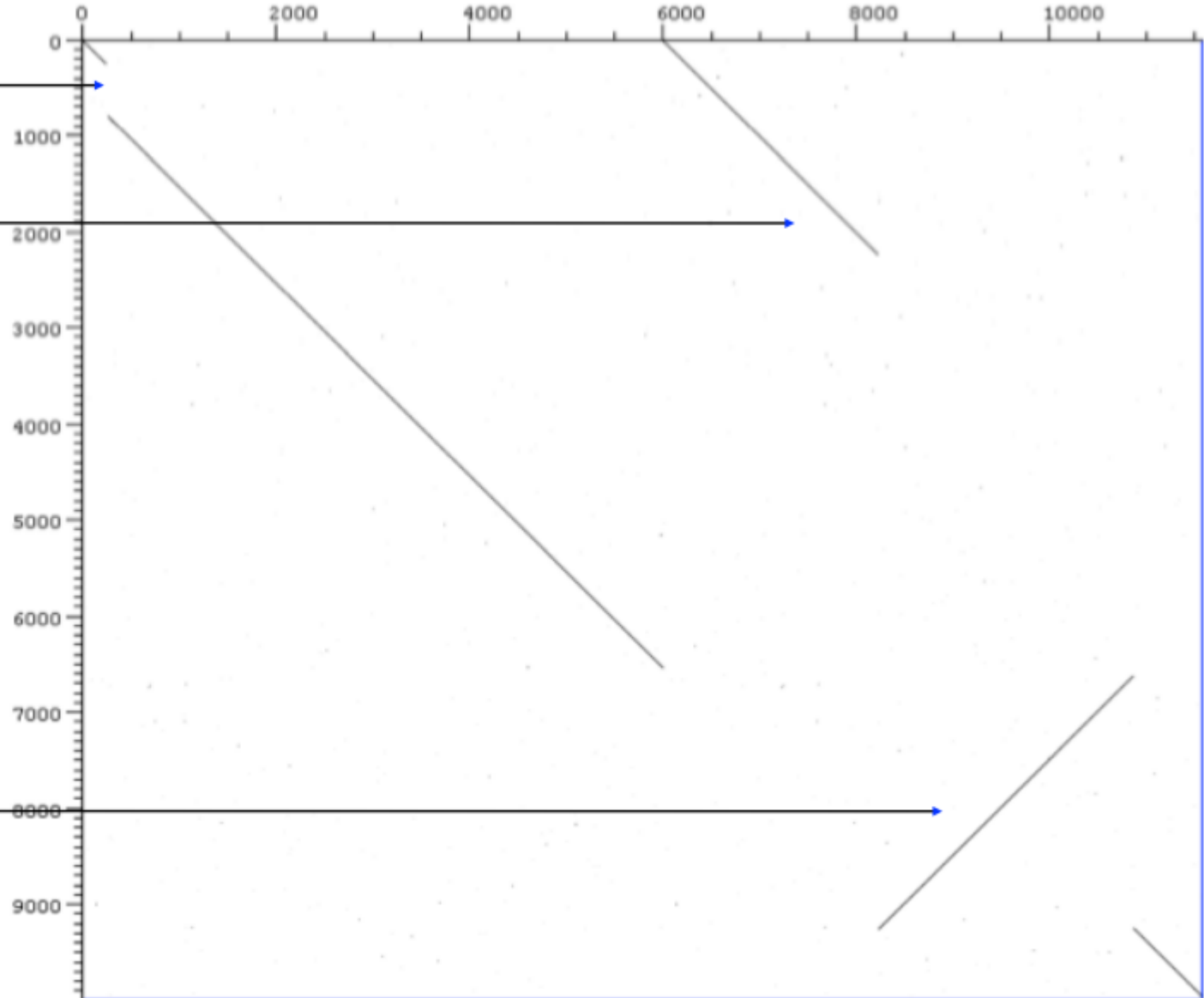


❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

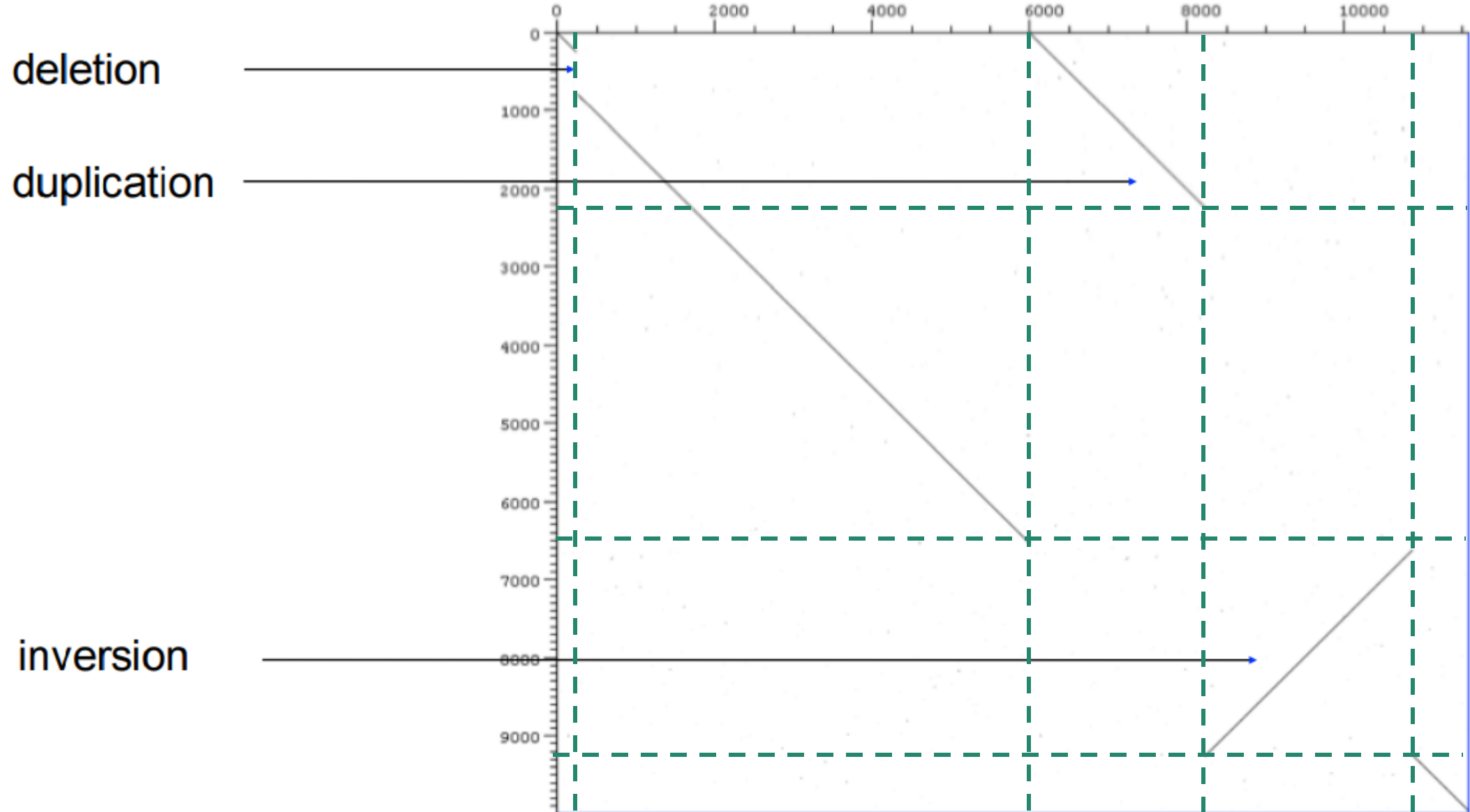
deletion

duplication

inversion



Как читать такую карту



BLAST двух последовательностей

Nucleotide BLAST: Search nucleotide sequences

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastn suite

Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide 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 [Не выбран ... один файл ?](#)

Job Title

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

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

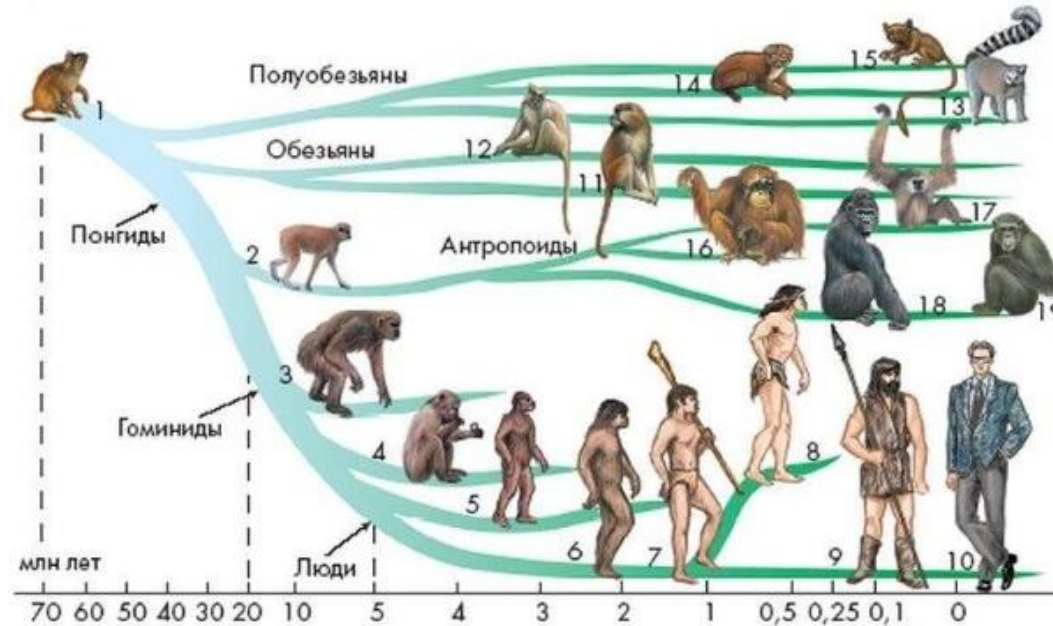
Organism Optional ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Гомология и сходство

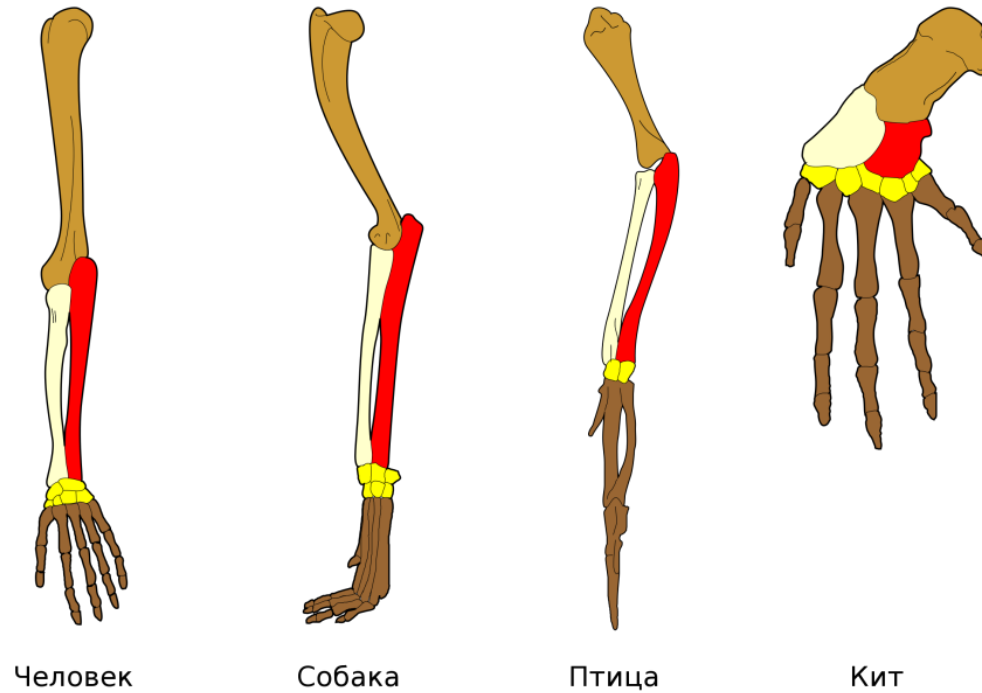
Общность происхождения



Последний общий предок ныне живущих обезьян.

- ❖ Гомоло́гия (в биологии) – сопоставимость частей сравниваемых биологических объектов, обусловленная общностью происхождения

Общность происхождения



Термин «гомология» употребляют для частей организмов.
Для целых организмов термин «гомология» не употребляют!

Гомология в молекулярной биологии

- ❖ Гомология того, что закодировано в геномах – генов, белков, РНК и просто участков ДНК
- ❖ Белки гомологичны, если их гены произошли из гена их общего предка.
- ❖ При каждом делении клетки ген дочерней клетки – копия гена материнской клетки
- ❖ Ошибки – мутации – в ДНК случаются. Самая распространенная мутация: из пары mC-G получается пара T-G (mC – метилированный цитозин). Такие неправильно спаренные основания возникают в процессе репликации.
- ❖ Пары T-G репарируются системой репарации неправильно спаренных оснований

Эволюция геномов (бактерий)

Половое размножение отсутствует

Мутации – наследственное изменение в нуклеотидной последовательности генома организма:

- ❖ Точечные – замена, делеция, вставка 1 нуклеотида
- ❖ Локальные – одновременное изменение короткого участка (делеция 5 нт, замена 2 нт рядом, вставка 6 нт)
- ❖ Крупные – делеции, дупликации, транслокации

Эволюция геномов (бактерий)

Половое размножение отсутствует

❖ Мутации

- ✓ Происходят случайно. С разной частотой*
- ✓ На мутации действует отбор:
 - Носители вредных и слабо вредных мутации удаляются из популяции
 - Полезные мутации чрезвычайно редки

❖ Локальные изменения накапливаются от поколения к поколению

- ✓ Пытаются по числу локальных изменений измерять время от потомков до последнего общего предка
- ✓ Крупные единовременные изменения закрепляются редко

**) У *Deinococcus radiodurans* частота повыше. Почему?*

Эволюция белков

Локальные изменения - результат небольших изменений в гене (замены аминокислот, делеции, вставки)

Большие изменения:

- ❖ Накопленные небольшие изменения
- ❖ Небольшие изменения гена ведущие к большим изменениям белка
 - ✓ Мутация стоп кодона => удлинение последовательности белка
 - ✓ Мутация кодона на стоп кодон
 - Гибель белка = псевдогенизация
 - Укорочение последовательности белка
 - Программируемый сдвиг рамки считывания
 - ✓ Мутация в сайте инициации (начала) трансляции – приобретение нового-старт-кодона
 - ✓ Делеция или вставка, не кратная 3 => сдвиг рамки считывания
- ❖ Крупные перестройки генома, затрагивающие гены (инверсии, транслокации, большие делеции или вставки)
- ❖ Закодированные в геноме перестройки для быстрого изменения последовательности белка (Ig, DGR)

Гомология в молекулярной биологии

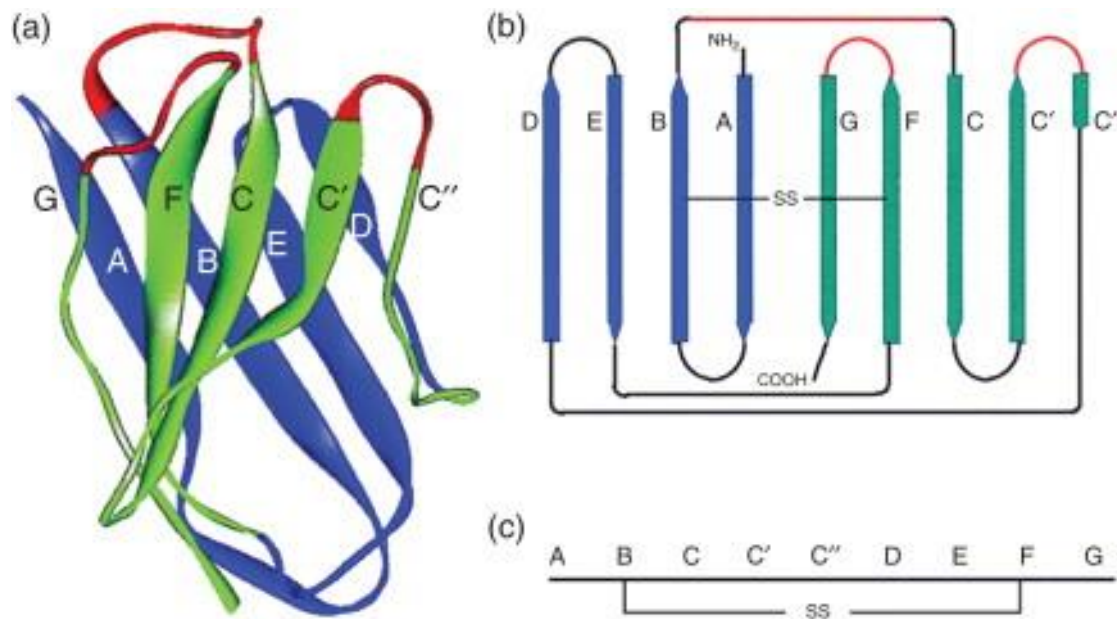
- ❖ О гомологии белков судят по сходству их последовательностей
- ❖ ГОМОЛОГИЧНЫЕ белки МОГУТ иметь отличающиеся функции
- ❖ Гомология  Сходство последовательностей

Гомология в молекулярной биологии

- ❖ Гомология  Сходство последовательностей
 - ✓ Последовательности могли сильно измениться в эволюции от общего предка
 - ✓ Есть системы быстрой эволюции белковых последовательностей (Иммуноглобулины, Diversity-generating retroelement (DGR) у вирусов и бактерий)
 - ✓ Дольше в эволюции сохраняется сходство пространственных структур

Иммуноглобулиновая укладка (Ig-like fold)

- ❖ Ig-like Fold встречается во многих белках одного организма, связанных с иммунитетом и не только, и у всех животных
- ❖ Последовательности Ig-like Fold могут сильно отличаться, детектируемое сходство только на коротких участках
- ❖ Ход полипептидной цепи и элементов вторичной структуры очень похожи



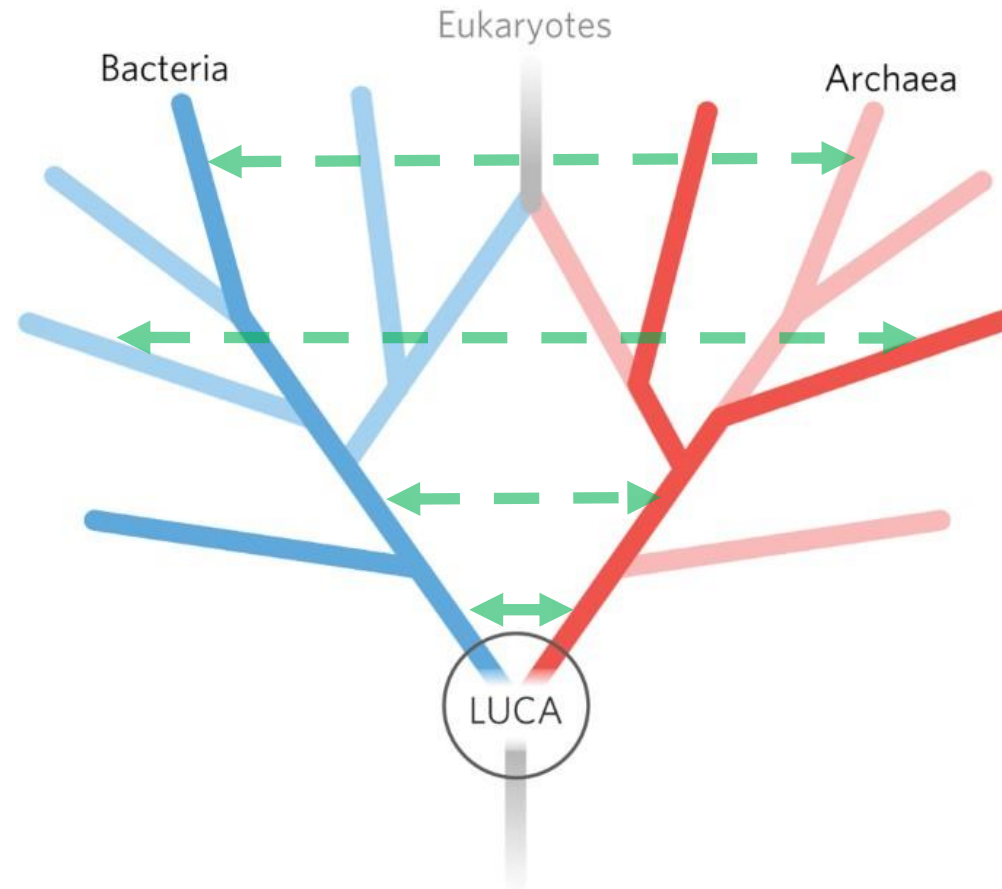
Гомология в молекулярной биологии

❖ Сходство последовательностей  гомология

Гомология в молекулярной биологии

- ❖ Сходство последовательностей  гомология
 - ✓ Бывает случайное совпадение – аналогия (повторы, простые участки)
 - ✓ Чем длиннее фрагмент последовательности, тем менее вероятно совпадение

LUCA – Last Universal Common Ancestor



- Обоснование на молекулярном уровне было сформулировано в начале 2000х (Aravind et al, 2002; Koonin, 2003 10.1038/nrmicro751)
- Были собраны все белки, гомологи которых встречаются во всех таксонах высокого порядка (Weiss et al, 2018 10.1038/nmicrobiol.2016.116)
- Возраст исчисляется миллиардами лет.

НЕ СЛЕДУЕТ ДУМАТЬ, LUCA БЫЛ чем-то одним
Всё выжившее – сложно устроено и очень разнообразно

Эволюционное выравнивание – выравнивание отражающее эволюцию последовательностей

предок	TATGCGAAT-GCCCTGAA	
сын	TATGC A AAT-GCCCTGAA	замена
внук	TATGC AA AT-GC T CTGAA	замена
правнук	TATGC AA AT C GCTCGGAA	вставка 1 п.н.
праправнук	TATGC AAA A C GCTCGGAA	замена
прапраправнук	TATGC AAA - C GCTCGGAA	делеция 1 п.н.
...	TATGC A T A- C GCTCGGAA	замена
...	TATGC A T A- C GC- - -GAA	делеция 3 п.н.

- Это эволюционное выравнивание
- Гомологичные нуклеотиды расположены друг под другом
- Так бывает только в экспериментах по изучению эволюции

предок

TATGCGAATGCCCTGAA



TATGCGAAT-GCCCTGAA

TATGCGAATG-CCCTGAA

TATGCGAATGC-CCTGAA

сын

TATGC**A**AAT-GCCCTGAA

TATGCC**A**ATG-CCCTGAA

TATGCCA**T**TGC-CCTGAA

внук

TATGC**A**AAT-GC**T**CTGAA

TATGCC**A**ATG-CC**T**TGAA

TATGCCA**T**TGC-CCT**A**AA

правнук

TATGC**A**AAT**C**GC**T**CGGAA

TATGCC**A**ATG**C**CC**T**GGAA

TATGCCA**T**TGC**A**CCG**A**AA

праправнук

TATGC**A**AA**A**C**G**CTCGGAA

TAT-**C**CAATG**C**CC**T**GGAA

T-**-**-CCAT**T**TGC**A**CCG**A**AA

прапраправнук

TATGC**A**AA-**C**GC**T**CGGAA

TAT-**C**CAATG**C**CC**T**GG**T**A

T-**-**-CCAT**T**TGC**A**CCG**A**A**T**

...

TATGC**A****T**A-**C**GC**T**CGGAA

TAT-**-**CAATG**C**CC**T**GG**T**A

T-**-**-**G**CA**T**TGC**A**CCG**A**A**T**

Потомок, которого
мы видим

TATGC**A****T**A-**C**GC-**-**-GAA

TAT-**-**CAATG**C**CC**T**GG**T**A

T-**-**-**G**CA**T**TGC**A**CCG**A**A**T**

Эволюционное выравнивание – выравнивание отражающее эволюцию последовательностей

- ❖ Выравниваем последовательности потомков относительно предка:

предок	TATGCGAAT-G-C-CCTGAA	} Это тоже эволюционное выравнивание!
потомок 1	TATGC ATA - CG -C- --- GAA	
потомок 2	TAT- -CAAT -G CC -CTGGTA	
потомок 3	T- ---GCATT -G-C ACCGAAT	

- ❖ В жизни мы решаем обратную задачу – построить эволюционное выравнивание, зная последовательности потомков

Гомологию последовательностей нуклеотидов и белков выводят из сходства последовательностей

Геномы и белки, как молекулы, определяются (почти *) однозначно) своей последовательностью

Поэтому их гомология определяется похожестью последовательностей

❖ Высокая ~~гомология~~ последовательностей – **НЕТ**

У гомологии нет степени, она есть у сходства!

❖ Высокое сходство последовательностей – **ДА**

Множественное выравнивание

```

                *           100           *
THIE_LACLS : AGVSEFIVNDDVELARELNADGIHIGQTDES SVSKVI
THIE_MANSM : FQVPFIVNDDVELALSIQADGIHVGQKDTAVETII
THIE_STRA3 : YQVPFIIDDDIDLVELIDADGLHIGQN DLPVDEAI
THIE_LISIN : YQVPFIINDDVALALEIGADGIHVGQNDEEIRQV
THIE_ANOFW : YNIPFIVNDDVDLALALQADGVHVGQED EVAERVI
THIE_GEOTN : YGVPFIVNDDVELAIAIDADGVHVGQDDEDARRVI
THIE_BACSU : AGVPFIVNDDVELALNLKADGIHIGQEDANAKEVI
THIE_BACA2 : AGIPFIINDDVELALRLEADGVHIGQDDADAEETI
THIE_OCEIH : FQIPFIINDDVDLAKQLDADGIHIGQDDQPVEVVI
THIE_STAAB : YNVPFIVNDDVSLAKEINADGIHVGQDDAKVKEI
THIE_STACT : YNVPFIVNDDVALAEEIDADGIHVGQDDEAVDDFI
              6pFI6lDD6 La  6 ADG6H6GQ D
```

Зачем нужно множественное выравнивание?

- ❖ Построение множественных выравниваний — необходимый этап решения многих задач молекулярной биологии
 - ✓ Эволюционные реконструкции
 - ✓ Предсказания функции
 - ✓ Предсказания структуры
 - ✓ Классификация
- ❖ Выравнивают и нуклеотидные, и аминокислотные последовательности

Смысл выравнивания последовательностей гомологичных белков

- ❖ Некоторые кодоны а.к.о. гомологичных белков потомков *произошли из одного кодона последнего общего предка* этих белков, или были deletированы в эволюции, или появились в результате вставки новых кодонов
- ❖ Цель программ множественного выравнивания *последовательностей гомологичных белков* – воспроизвести эволюционное выравнивание
- ❖ Это не всегда хорошо получается. Есть проблемы
- ❖ *Программы выравнивания основываются на сходстве последовательностей*, так как последовательности белков обычно подвержены стабилизирующему отбору и потому их последовательности изменяются медленно
- ❖ Сходство может появиться случайно (теор. вер.)

Вывод. Нужно учиться чему верить в выравнивании, а чему нет!

Пример хорошего выравнивания

```
11 DRREIRHIWDDWSSSF TDRRVAIVRAVFDDL FKHYP TSKALFERVK IDEPESGEF 66
8 DRHEVL DNWKGIWSAEFTGR RVAIGQAIFQELFALDPNAKGVFGRVNVDPKSEADW 63
8 DRREVQALWRSIWSAEDTGRRTLIGRL LFEELFEIDGATKGLFKRVNVDDTHSPEE 63
7 QR I KVKQQAQVYSVGES- -RTDFAIDVFNNFFRTNPD- RSLFN RVNGDNVYSPEF 59
9 QRLKVKQQA KAYGVGHE- -RVELGIALWKSMFAQDNDARDLFKRVHGEDVHSPAF 62
8 EGLKVKSEWGRAYGSGHD- -REAFSQA IWRATFAQVPESRSLFKRVHGGDDTSHPAF 61
6 QRFKVKHQWAEAFGTSHH- -RLDFGLKLWNSIFRDAPEIRGLFKRVDGDNAYS AEF 59
7 QRLKVKRQWAEAYGSGND- -REEFGHF I WTHVFKDAPSARDLFKRVRGDN IHTPAF 60

67 KSHLVRVANGLDLLINL LDDTLVLQSHLGHLADQHIQRKGVTKKEYFRG IGEAFA 120
64 KAHVIRVINGLDLAVNLL EDPKALQEELKHLARQHRERSGVKAVYFDEMEKALL 117
64 FAHVLRVVNGLDTLIGVLGDSDTLNSLIDHLAEQHKARAGFKTVYFKEFGKALN 117
60 KAHMVRVFAGFDILISVLDDKPVLDQALAHYAAFHKQF- GTIP- -FKAFGQTMF 110
63 EAHMARVFNGLD RVISSLTDEPVLNAQLEHLRQQHIKL- GITGHMFNLMRTGLA 115
62 I AHAERVLGGLDIAISTLDQPATLKEELDHLQVQHEGR- KIPDNYFDAFKTAIL 114
60 E AHAERVLGGLDMTISL LDDQAAFDAQLAHLKSQHAER- NIKADYYGVFVNELL 112
61 RAHATRVLGGLDMC IALLDDEGV LNTQLAHLASQHSSR- GVSAAQYDVVEHSVM 113
```

Пример плохого выравнивания

```
GSYLEVEGPKGKTTVYVTDLYPEGARGA-----L-DL---SPNAFRKIGNMKDGKINI-----KWRV
FSRYKLKYLDLDPKQSYGSAEE-----I-HT---ITRKGFSKEQPNAAKLLS-----QFKW
VKDTRAASEEELLQKAHFLQRMLSYGTTTAEVKSGYGLEKETELKQLRVAKKL-HE---SQPVDLVSTFMGAHAIPP-----EYQN
-----EFTDELAKQYF-----HA---VPKIAAIRDEFEPIGLN-----TLNN
VPEEWWDI I RESMPAYNS-----NY---TPTTRGMGQIVELFRSYP-----EVKR
GGIPQISLHLANPAFQSGHFKTPITNDQYKKILDSSTVEGKRLN-----AM---LSKIDGLQELENQGVVPV-----LFRP
-----EKFEAKLEEYNG-----NL---KRAAVELAKEWRSDKVLR-----KLEA
-----EE
YDPSTWGKKEPSGTQEEARARSQKNQKARVMVDIPANTTIVGSGTNAKVVGGNF-QI---KSDNVIIRNIEFQDAYDY-----FPQW
FIFTNCETICPPMTA-----HM---TDLQKKLKAENIDVRIIS-----FSVD
SALGINTDFSPVVDINNPNPNVIGVRSFSSNRELTSRLGLYTMKGLQRQDIAS-AL---KHFPGHGDTDVDSHYGLP-----LVSH
LSSDDPNSYHYFMNDR-----FF---QHIDSKPSRHFIPNGNAD-----DLEA
FSPKQIRARFPHL-----HI---GRSVHSLEEAVQAEKEDA-----DYVL
QNVKLLNIDADTLTD-----VF---EPGEVKRVYINESDPWPK-----KRHE
IEQAWEEGKQVCIPKC-----
FHLNKNEYTVPFVCGCRDLGEATRR-
IIVSGDSDQSPWVKKEMGRAAVACMK
-----SIMGLMSLAVSTGTEV-----TLIA
QLIKRMSDEGHIIGNHSFHHP-----DL---TTKTADQIQDELDSVNEE-----VYKI
---EKFEELYDHAAETVD-----TI---AERLLAIGGQPVATVKEY-----TEHA
-----EK---YRGLVLTSGKKGKIGKRLV-----YRHE
FSTFSVQSYNKRakeeAE-----KL---GGLTITSVESWYDEPKFV-----TYWV
SNLPEHTERYEIAQEHLDVVRGLWNSWEHDAFIHNKKTGQFFDQAKLHRLNHKGKYPQVEGPLNIGRSKQGEPPVVFQA-----GSSE
KGKSDFSLFSPLYEQLEIFNKRFIITYWGYQEHYPDALLDLFEPGVTVKVLDQLFAEL---KEAIPLVKQVTASGNKPDTSFITKAFPKEKQKELSLYFLQE
QFMEPLREFAAQGKPMFGTCA-----GL---ILLAKEIAGSDNPHLGLL-----NVVV
-----SIMGVMSLGIKGAIEI-----TISA
CKGANEVLYPEHAQALRE-----QL---APLATVITPNLFEASQLS-----GMDE
IDVYFPDSEISCLAQGFY-----AL---KAAELIKNGASSPEDIK-----ELEE
TYPLAIQIEEGIVFEDD-----QF---IVTAVSVIHGVEAFGYRV-----QEKD
TKDERYVKAARLRSLYGTLNRTSEGGFWHKDGYPYQMMLDGLYMGGPFALKYA-NL---KQETELFDQVVVLQESLMR-----KHTK
TIVDEKGRSFIDDVGEGDLWYFPSGLPHSIQALEEGAELLVFDGGSFSENSTF-QL---TDWLAHTPKEVIAANFGV-----TKEE
RNDIEIEVKDGELPST-----FV---PGRNLVFLSFASILAYQI-----GARH
GELVPDEVTIGIVKERLGK-----DD---CERGFLLDGFPRTPVAQAE-----ALEE
ADGIIHGQEDANAKEVRA-----AI---GDMILGVSAHTMSEVKQA-----EE
EFAELLEISEERKAF-----KP---SPTAYSYSHTMYRSVLSG-----NFAE
INMTGDGAEDTFQIPVTDGQLNLLVTEGKAGTAFTLSALKIKKLSQPVNTRTI-YV---GGDSTVCNYYPLNSSKQA-----GWGQ
```

Программа выравнивает всё, что Вы ей дадите

DNA (cytosine-5-)-methyltransferases

	10	20	30	40	50	60
<i>EJL77459.1</i>	GVDLVF	GGPPCQGF	SQIGMRR-	LDDE-NE	LYQQYTRIVAKLKPRVFLMENVPNL	ALMNKGH
<i>RXK67093.1</i>	DLDVVFG	GGPPCQGY	SQIGTRR-	LDDE-NE	LYLQYARIVEKQPRMFLMENVPNMVLLNKGH	
<i>OJY44288.1</i>	NVDLVFG	GGPPCQGY	SQIGTRD-	LHDPR-NR	LFEEFARVVATLKP	KLFLMENVPNL
<i>TRU90449.1</i>	NPEMIVG	SPPCQDF	SSAGKRNEGLGR-	-ANLTLTFAEIVTRVSPQWFVMENV	D--RIE	SKSK
<i>OXI46696.1</i>	GTDLVFG	GGPPCQGF	SQIGMRR-	LDDE-NE	LYKQYTRVVSTLRPRVFLMENVPNL	ALMNKGH
<i>AVZ30243.1</i>	EIDVVFG	GGPPCQGF	SLIGKRS-	FEDPR-NS	LVFHYIRLVLELSPKFFVIENVKGMTAGNHQA	
<i>AFZ12381.1</i>	DIIGFI	GGAPCPDFS	VGGKNRGSEGD	-GKLSASYIELICQQKPDFFLFENVKGLYKTKKHR		
<i>HCQ21462.1</i>	HIIGFI	GGPPCPDFS	VGGKNKGHLGDN-	GKLSSAYIELICQNL	PDFFLFENVKGLWRTTKHR	
<i>EDN77159.1</i>	SLIGFI	GGPPCPDFS	IAGKNKGKDGD	-GKLSLSYTNLIEMKPDFFLFENVKGLWRTARHR		
<i>SOD91684.1</i>	EVSLVV	GGAPCQPF	SNIGKKLGKNDER	NGDLFLFVRMVKGIQPEAFIFENVVGITQNKHSD		
<i>QCS48280.1</i>	NVVGFI	GGPPCPDFS	IGGKNRGRQGDH-	GKLSESYIDLIIQHQP	DDFFIFENVKGLYRTTKHR	
<i>SMB95934.1</i>	GLFGI	GGPPCPDFS	VGGKNRGENG	EQ-GRLSKVVDKILDLQPVFFLYENVPGLIRTAKHR		
<i>RUO38876.1</i>	SPVGF	IGGPPCPDFS	VGGKNRGHEGEN-	GRLTRTYVDGIIKYAP	DDFFIFENVKGLWRTKRHR	
<i>OIP70538.1</i>	TIDLIC	GGPPCQGF	STIGTND-	KKDHR-NFLFEFLRMVETFKPNFIILENVTGLLAKKNES		
<i>AFY60915.1</i>	NLVGFV	GGPPCPDFS	IGGKNKGQYGD	-GKLTKVYVDIIENQPDFFVFENVKGLWRTRSRHR		
<i>CUR30340.1</i>	DLIGFI	AGPPCPDFS	VGGKNRGKNGDQ-	GKLTA	CYVELICQQRPDFFVFENVKGLWSTTKKHR	
<i>TAK03971.1</i>	QAALVV	GGAPCQPF	SNLGSKRGTADSR-	GT	LFQDFIRIVKGV	RPKGFI
<i>AEE51071.1</i>	KVALVV	GGAPCQPF	SNIGKKEGENDAK	NGDLFLFVRMVKGIQPEAFIFENVAGIIQSKH	SK	
<i>RTR31666.1</i>	RLVG	FVGGPPCPDFS	VGGKNKGSEGEN-	GKLTRTYIDLIVKDN	PDYFI	ENVKGLWRTTRHR
<i>PTU64472.1</i>	NIDLVFG	GGPPCQGF	SQIGTRR-	LDDE-NE	LYKQYTRIVKTLKPRVFLMENVPNL	AMMNKGH

Продолжение того же выравнивания

	200	210	220	230	240	250	260	270	280
YIG-VPQDRKRVFIVGYREDLN	200	210	220	230	240	250	260	270	280
YIG-VAQDRERVVFYVGRKDLN									
YIG-VAQERKRVFYIGFRKDLEIKF									
YIG-VAQDRKRVFYIGFRKELNIN									
YIG-IPQQRDRLVLF									
YIG-VPQSRQRVFFIGLKSDRPLNQ									
CG-VPQLRKRTFVIGHRHGSIAD									
CG-VPQSRTRFSLIGKLNSEHNF									
YIG-VPQRRHRIIVGIRKDQD									
IG-AHHQRHRWFCLAIRKDYEP									
FGNIPQNRERIYIVGFRN									
FN-VPQNRERLYIIGIREDLIKNEE									
FG-IPQNRERVFCISILN									
FG-SSQARRRVFMISTLNEF									
FG-IPQNRERIYLVGF									
FG-LPQNRERIYIVGFDRKS									
FG-VPQNRERIYIVGFNKEK									
FG-VPQNRERIYIVGFHKS									
IFQ-VPQNRRLRVYIVGLDQSQPELT									
FG-VPQNRVRIYILGILGSKPKLT									
IFG-IPQKRKRFFYLVAFLNQ									
IFG-LPQRRERIIVIVGFHPDLG									
YIG-IPQKRERIYMICFRNDLN									
YIG-NAQRRRRVFIFGYKQDLNYSKAME									
YIG-TPQRRRKRAIRLNKKGTIWN									
IGG-TPQVRERVFITATLVPERMRDERIPRTETGEIDAEAIGPKPVATMNDRFP									
YIG-VAQNRDRVFIIIGIQQKLGVPD									
IFG-VAQNRDRVFIVGIIQQKLDLNG									
IYD-VAQKRERIVIIIGIREDLVK									
YIG-VSQLRPVLFVALKNEYTN									

Почему выравнивание неоднородно?

- ❖ Программа построила выравнивание неправильно?
- ❖ Неравномерная скорость мутаций в разных местах белка.
Почему?
 - ✓ Мутации в геноме происходят неравномерно – НЕТ. Мутации происходят случайно, им всё равно – где (в первом приближении).
 - ✓ Отбор решает какие мутации и даже крупные перестройки оставить, а какие запретить.
- ❖ Где выравнивание правильное (имеет шанс соответствовать эволюционному), а где – НЕТ

Хорошее выравнивание

		10	20	30	40	50	60
EJL77459.1	GVDLVF	GGPPCQGFSQ	IGMRR-LDDER-NE	LYQQYTRIVAKLKPRVFL	MENVPNLALMNK	GH	
RXK67093.1	DLDVVF	GGPPCQGYSQ	IGTRR-LDDER-NE	LYLQYARIVEKQRPRMFL	MENVPNMVL LNK	GH	
OJY44288.1	NVDLVF	GGPPCQGYSQ	IGTRD-LHDPR-NRL	FEEFARVVATLKPKLFL	MENVPNL LLLN	KG	H
TRU90449.1	NPEMIV	GSPPCQDFSSA	GKRNEGLGR--ANL	TLTFAEIVTRVSPQWF	VMENVD--RIE	KSK	
OXI46696.1	GTDLVF	GGPPCQGFSQ	IGMRR-LDDER-NE	LYKQYTRVVSTLRPRVFL	MENVPNLALMNK	GH	
AVZ30243.1	EIDVVF	GGPPCQGFSL	IGKRS-FEDPR-NS	LVFHYIRLVLELSPKFF	VIENVKGMTAGN	HQA	
AFZ12381.1	DIIGFI	GGAPCPDFSV	GGKNRGSEGDK-GK	LSASYIELICQQKPDF	FLFENVKGLYKTK	KHR	
HCQ21462.1	HIIGFI	GGPPCPDFSV	GGKNKGHLGDN-GK	LSSAYIELICQNLPDF	FLFENVKGLWRTT	KHR	
EDN77159.1	SLIGFI	GGPPCPDFSI	AGKNKGKDGDN-GK	LSLSYTNLI IEMKPDF	FLFENVKGLWRTA	RHR	
SOD91684.1	EVSLVV	GGAPCQPF	SNIGKKLGKNDER	NGDLFLEFVRMVKG	IQPEAFIFENVVGI	TQNKHSD	
QCS48280.1	NVVGFI	GGPPCPDFSI	GGKNRGRQGDH-GK	LSESYIDLIIQHQP	PDFFI FENVKGLY	RTKKHR	
SMB95934.1	GLFGII	GGPPCPDFSV	GGKNRGENGEGE	Q-GRLSKV FVDKI	LDLQPVFFLYENV	PGLIRTAKHR	
RUO38876.1	SPVGFI	GGPPCPDFSV	GGKNRGHEGEN-G	RLTRTYVDGIIKYA	PDFFI FENVKGLW	RTKRHR	
OIP70538.1	TIDLIC	GGPPCQGFST	IGTND-KKDH-NF	LFEEFLRMVETFK	PNFI ILENV	TGLLAKKNES	
AFY60915.1	NLVGFV	GGPPCPDFSI	GGKNKGQYGDN-GK	LTKVYVDII IENQ	PDFFVFENVKGLW	RTRSRHR	
CUR30340.1	DLIGFI	AGPPCPDFSV	GGKNRGKNGDQ-GK	LACYVELICQQR	PDFFVFENVKGLW	STKKHR	
TAK03971.1	QAALVV	GGAPCQPF	SNLGSKRG	TADSR-GTLFQDF	IRIVKGV	RPKGFIFENVEGL	TQDKHKG
AEE51071.1	KVALVV	GGAPCQPF	SNIGKKEGENDAK	NGDLFLEFVRMVKG	IQPEAFIFENVAG	I IQSKHSK	
RTR31666.1	RLVGFI	GGPPCPDFSV	GGKNKGSEGEN-GK	LTRTYIDLIVKDN	PDYFI FENVKGLW	RTTRHR	
PTU64472.1	NIDLVF	GGPPCQGFSQ	IGTRR-LDDER-NE	LYKQYTRIVKTLKPRVFL	MENVPNLAMMNK	GH	

Учитываем

- Наличие консервативных позиций
- Отсутствие гэпов

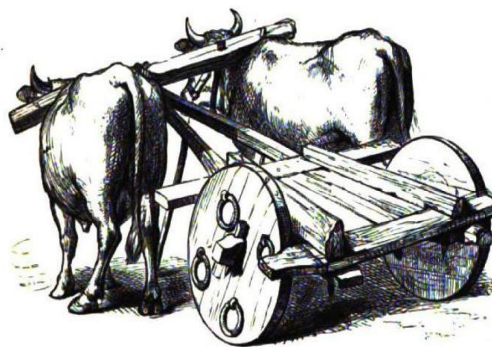
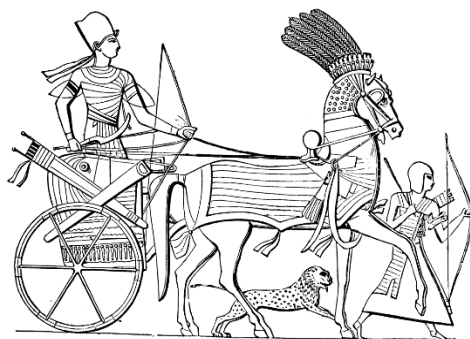
- ❖ Участки выравнивания (не отдельные колонки), со значительным сходством свидетельствуют об общности происхождения

Выравнивание тех же белков другой программой

	10	20	30	40	50	60	70	80
-GVPQDRKR VFI	---	VG	---	REDL	---	NL-KFEFPKP	---	LN--KKVTLRDAIGDL
-GVAQDRER VFY	---	VG	---	RKDL	---	NISNFEFPYP	---	ISEK--ERKYLKDSI WDL
-GVAQERKR VFY	---	IG	---	RKDL	---	EI-KFSFPKGSTVEDK	---	DKITLKDVI WDL
-GVAQDRKR VFY	---	IG	---	RKEL	---	NI-NYLPPIPHLI	---	KPTFKDVI WDL
-GIPQQRDR LVL	---	FAA	---	KQGV	---	KIIPPTH	---	TPENYRTVRDVI GSL
-GVPQSRQR VFF	---	IGL	---	KSDR	---	PLNQQILTPPSK	---	VIES--EYTSLEEAI SDL
-GVPQLRKRTFV	---	IGH	---	RHGS	---	ADLANVLQ	---	QRLAKQSLTVRDY
-GVPQSRTRFSL	---	IGK	---	LNSEH	---	NFLIPTLS	---	RKLSDKPMTVRDY
-GVPQRRHR I I I	---	VGI	---	RKDQ	---	DV-AFRVPEP	---	THKE--KYRTASEALADI
-GAHHQRHRWFC	---	LAI	---	RKDYEP E E I I	---	VSVNATKFDWENNE	---	PPCQVDNKS YENSTLVRL
GNIPQNRERI YI	---	VG	---	RNIE	---	HYKNFNFMP	---	QPLTLTIKDM
-NVPQNRERLYI	---	IGI	---	REDLIKNE	---	EWSLDFKRKDI LQKGKQRLVELDIKSFNFRWT	---	AQSAATKRLKDL
-GIPQNRER VFC	---	ISI	---	LNPN	---	DFTFPQK	---	QNLTLSMNDL
-GSSQARRRVFM	---	IST	---		---	LNFEVELPKGD	---	KKPKSIKKV
-GIPQNRERI YL	---	VG	---	LNHDV	---	DFRFPQP	---	IGQATAVGD I
-GLPQNRERI YI	---	VG	---	DRKSI S	---	NYSDFQMPTP	---	LQEKTRVGN I
-GVPQNRERI YI	---	VG	---	NKEKVR	---	NHEHFTFPTP	---	LKTKTRVGD I
-GVPQNRERI YI	---	VG	---	HKST	---	GVNSFSYPE	---	PLDKIVTFADI
-QVPQNRRLRVYI	---	VGLD	---	QSQPELTITSHI	---	GATDS	---	HKFKQ--LSNQASL--FD--TNKIMLV RDI
-GVPQNRVR IYI	---	LGIL	---	GSKPKLTLTSNV	---	GAADS	---	HKYK--NEQISL--FD--ESYATVKDI
-GIPQKRKR FYL	---	VA	---	LNQNI	---	HFEFPKP	---	PMISKDIGEV
-GLPQRER IVI	---	VG	---	HPDL	---	GINDFSFPK	---	GNPDNKVP INAI
-GIPQKRERI YM	---	ICF	---	RNDL	---	NIQNFQFPKP	---	FELNTFVKDL
-GNAQRRRR VFI	---	FGY	---	KQDLNYSKAMEESPLD	---	KI IYHN	---	GLFAEAFPIEDYANKNRVNRTHITHDIVDI
-GTPQRRKRAI IRLNKKGT	---	IWNL	---		---	PLKQNI	---	SVEQAIGNL
-GTPQVRER VFI	---	TATLVPERMRDERIPRTET	---	GEIDA	---	EAIGPK	---	PVATMNDRFP I KK--GGTELFHPGDRKSGWNL
-GVAQNRDR VFI	---	IGI	---	QKQL	---	GVPDFSFP EYSESEQRLYDILDNLQTPSII	---	
-GVAQNRDR VFI	---	VGI	---	QKQL	---	DLNGFSFP EYAESDQRLYHILDNLEAPETK	---	
-GVAQNRDR VFI	---	VGI	---	QKQL	---	DLNGFSFP EYTESEQRLYHILDNLEVPETK	---	
-DVAQKRER IVI	---	IGI	---	REDLVKE	---	QKYPFRFPLA	---	QVYKPV LKDV
-GVSQLRPRVLF	---	VAL	---	KNEY	---	TNFFKWPEP	---	NSEQPKTVGELLFDL

Консервативное значит важное

- ❖ Консервативное - то, что длительно существует в эволюции, с несущественными изменениями



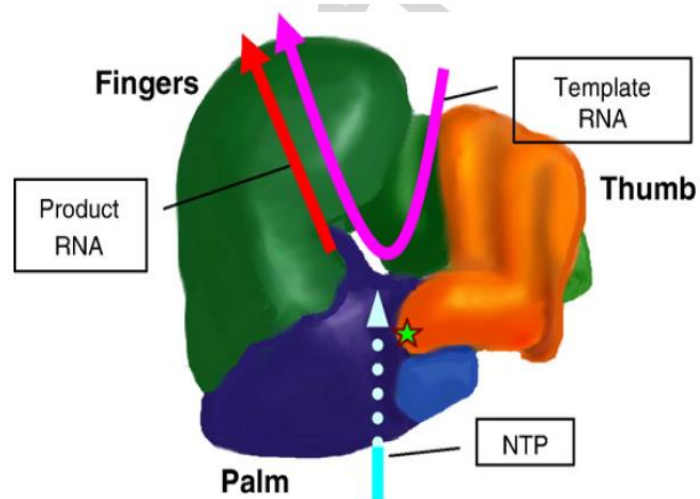
- ❖ У белков всё то же самое
- ❖ Сохраняющееся в эволюции (консервативное) – важно

Выравнивание РНК зависима РНК полимераз (RdRP)

```

*      320      *      340      *      360      *      380      *      400      *      420      *      440      *      460
'KTMIRFGDVGLDLDFSAFASLSPFMIREA...GRIMSELS...GTPSHFGTALINTIISKHLLYNCCY...HVCGSMPSGSGCTALINSIINNVLYYVFSKIFGKSPVFF.....CQALKILC.YGDDVLIIVFSRDV
SVAMQG.FERVYDVLDYSNEDSTHSVAMFRLL...A...EEFF.TPENGFDPLTREYLESIAISTHAFEEKRF.....LITGGLPSGCAATSMINTIMNNIIRAGLYLTYNKFEFDD.....VKVLS.YGDDLLVATNYQL
STHFAQ.YKNVWDVDYSAEDANHCSTAMNIMFEEVTEFG.....FHPNAEWILKTLVNTTEHAYENKRI.....VVEGCMPSGCSATSIINTILNNIYVLYALRRHYEGVELDT.....YTMIS.YGDDIVVASDYDL
.....WSLCVATIVSDHDTFWPGWLRDLICDELINMGYA.PWVVKLFETSLKLPVYVGAFAPEQGHITLLGDPSNPDLEVGLSSGQGATDLMGTILMSITYLVMQLDHTAPHLNSRIKCMPSACRFLDSYWQGHEEIROIS.KSDAILGWTGKR
RRLEPE.NWVYCDAGSQEDSSLTPYLINAV...LTIIRSTYMEDWDVGLQMLRNLYTEIVYTPISTPDGTIV.....KKFRGNNSGQSTVDNSIMVVIAMHYALIKECVEFEEID.....STCVFFV.NGDDLLIAVNPEK
IDKLNRPGLWLGSGDGRDSSIDPFFFDVV..KTRKHFL..PSEHHKAIDLIYDEILNTTICLANGMVI.....KKNVGTQR.QPSTVVDNTLVLMTAFLYAYIHKTGDRELAL.....LNERFIFVC.NGDDNKFAISPQF
VLSIASFSPYGFNGDFANFDGMFHPSSFSMV..SEIANIFY...GNFLSTERDNLRLTNRFSLMKGAIL.....RVPGGPGSGGFEMTVIENSFINLFYIQA WIMLARFNGRQDISH.....PCNFPKYVRACV.YGDDNTVAIKMEV
VARMKEKGNDVLCDDYSEDGLLSKQVMDVI..ASINELC.GGEDQLKNARRNLLMACCSRLAICKNTVW.....RVECGIPSGGFEMTVIENSFINETILIRYHYKKLMREQQAPELMV.....QSFDKLIIGLVT.YGDDNTISVNAVY
AEHAK.YKNHFDADYIANDSTQNRQIMTES..FSIMSRLT...ASPELAEVVAQDLLAPSEMDVGDYVI.....RVKEGLPSGFPCTSQVNSINHWTITLCAISEATGLSPDVV.....QSMSYFSFYGDDEIVSTDIDF
NNLTSKASDFLCDDYSKNDSTMSPCVVRIA..IDLADCC...EQTELTKSVVLTIKSHFMTILEAMIV.....QTKRGLPSGMPPTSVINSTCHWLWSAAVYKSCAEIGLHCS.....NLYELAPFYT.YGDDGVYAMTFMM
IQRIKS.AAKVYAVDYSKNDSTQSPRVSAA..IDLRYFS...DRSPIVDSAANTLKSPPIAIFNGVAV.....KVSSGLPSGMPITSVINSLNHCLYVGCAILQSLEARGVPVTW.....NLFSTFCMMT.YGDDGVYMFMMF
'KRLEKPKHDRVCDYSKNDSTQPPKVTQS..IDLRHET....DKSPIVDSACATLKSNPIGIFNGVAF.....KVAGGLPSGMPITSVINSLNHCLMGSAAVKALED SGVRVTW.....NIFDSMDLFT.YGDDGVYIVPPLI
D      D                                     g  sg  T  n3                                     gDD

```



На каких участках выравнивание правильное – совпадает с эволюционным?

Множественное дает больше информации, чем парное

```

      *      100      *      120      *      140      *      160
THIE_LACLS : AGVSFIVNDDVELARELNADGIHIGQTDSESVSKVREKVGQEMWLGLSVTKADELKTAQ-SSGADYLGIGPIYPTNSKND : 14
THIE_MANSM : FQVFFIVNDDVELALSIQADGIHVGQKDTAVETILRNTRNKPIIIGLSINTLAQALANKDRQIDIYFGVGPIFPTNSKADH : 15
THIE_STRA3 : YQVFFIIDDDIDLVELIDADGLHIGQNDLPVDEARRRLPDKI-IGLSVSTMAEYQKSQ-LSVVDYIIGIGFENPTQSKADA : 14
THIE_LISIN : YQVFFIINDDDVALALEIGADGIHVGQNDDEEIRQVIASCAGKMKIIGLSVHSVSEAEAEERLGSVDYIIGVGPIFPTISKADA : 14
THIE_ANOFW : YNIPFFIVNDDVDLALALQADGVHVGQDEVAERVRDRIGDKY-LGVSVHNLNEVKKAL-AACADYVGLGPIFPTVSKEDA : 14
THIE_GEOTN : YGVFFIVNDDVELAIAIDADGVHVGQDDEARRVREKIGDKI-LGVSAHNVVEBARAAI-EAGADYIGVGPIYPTRSKDDA : 14
THIE_BACSU : AGVFFIVNDDVELALNLKADGIHIGQEDANAKEVRAAIGDMI-LGVSHTMSEVKQAE-EDGADYVGLGPIYPTETKKDT : 14
THIE_BACA2 : AGIPFFIINDDDVELALRLBADGVHIGQDDADAETRAAIGDMI-LGVSAHNVSEVKRAE-AAGADYVGMGPVYPTETKKDA : 14
THIE_OCEIH : FQIPFFIINDDDVLAQLDADGIHIGQDDQPEVVRKQF PNKI-IGLSISTNNELNQSP-LDLVDYIIGVGPIEDTNTKEDA : 14
THIE_STAAB : YNVFFIVNDDVSLAKEINADGIHVGQDDAKVKEIAQYFTDKI-IGLSISDLGEYAKSD-LTHVDYIIGVGPIYPTPSKHDA : 14
THIE_STACT : YNVFFIVNDDVALAEEDADGIHVGQDDEAVDDFNNRFEKGI-IGLSIGNLEEELNASD-LTYVDYIIGVGPIFATPSKDDA : 14
      6pFI61DD6 La 6 ADG6H6GQ D      6G6S      2      DY G6GP pT 3K Da

      *      180      *      200      *      220      *      240
THIE_LACLS : AKHIGIKDLR-LMLLENQLPIVGIGGITQDSLTELISAIGLDGLAVISLLTEAENPKKVAQMIRQKITKNG~~~~~ : 21
THIE_MANSM : SPIVGMNFIRQIRQLGIDKPCVAIGGIKEESAAILRRLGADGVAVISAIHSHSVNIANTVKTLAQK~~~~~ : 22
THIE_STRA3 : KPAVGNRTTKAVREINQDIPITVAIGGITSDFVHDIIESGADGIAVISAIKANHIVDATRQLRYEVEKALVNRQKRSDVI : 22
THIE_LISIN : EPVSGTAILEEIRRAGIKDPIVGIGGINETNSAEVLTAGADGVSVISAITRSEDQCSVIKQLKNPGSPS~~~~~ : 21
THIE_ANOFW : KQACGLTMIEHIRAHEKRVPLVAIGGITEQTAKQVIEAGADGIAVISAI CRAEHIYEQTKRRLYEMVMRAKQKGDR~~~~ : 21
THIE_GEOTN : NEAOGPGILRHLREQGITIPITVAIGGITADNTRAVIEAGADGVSVISAIASAPEPKAAAAALATAVREANL---R~~~~ : 22
THIE_BACSU : RAVQGVSLIEAVRRQGISIPITVGIGGITIDNAAPVIAQAGADGVSMISAISQAEDPESAARKFREEIQTYKTG--R~~~~ : 22
THIE_BACA2 : EAVQGVTLIEEVRQGITIPITVGIGGITADNAAPVIEAGADGVSMISAISQAEDPKAAARKFSEEIRRSKAGLSR~~~~ : 22
THIE_OCEIH : KTAVGLEWISLKKQHPSLPIVAIGGINTTNAQEIIQAGADGVSFISAITETDHLQAVQRL~~~~~ : 20
THIE_STAAB : HTPVGPEMIATFKEMNPQLPIVAIGGINTSNVAPIVEAGANGISVISAIKSENIEKTVNRFKDFFN~~~~~ : 21
THIE_STACT : SEPVGPKMIETLRKEVGDLPIVAIGGISLDNVQEVAKTSADGVSVISAIARS PHVTETVHKFLQYFK~~~~~ : 21
      G      P V IGGI      6 galG6 ISa6

      80      *      100      *      120      *      140      *
THIE_LACLS : VSFIVNDDVELARELNADGIHIGQTDSESVSKVREKVGQEMWLGLSV-TKADELKTAQSSGADYLGIGPIYPTNSKND :
THIE_MANSM : VFFIVNDDVELALSIQADGIHVGQKDTAVETILRNTRNKPIIIGLSINTLAQALANKDRQIDIYFGVGPIFPTNSKAD :
      V FIVNDDVELA 6 ADGIH6GQ D V 6      6GLS6 T A L      DY G6GPI5PTNSK D

      160      *      180      *      200      *      220
THIE_LACLS : AAKHIG---TKDLRLMLLENQLPIVGIGGITQDSLTELISAIGLDGLAVISLLTEAENPKKVAQMIRQKITKNG : 218
THIE_MANSM : HSPIVGMNFIRQIRQLGIDK--PCVAIGGIKEESAAILRRLGADGVAVISAIHSHSVNIANTVKTLAQK----- : 220
      6G T4 6R 6 6 P V IGGI 2 S L 6G DG6AVIS 63 N      6 QK

```

- ❖ Множественные выравнивания – глобальные
- ❖ Для всякого ли множества последовательностей имеет смысл строить такое выравнивание?

JalView – редактор выравниваний

Полезные умения в Jalview

Действие	меню	подменю	варианты	комментарии
Импорт последовательно стей	File	Add sequences	- Из файла - From textbox - url	
Команды относятся к выделенному	Ctrl-A			
Выравнивание	Web services	Alignment	Выбираете программу	Mafft быстро работает Результат в новом окне
Раскраска	Color	Clustal и By conservation	Conservation threshold	Меняя порог, можно увидеть консервативные участки
Создание групп по выделенным колонкам	Select Calculate	Make groups for selection sort	by groups	Удобно для поиска гомологичных последовательностей Меньше колонок – больше группы
Поиск	select	find	Пишете посл. Или паттерн	Пример [FY].[GA].{1,2}[GA] "." – любая буква {от,до} раз [FY] – и F, и Y годятся

Действие	меню	подменю	варианты	комментарии
Выделение прямоугольного блока в отдельное окно	Мышкой выделяете прямоугольный блок	Правой кнопкой selection	Output to text block	Сразу new windows Текстовое окно можно закрыть
	Edit	Remove empty columns		После создания окна из выделенного блока
	Edit	Remove all gaps		Для перевыравнивания
Сократить выравнивание за счёт удаления почти идентичных последовательностей	Edit	Remove redundancy	Поставить порог сходства	Из высокосходных последовательностей оставить одну
Сохранение выравнивания				
Сохранение проекта со всеми окнами				

**Научимся искать в выравнивании блоки, в
которых выравнивание имеет высокий шанс быть
эволюционным**

В JalView загружаем выравнивание

File => Fetch Sequence => Выбираем PFAM в Select Database
=> **PF00284**

Определения

- ❖ **Блок** определяется подмножеством последовательностей и участком от позиции S (start) до позиции E (end) выравнивания.
- ❖ Блок задаётся:
 - ✓ перечислением последовательностей;
 - ✓ участком позиции от S-той до E-той подряд.

PF00284

Luminal portion of Cytochrome b559, alpha (gene psbE) subunit

	10	20	30
K9Z0V0_CYAAP/42-80	L A Y D V F G T P R P D E Y F T Q T R Q E L P I I N D R Y E A K N Q I E Q F N		
K9SST0_9SYNE/42-80	L A Y D I F G T P R P N E Y Y T N D R Q E A P I L R D R Y D A N Q E I P N Y N		
K9SGD2_9CYAN/42-80	L A Y D V F G T P R P N E Y Y T L D R S D A P V L L D R F N A I D D I E D F N		
K9Y0S0_STAC7/42-80	L A Y D V F G T P R P N E Y F T Q D R L E L P I I N D R Y S A S E Q I K E F N		
PSBE_RIPO1/42-80	L A Y D V F G T P R P D Q Y F T Q E R Q E L P I I S D R Y K A S Q Q I Q E F N		
PSBE_SYNY3/42-80	L A Y D A F G T P R P D E Y F T Q T R Q E L P I L Q E R Y D I N Q E I Q E F N		
PSBE_GLOC7/42-80	L A Y D V F G T P R P D E Y Y T Q E R L E L P I L K D R Y N T D Q Q I Q E F N		
K9W859_9CYAN/42-80	L A Y D V F G T P R P N E Y F T Q D R V E A P I L S D R Y E A K Q Q I Q Q F L		
PSBE_SYNE7/43-81	L A Y D A F G T P R P N E Y F T Q D R T E V P I V S D R Y S A K Q Q V D R F S		
PSBE_ACAM1/42-80	L A Y D V F G T P R P N E Y F D Q A R Q G L P L V T D R Y E G K Q Q I D E F T		
C7B2G7_SELML/42-80	L A Y D V S G S P R P N E Y S T E S R R D I P L I T G R F D S S E Q I D E F T		
PSBE_PHATG/43-81	L A Y D V F G T P R P N E Y F T Q D R Q Q I P L V N D R F S A K Q E L E D L T		
PSBE_OSTTA/42-79	L A Y D V F G T P R P N E Y F T E E R Q E L P L I S D R Y G A L A Q L D D L -		
PSBE_CHLRE/42-79	L A Y D V F G T P R P N E Y F T E D R Q E A P L I T D R F N A L E Q V K K L -		
PSBE_SYNJB/42-80	L A Y D I F G T P R P N E Y F T A E R Q E L P I V S D R F N A L E E L E R L T		
K9TLB2_9CYAN/43-80	L A Y D A F G T P R P N E Y F S Q E R L E L P I V Q D R F N A K D Q I D D L -		
K9S2L1_9CYAN/42-80	L A Y D A F G T P R P D E Y F T Q Q R Q E L P I V M D R F G A K Q Q V E E F I		
K9YE64_HALP7/42-80	L A Y D A F G T P R P D E Y F T Q Q R Q E L P I I S D R F Q A R E E V E E F N		
K9R7L0_9CYAN/42-80	L A Y D A F G T P R P D T Y F T Q E R Q E V P L V N D R F E A K N Q V D V F T		
PSBE_THEVB/42-80	L A Y D V F G T P R P D S Y Y A Q E Q R S I P L V T D R F E A K Q Q V E T F L		
PSBE_GLOW/43-81	L A Y D V F G T P R P D E Y Y N S D N T K K P L V N K R F E A K Q Q L D E A T		
PSBE_SYNR3/43-81	L A Y D A F G T P R P D S Y F Q A A D V K A P V V S Q R Y E G K S Q L D T R L		
PSBE_PROMA/43-81	L A Y D A F G T P R P D T Y F Q A S E T K A P V V S Q R F E S K A Q L D L R L		

Последовательности 1-7
Позиции 1-13

- ❖ **Блок без гэпов** это блок, в выравнивании которого нет **НИ ОДНОГО** гэта

Блок с гэпами

Блок без гэпов

Определения

- ❖ **Функционально консервативная колонка** – все а.к.о. в колонке принадлежат одной группе по свойствам.

	11		30														
K9Z0V0_CYAAP/42-80	P	I	I	N	D	R	Y	E	A	K	N	Q	I	E	Q	F	N
K9SST0_9SYNE/42-80	P	I	L	R	D	R	Y	D	A	N	Q	E	I	P	N	Y	N
K9SGD2_9CYAN/42-80	P	V	L	L	D	R	F	N	A	I	D	D	I	E	D	F	N
K9Y0S0_STAC7/42-80	P	I	I	N	D	R	Y	S	A	S	E	Q	I	K	E	F	N
PSBE_RIPO1/42-80	P	I	I	S	D	R	Y	K	A	S	Q	Q	I	Q	E	F	N
PSBE_SYNY3/42-80	P	I	L	Q	E	R	Y	D	I	N	Q	E	I	Q	E	F	N
PSBE_GLOC7/42-80	P	I	L	K	D	R	Y	N	T	D	Q	Q	I	Q	E	F	N
PSBE_OSTTA/42-79	P	L	I	S	D	R	Y	G	A	L	A	Q	L	D	D	L	-
K9W859_9CYAN/42-80	P	I	L	S	D	R	Y	E	A	K	Q	Q	I	Q	Q	F	L
PSBE_SYNE7/43-81	P	I	V	S	D	R	Y	S	A	K	Q	Q	V	D	R	F	S
PSBE_ACAM1/42-80	P	L	V	T	D	R	Y	E	G	K	Q	Q	I	D	E	F	T
C7B2G7_SELML/42-80	P	L	I	T	G	R	F	D	S	S	E	Q	I	D	E	F	T
PSBE_PHATG/43-81	P	L	V	N	D	R	F	S	A	K	Q	E	L	D	L	T	
PSBE_CHLRE/42-79	P	L	I	T	D	R	F	N	A	L	E	Q	V	K	K	L	-
PSBE_SYNJB/42-80	P	I	V	S	D	R	F	N	A	L	E	E	L	E	R	L	T
K9TLB2_9CYAN/43-80	P	I	V	Q	D	R	F	N	A	K	D	Q	I	D	D	L	-
K9S2L1_9CYAN/42-80	P	I	V	M	D	R	F	G	A	K	Q	Q	V	E	E	F	I
K9YE64_HALP7/42-80	P	I	I	S	D	R	F	Q	A	R	E	E	V	E	E	F	N
K9R7L0_9CYAN/42-80	P	L	V	N	D	R	F	E	A	K	N	Q	V	D	V	F	T
PSBE_THEVB/42-80	P	L	V	T	D	R	F	E	A	K	Q	Q	V	E	T	F	L
PSBE_GLOWW/43-81	P	L	V	N	K	R	F	E	A	K	Q	Q	L	D	E	A	T
PSBE_SYNR3/43-81	P	V	V	S	Q	R	Y	E	G	K	S	Q	L	D	T	R	L
PSBE_PROMA/43-81	P	V	V	S	Q	R	F	E	S	K	A	Q	L	D	L	R	L
K9RTX4_SYNP3/42-80	P	V	V	S	D	R	F	E	S	K	Q	Q	I	N	E	F	L
K9ULI3_CHAP6/43-81	P	I	V	S	D	R	F	E	S	K	G	Q	L	D	K	F	V
K9UY34_9CYAN/42-79	P	I	V	N	N	R	F	D	A	K	K	Q	V	E	T	Y	-
PSBE_NOSS1/42-80	P	I	V	N	N	R	F	E	A	K	K	Q	V	E	Q	L	I
K9QJT8_9NOSQ/42-80	P	I	V	K	N	R	Y	E	A	K	Q	Q	V	E	K	F	I
K8GLK0_9CYAN/42-80	P	I	V	K	D	R	F	E	A	K	Q	Q	I	D	Q	F	I
K9VJB4_9CYAN/42-80	P	I	I	S	D	R	F	E	A	K	K	Q	V	D	Q	F	I
K9W2B5_9CYAN/42-80	P	I	V	S	D	R	F	E	S	K	Q	E	I	D	Q	F	L
K9TZN4_CHRTP/42-80	P	I	V	S	D	R	F	E	S	R	K	Q	I	E	E	F	I
K9XG15_9CHRO/42-80	P	I	I	G	D	R	Y	Q	A	K	Q	E	I	K	Q	F	I
	↑↑↑	↑↑		↑													

Определения

- ❖ **Блок достоверный**, т.е. можно считать, что выравнивание в нем соответствует эволюционному, если
 1. он без гэпов
 2. первая позиция блока консервативна или функционально консервативна в блоке (а не во всем выравнивании)
 3. то же условие на последнюю позицию блока
- ❖ Чем больше консервативных позиций в блоке, и чем чаще они идут в блоке, тем достоверность блока выше.

	10	20	30
K9Z0V0_CYAAP/42-80	LAYDVFGT	PRPDEYFTQT	RQELPIINDR
K9SST0_9SYNE/42-80	LAYDIFGT	PRPNEYYTND	RQEAPILRDR
K9SGD2_9CYAN/42-80	LAYDVFGT	PRPNEYYTLD	RSDAPVLLDR
K9Y0S0_STAC7/42-80	LAYDVFGT	PRPNEYFTQD	RLELPIINDR
PSBE_RIPO1/42-80	LAYDVFGT	PRPDQYFTQER	RQELPIISDR
PSBE_SYNY3/42-80	LAYDAFGT	PRPDEYFTQT	RQELPILQER
PSBE_GLOC7/42-80	LAYDVFGT	PRPDEYYTQER	RLELPIILKDR
K9W859_9CYAN/42-80	LAYDVFGT	PRPNEYFTQD	RVEAPILSDR
PSBE_SYNE7/43-81	LAYDAFGT	PRPNEYFTQD	RTEVPIVSDR
PSBE_ACAM1/42-80	LAYDVFGT	PRPNEYFDQA	RQGLPLVTD
C7B2G7_SELML/42-80	LAYDVSGS	PRPNEYSTES	RRDIPLITGR
PSBE_PHATC/43-81	LAYDVFGT	PRPNEYFTQD	RQQIPLVND
PSBE_OSTTA/42-79	LAYDVFGT	PRPNEYFTEER	RQELPLISDR
PSBE_CHLRE/42-79	LAYDVFGT	PRPNEYFTED	RQEAPLITDR
PSBE_SYNJB/42-80	LAYDIFGT	PRPNEYFTAER	RQELPIVSDR
K9TLB2_9CYAN/43-80	LAYDAFGT	PRPNEYFSQER	RLELPIVQDR
K9S2L1_9CYAN/42-80	LAYDAFGT	PRPDEYFTQQ	RQELPIVMDR
K9YE64_HALP7/42-80	LAYDAFGT	PRPDEYFTKQ	RQELPIISDR
K9R7L0_9CYAN/42-80	LAYDAFGT	PRPDYFTQER	RQEVPLVND
PSBE_THEVB/42-80	LAYDVFGT	PRPDSYQAQ	EQRSIPLVTD
PSBE_GLOWV/43-81	LAYDVFGT	PRPDEYYNSD	NTKKPLVNR
PSBE_SYNR3/43-81	LAYDAFGT	PRPDSYFQA	ADVKA
PSBE_PROMA/43-81	LAYDAFGT	PRPDYFQAS	ETKAPVVSQ
K9RTX4_SYNP3/42-80	LAYDIFGT	PRPDYFSQEQ	QAAPVVS
K9ULI3_CHAP6/43-81	LAYDAFGT	PRPDEYFTQT	RQELPIVSD
K9UY34_9CYAN/42-79	LAYDAFGT	PRPNEYFTQT	RQEVPIVNN
PSBE_NOSS1/42-80	LAYDVFGT	PRPDEYYTQA	RQELPIVNN
K9QJTB_9NOSO/42-80	LAYDVFGT	PRPNEYYTQT	RQEVPIVKN
K8GLK0_9CYAN/42-80	LAYDVFGT	PRPNEYYTST	RQQAPIVKD
K9VJB4_9CYAN/42-80	LAYDAFGT	PRPNQYFTPQ	REEVPIISD
K9W2B5_9CYAN/42-80	LAYDAFGT	PRPNEYFTQQ	RQEVPIVSD
K9TZN4_CHRTP/42-80	LAYDVFGT	PRPNEYYPQER	RQELPIVSD
K9XG15_9CHRO/42-80	LAYDVFGT	PRPNEYYPQER	RQELPIIGD

Пример достоверного блока

		10	20	30
K9Z0V0_CYAAP/42-80	LAYDVFGT	PRPDEYFTQT	RQELPI	INDRYEAKNQIEQFN
K9SST0_9SYNE/42-80	LAYDIFGT	PRPNEYYTND	RQEAPI	LRDRYDANQEIPNYN
K9SGD2_9CYAN/42-80	LAYDVFGT	PRPNEYYTLD	RSDAPV	LDLRFNAIDDIEDFN
K9Y0S0_STAC7/42-80	LAYDVFGT	PRPNEYFTQD	RLELPI	INDRYSASEQIKEFN
PSBE_RIPO1/42-80	LAYDVFGT	PRPDQYFTQER	RQELPI	ISDRYKASQQIQEFN
PSBE_SYNY3/42-80	LAYDAFGT	PRPDEYFTQT	RQELPI	LQERYDINQEQIEFN
PSBE_GLOC7/42-80	LAYDVFGT	PRPDEYFTQER	RLELPI	KDRYNTDQQIQEFN
K9W859_9CYAN/42-80	LAYDVFGT	PRPNEYFTQD	RVEAPI	ISDRYEAKQQIQQFL
PSBE_SYNE7/43-81	LAYDAFGT	PRPNEYFTQD	RTEVP	ISDRYSAKQQVDRFS
PSBE_ACAM1/42-80	LAYDVFGT	PRPNEYFDQA	RQGLPL	VTDRYEGKQQIDEFT
C7B2G7_SELML/42-80	LAYDVSGS	PRPNEYSTES	RRDIPL	ITGRFDSSEQIDEFT
PSBE_PHATC/43-81	LAYDVFGT	PRPNEYFTQD	RQQIPL	VNDRFSAKQELEDLT
PSBE_OSTTA/42-79	LAYDVFGT	PRPNEYFTEER	RQELPL	ISDRYGALAQLDDL
PSBE_CHLRE/42-79	LAYDVFGT	PRPNEYFTED	RQEAPL	ITDRFNALEQVKKL
PSBE_SYNJB/42-80	LAYDIFGT	PRPNEYFTAER	RQELPI	ISDRFNALEELERLT
K9TLB2_9CYAN/43-80	LAYDAFGT	PRPNEYFSQER	RLELPI	VQDRFNAKDQIDDL
K9S2L1_9CYAN/42-80	LAYDAFGT	PRPDEYFTQQ	RQELPI	VMDRFQAKQQVEEF
K9YE64_HALP7/42-80	LAYDAFGT	PRPDEYFTKQ	RQELPI	ISDRFQAREEVEEFN
K9R7L0_9CYAN/42-80	LAYDAFGT	PRPDYFTQER	RQEVPL	VNDRFEAKNQVDVFT
PSBE_THEVB/42-80	LAYDVFGT	PRPDSYQAQEQ	RSIPL	VTDREFEAKQQVETFL
PSBE_GLOWV/43-81	LAYDVFGT	PRPDEYNSDNT	KKPL	VNKRFEAKQQQLDEAT
PSBE_SYNR3/43-81	LAYDAFGT	PRPDSYFQAADV	KAPV	VVSQRFEYEGKSQQLDTRL
PSBE_PROMA/43-81	LAYDAFGT	PRPDYFQASET	KAPV	VVSQRFESKAQLDLRL
K9RTX4_SYNP3/42-80	LAYDIFGT	PRPDYFSQEQQA	APV	VSDRFESKQQINEFL
K9ULI3_CHAP6/43-81	LAYDAFGT	PRPDEYFTQTR	RQELPI	VSDRFESKGQLDKFV
K9UY34_9CYAN/42-79	LAYDAFGT	PRPNEYFTQTR	RQEVPI	VNNRFDKAKQVETY
PSBE_NOSS1/42-80	LAYDVFGT	PRPDEYYTQAR	RQELPI	VNNRFEAKKQVEQLI
K9QJTB_9NOSQ/42-80	LAYDVFGT	PRPNEYYTQTR	RQEVPI	VKNRYEAKQQVEKFI
K8GLK0_9CYAN/42-80	LAYDVFGT	PRPNEYYTSTR	RQAPV	IKDRFEAKQQIDQFI
K9VJB4_9CYAN/42-80	LAYDAFGT	PRPNQYFTPQ	REEVPI	ISDRFEAKKQVDQFI
K9W2B5_9CYAN/42-80	LAYDAFGT	PRPNEYFTQQR	RQEVPI	VSDRFESKQEIQQFL
K9TZN4_CHRTP/42-80	LAYDVFGT	PRPNEYYPQER	RQELPI	VSDRFESRKQIEEFI
K9XG15_9CHRO/42-80	LAYDVFGT	PRPNEYYPQER	RQELPI	IGDRYQAKQEIQQFI

Пример достоверного блока

Определения

- ❖ Достоверный блок **максимален** если
 - ✓ нельзя добавить последовательность к блоку так, чтобы не уменьшить его достоверность
 - ✓ нельзя расширить его с N или C конца так, чтобы не уменьшить его достоверность

		10	20	30	
K9Z0V0_CYAAP/42-80	LAYD	VFGT	PRPD	EYFTQT	RQELPIINDRYEAKNQIEQFN
K9SST0_9SYNE/42-80	LAYD	IFGT	PRPNE	YNTNDR	QEAPILRDRYDANQEIPNYN
K9SGD2_9CYAN/42-80	LAYD	VFGT	PRPNE	YTLDR	SDAFVLLDRFNAIDDIEDFN
K9Y0S0_STAC7/42-80	LAYD	VFGT	PRPNE	YFTQDR	LELPINDRYSASEQIKEFN
PSBE_RIPO1/42-80	LAYD	VFGT	PRPD	QYFTQER	QELPIISDRYKASQQIQEFN
PSBE_SYNY3/42-80	LAYD	AFGT	PRPD	EYFTQT	RQELPILQERYDINQEIQEFN
PSBE_GLOC7/42-80	LAYD	VFGT	PRPDE	YNTQER	LELPILKDRYNTDQQIQEFN
K9W859_9CYAN/42-80	LAYD	VFGT	PRPNE	YFTQDR	VEAPILSDRYEAKQQIQQFL
PSBE_SYNE7/43-81	LAYD	AFGT	PRPNE	YFTQDR	TEVPIVSDRYSAKQQVDRFS
PSBE_ACAM1/42-80	LAYD	VFGT	PRPNE	YFDQAR	QGLPLVTDRYEGKQQIDEFT
C7B2G7_SELML/42-80	LAYD	VSGS	PRPNE	YSTES	RRDIPLITGRFDSSEQIDEFT
PSBE_PHATC/43-81	LAYD	VFGT	PRPNE	YFTQDR	QQIPLVNDRFSAKQLEDLT
PSBE_OSTTA/42-79	LAYD	VFGT	PRPNE	YFTEER	QELPLISDRYGALAQLDDL
PSBE_CHLRE/42-79	LAYD	VFGT	PRPNE	YFTEDR	QEAPLITDRFNALEQVKKL
PSBE_SYNJB/42-80	LAYD	IFGT	PRPNE	YFTAER	QELPIVSDRFNALEELERLT
K9TLB2_9CYAN/43-80	LAYD	AFGT	PRPNE	YFSQER	LELPIVQDRFNAKDQIDDL
K9S2L1_9CYAN/42-80	LAYD	AFGT	PRPDE	YFTQQR	RQELPIVMDRFSAKQQVEEF
K9YE64_HALP7/42-80	LAYD	AFGT	PRPDE	YFTKQR	RQELPIISDRFQAREEVEEFN
K9R7L0_9CYAN/42-80	LAYD	AFGT	PRPD	TYFTQER	QEVPLVNDRFEAKNQVDVFT
PSBE_THEVB/42-80	LAYD	VFGT	PRPDS	YQAQEQ	RSIPLVTDRF EAKQQVETFL
PSBE_GLOWV/43-81	LAYD	VFGT	PRPDE	YNSDNT	KKPLVNKRF EAKQQQLDEAT
PSBE_SYNR3/43-81	LAYD	AFGT	PRPDS	YFQAADV	KAPVVSQRYEGKSQQLDTRL
PSBE_PROMA/43-81	LAYD	AFGT	PRPD	TYFQASET	KAPVVSQRFESKAQLDLRL
K9RTX4_SYNP3/42-80	LAYD	IFGT	PRPD	TYFSQEQ	QAAPVVSDRFESKQQINEFL
K9ULI3_CHAP6/43-81	LAYD	AFGT	PRPDE	YFTQTR	RQELPIVSDRFESKGQLDKFV
K9UY34_9CYAN/42-79	LAYD	AFGT	PRPNE	YFTQTR	QEVPIVNRRFDAKKQVETY
PSBE_NOSS1/42-80	LAYD	VFGT	PRPDE	YNTQAR	QELPIVNRRFEAKKQVEQLI
K9QJTB_9NOSQ/42-80	LAYD	VFGT	PRPNE	YNTQTR	QEVPIVKNRYEAKQQVEKFI
K8GLK0_9CYAN/42-80	LAYD	VFGT	PRPNE	YNTSTR	QQAPIVKDRFEAKQQIDQFI
K9VJB4_9CYAN/42-80	LAYD	AFGT	PRPN	QYFTPQ	REEVPIISDRFEAKKQVDQFI
K9W2B5_9CYAN/42-80	LAYD	AFGT	PRPNE	YFTQQR	RQEVPIVSDRFESKQIDQFL
K9TZN4_CHRTP/42-80	LAYD	VFGT	PRPNE	YPQER	QELPIVSDRFESRKQIEEFI
K9XG15_9CHRO/42-80	LAYD	VFGT	PRPNE	YPQER	QELPIIGDRYQAKQEIKQFI

Это максимальный достоверный блок?

	10										20										30																	
K9Z0V0_CYAAP/42-80	L	A	D	V	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	I	N	D	R	Y	E	A	K	N	Q	I	E	Q	F	N
K9SST0_9SYNE/42-80	L	A	D	I	F	G	T	P	R	P	N	E	Y	T	N	D	R	Q	E	A	P	I	L	R	D	R	Y	D	A	N	Q	E	I	P	N	Y	N	
K9SGD2_9CYAN/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	T	L	D	R	S	D	A	P	V	L	L	D	R	F	N	A	I	D	D	I	E	D	F	N	
K9Y0S0_STAC7/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	L	E	L	P	I	I	N	D	R	Y	S	A	S	E	Q	I	K	E	F	N
PSBE_RIPO1/42-80	L	A	D	V	F	G	T	P	R	P	D	Q	Y	F	T	Q	E	R	Q	E	L	P	I	I	S	D	R	Y	K	A	S	Q	Q	I	Q	E	F	N
PSBE_SYNY3/42-80	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	L	Q	E	R	Y	D	I	N	Q	E	I	Q	E	F	N
PSBE_GLOC7/42-80	L	A	D	V	F	G	T	P	R	P	D	E	Y	T	Q	E	R	L	E	L	P	I	L	K	D	R	Y	N	T	D	Q	Q	I	Q	E	F	N	
K9W859_9CYAN/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	V	E	A	P	I	L	S	D	R	Y	E	A	K	Q	Q	I	Q	Q	F	L
PSBE_SYNE7/43-81	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	T	E	V	P	I	V	S	D	R	Y	S	A	K	Q	Q	V	D	R	F	S
PSBE_ACAM1/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	D	Q	A	R	Q	G	L	P	L	V	T	D	R	Y	E	G	K	Q	Q	I	D	E	F	T
C7B2G7_SELML/42-80	L	A	D	V	S	G	S	P	R	P	N	E	Y	S	T	E	S	R	R	D	I	P	L	I	T	G	R	F	D	S	S	E	Q	I	D	E	F	T
PSBE_PHATC/43-81	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	Q	Q	I	P	L	V	N	D	R	F	S	A	K	Q	E	L	E	D	L	T
PSBE_OSTTA/42-79	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	E	E	R	Q	E	L	P	L	I	S	D	R	Y	G	A	L	A	Q	L	D	D	L	-
PSBE_CHLRE/42-79	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	E	D	R	Q	E	A	P	L	I	T	D	R	F	N	A	L	E	Q	V	K	K	L	-
PSBE_SYNJB/42-80	L	A	D	I	F	G	T	P	R	P	N	E	Y	F	T	A	E	R	Q	E	L	P	I	V	S	D	R	F	N	A	L	E	E	L	E	R	L	T
K9TLB2_9CYAN/43-80	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	S	Q	E	R	L	E	L	P	I	V	Q	D	R	F	N	A	K	D	Q	I	D	D	L	-
K9S2L1_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	Q	R	Q	E	L	P	I	V	M	D	R	F	G	A	K	Q	Q	V	E	E	F	I
K9YE64_HALP7/42-80	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	K	Q	R	Q	E	L	P	I	I	S	D	R	F	Q	A	R	E	E	V	E	E	F	N
K9R7L0_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	D	T	Y	F	T	Q	E	R	Q	E	V	P	L	V	N	D	R	F	E	A	K	N	Q	V	D	V	F	T
PSBE_THEVB/42-80	L	A	D	V	F	G	T	P	R	P	D	S	Y	A	Q	E	Q	R	S	I	P	L	V	T	D	R	F	E	A	K	Q	Q	V	E	T	F	L	
PSBE_GLOW/43-81	L	A	D	V	F	G	T	P	R	P	D	E	Y	N	S	D	N	T	K	K	P	L	V	N	K	R	F	E	A	K	Q	Q	L	D	E	A	T	
PSBE_SYNR3/43-81	L	A	D	A	F	G	T	P	R	P	D	S	Y	F	Q	A	A	D	V	K	A	P	V	V	S	Q	R	Y	E	G	K	S	Q	L	D	T	R	L
PSBE_PROMA/43-81	L	A	D	A	F	G	T	P	R	P	D	T	Y	F	Q	A	S	E	T	K	A	P	V	V	S	Q	R	F	E	S	K	A	Q	L	D	L	R	L
K9RTX4_SYNP3/42-80	L	A	D	I	F	G	T	P	R	P	D	T	Y	F	S	Q	E	Q	Q	A	A	P	V	V	S	D	R	F	E	S	K	Q	Q	I	N	E	F	L
K9ULI3_CHAP6/43-81	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	V	S	D	R	F	E	S	K	G	Q	L	D	K	F	V
K9UY34_9CYAN/42-79	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	T	R	Q	E	V	P	I	V	N	N	R	F	D	A	K	K	Q	V	E	T	Y	-
PSBE_NOSS1/42-80	L	A	D	V	F	G	T	P	R	P	D	E	Y	T	Q	A	R	Q	E	L	P	I	V	N	N	R	F	E	A	K	K	Q	V	E	Q	L	I	
K9QJT8_9NOSO/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	T	Q	T	R	Q	E	V	P	I	V	K	N	R	Y	E	A	K	Q	Q	V	E	K	F	I	
K8GLK0_9CYAN/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	T	S	T	R	Q	Q	A	P	I	V	K	D	R	F	E	A	K	Q	Q	I	D	Q	F	I	
K9VJB4_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	N	Q	Y	F	T	P	Q	R	E	E	V	P	I	I	S	D	R	F	E	A	K	K	Q	V	D	Q	F	I
K9W2B5_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	Q	R	Q	E	V	P	I	V	S	D	R	F	E	S	K	Q	E	I	D	Q	F	L
K9TZN4_CHRTP/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	P	Q	E	R	Q	E	L	P	I	V	S	D	R	F	E	S	R	K	Q	I	E	E	F	I	
K9XG15_9CHRO/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	P	Q	E	R	Q	E	L	P	I	I	G	D	R	Y	Q	A	K	Q	E	I	K	Q	F	I	

Или это?

		10	20	30																																		
K9Z0V0_CYAAP/42-80	L	A	D	V	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	I	N	D	R	Y	E	A	K	N	Q	I	E	Q	F	N
K9SST0_9SYNE/42-80	L	A	D	I	F	G	T	P	R	P	N	E	Y	T	N	D	R	Q	E	A	P	I	L	R	D	R	Y	D	A	N	Q	E	I	P	N	Y	N	
K9SGD2_9CYAN/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	T	L	D	R	S	D	A	P	V	L	L	D	R	F	N	A	I	D	D	I	E	D	F	N	
K9Y0S0_STAC7/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	L	E	L	P	I	I	N	D	R	Y	S	A	S	E	Q	I	K	E	F	N
PSBE_RIPO1/42-80	L	A	D	V	F	G	T	P	R	P	D	Q	Y	F	T	Q	E	R	Q	E	L	P	I	I	S	D	R	Y	K	A	S	Q	Q	I	Q	E	F	N
PSBE_SYNY3/42-80	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	L	Q	E	R	Y	D	I	N	Q	E	I	Q	E	F	N
PSBE_GLOC7/42-80	L	A	D	V	F	G	T	P	R	P	D	E	Y	T	Q	E	R	L	E	L	P	I	L	K	D	R	Y	N	T	D	Q	Q	I	Q	E	F	N	
K9W859_9CYAN/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	V	E	A	P	I	L	S	D	R	Y	E	A	K	Q	Q	I	Q	Q	F	L
PSBE_SYNE7/43-81	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	T	E	V	P	I	V	S	D	R	Y	S	A	K	Q	Q	V	D	R	F	S
PSBE_ACAM1/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	D	Q	A	R	Q	G	L	P	L	V	T	D	R	Y	E	G	K	Q	Q	I	D	E	F	T
C7B2G7_SELML/42-80	L	A	D	V	S	G	S	P	R	P	N	E	Y	S	T	E	S	R	R	D	I	P	L	I	T	G	R	F	D	S	S	E	Q	I	D	E	F	T
PSBE_PHATC/43-81	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	Q	Q	I	P	L	V	N	D	R	F	S	A	K	Q	E	L	E	D	L	T
PSBE_OSTTA/42-79	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	E	E	R	Q	E	L	P	L	I	S	D	R	Y	G	A	L	A	Q	L	D	D	L	-
PSBE_CHLRE/42-79	L	A	D	V	F	G	T	P	R	P	N	E	Y	F	T	E	D	R	Q	E	A	P	L	I	T	D	R	F	N	A	L	E	Q	V	K	K	L	-
PSBE_SYNJB/42-80	L	A	D	I	F	G	T	P	R	P	N	E	Y	F	T	A	E	R	Q	E	L	P	I	V	S	D	R	F	N	A	L	E	E	L	E	R	L	T
K9TLB2_9CYAN/43-80	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	S	Q	E	R	L	E	L	P	I	V	Q	D	R	F	N	A	K	D	Q	I	D	D	L	-
K9S2L1_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	Q	R	Q	E	L	P	I	V	M	D	R	F	G	A	K	Q	Q	V	E	E	F	I
K9YE64_HALP7/42-80	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	K	Q	R	Q	E	L	P	I	I	S	D	R	F	Q	A	R	E	E	V	E	E	F	N
K9R7L0_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	D	T	Y	F	T	Q	E	R	Q	E	V	P	L	V	N	D	R	F	E	A	K	N	Q	V	D	V	F	T
PSBE_THEVB/42-80	L	A	D	V	F	G	T	P	R	P	D	S	Y	A	Q	E	Q	R	S	I	P	L	V	T	D	R	F	E	A	K	Q	Q	V	E	T	F	L	
PSBE_GLOWV43-81	L	A	D	V	F	G	T	P	R	P	D	E	Y	N	S	D	N	T	K	K	P	L	V	N	K	R	F	E	A	K	Q	Q	L	D	E	A	T	
PSBE_SYNR3/43-81	L	A	D	A	F	G	T	P	R	P	D	S	Y	F	Q	A	A	D	V	K	A	P	V	V	S	Q	R	Y	E	G	K	S	Q	L	D	T	R	L
PSBE_PROMA/43-81	L	A	D	A	F	G	T	P	R	P	D	T	Y	F	Q	A	S	E	T	K	A	P	V	V	S	Q	R	F	E	S	K	A	Q	L	D	L	R	L
K9RTX4_SYNP3/42-80	L	A	D	I	F	G	T	P	R	P	D	T	Y	F	S	Q	E	Q	Q	A	A	P	V	V	S	D	R	F	E	S	K	Q	Q	I	N	E	F	L
K9ULI3_CHAP6/43-81	L	A	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	V	S	D	R	F	E	S	K	G	Q	L	D	K	F	V
K9UY34_9CYAN/42-79	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	T	R	Q	E	V	P	I	V	N	N	R	F	D	A	K	K	Q	V	E	T	Y	-
PSBE_NOSS1/42-80	L	A	D	V	F	G	T	P	R	P	D	E	Y	T	Q	A	R	Q	E	L	P	I	V	N	N	R	F	E	A	K	K	Q	V	E	Q	L	I	
K9QJT8_9NOSQ/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	T	Q	T	R	Q	E	V	P	I	V	K	N	R	Y	E	A	K	Q	Q	V	E	K	F	I	
K8GLK0_9CYAN/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	T	S	T	R	Q	Q	A	P	I	V	K	D	R	F	E	A	K	Q	Q	I	D	Q	F	I	
K9VJB4_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	N	Q	Y	F	T	P	Q	R	E	E	V	P	I	I	S	D	R	F	E	A	K	K	Q	V	D	Q	F	I
K9W2B5_9CYAN/42-80	L	A	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	Q	R	Q	E	V	P	I	V	S	D	R	F	E	S	K	Q	E	I	D	Q	F	L
K9TZN4_CHRTP/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	P	Q	E	R	Q	E	L	P	I	V	S	D	R	F	E	S	R	K	Q	I	E	E	F	I	
K9XG15_9CHRO/42-80	L	A	D	V	F	G	T	P	R	P	N	E	Y	P	Q	E	R	Q	E	L	P	I	I	G	D	R	Y	Q	A	K	Q	E	I	K	Q	F	I	

Функционально консервативная колонка
Раскраска 100% консервативности, поэтому
она не покрашена

А, может, это максимально достоверный блок?

	10	20	30	
K9Z0V0_CYAAP/42-80	LAYDVFGT	PRPDEYFTQT	RQELPIINDRYEAKNQI	EQFN
K9SST0_9SYNE/42-80	LAYDIFGT	PRPNEYYTND	RQEAPILRDRYDANQE	PNYN
K9SGD2_9CYAN/42-80	LAYDVFGT	PRPNEYYTLD	RSDAPVLLDRFNAIDD	EDFN
K9Y0S0_STAC7/42-80	LAYDVFGT	PRPNEYFTQD	RLELPIINDRYSASEQ	KEFN
PSBE_RIPO1/42-80	LAYDVFGT	PRPDQYFTQE	RQELPIISDRYKASQQ	IEFN
PSBE_SYNY3/42-80	LAYDAFGT	PRPDEYFTQT	RQELPILQERYDINQE	IEFN
PSBE_GLOC7/42-80	LAYDVFGT	PRPDEYFTQE	RLELPIILKDRYNTDQQ	IEFN
K9W859_9CYAN/42-80	LAYDVFGT	PRPNEYFTQD	RVEAPILSDRYEAKQQ	IQFL
PSBE_SYNE7/43-81	LAYDAFGT	PRPNEYFTQD	RTEVPIVSDRYSAKQQV	DRFS
PSBE_ACAM1/42-80	LAYDVFGT	PRPNEYFDQA	RQGLPLVTDREYEGKQQ	IEFT
G7B2G7_SELML/42-80	LAYDVSGS	PRPNEYSTES	RRDIPLITGRFDSSEQ	IEFT
PSBE_PHATC/43-81	LAYDVFGT	PRPNEYFTQD	RQQIPLVNDRFSAKQEL	EDLT
PSBE_OSTTA/42-79	LAYDVFGT	PRPNEYFTEER	RQELPLISDRYGALAQL	ODL
PSBE_CHLRE/42-79	LAYDVFGT	PRPNEYFTED	RQEAPLITDRFNALEQV	KKL
PSBE_SYNJB/42-80	LAYDIFGT	PRPNEYFTAER	RQELPIVSDRFNALEEL	ERLT
K9TLB2_9CYAN/43-80	LAYDAFGT	PRPNEYFSQER	RLELPIVQDRFNAKDQ	IODL
K9S2L1_9CYAN/42-80	LAYDAFGT	PRPDEYFTQQ	RQELPIVMDRFQAKQQV	EEFI
K9YE64_HALP7/42-80	LAYDAFGT	PRPDEYFTKQ	RQELPIISDRFQAREEV	EEFN
K9R7L0_9CYAN/42-80	LAYDAFGT	PRPDTYFTQER	QEVPLVNDRFEAKNQV	OVFT
PSBE_THEVB/42-80	LAYDVFGT	PRPDSYAAQEQR	SIPLVTDREFAKQQVET	FL
PSBE_GLOV/43-81	LAYDVFGT	PRPDEYYNSDNT	KKPLVNKRFEAKQQQLD	EAT
PSBE_SYNR3/43-81	LAYDAFGT	PRPDSYFQAADV	KAPVVSQRYEGKSQLD	TRL
PSBE_PROMA/43-81	LAYDAFGT	PRPDTYFQASET	KAPVVSQRFESKAQLD	LRL
K9RTX4_SYNP3/42-80	LAYDIFGT	PRPDTYFSQEQQA	APVVSDRFESKQQINE	FL
K9ULI3_CHAP6/43-81	LAYDAFGT	PRPDEYFTQTR	QELPIVSDRFESKGQLD	KFV
K9UY34_9CYAN/42-79	LAYDAFGT	PRPNEYFTQTR	QEVPIVNNRFDKAKQV	ETY
PSBE_NOSS1/42-80	LAYDVFGT	PRPDEYTTQAR	QELPIVNNRFEAKKQVE	QLI
K9QJT8_9NOSO/42-80	LAYDVFGT	PRPNEYTTQTR	QEVPIVKNRYEAKQQVE	KFI
K8GLK0_9CYAN/42-80	LAYDVFGT	PRPNEYTTSTR	QQAPIVKDRFEAKQQID	QFI
K9VJB4_9CYAN/42-80	LAYDAFGT	PRPNQYFTPQRE	EVPIISDRFEAKKQVD	QFI
K9W2B5_9CYAN/42-80	LAYDAFGT	PRPNEYFTQQR	QEVPIVSDRFESKQEI	DQFL
K9TZN4_CHRTP/42-80	LAYDVFGT	PRPNEYYPQER	QELPIVSDRFESRKQIE	EFI
K9XG15_9CHRO/42-80	LAYDVFGT	PRPNEYYPQER	QELPIIGDRYQAKQEI	KQFI

Можно ориентироваться на следующее:

- Не содержит гэпов.
- Длина - не менее 4 колонок.
- Первая и последняя колонки - абсолютно консервативны или абсолютно функционально консервативны.
- Процент абсолютно консервативных колонок больше ... 35%
- Не содержит более 10 колонок, не являющихся абсолютно консервативными или абсолютно функционально консервативными, подряд.
- Не может быть расширен без нарушения предыдущих свойств.

Но точных критериев нет. Нужен опыт.

	10	20	30
K9Z0V0_CYAAP/42-80	LAYDVFGT	PRPDEYFTQT	RQELPIINDRYEAKNQI
K9SST0_9SYNE/42-80	LAYDIFGT	PRPNEYTND	RQEAPILRDRYDANQEI
K9SGD2_9CYAN/42-80	LAYDVFGT	PRPNEYTLD	RSDAPVLLDRFNAIDDI
K9Y0S0_STAC7/42-80	LAYDVFGT	PRPNEYFTQD	RLELPIINDRYSASEQI
PSBE_RIPO1/42-80	LAYDVFGT	PRPDQYFTQER	QELPIISDRYKASQQI
PSBE_SYNY3/42-80	LAYDAFGT	PRPDEYFTQT	RQELPILQERYDINQEI
PSBE_GLOC7/42-80	LAYDVFGT	PRPDEYFTQER	LELPIILKDRYNTDQQI
K9W859_9CYAN/42-80	LAYDVFGT	PRPNEYFTQD	RVEAPILSDRYEAKQQI
PSBE_SYNE7/43-81	LAYDAFGT	PRPNEYFTQD	RTEVPIVSDRYSAKQQV
PSBE_ACAM1/42-80	LAYDVFGT	PRPNEYFDQAR	QGLPLVTDRYEGKQQI
C7B2G7_SELML/42-80	LAYDVSGS	PRPNEYSTES	RRDIPLITGRFDSSEI
PSBE_PHATC/43-81	LAYDVFGT	PRPNEYFTQD	RQQIPLVNDRFSAKQEL
PSBE_OSTTA/42-79	LAYDVFGT	PRPNEYFTEER	QELPLISDRYGALAQL
PSBE_CHLRE/42-79	LAYDVFGT	PRPNEYFTEDR	QEAPLITDRFNALEQV
PSBE_SYNJB/42-80	LAYDIFGT	PRPNEYFTAER	QELPIVSDRFNALEEL
K9TLB2_9CYAN/43-80	LAYDAFGT	PRPNEYFSQER	LELPIVQDRFNAKDQI
K9S2L1_9CYAN/42-80	LAYDAFGT	PRPDEYFTQQ	RQELPIVMDRFQAKQQV
K9YE64_HALP7/42-80	LAYDAFGT	PRPDEYFTKQ	RQELPIISDRFQAREEV
K9R7L0_9CYAN/42-80	LAYDAFGT	PRPDYFTQER	QEVPLVNDRFQAKNQV
PSBE_THEVB/42-80	LAYDVFGT	PRPDSYQAQEQ	RSIPLVTDRFQAKQQV
PSBE_GLOWV/43-81	LAYDVFGT	PRPDEYNSDNT	KKPLVNRKFQAKQQL
PSBE_SYNR3/43-81	LAYDAFGT	PRPDSYFQAADV	KAPVVSQRYEGKSQ
PSBE_PROMA/43-81	LAYDAFGT	PRPDYFQASET	KAPVVSQRFESKAQ
K9RTX4_SYNP3/42-80	LAYDIFGT	PRPDYFSQEQQA	APVVSDFESKQQI
K9ULI3_CHAP6/43-81	LAYDAFGT	PRPDEYFTQTR	QELPIVSDRFESKGQ
K9UY34_9CYAN/42-79	LAYDAFGT	PRPNEYFTQTR	QEVPIVNNRFDAKKQ
PSBE_NOSS1/42-80	LAYDVFGT	PRPDEYYTQAR	QELPIVNNRFQAKKQ
K9QJTB_9NOSQ/42-80	LAYDVFGT	PRPNEYYTQTR	QEVPIVKNRYEAKQ
K8GLK0_9CYAN/42-80	LAYDVFGT	PRPNEYYTSTR	QAPIVKDRFEAKQ
K9VJB4_9CYAN/42-80	LAYDAFGT	PRPNQYFTPQRE	EVPIISDRFEAKKQ
K9W2B5_9CYAN/42-80	LAYDAFGT	PRPNEYFTQQR	QEVPIVSDRFESKQ
K9TZN4_CHRTP/42-80	LAYDVFGT	PRPNEYYPQER	QELPIVSDRFESRKQ
K9XG15_9CHRO/42-80	LAYDVFGT	PRPNEYYPQER	QELPIIGDRYQAKQ

- Для выделения блоков в выравнивании очень полезен JalView
- Практическое задание на семинаре – освоить функции Create groups from selection и sort



Спасибо за внимание!