

Множественное выравнивание последовательностей белков

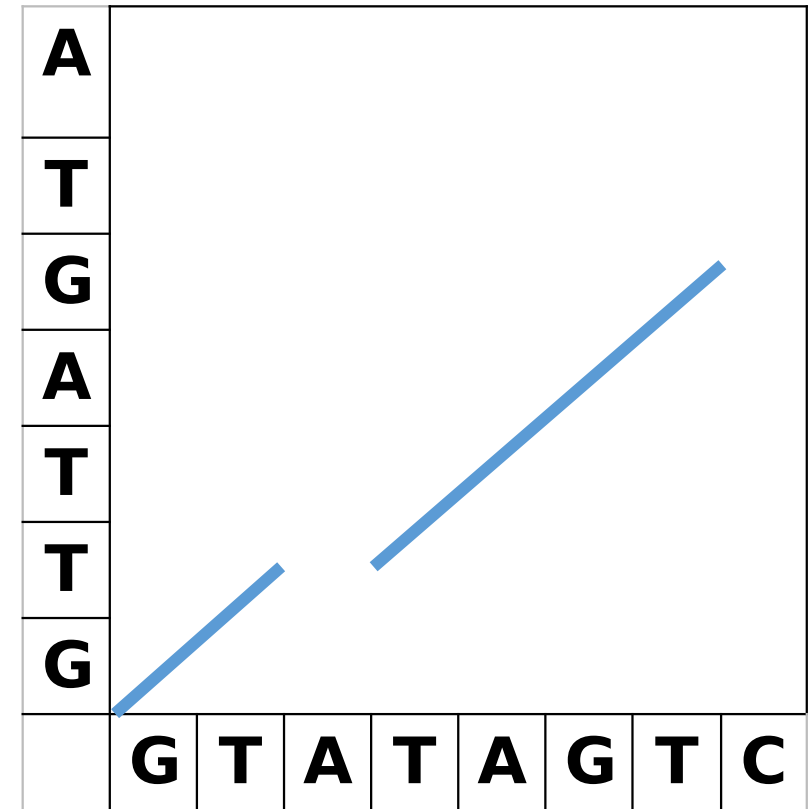
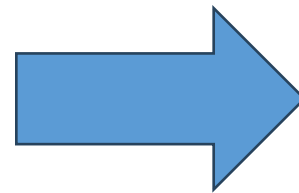
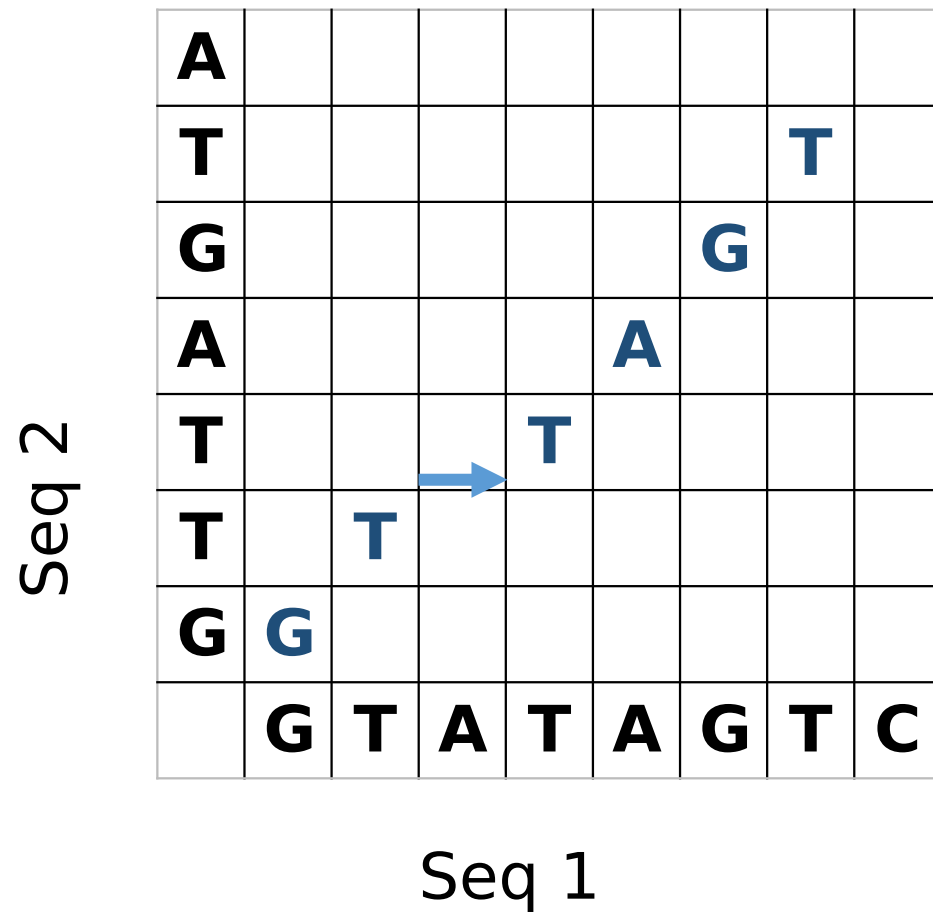
Практикум 11: парное выравнивание – карта локального сходства. Обязательная для зачёта КР по рассказанному!

Гомология и сходство последовательностей

Jalview - редактор выравниваний. приёмы

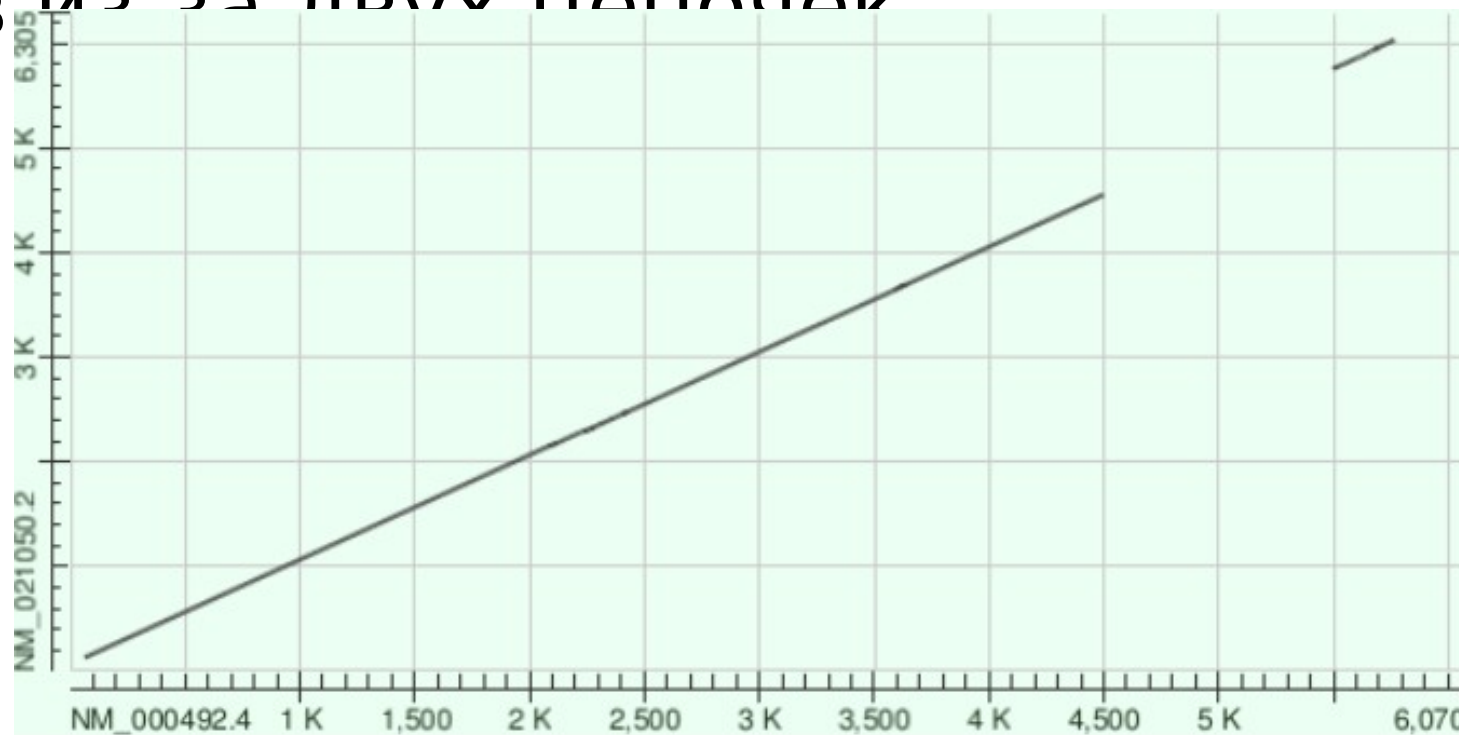
Блоки достоверного выравнивания как замена карты локального
сходства

Парное выравнивание как матрица. Карта ЛОКАЛЬНОГО СХОДСТВА **Dotplot**



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

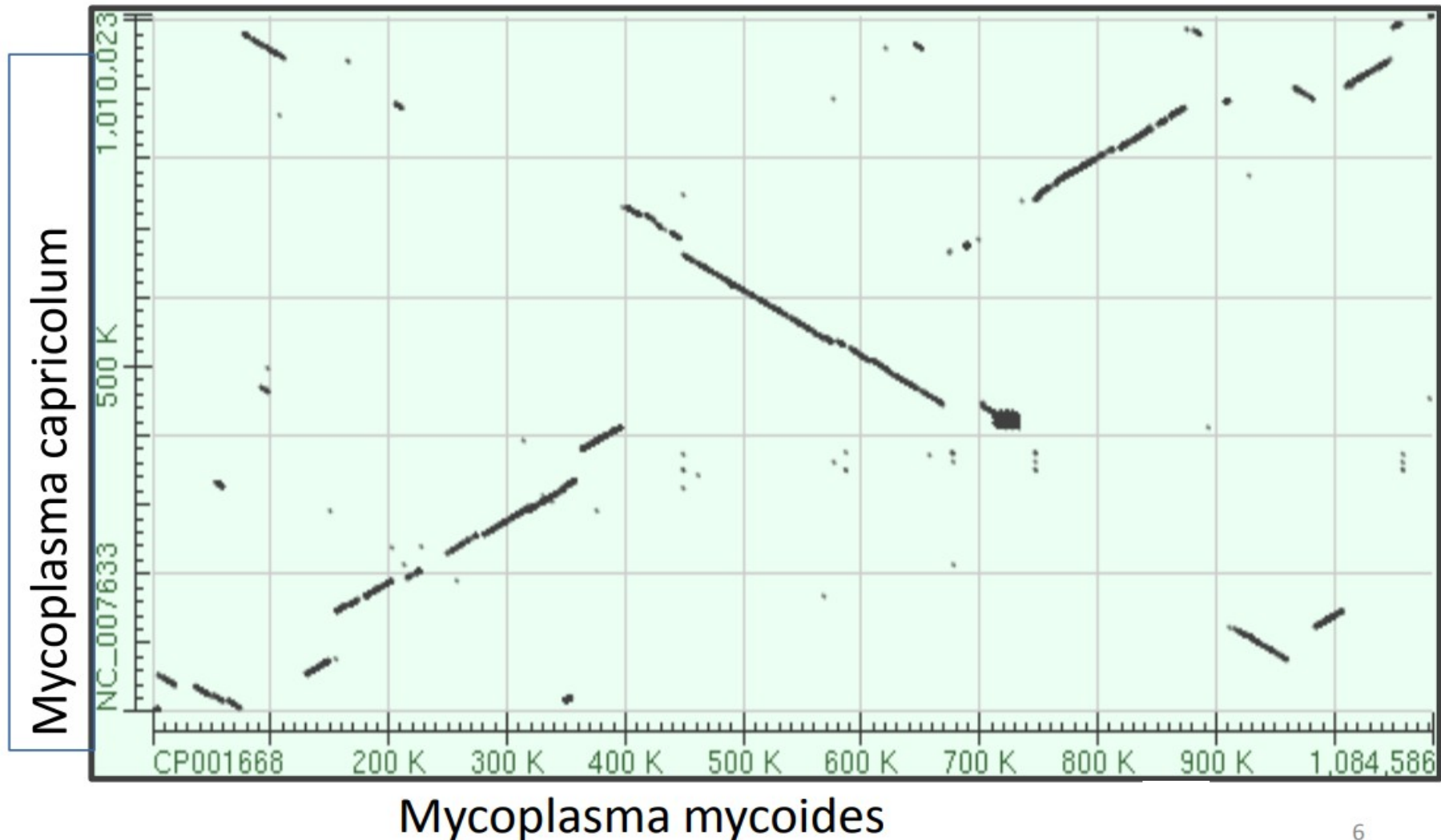
Карта локального сходства. у выравнивания
ДНК (например, геномов) **есть отличие от**
выравнивания последовательностей
белков из-за двух цепочек



Отрезки отображают наличие локального выравнивания
 Как отобразить Сходство прямой цепочки одной
 последовательности ДНК с комплементарной цепочкой
 второй последовательности?

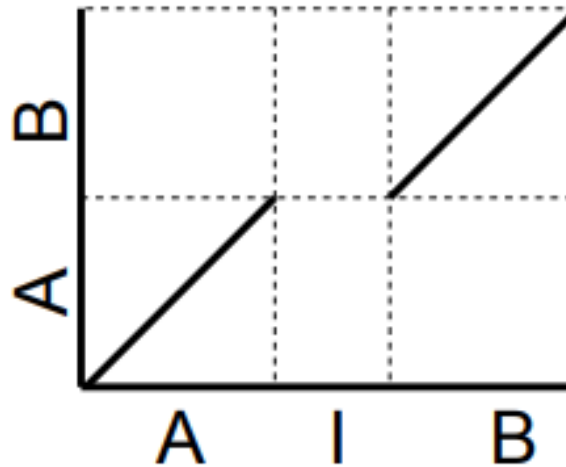
Карта локального сходства ДВУХ маленьких геномов одного рода *Mycoplasma*.

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



Делеция или вставка

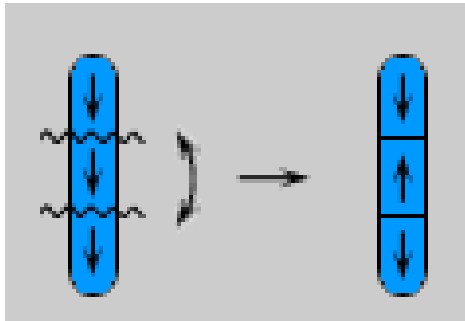
R: AIB
Q: AB



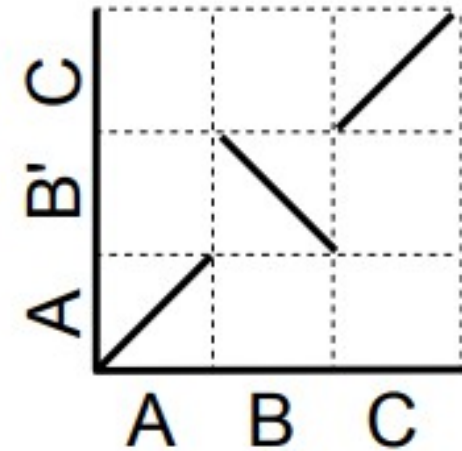
- <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

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. Упражнение нарисуйте карту л.с.

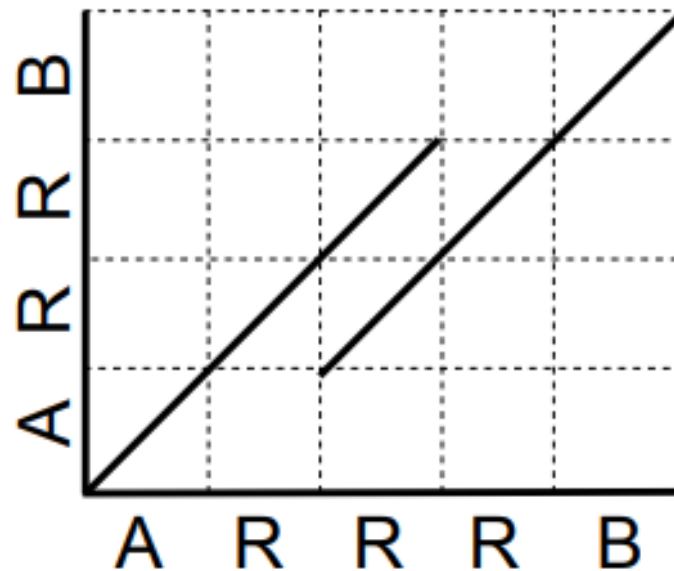
3'

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								
	5'	A	T	G	A	C	G	G	C	A
		T	A	C	T	G	C	C	G	T

3'

Дубликация. Может оказаться в КР. Чего не хватает!!!

R: ARRRB
Q: ARRB



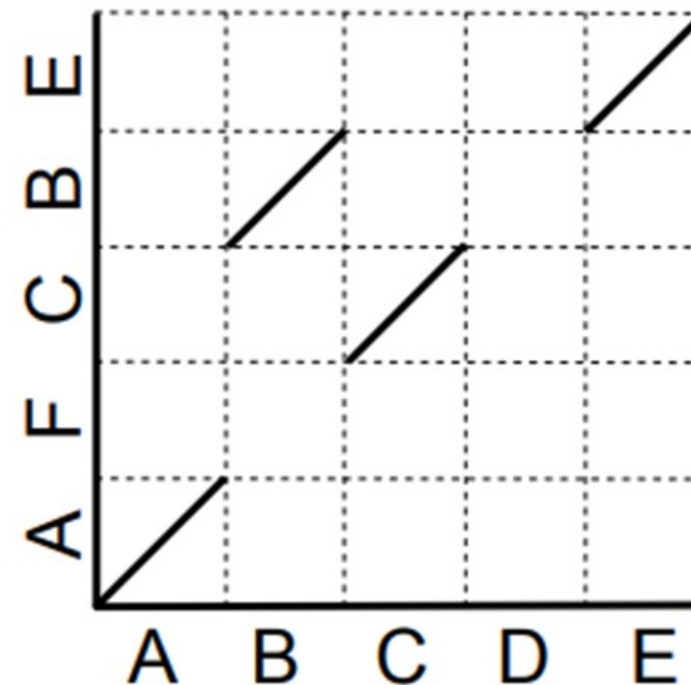
- <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

Транслокация

Транслокация –
перенесение фрагмента в
другое место.

С сохранением ориентации
или нет.

R: ABCDE
Q: AFCBE



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

BLAST двух последовательностей

The screenshot shows the NCBI BLAST website interface. The browser address bar displays the URL: `https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome`. The page header includes the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". A "Log in" button is visible in the top right corner.

The main heading is "BLAST® » blastn suite". Below this, there are navigation links: "Home", "Recent Results", "Saved Strategies", and "Help".

The "Standard Nucleotide BLAST" section is active, with tabs for "blastn", "blastp", "blastx", "tblastn", and "tblastx". Below the tabs, a note states: "BLASTN programs search nucleotide databases using a nucleotide query. more...". There are "Reset page" and "Bookmark" buttons.

The "Enter Query Sequence" section contains the following fields:

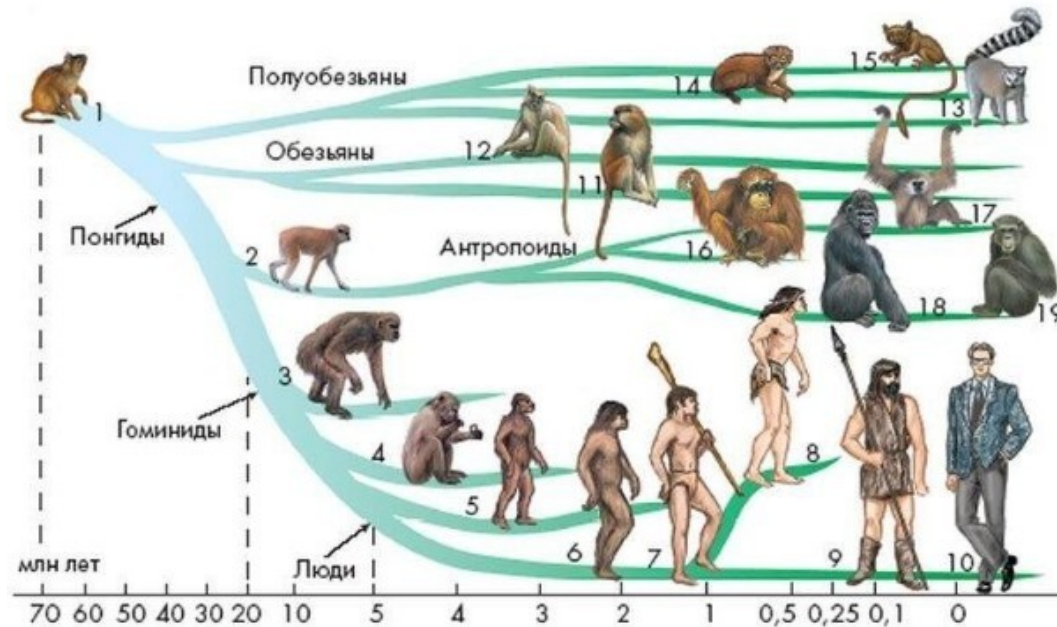
- "Enter accession number(s), gi(s), or FASTA sequence(s)" with a "Clear" link.
- "Query subrange" with "From" and "To" input fields.
- "Or, upload file" with a "Выбор файла" button and the text "Не выбран ... один файл".
- "Job Title" with a text input field and the prompt "Enter a descriptive title for your BLAST search".
- A checkbox labeled "Align two or more sequences" which is currently unchecked. A red arrow points to this checkbox.

The "Choose Search Set" section includes:

- "Database" with radio buttons for "Standard databases (nr etc.)", "rRNA/ITS databases", "Genomic + transcript databases", and "Betacoronavirus". A dropdown menu shows "Nucleotide collection (nr/nt)".
- "Organism" (Optional) with a text input field, an "exclude" checkbox, and an "Add organism" button. Below the input field, it says: "Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown".
- "Exclude" with checkboxes for "Models (XM/XP)" and "Uncultured/environmental sample sequences".

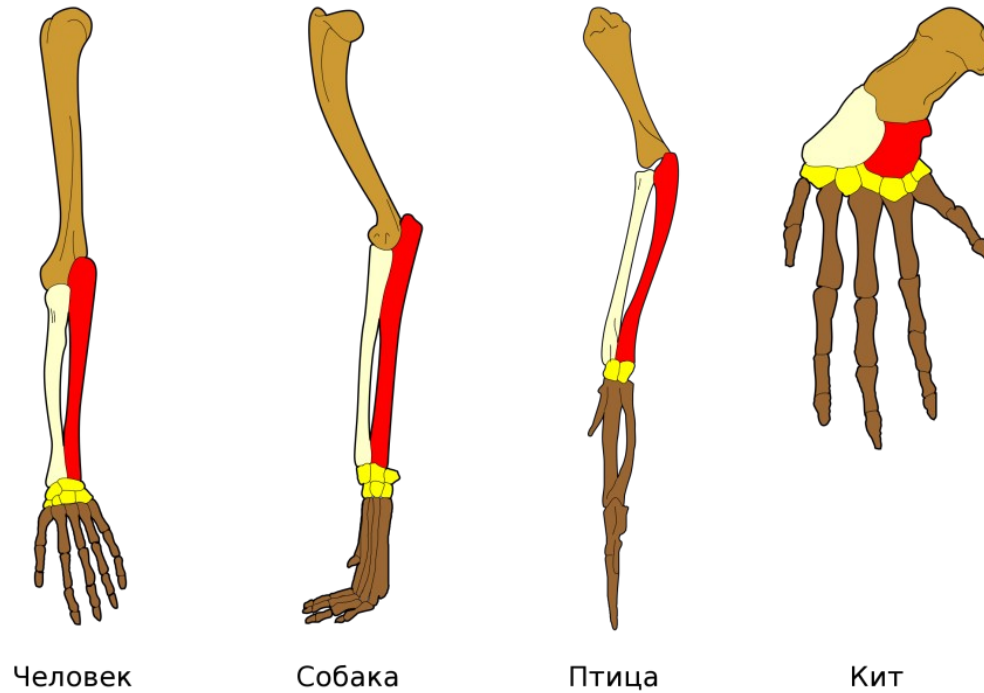
Гомология и сходство

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



Последний общий предок ныне живущих обезьян.

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, иногда неправильно)))
- ВАЖНО – Мутации происходят случайно и в случайных местах генома. Они не знают где гены, где сигналы управляющие экспрессией генов.
- Отбор удаляет вредные мутации, слабовредные, оставляет нейтральные мутации и полезные, это нечасто но бывает.

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, вставки в результате горизонтального переноса ДНК, дупликации, транслокации, инверсии

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

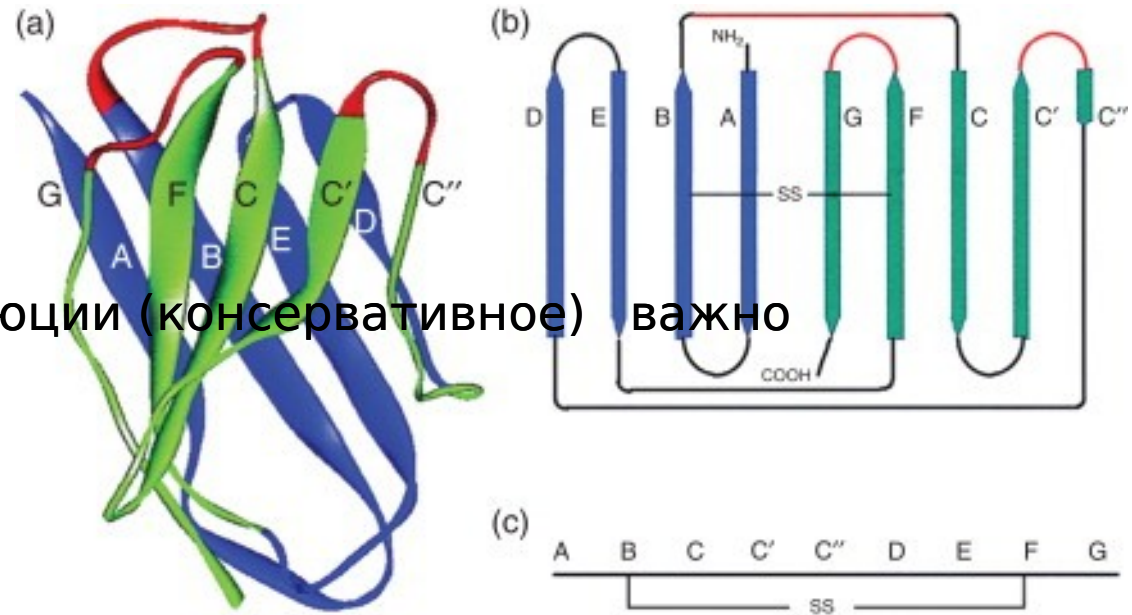


Гомология в молекулярной биологии

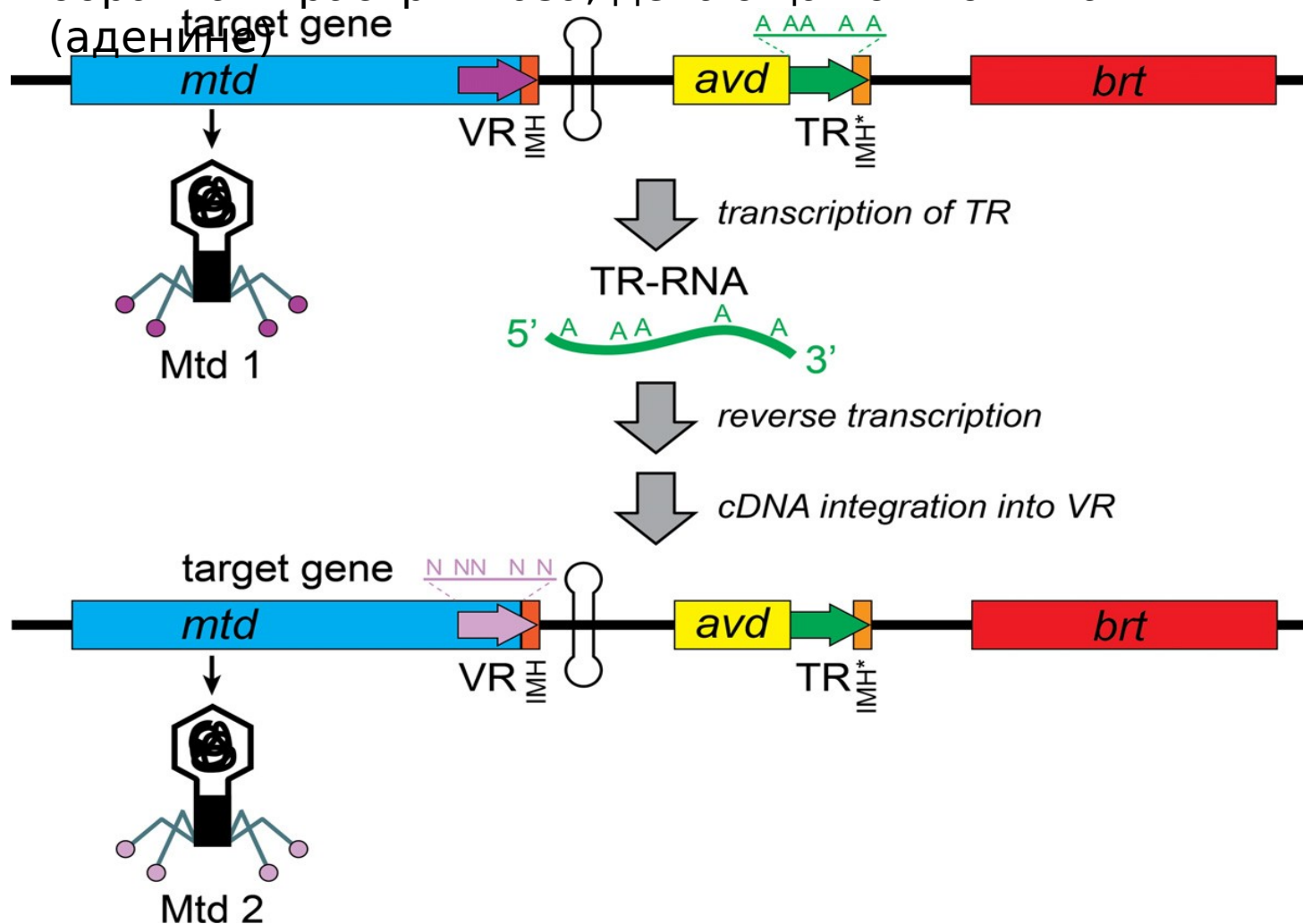
- Гомология  Сходство последовательностей
 - Последовательности могли сильно измениться в эволюции от общего предка
 - Есть системы быстрой эволюции белковых последовательностей (Иммуноглобулины, Diversity-generating retroelement (DGR) у вирусов и бактерий)
 - Дольше в эволюции сохраняется сходство пространственных 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. Сохраняющееся в эволюции (консервативное) важно
- Последовательности Ig-like Fold могут сильно отличаться, детектируемое сходство только на коротких участках
- Ход полипептидной цепи и элементов вторичной структуры очень похожи



В основе гипервариабельности (в известной системе DGR) лежит обратная транскриптаза, делающая ошибки на А (аденине)



Wu L et al. Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. Nucleic Acids Res. 2018 Jan 9;46(1):11-24

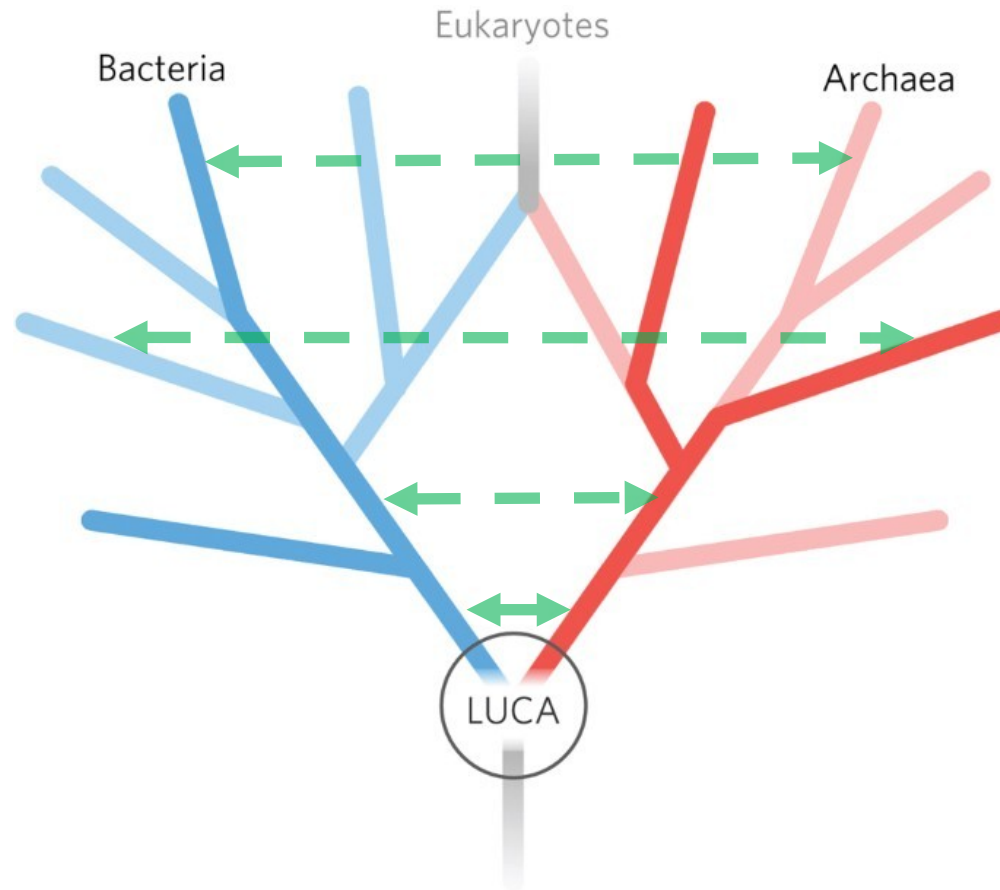
Гомология в молекулярной биологии

- Сходство последовательностей  гомология

Гомология в молекулярной биологии

- Сходство последовательностей  гия
 - Бывает случайное совпадение – аналогия (повторы, простые участки)
 - Чем длиннее фрагмент последовательности, тем менее вероятно совпадение
 - Сходство последовательностей белков более значимо, чем последовательностей их генов из-за синонимических мутаций.
 - Чем больше идентичных или функционально сходных аминокислотных остатков, тем менее вероятно совпадение

LUCA – Last Universal Common Ancestor



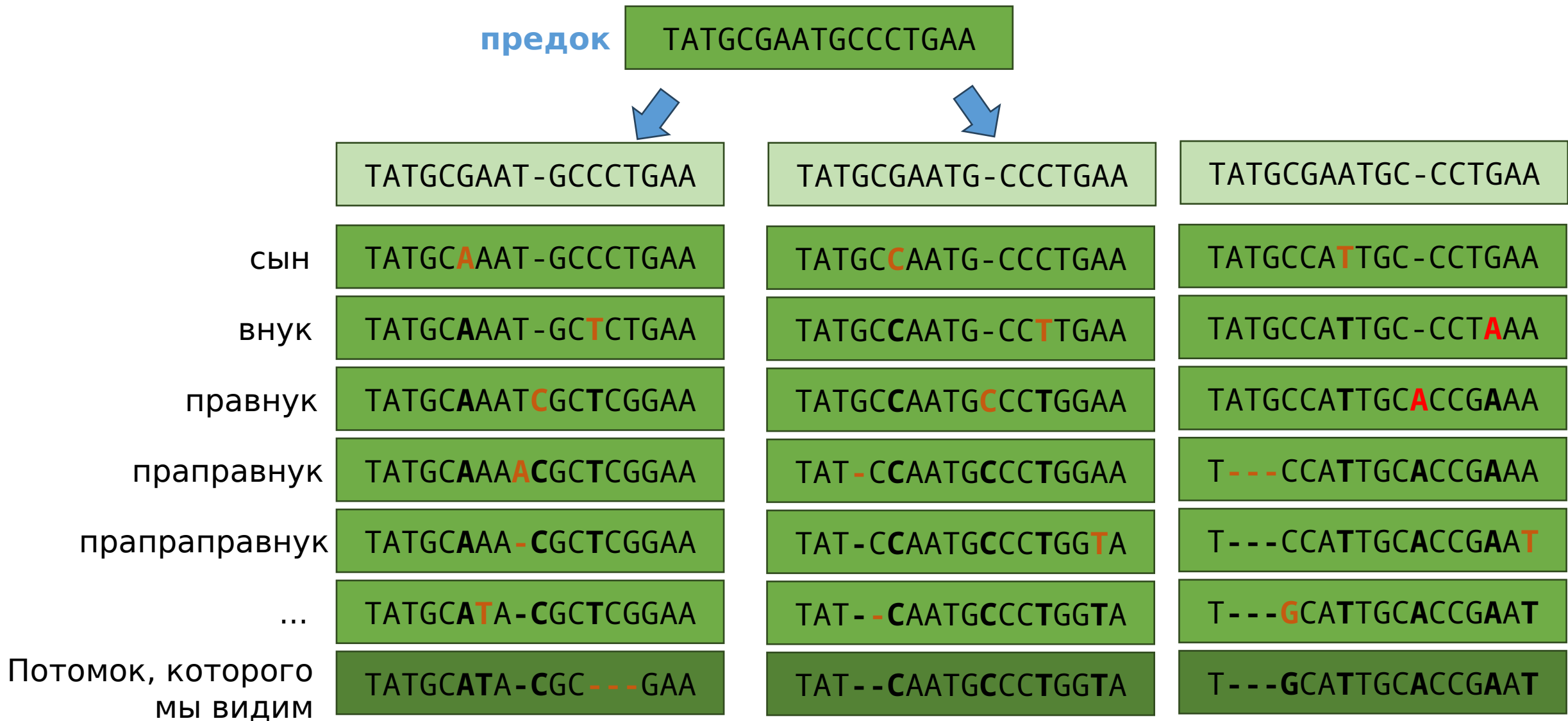
- Обоснование на молекулярном уровне было сформулировано в начале 2000х (Aravind et al, 2002; Koonin, 2003 10.1038/nrmicro751)
- Были собраны все белки, гомологи которых встречаются во всех таксонах высокого порядка (Weiss et al, 2018 10.1038/nmicrobiol.2016.116)
- Возраст исчисляется миллиардами лет.

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

Эволюционное выравнивание – выравнивание отражающее эволюцию последовательностей

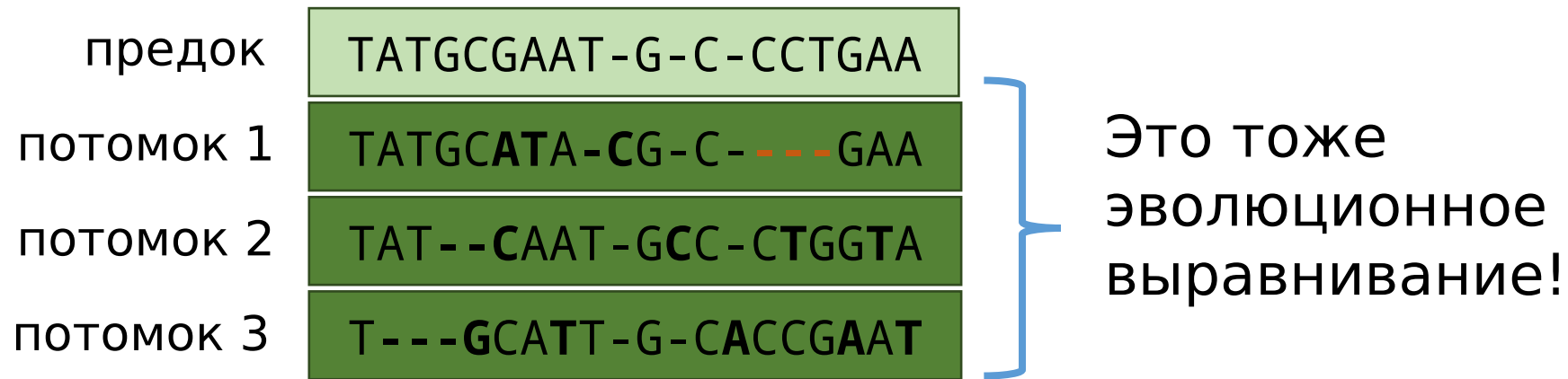
предок	TATGCGAAT-GCCCTGAA	
сын	TATGCA ^A AAT-GCCCTGAA	замена
внук	TATGCA ^A AAT-GC ^T CTGAA	замена
правнук	TATGCA ^A AAT ^C GCTCGGAA	вставка 1 п.н.
праправнук	TATGCA ^A AA ^A CGCTCGGAA	замена
прапраправнук	TATGCA ^A AA- ^C GCTCGGAA	делеция 1 п.н.
...	TATGC ^A TA- ^C GCTCGGAA	замена
...	TATGCATA- ^C GC- - -GAA	делеция 3 п.н.

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



Эволюционное выравнивание – выравнивание отражающее эволюцию последовательностей

- Выравниваем последовательности потомков относительно предка:



- 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 : AGVSEFIVNDDVELARELNADGIHIGQTDDESVSKVF
THIE_MANSM : FQVPFIVNDDVELALSIQADGIHVGQKDTAVETII
THIE_STRA3 : YQVPFIIDDDIDLVELIDADGLHIGQNDLPVDEAF
THIE_LISIN : YQVPFIINDDVALALEIGADGIHVGQNDEEIRQVI
THIE_ANOFW : YNIPFIVNDDVDLALALQADGVHVGQDEDEVAERVI
THIE_GEOTN : YGVPFIVNDDVELAIAIDADGVHVGQDDEDARRVI
THIE_BACSU : AGVPFIVNDDVELALNLKADGIHIGQEDANAKEVI
THIE_BACA2 : AGIPFIINDDVELALRLEADGVHIGQDDADAETI
THIE_OCEIH : FQIPFIINDDVDLAKQLDADGIHIGQDDQPVEVVI
THIE_STAAB : YNVPFIVNDDVSLAKEINADGIHVGQDDAKVKEI/
THIE_STACT : YNVPFIVNDDVALAEEIDADGIHVGQDDEAVDDFT
                6pFI61DD6 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 *делетированы* 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 DRHEVLDNWKGIWSAEFTGRRVAIGQAIFQELFALDPNAKGVFGRVNVDPKSEADW 63
8 DRREVQALWRSIWSAEDTGRRTLIGRL LFEELFEIDGATKGLFKRVNVDDTHSPEE 63
7 QRIVKVKQWQAQVYSVGES--RTDFAIDVFNNFFRTNPD-RSLFNRVNGDNVYSPEF 59
9 QRLKVKQWAKAYGVGHE--RVELGIALWKSMFAQDNDARDLFKRVHGEDVHSPAF 62
8 EGLKVKSEWGRAYGSGHD--REAFSQA IWRATFAQVPESRSLFKRVHGDDTSHPAF 61
6 QRFKVKHQWAEAFGTSHH--RLDFGLKLWNSIFRDAPEIRGLFKRVDGDNAYSAEF 59
7 QRLKVKRQWAEAYGSGND--REEFGHF IwthvfkDAPSARDLFKRVRGDNIHTPAF 60

67 KSHLVRVANGLDLLINLLDDTLVLQSHLGHLADQHIQRKGVTKKEYFRGIGEAFA 120
64 KAHVIRVINGLDLAVNLLDPKALQEELKHLARQHRERSGVKAVYFDEMEKALL 117
64 FAHVLRVVNGLDTLIGVLGDSDTLNSLIDHLAEQHKARAGFKTVYFKEFGKALN 117
60 KAHMVRVFAFGDILISVLDDKPVLDQALAHYAAFHKQF-GTIP--FKAFGQTMF 110
63 EAHMARVFNGLDRLVSSLTDEPVLNAQLEHLRQQHIKL-GITGHMFNLMRTGLA 115
62 IAHAEVLGGLDIAISTLDQPATLKEELDHLQVQHEGR-KIPDNYFDAFKTAIL 114
60 EAHAEVLGGLDMTISLLDDQAAFDAQLAHLKSQHAER-NIKADYYGVFVNELL 112
61 RAHATRVLGGLDMCIALLDDEGV LNTQLAHLASQHSSR-GVSAAQYDVVEHSVM 113
```


Пример плохого выравнивания

```
GSYLEVEGPKGKTTVYVTDLYPEGARGA-----L-DL---SPNAFRKIGNMKDGKINI-----KWRV
FSRYKLKYLDLDPKQSYGSAEE-----I-HT---ITRKGFSKEQPNAAKLLS-----QFKW
VKDTRAASEEELLQKAHFLQRMLSYGTTTAEVKSGYGLEKETELKQLRVAKKL-HE---SQPVDLVSTFMGAHAIPP-----EYQN
-----EFTDELAKQYF-----HA---VPKIAARIRDEFEPIGLN-----TLNN
VPEEWDI I RESMPAYNS-----NY---TPPTRGMGQIVELFRSYP-----EVKR
GGIPQISLHLANPAFQSGHFKTPITNDQYKKILDSSTVEGKRLN-----AM---LSKIDGLQELENQGVVPV-----LFRP
-----EKFEAKLEEYNG-----NL---KRAAVELAKEWRSDKVLRL-----KLEA
-----EE
YDPSTWGKKEPSGTQEEARARSQKNQKARVMVDIPANTTIVGSGTNAKVVGGNF-QI---KSDNVIIRNIEFQDAYDY-----FPQW
FIFTNCETICPPMTA-----HM---TDLQKKLKAENIDVRIIS-----FSVD
SALGINTDFSPVVDINNPNPNVIGVRSFSSNRELTSRLGLYTMKGLQRQDIAS-AL---KHFPGHGDTDVDSHYGLP-----LVSH
LSSDDPN SYHYMNDR-----FF---QHIDSKPSRHFIPNGNAD-----DLEA
FSPKQIRARFPHL-----HI---GRSVHSLEEAVQAEKEDA-----DYVL
QNVKLLNIDADTLTD-----VF---EPGEVKRVYINFSDPWPK-----KRHE
IEQAWEEGKQVCIPKC-----
FHLNKN EYTVPFVCGCF-----
IIVSGDSQSPWVKKEN-----
-----SIMGLMSLAVSTGTEV-----TLIA
QLIKRMSDEGHIIGNHSFHHP-----DL---TTKTADQIQDELDVNEE-----VYKI
---EKFEELYDHAAETVD-----TI---AERLLAIGGQPVATVKEY-----TEHA
-----EK---YRGLVLTSGKKGKIGKRLV-----YRHE
FSTFSVQSYNKRKEEAE-----KL---GGLTITSVESWYDEPKFV-----TYWV
SNLPEHTERYEIAQEHLDVVRGLWNSWEHDAFIHNKKTGQFFDQAKLHRLNHKGKYFQVEGPLNIGRSKQGEPPVVFQA-----GSSE
KGKSDFSLFSPLYEQLEIFNKRFITYWGYQEHPYDALLDLFEPGVTVKVLDQLFAEL---KEAIIPLVKQVTASGNKPDTSFITKAFPKEKQKELSLYFLQE
QFMEPLREFAAQGKPMFGTCA-----GL---IILAKEIAGSDNPHLGLL-----NVVV
-----SIMGVMSLGIKGAIEI-----TISA
CKGANEVLYPEHAQALRE-----QL---APLATVITPNLFEASQLS-----GMDE
IDVYFPDSEISCLAQGFY-----AL---KAAELIKNGASSPEDIK-----ELEE
TYPLAIQEIEEGIVFEDD-----QF---IVTAVSVIHGVEAFGYRV-----QEKD
TKDERYVKAARLRSLYGTNLNRTSEGGFWHKDGYPYQMMLDGLYMGGPFALKYA-NL---KQETELFDQVVVLQESLMR-----KHTK
TIVDEKGRSFIDDVGEGDLWYFPSGLPHSIALEEGAELLLVFDDGSFSENSTF-QL---TDWLAHTPKEVIAANFGV-----TKEE
RNDIEIEVKDGELPST-----FV---PGRNLVFLSFASILAYQI-----GARH
GELVPDEVTIGIVKERLGK-----DD---CERGFLLDGFPRTPVAQAE-----ALEE
ADGIIHIGQEDANAKEVRA-----AI---GDMILGVSAHTMSEVKQA-----EE
EFAELLEISEERKAF-----KP---SPTAYSYTSHMYRSVLSG-----NFAE
INMTGDGAEDTFQIPVTDGQLNLLVTEGKAGTFTLSALKIKKLSQPVNTNRTI-YV---GGDSTVCNYYPLNSSKQA-----GWGQ
```

Программа выровняет всё, что Вы ей дадите

DNA (cytosine-5-)-methyltransferases

	10	20	30	40	50	60
<i>EJL77459.1</i>	GVDLVFGGPPCQGF	SIQIGMRR-LDDER-	NE	LYQQYTRIVAKLKPRV	FLMENVPNL	ALMNKGH
<i>RXK67093.1</i>	DLDVVFGGPPCQGY	SIQIGTRR-LDDER-	NE	LYLQYARIVEKQRP	RMFLMENVPNMV	LLNKGH
<i>OJY44288.1</i>	NVDLVFGGPPCQGY	SIQIGTRD-LHDPR-	NRL	FEEFARVVATLKP	KLFLMENVPNL	LLLNKGH
<i>TRU90449.1</i>	NPEMIVGSPPCQDF	SSAGKRNEGLGR-	-ANL	TLTFAEIVTRVSP	QWFMENVVD-	-RIEKS
<i>OXI46696.1</i>	GTDLVFGGPPCQGF	SIQIGMRR-LDDER-	NE	LYKQYTRVVSTLR	PRVFLMENVPNL	ALMNKGH
<i>AVZ30243.1</i>	EIDVVFGGPPCQGF	SLIGKRS-FEDPR-	NS	LVFHYIRLVLELS	PKFFVIENVKGM	TAGNHQA
<i>AFZ12381.1</i>	DIIGFIGGAPCPDF	SVGGKNRGSEGDK-	GK	LSASYIELICQQK	PDFFLFENVKGL	YKTKKHR
<i>HCQ21462.1</i>	HIIGFIGGPPCPDF	SVGGKNKGHLGDN-	GK	LSASYIELICQNL	PDFFLFENVKGL	WRTTKHR
<i>EDN77159.1</i>	SLIGFIGGPPCPDF	SIAGKNKGKDGDN-	GK	LSLSYTNLIEMK	PDFFLFENVKGL	WRTARHR
<i>SOD91684.1</i>	EVSLVVGGAPCQPF	SNIGKKLGKNDER	NGDL	LEFVRMVKGIQ	PEAFIFENVVG	ITQNKHSD
<i>QCS48280.1</i>	NVVGFIGGPPCPDF	SIIGGKNRGGRQGDH-	GK	LSSESYIDLIIQH	QPDFFLFENVKGL	YRTTKKHR
<i>SMB95934.1</i>	GLFGFIGGPPCPDF	SVGGKNRGENGEG-	GR	LSKVFDKILDLQ	PVFFLYENVPGL	IRTAKHR
<i>RUO38876.1</i>	SPVGFIGGPPCPDF	SVGGKNRGHEGEN-	GR	LTRYVDGIIKYA	PDFFLFENVKGL	WRTKRHR
<i>OIP70538.1</i>	TIDLICGGPPCQGF	STIGTND-KKDH-	NF	LFEEFLRMVETFK	PNFIILENVTGL	LAKKNES
<i>AFY60915.1</i>	NLVGFVGGPPCPDF	SIIGGKNKGQYGDN-	GK	LTKVYVDIIENQ	PDFFLFENVKGL	WRTSRHR
<i>CUR30340.1</i>	DLIGFIGGPPCPDF	SVGGKNRGKNGDQ-	GK	LACYVELICQQR	PDFFLFENVKGL	WSTTKKHR
<i>TAK03971.1</i>	QAALVVGGAPCQPF	SNLGSKRGTADSR-	GT	LFQDFIRIVKGV	RPKGFI FENVEGL	TQDKHKG
<i>AEE51071.1</i>	KVALVVGGAPCQPF	SNIGKKEGENDAK	NGDL	LEFVRMVKGIQ	PEAFIFENVAG	IIQSKHSK
<i>RTR31666.1</i>	RLVGFIGGPPCPDF	SVGGKNKGSEGEN-	GK	LTRYIDLI VKDN	PDYFI FENVKGL	WRTTRHR
<i>PTU64472.1</i>	NIDLVFGGPPCQGF	SIQIGTRR-LDDER-	NE	LYKQYTRIVKTLK	PRVFLMENVPNL	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-VAQDRRERVFYVGRKDLN									
YIG-VAQERKRVFYIGFRKDLEIKF									
YIG-VAQDRKRVFYIGFRKELNIN									
YIG-IPQQRDRLVLF									
YIG-VPQSRQRVFFIGLKSDRPLNQQ									
CG-VPQLRKRTFVIGHRHGSIAD									
CG-VPQSRTRFSLIGKLNSEHNF									
YIG-VPQRRHRIIVGIRKDQD									
IG-AHHQRHRWFCLAIRKDYEP									
FGNIPQNRERIYIVGFRN									
FN-VPQNRERLYIIGIREDLIKNEE									
FG-IPQNRERVFICISILN									
FG-SSQARRRVFMISTLNEF									
FG-IPQNRERIYLVGF									
FG-LPQNRERIYIVGFDRKS									
FG-VPQNRERIYIVGFNKEK									
FG-VPQNRERIYIVGFHKS									
IFQ-VPQNRRLRVYIVGLDQSQPELT									
FG-VPQNRVRIYILGILGSKPKLT									
IFG-IPQKRKR FYLVAF									
IFG-LPQRRERIIVVGFHPDLG									
YIG-IPQKRERIYMICFRNDLN									
YIG-NAQRRRRVFIFGYKQDLNYSKAME									
YIG-TPQRRKRAIRLNKKGTIWN									
IGG-TPQVRERVFITATLVPERMRDERIPRTETGEIDAEAIGPKPVATMNDRFP									
YIG-VAQNRDRVFIIIGIQQKLGVPD									
IFG-VAQNRDRVFIVGIIQQKLDLNG									
YID-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	LYQQYTRIVAKLKPRVFLMENVPNL	ALMNKGH		
RXK67093.1	DLDVVF	GGPPCQGYSQ	IGTRR-LDDER-NE	LYLQYARIVEKQRPRMFLMENVPNM	VLLNKGH		
OJY44288.1	NVDLVF	GGPPCQGYSQ	IGTRD-LHDPR-NRL	FEEFARVVATLKPKLFLMENVPNL	LLL NKGH		
TRU90449.1	NPEMIV	GSPPCQDFSS	AGKRNEGLGR--ANL	TLTFAEIVTRVSPQWFVMENV	D--RIE	SK	
OXI46696.1	GTDLVF	GGPPCQGFSQ	IGMRR-LDDER-NE	LYKQYTRVVSTLRPRVFLMENVPNL	ALMNKGH		
AVZ30243.1	EIDVVF	GGPPCQGFS	LIGKRS-FEDPR-NS	LVFHYIRLVLELSPKFFVIENV	KGMTAGNHQA		
AFZ12381.1	DIIGFI	GGAPCPDFS	VGKNRGSEGDK-GK	LSASYIELICQQKPDFFLFENV	KGLYKTKKHR		
HCQ21462.1	HIIGFI	GGPPCPDFS	VGKKNKGHLGDN-GK	LSSAYIELICQNLPDFFLFENV	KGLWRTTKHR		
EDN77159.1	SLIGFI	GGPPCPDFS	IAGKKNKGKDGDN-GK	LSLSYTNLI IEMKPDFFLFENV	KGLWRTARHR		
SOD91684.1	EVSLVV	GGAPCQPF	SNIGKKLGKNDERNGDL	FLEFVRMVKGIQPEAFIFENV	VGITQNKHSD		
QCS48280.1	NVVGFI	GGPPCPDFS	IGKNRGRQGDH-GK	LSESYIDLIIQHQPDFFI	FENVKGLYRTKKHR		
SMB95934.1	GLFGII	GGPPCPDFS	VGKNRGENGEQ-GR	LSKVFDKILDLQPVFFLYENV	PGLIRTAKHR		
RUO38876.1	SPVGFI	GGPPCPDFS	VGKNRGHEGEN-GR	LTRTYVDGIIKYAPDFFI	FENVKGLWRTKRHR		
OIP70538.1	TIDLIC	GGPPCQGFST	IGTND-KKDH-NF	LFEEFLRMVETFKPNFI	I LENV	TGLLAKKNES	
AFY60915.1	NLVGFV	GGPPCPDFS	IGKKNKGQYGDN-GK	LTKVYVDII IENQPDFFV	FENVKGLWRTRSR		
CUR30340.1	DLIGFI	AGPPCPDFS	VGKNRGKNGDQ-GK	LACYVELICQQRPDFFV	FENVKGLWSTKKHR		
TAK03971.1	QAALVV	GGAPCQPF	SNLGSKRG TADSR-GT	LFQDFIRIVKGV	RPKGFIFENV	EGLTQDKHKG	
AEE51071.1	KVALVV	GGAPCQPF	SNIGKKEGENDAKNGDL	FLEFVRMVKGIQPEAFIFENV	VAGIIQSKHSK		
RTR31666.1	RLVGFI	GGPPCPDFS	VGKKNKGSEGEN-GK	LTRTYIDLIVKDNPDYFI	FENVKGLWRTTRHR		
PTU64472.1	NIDLVF	GGPPCQGFSQ	IGTRR-LDDER-NE	LYKQYTRIVKTLKPRVFLMENVPNL	AMMNKGH		

Учитываем

- Наличие консервативных позиций
- Отсутствие гэпов

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

200	210	220	230	240	250	260	270
YVG-VPQDRKRVFIVGYREDLNLK-----				FEFPKPLNKKVTLRD-----			AIGDL
YVG-VAQDRERVFYVGFRKDLNLSN-----				FE-FPYPISEKERKYLKD-----			SIWDL
YVG-VAQERKRVFYIGFRKDLEIKF-----				SFPKGSTVEDKDKITLKD-----			VIWDL
YVG-VAQDRKRVFYIGFRKELNIN-----				YLPPIPHLIKPTFKD-----			VIWDL
YVG-IPQQRDLVLF AAKQG-----				VIKIIPPTHTPENYR-----			TVRDVIGSL
YVG-VPQSRQRVFFIGLKSDRPLNQQ-----			ILTP-----	PSKVIESEYTSLEEAI SDLPVIE-----			AGEGGEVQDY
CG-VPQLRKRTFVIGHRHGSIAD-----				LANVLQQRLAKQSL-----			TVRDY
CG-VPQSRTRFSLIGKLNSEHNF-----				L IPTLSRKLSDKPM-----			TVRDY
YVG-VPQRRHRI I I VGI RKDQD-----				VAFRVPEPTHKEKYR-----			TASEALADI
IG-AHHQRHRWFCLAIRKDYEPEE-----				I I VSVNATKFDWENNEPPCQVDNK-----			SYENSTLVRL
FGNIPQNRERIYIVGFRN-----				IEHYKNFNFPMPQP-----			LTLTIKDM
FN-VPQNRERLYIIGIREDLIKNEE-----			WSLDFKRKDI LQKGKQRLVELDI KSFNFRWTAQ-----				SAATKRLKDL
IFG-IPQNRERVFCISILN-----				PNEDFTFPQKQ-----			NLTLSMNDL
IFG-SSQARRRVFMI STLNEF-----				VELPKGDKKPKS-----			IKKVLNKI
IFG-IPQNRERIYLVGF-----				LNHDVDFRFPQP-----			IGQATAVGDI
IFG-LPQNRERIYIVGFDRKS-----				ISNYSDFQMPTP-----			LQEKTRVGN I
IFG-VPQNRERIYIVGFNKEK-----				VRNHEHFTFPTP-----			LKTKTRVGDI
IFG-VPQNRERIYIVGFHKS-----				TGVNSFSYPEP-----			LDKIVTFADI
IFQ-VPQNRRLRVYIVGLDQSQPELT-----				ITSHIGATDSHKFKQLSNQASLFD-----			TNKIMLVRDI
IFG-VPQNRVRIYILGILGSKPKLT-----				LTSNVGAADSHKYK-NEQISLFD-----			ES-YATVKDI
IFG-IPQKRKRFYLVAFNLQN-----				IHFEPKP-----			PMISKDIGEV
IFG-LPQRRERIVIVGFHPDLG-----				INDFSFPKGN-----			PDNKVPINAI
YVG-IPQKRERIYMICFRNDLN-----				IQNFQFPKP-----			FELNTFVKDL
YVG-NAQRRRRVFI FG YKQDLNYSKAME-----			ESPLDKI IYHNGLFAEAFPI EDYANKNR-----				VNRTHITHDIVDI
YVG-TPQRRKRAI IRLNKKGT IWN-----				LPLKQNI VSV EQ-----			AIGNLPSI
IGG-TPQVRERVFI TATLVPERMRDERIPRTETGEI DAEAIGPKPVATMNDRFPI KKGTEL FHPGDRKSGWNLLTSGI IREG							
YVG-VAQNRDRVFIIGIQQKLGVPD-----				FSFPEYSESEQRLYDILDNLQTPSII-----			PESLPIQRNL
IFG-VAQNRDRVFI VGIQQKLDLNG-----				FSFPEYAFSDQRLYHILDNL EAPETK-----			LESIP IQRNL
IYD-VAQKRERIV IIGIREDLVK-----				EQKYPFRFPLAQ-----			VYKPVLKDV
YVG-VSQLRPVLFVALKNEYTN-----				FFKWPEPNSEQPK-----			TVGELLFDL

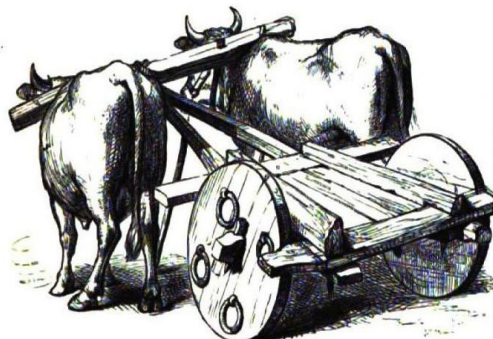
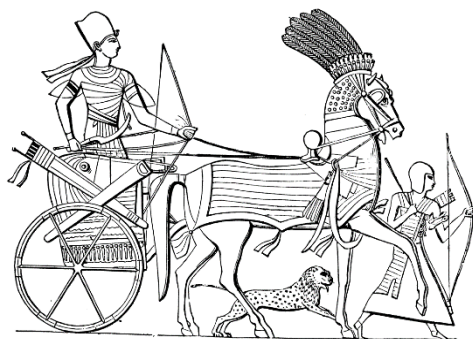
Жёлтый – есть выравнивания отдельных последовательностей (блоки);
 между блоками подгонка программы, гомологии по а.к.о. по колонкам нет

Выравнивание тех же белков другой программой

	10	20	30	40	50	60	70	80
-GVPODRKR VFI	---	VGY	---	REDL	---	NL-KFEFPKP	---	LN--KKVTLRDAIGDL
-GVAQDRERVFY	---	VG	---	RKDL	---	NISNFEFPYP	---	ISEK--ERKYLKDSIWDL
-GVAQERKR VFY	---	IGF	---	RKDL	---	EI-KFSFPKGSTVEDK	---	DKITLKDVIWDL
-GVAQDRKR VFY	---	IGF	---	RKEL	---	NI-NYLPPIPHLI	---	KPTFKDVIWDL
-GIPQQRDR LVL	---	FAA	---	KQGV	---	KIIPPTH	---	TPENYRTVRDVI GSL
-GVPSRQR VFF	---	IGL	---	KSDR	---	PLNQILTPPSK	---	VIES--EYTSLEEAI SDL
-GVPQLRKRTFV	---	IGH	---	RHGS	---	ADLANVLQ	---	QRLAKQSLTVRDY
-GVPSRTRFSL	---	IGK	---	LNSEH	---	NFLIPTLS	---	RKLSDKPMTVRDY
-GVPQRHRRII	---	VGI	---	RKDQ	---	DV-AFRVPEP	---	THKE--KYRTASEALADI
-GAHHQRHRWFC	---	LAI	---	RKDYEP E E I	---	VSVNATKFDWENNE	---	PPCQVDNKS YENSTLVRL
GNIPQNRERIYI	---	VG	---	RNIE	---	HYKNFNFPMP	---	QPLTLTIKDM
-NVPQNRERLYI	---	IGI	---	REDLIKNE	---	EWSLDFKRKDI LQKGKQRLVELDIKSFNFRWT	---	AQSAATKRLKDL
-GIPQNRERVFC	---	ISI	---	LNPN	---	DFTFPQK	---	QNLTLSMNDL
-GSSQARRRVFM	---	IST	---		---	LNEFVELPKGD	---	KKPKSIKKV
-GIPQNRERIYL	---	VG	---	LNHDV	---	DFRFPQP	---	IGQATAVGD I
-GLPQNRERIYI	---	VG	---	DRKSI	---	NYSDFQMPTP	---	LQEKTRVGN I
-GVPQNRERIYI	---	VG	---	NKEKVR	---	NHEHFTFPTP	---	LKTKTRVGD I
-GVPQNRERIYI	---	VG	---	HKST	---	GVNSFSYPE	---	PLDKIVTFADI
-QVPQNRRLRVYI	---	VGLD	---	QSQPELTITSHI	---	GATDS	---	HKFKQ--LSNQASL--FD--TNKIMLVRI
-GVPQNRVRIYI	---	LGIL	---	GSKPKLTLTSNV	---	GAADS	---	HKYK--NEQISL--FD--ESYATVKDI
-GIPQKRKR FYL	---	VA	---	LNQNI	---	HFEFPKP	---	PMISKDIGEV
-GLPQRERIRVI	---	VG	---	HPDL	---	GINDFSFPK	---	GNPDNKVP INAI
-GIPQKRERIYM	---	ICF	---	RNDL	---	NIQNFQFPKP	---	FELNTFVKDL
-GNAQRRRRVFI	---	FGY	---	KQDLNYSKAMEESPLD	---	KI IYHN	---	GLFAEAFPIEDYANKNRVNRTHITHDIVDI
-GTPQRKRRAIIRLNKKGT	---		---	IWNL	---	PLKQNI	---	SVEQAIGNL
-GTPQVRERVFI	---	TATLVPERMRDERIPRTET	---	GEIDA	---	EAIGPK	---	PVATMNDRFPIKK--GGTELFHPGDRKSGWNL
-GVAQNRDRVFI	---	IGI	---	QKQL	---	GVPDFSFP EYSESEQRLYDILDNLQTPSII	---	
-GVAQNRDRVFI	---	VGI	---	QKQL	---	DLNGFSFP EYAESDQRLYHILDNLEAPETK	---	
-GVAQNRDRVFI	---	VGI	---	QKQL	---	DLNGFSFP EYTESEQRLYHILDNLEVPETK	---	
-DVAQKRERIRVI	---	IGI	---	REDLVKE	---	QKYPFRFPLA	---	QVYKPV LKDV
-GVSQLRPRVLF	---	VAL	---	KNEY	---	TNFFKWPEP	---	NSEQPKTVGELLFDL

Консервативное значит важное

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



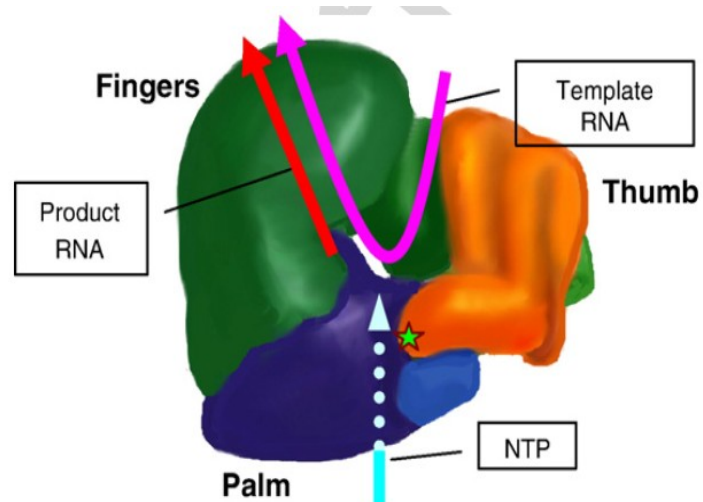
Сохраняющееся в эволюции
(консервативное) важно

Выравнивание РНК зависима РНК полимераз (RdRP)

```

*      320      *      340      *      360      *      380      *      400      *      420      *      440      *      460
'KTMIRFGDVGGLDIFSAFASLSPFMIREA...GRIMSELS...GTPSHFGTALINTIISKHLLYNCCY...HVCGSMPSGSPCTALINSIINNVLYYVFSKIFGKSPVFF.....CQALKILC.YGDDVLIIVFSRDV
SVAMQG.FERVYDVLYSNEDSTHSVAMFRLI...A...EEFF.TPENGFDPLTREYLESIAISTHAFEEKRF.....LITGGLPSGCAATSMINTIMNIIIRAGLYLYTKNFEEDD.....VKVLS.YGDDLVATNYQL
STHFAQ.YKNVWDVDYSAEDANHCSTAMNIMFEEVTEFG.....FHPNAEWILKTLVNTTEHAYENKRI.....VVEGGMPSGCSATSIINTILNNTIYVLYALRRHYEGVELDT.....YTMIS.YGDDIVASDYDL
.....WSLCVATIVSDHDTFWPGWLRDLICDELINMGYA.PWVVKLFETSLKLPVYVGAFAPEQGHITLLGDPSNPDLEVGLSSGQGATDLMGTILMSITILVMQLDHTAPHLNSRIKMPKSACRFLDSYWQGHEEIROIS.KSDAILGWTKGR
RRRIPE.NWVYCDAGSQEDSSLTPYLINAV...LTIRSTYMEDWDVGLQMLRNLYTEIVYTPISTPDGTIV.....KKFRGNNSGQSTVVDNSIMVVIAMHYALIKECVEFEEID.....STCVFFV.NGDDLIIVAVNPEK
IDKLNRPGLWHGSGDGRDSSIDPFFFDVV..KTRKHFL..PSEHHKAIDLIYDEILNTTICLANGMVI.....KKNVGTQR.QPSTVVDNTLVIMTAFLYAYIHKTGDRELAL.....LNERFIFVC.NGDDNKFAISPQF
VLSIASFSPYGFNGDFANFDGMFHSSFSMV..SEIANIFY...GNFLSTERDNLRLTNRFSLMKGAIL.....RVPGGSPSGFEMTVIENSFINLFYIQAWIMLARFNGRQDISH.....PCNFPKYVRACV.YGDDNTVAIKMEV
VARMKEKGNDVLCDDYSEDGLLSKQVMDVI..ASVINELC.GGEDQLKNARRNLLMACCSRLAICKNTVW...RVECGIPSGFEMTVIENSFINLFYIQAWIMLARFNGRQDISH.....QSFDKLIGLVT.YGDDNTISVNAVY
AEHAK.YKNHFDADYIANDSTQNRQIMTES..FSIMSRLT...ASPELAEVVAQDLAPSEMDVG DYVI.....RVKEGLPSGFPCTSQVNSINHWTITLCAISEATGLSPDVV.....QSMSYFSFYGDDEIVSTDIDF
NNLTSKASDFLCIDYSKNDSTMSPCVRIA..IDLADCC...EQTELTKSVVLTIKSHFMTILEAMIV.....QTKRGLPSGMPHTSVINSICHWLWSAAVYKSCAEIGLHCS.....NLYELAPFYT.YGDDGVYAMTFMM
IQRIKS.AAKVYAVDYSKNDSTQSPRVSAAAS..IDLRYFS...DRSPIVDSAANTLKSPPIAIFNGVAV.....KVSSGLPSGMPHTSVINSINHCLYVGCAILQSLEARGVPVTW.....NLFSTFCMMT.YGDDGVYMFMMF
'KRLEKPKHDRVCDYSKNDSTQPPKVTQS..IDLRHET...DKSPIVDSACATLKSNPIGIFNGVAF.....KVAGGLPSGMPHTSVINSINHCLYVGCAILQSLEARGVPVTW.....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 : FQVPFIVNDDVELALSIQADGIHVGQKDTAVETILRNTRNKPIIIGLSINTLAQALANKDRQDIDYFGVGPIFPTNSKADH : 15
THIE_STRA3 : YQVPFIIDDDIDLVELIDADGLHIGQNDLPVDEARRRLPDKI-IGLSVSTMAEYQKSQ-LSVVDYIGIGBFNPTQSKADA : 14
THIE_LISIN : YQVPFIINDDDVALALEIGADGIHVGQNDDEIRQVIASCAGKMKIIGLSVHSVSEAEAEERLGSVDYIGVGPIFPTISKADA : 14
THIE_ANOFW : YNIPFIVNDDVDLALALQADGVHVGQDEDEVAERVRDRIGDKY-LGVSVHNLNEVKKAL-AACADYVGLGPIFPTVSKEDA : 14
THIE_GEOTN : YGVFPFIVNDDVELAIAIDADGVHVGQDDEARRVREKIGDKI-LGVSAHNVVEEARAAT-EAGADYIGVGPIYPTRSKDDA : 14
THIE_BACSU : AGVPFIVNDDVELALNLKADGIHIGQEDANAKEVRAAIGDMI-LGVSHTMSEVKQAE-EDGADYVGLGPIYPTETKKDT : 14
THIE_BACA2 : AGIPFIINDDDVELALRLBADGVHIGQDDADAETRAAIGDMI-LGVSAHNVSEVKRAE-AAGADYVGMGPVYPTETKKDA : 14
THIE_OCEIH : FQIPFIINDDDVDLAKQLDADGIHIGQDDQPVEVVRKQFPNKI-IGLSISTNNELNQSP-LDLVDYIGVGPIEDTNTKEDA : 14
THIE_STAAB : YNVPFIVNDDVSLAKEINADGIHVGQDDAKVKEIAQYFTDKI-IGLSISDLGEYAKSD-LTHVDYIGVGPIYPTPSKHDA : 14
THIE_STACT : YNVPFIVNDDVALAEEDADGIHVGQDDEAVDDFNRRFEGKI-IGLSIGNLEEELNASD-LTYVDYIGVGPIFATPSKDDA : 14
      6pFI6LDD6 La 6 ADG6H6GQ D      6G6S      2      DY G6GP pT 3K Da

      *      180      *      200      *      220      *      240
THIE_LACLS : AKHIGIKDLR-LMLLENQLPIVIGGGITQDSLTELSAICLDGLAVISLLTEAENPKKVAQMIRQKITKNG~~~~~ : 21
THIE_MANSM : SPIVGMNFIRQIRQLGIDKPCVAIGGIKEESAAILRRLGADGVAVISAISHSVNIANTVKTLAQK~~~~~ : 22
THIE_STRA3 : KPAVGNRTTKAVREINQDIPVVAIGGITSDFVHDIIESGADGIAVISAIKANHIVDATRQLRYEVEKALVNRQKRSDVI : 22
THIE_LISIN : EPVSGTAILEEIRRAGIKDPIVIGGGINETNSAEVLTAGADGVSVISAITRSEDQCSVIKQLKNPGSPS~~~~~ : 21
THIE_ANOFW : KQACGLTMIEHIRAHEKRVPLVAIGGITEQTAKQVIEAGADGIAVISAIKRAEHIYEQTKRLYEMVMRAKQKGDR~~~~ : 21
THIE_GEOTN : NEAOGPGILRHLREQGITIPVVAIGGITADNTRAVIEAGADGVSVISAIASAPEPKAAAAALATAVREANL---R~~~~ : 22
THIE_BACSU : RAVQGVSLIEAVRRQGISIPVIGGGITIDNAAPVIAQAGADGVSMISAISQAEDPESAARKFREEIQTYKTG--R~~~~ : 22
THIE_BACA2 : EAVQGVTLIEEVRQGITIPVIGGGITADNAAPVIEAGADGVSMISAISQAEDPKAAARKFSEEIRRSKAGLSR~~~~ : 22
THIE_OCEIH : KTAVGLEWISLKKQHPSLPIVVAIGGINTTNAQEIIQAGADGVSFISAITETDHLQAVQRL~~~~~ : 20
THIE_STAAB : HTPVGPEMIATFKEMNPQLPIVVAIGGINTSNVAPIVEAGANGISVISAIKSENIEKTVNRFKDFFN~~~~~ : 21
THIE_STACT : SEPVGPKMIETLRKEVGDLPIVVAIGGISLDNVQEVAKTSADGVSVISAIARSPhVTETVHKFLQYFK~~~~~ : 21
      G      P V IGGI      6 galG6 ISa6

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

      160      *      180      *      200      *      220
THIE_LACLS : AAKHIG---IKDLRLMLLENQLPIVIGGGITQDSLTELSAICLDGLAVISLLTEAENPKKVAQMIRQKITKNG : 218
THIE_MANSM : HSPVGMNFIRQIRQLGIDK--PCVAIGGIKEESAAILRRLGADGVAVISAISHSVNIANTVKTLAQK----- : 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

Демонстрация

Полезные умения в Jalview

Действие	меню	подменю	варианты	комментарии
Импорт последовательностей	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 прямоугольн ый 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	LAYDVFGT	PRPDEYFTQTRQEL	PIINDRYEAKNQIEQFN
K9SST0_9SYNE/42-80	LAYDIFGT	PRPNEYTNDRQEAP	PILRDYDANQEI PNYN
K9SGD2_9CYAN/42-80	LAYDVFGT	PRPNEYTLD RSDAP	VLLDRFNAIDDI EDFN
K9Y0S0_STAC7/42-80	LAYDVFGT	PRPNEYFTQDRLEL	PIINDRYSASEQIKEFN
PSBE_RIPO1/42-80	LAYDVFGT	PRPDQYFTQERQEL	PIISDRYKASQQIQEFN
PSBE_SYNY3/42-80	LAYDAFGT	PRPDEYFTQTRQEL	PILQERYDINQEQEFN
PSBE_GLOC7/42-80	LAYDVFGT	PRPDEYFTQERLEL	PILKDRYNTDQQIQEFN
K9W859_9CYAN/42-80	LAYDVFGT	PRPNEYFTQDRVEAP	PILSDRYEAKQQIQQFL
PSBE_SYNE7/43-81	LAYDAFGT	PRPNEYFTQDRTEV	PIVSDRYSAKQQVDRFS
PSBE_ACAM1/42-80	LAYDVFGT	PRPNEYFDQARQGL	PLVTD RYEGKQQIDEFT
C7B2G7_SELML/42-80	LAYDVSGS	PRPNEYSTESRRDI	PLITGRFDSSEQIDEFT
PSBE_PHATG/43-81	LAYDVFGT	PRPNEYFTQDRQQI	PLVND RFSAKQELEDLT
PSBE_OSTTA/42-79	LAYDVFGT	PRPNEYFTTEERQEL	PLISDRYGALAQLDDL
PSBE_CHLRE/42-79	LAYDVFGT	PRPNEYFTEDRQEAP	PLITDRFNALEQVKKL
PSBE_SYNJB/42-80	LAYDIFGT	PRPNEYFTAERQEL	PIVSD RFNAL EELERLT
K9TLB2_9CYAN/43-80	LAYDAFGT	PRPNEYFSQERLEL	PIVQDRFNAKDQIDDL
K9S2L1_9CYAN/42-80	LAYDAFGT	PRPDEYFTQQRQEL	PIVMD RFGAKQQVEEFI
K9YE64_HALP7/42-80	LAYDAFGT	PRPDEYFTKQRQEL	PIISDRFQAREEVEEFN
K9R7L0_9CYAN/42-80	LAYDAFGT	PRPDYFTQERQEV	PLVND RFEAKNQVDVFT
PSBE_THEVB/42-80	LAYDVFGT	PRPDSYQAQEQRSI	PLVTD RFEAKQQVETFL
PSBE_GLOW/43-81	LAYDVFGT	PRPDEYNSDNTKK	PLVNKRFEAKQQLD EAT
PSBE_SYNR3/43-81	LAYDAFGT	PRPDSYFQAADV KAP	PVVSQR YEGKSQLDTRL
PSBE_PROMA/43-81	LAYDAFGT	PRPDYFQASETKAP	PVVSQR FESKAQLDLRL

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 E 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_GLOWV/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	LAYD	VFGT	PRPDEY	FTQTRQEL	PIINDRYEAKNQIEQFN
K9SST0_9SYNE/42-80	LAYD	IFGT	PRPNEY	YTNDRQEAP	ILRDRYDANQEIPNYN
K9SGD2_9CYAN/42-80	LAYD	VFGT	PRPNEY	YTLDRSDAP	VLLDRFNAIDDIEDFN
K9Y0S0_STAC7/42-80	LAYD	VFGT	PRPNEY	FTQDRLEL	PIINDRYASASEQIKEFN
PSBE_RIPO1/42-80	LAYD	VFGT	PRPDQY	FTQERQEL	PIISDRYKASQQIQEFN
PSBE_SYNY3/42-80	LAYD	AFGT	PRPDEY	FTQTRQEL	PI LQERYDINQEQEFN
PSBE_GLOC7/42-80	LAYD	VFGT	PRPDEY	YTQERLEL	PI LKDRYNTDQQIQEFN
K9W859_9CYAN/42-80	LAYD	VFGT	PRPNEY	FTQDRVEAP	ILSDRYEAKQQIQQFL
PSBE_SYNE7/43-81	LAYD	AFGT	PRPNEY	FTQDRTEVP	IVSDRYSAKQQVDRFS
PSBE_ACAM1/42-80	LAYD	VFGT	PRPNEY	FDQARQGL	PLVTDRYEGKQQIDEFT
C7B2G7_SELML/42-80	LAYD	VSGS	PRPNEY	STESRRDI	PLITGRFDSSEQIDEFT
PSBE_PHATC/43-81	LAYD	VFGT	PRPNEY	FTQDRQQI	PLVNDRFSAKQEELEDLT
PSBE_OSTTA/42-79	LAYD	VFGT	PRPNEY	FTTEERQEL	PLISDRYGALAQLDDL
PSBE_CHLRE/42-79	LAYD	VFGT	PRPNEY	FTEDRQEAP	PLITDRFNALEQVKKL
PSBE_SYNJB/42-80	LAYD	IFGT	PRPNEY	FTAERQEL	PIVSDRFNALEELERLT
K9TLB2_9CYAN/43-80	LAYD	AFGT	PRPNEY	FSQERLEL	PIVQDRFNAKDQIDDL
K9S2L1_9CYAN/42-80	LAYD	AFGT	PRPDEY	FTQQRQEL	PIVMDRF GAKQQVEEF I
K9YE64_HALP7/42-80	LAYD	AFGT	PRPDEY	FTKQRQEL	PIISDRFQAREEVVEEFN
K9R7L0_9CYAN/42-80	LAYD	AFGT	PRPDY	FTQERQEV	PLVNDRF EAKNQVDVFT
PSBE_THEVB/42-80	LAYD	VFGT	PRPDSY	YAQEQRSI	PLVTDRF EAKQQVETFL
PSBE_GLOWV/43-81	LAYD	VFGT	PRPDEY	YNSDNTKK	PLVNKRF EAKQQQLDEAT
PSBE_SYNR3/43-81	LAYD	AFGT	PRPDSY	FQAADVKA	PVVSQRYEGKSQQLDTRL
PSBE_PROMA/43-81	LAYD	AFGT	PRPDY	FQASETKA	PVVSQRFESKAQLDLRL
K9RTX4_SYNP3/42-80	LAYD	IFGT	PRPDY	FSQEQQAAP	PVVSDRFESKQQINEFL
K9ULI3_CHAP6/43-81	LAYD	AFGT	PRPDEY	FTQTRQEL	PIVSDRFESKGQLDKFV
K9UY34_9CYAN/42-79	LAYD	AFGT	PRPNEY	FTQTRQEV	PIVNNRFDAKKQVETY
PSBE_NOSS1/42-80	LAYD	VFGT	PRPDEY	YTQARQEL	PIVNNRF EAKKQVEQL I
K9QJT8_9NOSQ/42-80	LAYD	VFGT	PRPNEY	YTQTRQEV	PIVKNRYEAKQQVEKFI
K8GLK0_9CYAN/42-80	LAYD	VFGT	PRPNEY	YTSTRQQA	PIVKDRFEAKQQIDQFI
K9VJB4_9CYAN/42-80	LAYD	AFGT	PRPNQY	FTPQREEVP	PIISDRFEAKKQVDQFI
K9W2B5_9CYAN/42-80	LAYD	AFGT	PRPNEY	FTQQRQEV	PIVSDRFESKQEI DQFL
K9TZN4_CHRTP/42-80	LAYD	VFGT	PRPNEY	YPQERQEL	PIVSDRFESRKQIEEFI
K9XG15_9CHRO/42-80	LAYD	VFGT	PRPNEY	YPQERQEL	PIIGDRYQAKQEIKQFI

Пример достоверного блока

		10	20	30
K9Z0V0_CYAAP/42-80	LAYDVFG	GTPRPDEY	FTQTRQEL	PIINDRYEAKNQIEQFN
K9SST0_9SYNE/42-80	LAYDIFG	TTPRPNEY	YTNDRQEAP	ILRDRYDANQEIPNYN
K9SGD2_9CYAN/42-80	LAYDVFG	GTPRPNEY	YTLD RSDAPV	LDRFNAIDDIEDFN
K9Y0S0_STAC7/42-80	LAYDVFG	GTPRPNEY	FTQDRLEL	PIINDRYSASEQIKEFN
PSBE_RIPO1/42-80	LAYDVFG	GTPRPDQY	FTQERQEL	PIISDRYKASQQIQEFN
PSBE_SYNY3/42-80	LAYDAFG	GTPRPDEY	FTQTRQEL	PI_LQERYDINQEIQEFN
PSBE_GLOC7/42-80	LAYDVFG	GTPRPDEY	YTQERLEL	PI_LKDRYNTDQQIQEFN
K9W859_9CYAN/42-80	LAYDVFG	GTPRPNEY	FTQDRVEAP	ILSDRYEAKQQIQQFL
PSBE_SYNE7/43-81	LAYDAFG	GTPRPNEY	FTQDRTEVP	IVSDRYSAKQQVDRFS
PSBE_ACAM1/42-80	LAYDVFG	GTPRPNEY	FDQARQGL	PLVTD RYEGKQQIDEFT
C7B2G7_SELML/42-80	LAYDVSG	SPRPNEY	STESRRD	IPLITGRFDSSEQIDEFT
PSBE_PHATC/43-81	LAYDVFG	GTPRPNEY	FTQDRQQI	PLVND RFS AKQELEDLT
PSBE_OSTTA/42-79	LAYDVFG	GTPRPNEY	FTTEERQEL	PLISDRYGALAQLDDL
PSBE_CHLRE/42-79	LAYDVFG	GTPRPNEY	FTEDRQEAP	PLITDRFNALQVKKL
PSBE_SYNJB/42-80	LAYDIFG	TTPRPNEY	FTAERQEL	PIVSD RFNALLEELERLT
K9TLB2_9CYAN/43-80	LAYDAFG	GTPRPNEY	FSQERLEL	PIVQDRFNAKDQIDDL
K9S2L1_9CYAN/42-80	LAYDAFG	GTPRPDEY	FTQQRQEL	PIVMDRFGAKQQVEEFI
K9YE64_HALP7/42-80	LAYDAFG	GTPRPDEY	FTKQRQEL	PIISDRRFQAREEEVEFN
K9R7L0_9CYAN/42-80	LAYDAFG	GTPRPDTY	FTQERQEV	PLVND RFEAKNQVDVFT
PSBE_THEVB/42-80	LAYDVFG	GTPRPDSY	YAEQQR	SIPLVTD RFEAKQQVETFL
PSBE_GLOW/43-81	LAYDVFG	GTPRPDEY	YNSDNTKK	PLVNR RFEAKQQLD EAT
PSBE_SYNR3/43-81	LAYDAFG	GTPRPDSY	FQAADVKA	PVVSR RYEGKSQLDTRL
PSBE_PROMA/43-81	LAYDAFG	GTPRPDTY	FQASETKA	PVVSR RFESKAQLDLRL
K9RTX4_SYNP3/42-80	LAYDIFG	TTPRPDTY	FSQEQQAA	PVVSD RFESKQQINEFL
K9ULI3_CHAP6/43-81	LAYDAFG	GTPRPDEY	FTQTRQEL	PIVSD RFESKGQLDKFV
K9UY34_9CYAN/42-79	LAYDAFG	GTPRPNEY	FTQTRQEV	PIVNNRFD AKKQVETY
PSBE_NOSS1/42-80	LAYDVFG	GTPRPDEY	YTQARQEL	PIVNNRFEAKKQVEQLI
K9QJT8_9NOSQ/42-80	LAYDVFG	GTPRPNEY	YTQTRQEV	PIVKNRYEAKQQVEKFI
K8GLK0_9CYAN/42-80	LAYDVFG	GTPRPNEY	YTSTRQQA	PIVKDRFEAKQQIDQFI
K9VJB4_9CYAN/42-80	LAYDAFG	GTPRPNQY	FTPQREEV	PIISDRFEAKKQVDQFI
K9W2B5_9CYAN/42-80	LAYDAFG	GTPRPNEY	FTQQRQEV	PIVSD RFESKQEIDQFL
K9TZN4_CHRTP/42-80	LAYDVFG	GTPRPNEY	YYPQERQEL	PIVSD RFESRKQIEEFI
K9XG15_9CHRO/42-80	LAYDVFG	GTPRPNEY	YYPQERQEL	PIIGDRYQAKQEIKQFI

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	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_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	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?

А если бы тут было много неконсервативных колонок без гэпов?



	10	20	30
K9Z0V0_CYAAP/42-80	LAYDVFGT	PRPDEYFTQT	RQELPIINDRYEAKNQIEQFN
K9SST0_9SYNE/42-80	LAYDIFGT	PRPNEYYTND	RQEAPILRDRIYDANQEIPNYN
K9SGD2_9CYAN/42-80	LAYDVFGT	PRPNEYYTLD	RSDAPVLLDRFNAIDDIEDFN
K9Y0S0_STAC7/42-80	LAYDVFGT	PRPNEYFTQD	RLELPIINDRYSASEQIKEFN
PSBE_RIPO1/42-80	LAYDVFGT	PRPDQYFTQE	RQELPIISDRYKASQQIQEFN
PSBE_SYNY3/42-80	LAYDAFGT	PRPDEYFTQT	RQELPILQERYDINQEQEFN
PSBE_GLOC7/42-80	LAYDVFGT	PRPDEYFTQE	RLELPIILKDRYNTDQQIQEFN
K9W859_9CYAN/42-80	LAYDVFGT	PRPNEYFTQD	RVEAPILSDRYEAKQQIQQFL
PSBE_SYNE7/43-81	LAYDAFGT	PRPNEYFTQD	RTEVPIVSDRYSAKQQVDRFS
PSBE_ACAM1/42-80	LAYDVFGT	PRPNEYFDQA	RQGLPLVTDRIYEGKQQIDEFT
G7B2G7_SELML/42-80	LAYDVSGS	PRPNEYSTES	RRIPLITGRFDSSEQIDEFT
PSBE_PHATC/43-81	LAYDVFGT	PRPNEYFTQD	RQQIPLVNDRIYSAKQLEDLT
PSBE_OSTTA/42-79	LAYDVFGT	PRPNEYFTEER	RQELPLISDRYGALAQLDDL
PSBE_CHLRE/42-79	LAYDVFGT	PRPNEYFTED	RQEAPLITDRFNALEQVKKL
PSBE_SYNJB/42-80	LAYDIFGT	PRPNEYFTAER	RQELPIVSDRIYSAKQVEEFN
K9TLB2_9CYAN/43-80	LAYDAFGT	PRPNEYFSQER	RLELPIVQDRFNAKDQIDDL
K9S2L1_9CYAN/42-80	LAYDAFGT	PRPDEYFTQQ	RQELPIVMDRIYSAKQVEEFN
K9YE64_HALP7/42-80	LAYDAFGT	PRPDEYFTKQ	RQELPIISDRYSAKQVEEFN
K9R7L0_9CYAN/42-80	LAYDAFGT	PRPDTYFTQER	RQEVPLVNDRIYSAKQVDFVT
PSBE_THEVB/42-80	LAYDVFGT	PRPDSYYAQEQ	RSIPLVTDRIYSAKQVETFL
PSBE_GLOV/43-81	LAYDVFGT	PRPDEYYNSDNT	KKPLVNRRIYSAKQQLDEAT
PSBE_SYNR3/43-81	LAYDAFGT	PRPDSYFQAADV	KAPVVSQRIYEGKSQDLTRL
PSBE_PROMA/43-81	LAYDAFGT	PRPDTYFQASE	TKAPVVSQRFESKAQLDLRL
K9RTX4_SYNP3/42-80	LAYDIFGT	PRPDTYFSQEQQA	APVVSDFESKQQINEFL
K9ULI3_CHAP6/43-81	LAYDAFGT	PRPDEYFTQTR	RQELPIVSDRIYSAKQQLDKFV
K9UY34_9CYAN/42-79	LAYDAFGT	PRPNEYFTQTR	RQEVPIVNNRIYDAKKQVETY
PSBE_NOSS1/42-80	LAYDVFGT	PRPDEYFTQAR	RQELPIVNNRIYSAKQVEQLI
K9QJT8_9NOSO/42-80	LAYDVFGT	PRPNEYYTQTR	RQEVPIVKNRIYSAKQVEKFI
K8GLK0_9CYAN/42-80	LAYDVFGT	PRPNEYYTSTR	QQAPIVKDRFIYSAKQVIDQFI
K9VJB4_9CYAN/42-80	LAYDAFGT	PRPNQYFTPQRE	EVPIISDRFIYSAKQVDDQFI
K9W2B5_9CYAN/42-80	LAYDAFGT	PRPNEYFTQQR	RQEVPIVSDRIYSAKQVIDQFL
K9TZN4_CHRTP/42-80	LAYDVFGT	PRPNEYYPQER	RQELPIVSDRIYSAKQVEEFN
K9XG15_9CHRO/42-80	LAYDVFGT	PRPNEYYPQER	RQELPIIGDRYQAKQEI KQFI

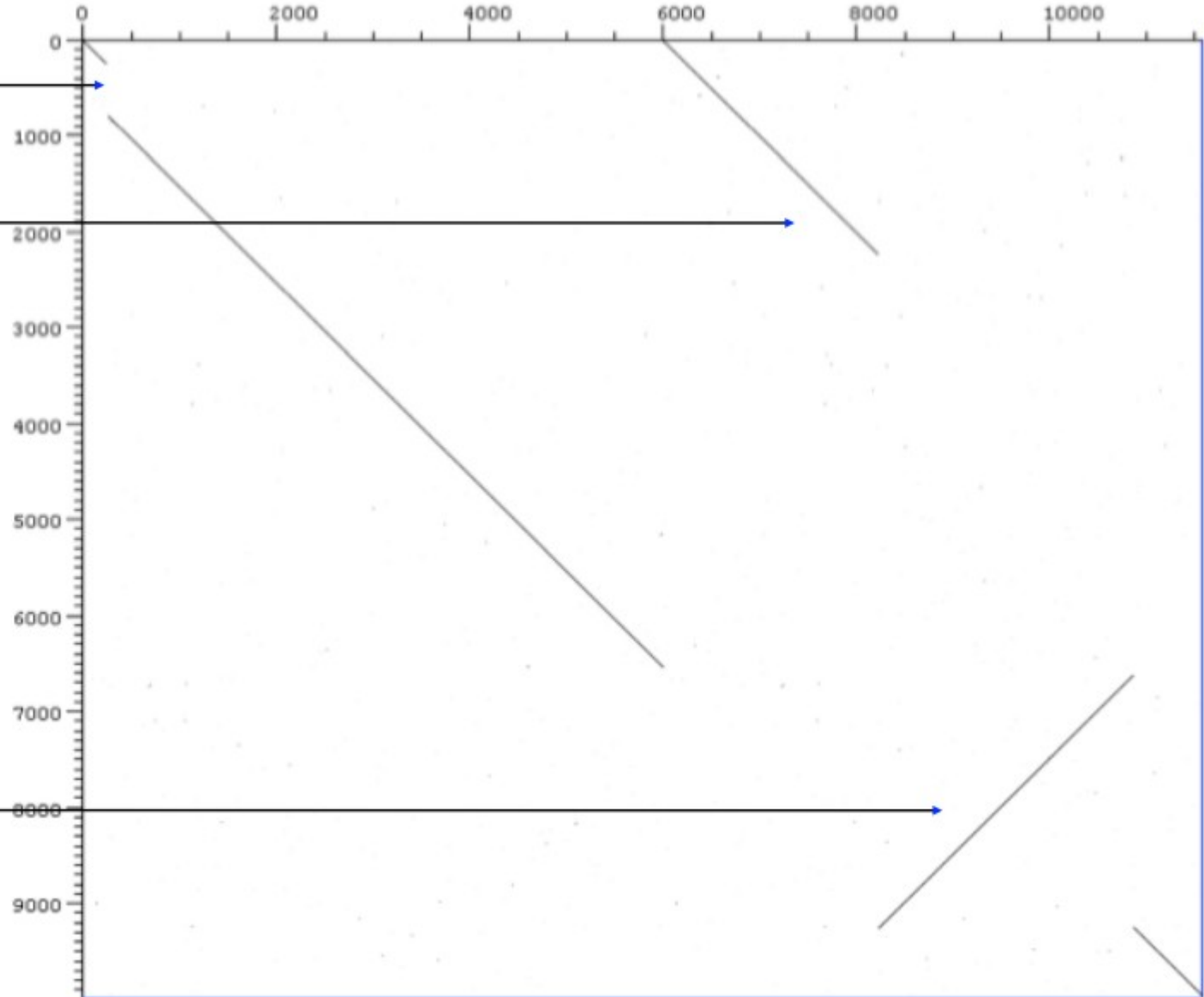


Как- то
так

deletion

duplication

inversion



Выравнивание как матрица

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

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

Выравнивание как матрица

Seq 1

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: || |||||

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---