



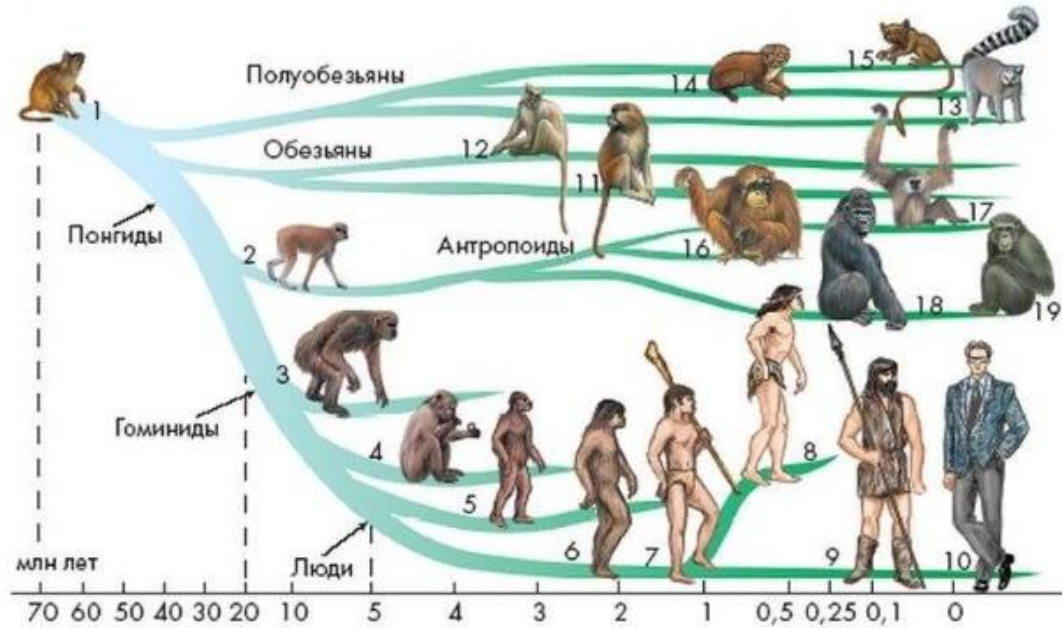
Гомология и множественное выравнивание Домены белков. Pfam.

Алешина Юлия Александровна
Спирин Сергей Александрович

21 апреля 2026

По материалам Алексеевского А.В.


Общность происхождения



Последний общий предок ныне живущих обезьян.

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

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

Гомологию последовательностей нуклеотидов и белков выводят из сходства последовательностей

ДНК, РНК и белки, как молекулы, определяются (почти однозначно) своей последовательностью

Схожесть последовательностей — признак гомологии
(для белков существует более надёжный признак: сходство 3D-структур)

❖ ~~“Высокая гомология последовательностей”~~

У гомологии нет степени, она есть у сходства!

❖ “Высокое сходство последовательностей” — **ДА**

Множественное выравнивание

```

                *           100           *
THIE_LACLS : AGVSEFIVNDDVELARELNADGIHIGQTD E SVSKVI
THIE_MANSM : FQVPFIVNDDVELALSIQADGIHVGQKDTAVETI I
THIE_STRA3 : YQVPFIIDDDIDLVELIDADGLHIGQN DLPVDEAI
THIE_LISIN : YQVPFIINDDVALALEIGADGIHVGQN DDEEIRQV
THIE_ANOFW : YNIPFIVNDDVDLALALQADGVHVGQ EDEVAERVI
THIE_GEOTN : YGVPFIVNDDVELAIAIDADGVHVGQ DDEDARRVI
THIE_BACSU : AGVPFIVNDDVELALNLKADGIHIGQ E DANAKEVI
THIE_BACA2 : AGIPFIINDDVELALRLEADGVHIGQ D DADAETI
THIE_OCEIH : FQIPFIINDDVDLAKQLDADGIHIGQ D DQPVEVVI
THIE_STAAB : YNVPFIVNDDVSLAKEINADGIHVGQ D DAKVKEI I
THIE_STACT : YNVPFIVNDDVALAEEIDADGIHVGQ D DDEAVDDFI
                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.о. гомологичных белков потомков произошли из одного кодона последнего общего предка этих белков, другие 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. Бывают 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

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 QR I KVKQQAQVYSVGES - - RTDF A I DVFNNF FRTNPD - RSLFN RVNGDNVYSPEF 59
9 QRLKVKQQAQVYSVGHE - - RVELG I ALWKS MFAQDNDARDLFKRVHGEDVHSPAF 62
8 EGLKVKSEWGRAYGSGHD - - REAF SQA IWRATFAQVPESRSLFKRVHGD DTSHPAF 61
6 QRFKVKHQWAEAFGTSHH - - RLDFGLKLWNS IFRDAPEIRGLFKRVDGDNAYS AEF 59
7 QRLKVKRQWAEAYGSGND - - REEFGHF I WTHVF KDAPSARDLFKRVRGDNIHTPAF 60

67 KSHLVRVANGLDLLINLLDDTLVLQSHLGHLADQHIQRKGVTKKEYFRGIGEAFA 120
64 KAHVIRVINGLDLAVNLLDPKALQEELKHLARQHRERSGVKAVYFDEMEKALL 117
64 FAHVLRVVNGLDTLIGVLGDSDTLNSLIDHLAEQHKARAGFKTVYFKEFGKALN 117
60 KAHMVRVFAFDILISVLDDKPVLDQALAHYAAFHKQF-GTIP--FKAFGQTMF 110
63 EAHMARVFNGLD RVISSLTDEPVLNAQLEHLRQQHIKL-GITGHMFNLMRTGLA 115
62 I AHAERVLGGLDIAISTLDQPATLKEELDHLQVQHEGR-KIPDNYFDAFKTAIL 114
60 E AHAERVLGGLDMTISLLDDQAAFDAQLAHLKSQHAER-NIKADYYGVFVNELL 112
61 RAHATRVLGGLDMCIALLDDEGVLNTQLAHLASQHSSR-GVSAAQYDVVEHSVM 113
```

Пример плохого выравнивания

```
GSYLEVEGPKGKTTVYVTDLYPEGARGA-----L-DL---SPNAFRKIGNMKDGKINI-----KWRV
FSRYKLYLDDPKQSYGSAEE-----I-HT---ITRKGFSKEQPNAAKLLS-----QFKW
VKDTRAASEEELLQKAHFLQRMLSYGTTTAEVKSGYGLEKETELKQLRVAKKL-HE---SQPVDLVSTFMGAHAIPP-----EYQN
-----EFTDELAKQYF-----HA---VPK|ARA|RDEFEP|GLN-----TLNN
VPEEWDI|RESMPAYNS-----NY---TPTRGMGQ|VELFRSYP-----EVKR
GG|PQ|SLHLANPAFQSGHFKTP|TNDQYK|LDSSTVEGKRLN-----AM---LSK|ADGLQELENQGVV-----LFRP
-----EKFEAKLEEYNG-----NL---KRAAVELAKEWRSQKVL-----KLEA
-----EE
YDPSTWGGKKEPSGTQEEARARSQKNQKARVMVD|PANTT|VGSGTNAKVVGGNF-QI---KSDNV|IRN|EFQDAYDY-----FPQW
F|FTNCET|CPPMTA-----HM---TDLQK|KAEN|DVR|IS-----FSVD
SALG|INTDFSPVVD|NNNPDPNV|GVRSFSSNRELTSRLGLYTMKGLQRQD|AS-AL---KHFPGHGDTD|VDSHYGLP-----LVSH
LSSDDPNSYHFYMNDR-----FF---QH|DSKPSRHF|PNGNAD-----DLEA
FSPKQ|RARFPHL-----HI---GRSVHSL|EEAVQAEKEDA-----DYVL
QNVKLLN|DADTLTD-----VF---EPGEV|KRVYI|NESDPWPK-----KRHE
I|EQAWEEGKQVC|PKC-----
FHLNKNEYTVPFVCGCRDLGEATRR-
I|VSGDSDQSPWVKKEMGRAAVACMK
-----S|IMGLMSLAVSTGTEV-----TLTA
QL|KRMSDEGHI|GNHSFHHP-----DL---TTKTADQ|IQDELDSVNEE-----VYKI
---EKFEELYDHAETVD-----TI---AERLLA|IGQP|VATVKEY-----TEHA
-----EK---YRGLV|LTSK|GK|GKRLV-----YRHE
FSTFSVQSYNKRKKEEAE-----KL---GGLT|ITSVESWYDEPKFV-----TYWV
SNLPEHTERYE|AQEHLDVVRGLWNSWEHDAF|HNKKTGQFFDQAKLHRLNHKGKYFQVEGPLN|GRSKQGEPPVVFQA-----GSSE
KGKSDFLSPYLEQL|EFNKRF|ITYWGYQEHYPDALLDLFEPGVTVKVLDQLFAEL---KEA|IPLV|KQV|TASGNKPDTSF|TKAFPKEKQKELSLYFLQE
QFMEPLREFAAQGKPMFGTCA-----GL---I|L|AKE|IAGSDNPHLGLL-----NVVV
-----S|IMGVMSLGI|AKGAEI-----TISA
CKGANEVLYPEHAQALRE-----QL---APLATV|IPNL|FEASQLS-----GMDE
IDVYFPDSE|SCLAQGFY-----AL---KAAEL|KNGASSPEDI|K-----ELEE
TYPLA|QE|EEG|VFEDD-----QF---I|VTAVSV|HGVEAFGYRV-----QEKD
TKDERYVKAARLRSLYGT|LNRT|SEGGFWHKDGYPYQMMLDGLYMGPPFALKYA-NL---KQETELFDQVV|LQESLMR-----KHTK
T|VDEKGRSF|IDVGEGLWYF|PSGLPHS|IQALEEGAEL|LVFDDGSF|SENSTF-QL---TDWLAHTPKEV|IAANFGV-----TKEE
RND|E|E|VKD|GELPST-----FV---PGRNLV|FLSFAS|LAYQI-----GARH
GELVPDEVT|G|V|KERL|GK-----DD---CERGFLLDGF|PRTVAQAE-----ALEE
ADG|H|IGQEDANAKEVRA-----AI---GDM|L|GVS|AHTMSEVKQA-----EE
EFAELLE|SEEERKAF-----KP---SPTAYS|YTS|HMYRSVLSG-----NFAE
INMTGDGAEDTFQ|IPVTDGQLNLLVTEGKAGTFTLSALK|IKKLSQPV|TNRTI-YV---GGDSTVCN|YPLN|SSKQA-----GWGQ
```

Программа выровняет всё, что вы ей дадите

DNA (cytosine-5-)-methyltransferases

| | 10 | 20 | 30 | 40 | 50 | 60 |
|-------------------|--------|----------|-----------|-----------|---------------|--------------------------|
| <i>EJL77459.1</i> | GVDLVF | GGPPCQGF | SQIGMRR- | LDDER-NE | LYQQYTRIVAKL | KPRVFLMENVPNLALMNKGH |
| <i>RXK67093.1</i> | DLDVVF | GGPPCQGY | SQIGTRR- | LDDER-NE | LYLQYARIVEKQR | PRMFLMENVPNMVL LNKGH |
| <i>OJY44288.1</i> | NVDLVF | GGPPCQGY | SQIGTRD- | LHDPR-NR | LFEEFARVVATL | KPKLFLMENVPNL LLLNKGH |
| <i>TRU90449.1</i> | NPEMIV | GSPPCQDF | SSAGKRNE | GLGR--ANL | TLTFAEIVTRV | SPQWFMENVD- - -RIEKSK |
| <i>OXI46696.1</i> | GTDLVF | GGPPCQGF | SQIGMRR- | LDDER-NE | LYKQYTRVVSTLR | PRVFLMENVPNLALMNKGH |
| <i>AVZ30243.1</i> | EIDVVF | GGPPCQGF | SLIGKRS- | FEDPR-NS | LVFHYIRLVLE | LSPKFFVIENVKGMTAGNHQA |
| <i>AFZ12381.1</i> | DIIGFI | GGAPCPDF | SVGGKNRG | SEGDK-GK | LSASYIELICQ | QKPDFFLFENVKGLYKTKKHR |
| <i>HCQ21462.1</i> | HIIGFI | GGPPCPDF | SVGGKNKG | HLDGN-GK | LSSAYIELICQ | NLPDFFLFENVKGLWRTTKHR |
| <i>EDN77159.1</i> | SLIGFI | GGPPCPDF | SVAGKNKG | KDGDN-GK | LSLSYTNLI | IEMKPDFFLFENVKGLWRTARHR |
| <i>SOD91684.1</i> | EVSLVV | GGAPCQPF | SNIGKKLG | KNDERNGD | LFLEFVRMVK | GIQPEAFIFENVVGITQNKHSD |
| <i>QCS48280.1</i> | NVVGFI | GGPPCPDF | SVGGKNRGR | QGDH-GK | LSESYIDLII | IQHQPDFFI FENVKGLYRTKKHR |
| <i>SMB95934.1</i> | GLFGFI | GGPPCPDF | SVGGKNRG | ENGEQ-GR | LSKVFVDKI | LDLQPVFFLYENVPGLIRTAKHR |
| <i>RUO38876.1</i> | SPVGF | GGPPCPDF | SVGGKNRG | HEGEN-GR | LTRTYVDGI | IKYAPDFFI FENVKGLWRTKRHR |
| <i>OIP70538.1</i> | TIDLIC | GGPPCQGF | STIGTND- | KKTHR-NF | LFEFFLRMVET | FKPNFIILENVTGLLAKKNES |
| <i>AFY60915.1</i> | NLVGFV | GGPPCPDF | SVGGKNKG | QYGDN-GK | LTKVYVDII | IENQPDFFVFENVKGLWRTRSRHR |
| <i>CUR30340.1</i> | DLIGFI | AGPPCPDF | SVGGKNRG | KNGDQ-GK | LTAcyVELIC | QQRPDFFVFENVKGLWSTKKHR |
| <i>TAK03971.1</i> | QAALVV | GGAPCQPF | SNLGSKRGT | ADSR-GT | LFQDFIRIVK | GVRPKGFI FENVEGLTQDKHKG |
| <i>AEE51071.1</i> | KVALVV | GGAPCQPF | SNIGKKEG | ENDAKNGD | LFLEFVRMVK | GIQPEAFIFENVAGIIQSKHSK |
| <i>RTR31666.1</i> | RLVGFV | GGPPCPDF | SVGGKNKG | SEGEN-GK | LTRTYIDLIV | KDNPDYFI FENVKGLWRTTRHR |
| <i>PTU64472.1</i> | NIDLVF | GGPPCQGF | SQIGTRR- | LDDER-NE | LYKQYTRIVKTL | KPRVFLMENVPNLAMMNKGH |

Продолжение того же выравнивания

```
      200      210      220      230      240      250      260      270      280
IYG-VPQDRKRVFIVGYREDLNLK-----FEFPKPLNKKVTLRD-----AIGDLPE-F
IYG-VAQDRERVFYVGFVKDLNLSN-----FE-FYPPISEKERKYLKD-----SIWDLKDN/
IYG-VAQERKRVFYIGFRKDLEIKF-----SFPKGSTVEDKDKITLKD-----VIWDLQDT/
IYG-VAQDRKRVFYIGFRKELNIN-----YLPPIPHLIKPTFKD-----VIWDLKDNF
IYG-IPQQRDRLVLFAAKQG-----VIKIIPPTHTPENYR-----TVRDVIGSLATN\
IYG-VPQSRQRVFFIGLKSDRPLNQQ-----ILT-----PSKVI ESEYTSLEEAISDLPVIE-----AGEGGEVQDYPVAF
CG-VPQLRKRTFVIGHRHGSIAD-----LANVLQQRLAKQSL-----TVRDYFG--
CG-VPQSRTRFSLIGKLNSEHNF-----LIPTLSRKLSDKPM-----TVRDYLG--
IYG-VPQRRHRIIIVGIRKDQD-----VAFRVPEPTHKEKYR-----TASEALADIPED/
IG-AHHQRHRWFCLAIRKDYEPPE-----IIVSVNATKFDWENNEPPCQVDNK-----SYENSTLVRLAGYS
IFGNI PQNRERIYIVGFRN-----IEHYKNFNFPMPQP-----LTLTIKDMINLS
IFN-VPQNRERLYIIGIREDLIKNEE-----WSLDFKRKDI LQKGKQLVELDIKSFNFRWTAQ-----SAATKRLKDLLLEE\
IFG-IPQNRERVFCISILN-----PNEDFTFPQKQ-----NLTL SMNDLLEEM
IFG-SSQARRRVFMISTLNEF-----VELPKGDKKPKS-----IKKVLNKIVSE.
IFG-IPQNRERIYLVGF-----LNHDVDFRFPQP-----IGQATAVGDI LEA.
IFG-LPQNRERIYIVGFDRKS-----ISNYSDFQMPTP-----LQEKTRVGNILES.
IFG-VPQNRERIYIVGFNKEK-----VRNHEHFTFPTP-----LKT KTRVGDILEK.
IFG-VPQNRERIYIVGFHKS-----TGVNSFSYPEP-----LDKIVTFADIREEF
IFQ-VPQNRRLRVYIVGLDQSQPELT-----ITSHIGATDSHKFKQLSNQASLFD-----TNKIMLVRDILED.
IFG-VPQNRVRIYILGILGSKPKLT-----LTSNVGAADSHKYK--NEQISLFD-----ES-YATVKDILED.
IFG-IPQKRKR FYLVAF LNQN-----IHFEFPKP-----PMISKDIGEVLES.
IFG-LPQRRERIVIVGFHPDLG-----INDFSFPKGN-----PDNKVPINAILEH.
IYG-IPQKRERIYMICFRNDLN-----IQNFQFPKP-----FELNTFVKDLLLPI
IYG-NAQRRRRVFI FGYKQDLNYSKAME-----ESPLDKI IYHNGLFAEAFP IEDYANKNR-----VNRTHITHDIVDISDNI
IYG-TPQRRKRAIIRLNKKGTIWN-----LPLKQNI VSVEQ-----AIGNLPSIESG\
IGG-TPQVRERVFITATLVPERMRDERIPRTETGEIDAEAIGPKPVATMNDRFP I KKGTEL FHPGDRKSGWNLLTSGI IREGDPEF
IYG-VAQNRDRVFIIGIQQKLGVPD-----FSFPEYSESEQRLYDILDNLQTPSII-----PESLPIQRNLFGEF
IFG-VAQNRDRVFI VGIQQKLDLNG-----FSFPEYAESDQRLYHILDNLEAPETK-----LESIP IQRNLFGEF
IYD-VAQKRERIVIIGIREDLVK-----EQYPPFRFPLAQ-----VYKPVLKDV LKDY
IYG-VSQLRPRVLFVALKNEYTN-----FFKWPEPNSEQPK-----TVGELLFDLMSE.
```

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.
- ❖ Задача пользователя — разобраться, где выравнивание правильное (имеет шанс соответствовать эволюционному

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-NR | LFEEFARVVATLKPFLMENVPNL | LLLNKGH | |
| TRU90449.1 | NPEMIV | GSPPCQDFSS | AGKRNEGLGR--ANLTLTFAEIVTRVSPQWFVMENV | D---RIEKSK | | |
| OXI46696.1 | GTDLVF | GGPPCQGFSQ | IGMRR-LDDER-NE | LYKQYTRVVSTLRPRVFLMENVPNL | ALMNKGH | |
| AVZ30243.1 | EIDVVF | GGPPCQGFSL | IGKRS-FEDPR-NS | LVFHYIRLVLELSPKFFVIENVKGMTAGNHQA | | |
| AFZ12381.1 | DIIGFI | GGAPCPDFSV | GGKNRGSEGDK-GKLSASYIELICQQKPDFFLFENVKGLYKTKKHR | | | |
| HCQ21462.1 | HIIGFI | GGPPCPDFSV | GGKNKGHLGDN-GKLSSAYIELICQNLPDFFLFENVKGLWRTTKHR | | | |
| EDN77159.1 | SLIGFI | GGPPCPDFSI | AGKNKGKGDGN-GKLSLSYTNLI IEMKPDFFLFENVKGLWRTARHR | | | |
| SOD91684.1 | EVSLVV | GGAPCQPF | SNIGKKLGNKDERNGDLFLEFVRMVKGIQPEAFIFENVVGI TQNKHSD | | | |
| QCS48280.1 | NVVGFI | GGPPCPDFSI | GGKNRGRQGDH-GKLSSESYIDLIIQHQPDFFI FENVKGLYRTKKHR | | | |
| SMB95934.1 | GLFGII | GGPPCPDFSV | GGKNRGENGEQ-GRLSKVFVDKILDLQPVFFLYENVPGLIRTAKHR | | | |
| RUO38876.1 | SPVGFI | GGPPCPDFSV | GGKNRGHEGEN-GRLTRTYVDGIIKYAPDFFI FENVKGLWRTKRHR | | | |
| OIP70538.1 | TIDLIC | GGPPCQGFST | IGTND-KKDR-NR | LFEEFLRMVETFKPNFI ILENTGLLAKKNES | | |
| AFY60915.1 | NLVGFV | GGPPCPDFSI | GGKNKGQYGDN-GKLTKVYVDII IENQPDFFVFENVKGLWRTRSRHR | | | |
| CUR30340.1 | DLIGFI | AGPPCPDFSV | GGKNRGKNGDQ-GKLTACYVELICQQRPDFFVFENVKGLWSTTKKHR | | | |
| TAK03971.1 | QAALVV | GGAPCQPF | SNLGSKRGTADSR-GTLFQDFIRIVKGVPRKGFIFENVEGLTQDKHKG | | | |
| AEE51071.1 | KVALVV | GGAPCQPF | SNIGKKEGENDAKNGDLFLEFVRMVKGIQPEAFIFENVAGIIQSKHSK | | | |
| RTR31666.1 | RLVGFI | GGPPCPDFSV | GGKNKGSEGEN-GKLTRTYIDLIVKDNPDYFI FENVKGLWRTTRHR | | | |
| PTU64472.1 | NIDLVF | GGPPCQGFSQ | IGTRR-LDDER-NE | LYKQYTRIVKTLKPRVFLMENVPNLAMMNKGH | | |

Учитываем:

- Наличие консервативных позиций
- Малое число гэпов

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

| 200 | 210 | 220 | 230 | 240 | 250 | 260 | 270 |
|--|-----|-----|-----|-----|-----|-----|-----|
| YG-VPQDRKRVFIVGYREDLNLK-----FEFPKPLNKKVTLRD-----AIGDL | | | | | | | |
| YG-VAQDRERVFYVGFVKDLNLSN-----FE-FYPYISEKERKYLKD-----SIWDL | | | | | | | |
| YG-VAQERKRVFYIGFRKDLEIKF-----SFPKGSTVEDKDKITLKD-----VIWDL | | | | | | | |
| YG-VAQDRKRVFYIGFRKELNIN-----YLPPIPHLIKPTFKD-----VIWDL | | | | | | | |
| YG-IPQQRDLVLFVAAKQG-----VIKIIPPTHTPENYR-----TVRDVIGSL | | | | | | | |
| YG-VPQSRQRVFFIGLKS DRPLNQQ-----ILTP-----PSKVI ESEYTSLEEAI SDLPVIE-----AGEGGEVQDY | | | | | | | |
| CG-VPQLRKRTFVIGHRHGS IAD-----LANVLQQRLAKQSL-----TVRDY | | | | | | | |
| CG-VPQSRTRFSLIGKLNSEHNF-----LIPTLSRKLSDKPM-----TVRDY | | | | | | | |
| YG-VPQRRHRIIVGIRKDQD-----VAFRVPEPTHKEKYR-----TASEALADI | | | | | | | |
| IG-AHHQRHRWFCLAIRKDY EPEE-----IIVSVNATKFDWENNEPPCQVDNK-----SYENSTLVRL | | | | | | | |
| FGNIPQNRERIYIVGFRN-----IEHYKNFNFPMPQP-----LTLTIKDM | | | | | | | |
| FN-VPQNRERLYIIGIREDLIKNEE-----WSLDFKRKDI LQKGKQRLVELDIKSFNFRWTAQ-----SAATKRLKDL | | | | | | | |
| FG-IPQNRERVFCISILN-----PNEDFTFPQKQ-----NLTL SMNDL | | | | | | | |
| FG-SSQARRRVFMI STLNEF-----VELPKGDKKPKS-----IKKVLNKI | | | | | | | |
| FG-IPQNRERIYLVGF-----LNHDVDFRFPQP-----IGQATAVGD I | | | | | | | |
| FG-LPQNRERIYIVGFDRKS-----ISNYSDFQMPTP-----LQEKTRVGN I | | | | | | | |
| FG-VPQNRERIYIVGFNKEK-----VRNHEHFTFPTP-----LKT KTRVGD I | | | | | | | |
| IFG-VPQNRERIYIVGFHKS-----TGVNSFSYPEP-----LDKIVTFADI | | | | | | | |
| IFQ-VPQNRRLRVYIVGLDQSQPELT-----ITSHIGATDSHKFKQLSNQASLFD-----TNKIMLVRDI | | | | | | | |
| IFG-VPQNRVRIYILGILGSKPKLT-----LTSNVGAADSHKYK-NEQISLFD-----ES-YATVKDI | | | | | | | |
| IFG-IPQKRKRFYLVAF LNQN-----IHFEF PKP-----PMISKDIGEV | | | | | | | |
| IFG-LPQRRERIVIVGFHPDLG-----INDFSFPKGN-----PDNKVPINA I | | | | | | | |
| YG-IPQKRERIYMICFRNDLN-----IQNFQFPKP-----FELNTFVKDL | | | | | | | |
| YG-NAQRRRRVFI FGYKQDLNYSKAME-----ESPLDKI IYHNGLFAEAFPI EDYANKNR-----VNRTHITHDIVDI | | | | | | | |
| YG-TPQRRKRAIIRLNKKGTIWN-----LPLKQNI VSVEQ-----AIGNLPSI | | | | | | | |
| IGG-TPQVRERVFITATLVPERMRDERIPRTETGEIDAEAIGPKPVATMNDRFP I KKGTEL FHPGDRKSGWNLLTSGI IREG | | | | | | | |
| YG-VAQNRDRVFIIGIQKLGVPD-----FSFPEYSESEQRLYDILDNLQTPSII-----PESLPIQRNL | | | | | | | |
| IFG-VAQNRDRVFI VGIQKLDLNG-----FSFPEYAESDQRLYHILDNL EAPETK-----LESIP IQRNL | | | | | | | |
| IYD-VAQKRERIVIIIGIREDLVK-----EQKYPFRFPLAQ-----VYKPV LKDV | | | | | | | |
| YG-VSQLRPRVLFVALKNEYTN-----FFKWPEPNSEQPK-----TVGELLFDL | | | | | | | |

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

Выравнивание тех же белков другой программой

| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | |
|-----------------------|-----|-------|-----|--------------------|-----|-----|---|-----|------------------------|
| -GVPODRKR VFI | --- | VGY | --- | REDL | --- | --- | NL-KFEFPKP | --- | LN--KKVTLRDAIGDL |
| -GVAQDRER VFY | --- | VG | --- | RKDL | --- | --- | NISNFEFPYP | --- | I SEK--ERKYLKDSI WDL |
| -GVAQERKR VFY | --- | IGF | --- | RKDL | --- | --- | EI-KFSFPKGSTVEDK | --- | DKITLKDVI WDL |
| -GVAQDRKR VFY | --- | IGF | --- | RKEL | --- | --- | NI-NYLPP I PHLI | --- | ---KPTFKDVI WDL |
| -GIPQQRDR LVL | --- | FAA | --- | KQGV I | --- | --- | KI I PPTH | --- | TPENYRTVRDVI GSL |
| -GVPOSQRQ VFF | --- | IGL | --- | KSDR | --- | --- | PLNQI I LTPPSK | --- | VIES--EYTSLEEAI SDL |
| -GVPQLRKRT FV | --- | IGH | --- | RHGS I | --- | --- | ADLANVLQ | --- | ---QRLAKQSLTVRDY |
| -GVPOSRTRF SL | --- | IGK | --- | LNSEH | --- | --- | NFL I P T L S | --- | ---RKLSDKPM TVRDY |
| -GVPQRHRH R I I | --- | VGI | --- | RKDQ | --- | --- | DV-AFRVPEP | --- | THKE--KYRTASEALADI |
| -GAHQRRHR WFC | --- | LAI | --- | RKDY EPEE I I | --- | --- | VSVNATKFDWENNE | --- | PPCQVDNKS YENSTL VRL |
| GNIPQNRER IYI | --- | VG | --- | RNIE | --- | --- | HYKNFNFPMP | --- | ---QPLT L T I KDM |
| -NVPQNRER LYI | --- | IGI | --- | REDL I KNE | --- | --- | EWSLDFKRKDI LQK GKQRL VELD I KSFNFRWT | --- | ---AQSAATKRLKDL |
| -GIPQNRER VFC | --- | ISI | --- | LNPNE | --- | --- | DFTFPQK | --- | ---QNL T L S MNDL |
| -GSSQARRR VFM | --- | I S T | --- | | --- | --- | LNEFVELPKGD | --- | ---KKPKS I KKV |
| -GIPQNRER IYL | --- | VG | --- | LNHDV | --- | --- | DFRFPQP | --- | ---I GQATAV GDI |
| -GLPQNRER IYI | --- | VG | --- | DRKSI S | --- | --- | NYSDFQMPTP | --- | ---LQEKTRV GNI |
| -GVPQNRER IYI | --- | VG | --- | NKEKVR | --- | --- | NHEHFTFPTP | --- | ---LTKTRV GDI |
| -GVPQNRER IYI | --- | VG | --- | HKST | --- | --- | GVNSFSYPE | --- | ---PLDK I V T FADI |
| -QVPQNR LRVYI | --- | VGLD | --- | QSQPEL TITSHI | --- | --- | GATDS--HKFKQ--LSNQASL | --- | FD-----TNK I M L V RDI |
| -GVPQNRVR IYI | --- | LGIL | --- | GSKPKL TLTSNV | --- | --- | GAADS--HKYK--NEQ I S L | --- | FD-----ESYATV KDI |
| -GIPQKRKR FYL | --- | VAF | --- | LNQNI | --- | --- | HFEFPKP | --- | ---PMI SKDI GEV |
| -GLPQRER IVI | --- | VG | --- | HPDL | --- | --- | GINDFSFPK | --- | ---GNPDNKVPI NAI |
| -GIPQKRER IYM | --- | ICF | --- | RNDL | --- | --- | NIQNFQFPK | --- | ---FELNTFVKDL |
| -GNAQRRRR VFI | --- | FGY | --- | KQDLNYSKAMEESPLD | --- | --- | KI I YHN--GLFAEAFPI EDYANKNRVNRTHI THDI VDI | --- | |
| -GTPQRKRRA I IRLNKKGT | --- | | --- | IWNL | --- | --- | PLKQNI V | --- | ---SVEQAI GNL |
| -GTPQVRER VFI | --- | TATL | --- | VPERMRDER I PR TET | --- | --- | GEIDA--EAIGPK--PVATMNDRFP I KK | --- | GGTELFHPGDRKSGWNL |
| -GVAQNRDR VFI | --- | IGI | --- | QQKL | --- | --- | GVPDFSFP EYSESEQRLYDI LDNLQTPS I I | --- | |
| -GVAQNRDR VFI | --- | VGI | --- | QQKL | --- | --- | DLNGFSFP EYAESDQRLYHI LDNLEAPETK | --- | |
| -GVAQNRDR VFI | --- | VGI | --- | QQKL | --- | --- | DLNGFSFP EYTESEQRLYHI LDNLEVPETK | --- | |
| -DVAQKRER IVI | --- | IGI | --- | REDLVKE | --- | --- | QKYPFRFPLA | --- | ---QVYKPV LKDV |
| -GVSQLRPR VLF | --- | VAL | --- | KNEY | --- | --- | TNFFKWPEP | --- | ---NSEQPKT V GELL FDL |

Консервативное — значит важное

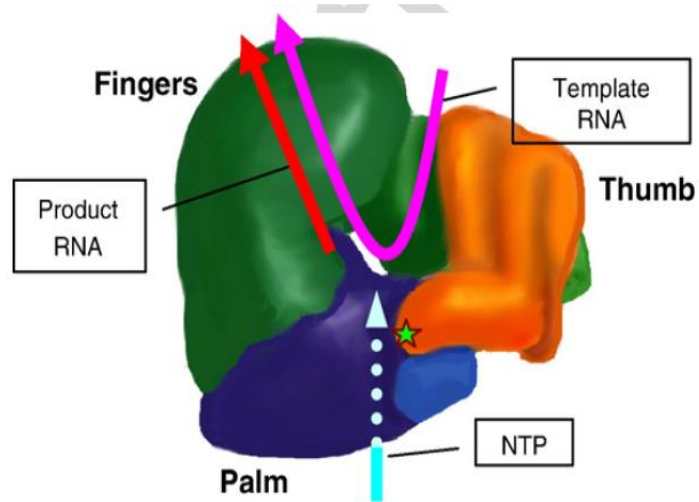
- ❖ Консервативное — то, что длительно существует в эволюции, с несущественными изменениями
- ❖ Это означает, что существенные изменения не имеют шансов закрепиться

Выравнивание РНК-зависимых РНК полимераз (RdRP) из разных вирусов

```

*      320      *      340      *      360      *      380      *      400      *      420      *      440      *      460
'KTMIRFGDVGGLDIDFSAFDASLSPFMIREA..GRIMSELS...GTPSHFGTALINTIIYSKHLIYNCCY.....HVCGSMPSGSPCTALINSIINNVLYYVFSKIFGKSPVFF.....CQALKILC.YGDDVIVFSDV
'VAMQG.FERVYDVIDYSNEDSTHSVAMFRLL..A...EEFF.TPENGFDPLTREYLESIAISTHAFEKRF.....LITGGLPSGCAATSMINTIMNIIIRAGLYLTYKNFEFDD.....VKVLS.YGDDLIVATNYQL
'YTHFAQ.YKNVWDVYSAFDANHCSDAMNIMFEEVFEFG.....FHPNAEWILKTLVNTTEHAYENKRI.....VVEGCMPSGCSATSIINTILNNTIYVLYALRRHYEGVELDT.....YTMIS.YGDDIVVASDYDL
.....WSLCVATIVSDHDTFWPGWLRDLICDELINMGYA.PWVVKLFETSLKLPVYVGAFAPEQGHTLLGDPSNPDLVGLSSGQGATDLMGTLMSITYLVMQLDHTAPHLNSRIKMPACRFLDSYWQGHEEIRQIS.KSDDAIIGWTKGR
.RRIPE.NWVYCDAGSQEDSSLTPLYLINAV..LTIIRSTYMEDWDVGLQMLRNLYTEIVYTPISTPDGTIV.....KKFRGNNSGQPSTVVDNSIMVVIAMHYALIKECVFEFEEID.....STCVFFV.NGDDLIVAVNPEK
IDKLNRPGLWLGSGDGRFDSSIDPFFFDVV..KTRKHFL..PSEHHKAIDLIDEILNTTICLANGMVI.....KKNVGTQR.QPSTVVDNTLVLMATAFLYAYIHKTDGRELAL.....LNERFIFVC.NGDDNKFAISPQF
VLSLASFSPYGFNGDFANFDGMFHSSFSMV..SEIANIFY...GNFLSTERDNLTRMLTNRFSIMKGAII.....RVPGGPGSGFEMTVIENSFINLFYIQAWIMLARFNQRQDISH.....PCNFPKYVRACV.YGDDNIVAIMEV
VARMKEKGNVDVLCQDYSSEFDGLLSKQVMDVI..ASVINELC.GGEDQLKNARRNLLMACCSRLAICKNTVW.....RVECGIPSGFEMTVIENSFINEITLIRYHYKLMREQQAPELMV.....QSFDKLIGLVT.YGDDNIVSNVAV
/AEHAK.YKNHFDADYIANDSTQNRQIMTES..FSIMSRLT...ASPELAEVVAQDLLAPSEMDVGDYVI.....RVKEGLPSGFPCTSQVNSINHWTITLCAISEATGLSPDVV.....QSMSYFSFYGDDEIVSTDIDF
'NNLTSKASDFLCDYSKWDSTMSPCVVRIA..IDLADCC...EQTELTKSVVLTIKSHEMTILAMIV.....QTKRGLPSGMPPTSVINSICHWILWSAAVYKSCAEIGLHCS.....NLYEDAPFYT.YGDDGYAMTFMM
IQRIKS.AAKVYAVDYSKWDSTQSPRVSAA..IDLRYFS...DRSPIVDSAANTLKSPPIAIFNGVAV.....KVSSGLPSGMPITSVINSINHCLYVGCAILQSLAARGVPVTW.....NLFSTFEMMT.YGDDGYMFFMMF
'KRLEKPKHRYCVDYSKWDSTQPPKVTQS..IDLRHFT...DKSPIVDSACATLKSNPIGIFNGVAF.....KVAGGLPSGMPITSVINSINHCLMVGSAVVKALEDSGVRVTW.....NIFDSMDLFT.YGDDGYIVPPLI
D      D
g      sg      T      n3
gDD

```



На каких участках выравнивание правильное – совпадает с эволюционным?

Множественное выравнивание дает больше информации, чем парное

```

      *           100           *           120           *           140           *           160
THIE_LACLS : AGVSFIVNDDVELARELNADGIHIGQTDESVSKVREKVGQEMWLGLSVTKADELKTAQ-SSGADYLGIGPIYPTNSKND : 14
THIE_MANSM : FQVFFIVNDDVELALSIQADGIHVGQKDTAVETILRNTRNKPIIGLSINTLAQALANKDRQDIDYFGVGPPIFPTNSKADH : 15
THIE_STRA3 : YQVFFIIDDIDLVELIDADGLHIGQNDLPVDEARRRLPDKI-IGLSVSTMAEYQKSQ-LSVVDYIGIGGFENPTQSKADA : 14
THIE_LISIN : YQVFFIINDDVALALEIGADGIHVGQNDDEEIRQVIASCAGKMKIGLSVHVSVEAEEAERLGSVDYIIGVGPPIFPTISKADA : 14
THIE_ANOFW : YNIPFIVNDDVDLALALQADGVHVGQDEVAERVRDRIGDKY-LGVSVHNLNEVKKAL-AACADYVGLGPIFPTVSKEDA : 14
THIE_GEOTN : YGVFFIVNDDVELAIAIDADGVHVGQDDEDARRVREKIGDKI-LGVSAHNVEBARAAI-EAGADYIGVGPIYPTRSKDDA : 14
THIE_BACSU : AGVFFIVNDDVELALNLKADGIHIGQEDANAKEVRAAIGDMI-LGVSAHTMSEVKQAE-EDGADYVGLGPIYPTETKKDT : 14
THIE_BACA2 : AGIPFIIINDDVELALRLBADGVHIGQDDADAEEETRAAIGDMI-LGVSAHNVSEVKRAE-AAGADYVGMGPVYPTETKKDA : 14
THIE_OCEIH : FQIPFIIINDDVLAKQLDADGIHIGQDDQPVEVVRKQF PNKI-IGLSISTNNELNQSP-LDLVDYIGVGPIEDTNTKEDA : 14
THIE_STAAB : YNVFFIVNDDVSLAKEINADGIHVGQDDAKVKEIAQYFTDKI-IGLSISDLGEYAKSD-LTHVDYIGVGPIYPTPSKHDA : 14
THIE_STACT : YNVFFIVNDDVALAEEDADGIHVGQDDEAVDDFNRRFEGKI-IGLSIGNLEELNASD-LTYVDYIGVGPIFATPSKDDA : 14
      6pFI61DD6 La 6 ADG6H6GQ D 6G6S 2 DY G6GP pT 3K Da

```

```

      *           180           *           200           *           220           *           240
THIE_LACLS : AKHIGIKDLR-LMLLENQLPIVGIGGITQDSLTELSAICLDGLAVISLLTEAENPKKVAQMIRQKITKNG~~~~~ : 21
THIE_MANSM : SPIVGMNFIRQIRQLGIDKPCVAIGGIKEESAAILRRLGADGVAVISSAISHSVNIANTVKTLAQK~~~~~ : 22
THIE_STRA3 : KPAVGNRTTKAVREINQDIPVAIGGITSDFVHDIIESGADGIAVISSAISKANHIVDATRQLRYEVEKALVNRQKRSDVI : 22
THIE_LISIN : EPVSGTAILEEIRRAGIKLPIVGIGGINETNSAEVLTAGADGVSVISAITRSEDCQSVIKQLKNPGSPS~~~~~ : 21
THIE_ANOFW : KQACGLTMIEHIRAHEKRVLPVAIGGITEQTAKQVIEAGADGIAVISSAICRAEHIYEQTKRLYEMVMRAKQKGR~~~~~ : 21
THIE_GEOTN : NEAQQPGILRHLREQGITIPIVAIGGITADNTRAVIEAGADGVSVISAIASAPEPKAAAAALATAVREANL---R~~~~~ : 22
THIE_BACSU : RAVQGVSLIEAVRRQGISIPIVGIGGITIDNAAPVIQAGADGVSMISSAISQAEDPESAARKFREEIQTYKTG--R~~~~~ : 22
THIE_BACA2 : EAVQGVTLIEEVRRQGITIPIVGIGGITADNAAPVIEAGADGVSMISSAISQAEDPKAAARKFSEIEIRRSKAGLSR~~~~~ : 22
THIE_OCEIH : KTAGLEWISLKKQHPSLPIVAIGGINTTNAQELIQAGADGVSFISAITETDHLLQAVQRL~~~~~ : 20
THIE_STAAB : HTPVGPEMIATFKEMNPQLPIVAIGGINTSNVAPIVEAGANGISVISSAISKSENIEKTVNRFKDFFN~~~~~ : 21
THIE_STACT : SEPVGPKMIETLRKEVGDLEPIVAIGGISLDNVQEVAKTSADGVSVISSAIARSPHVTETVHKFLQYFK~~~~~ : 21
      G P V IGGI 6 galG6 ISa6

```

```

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

```

```

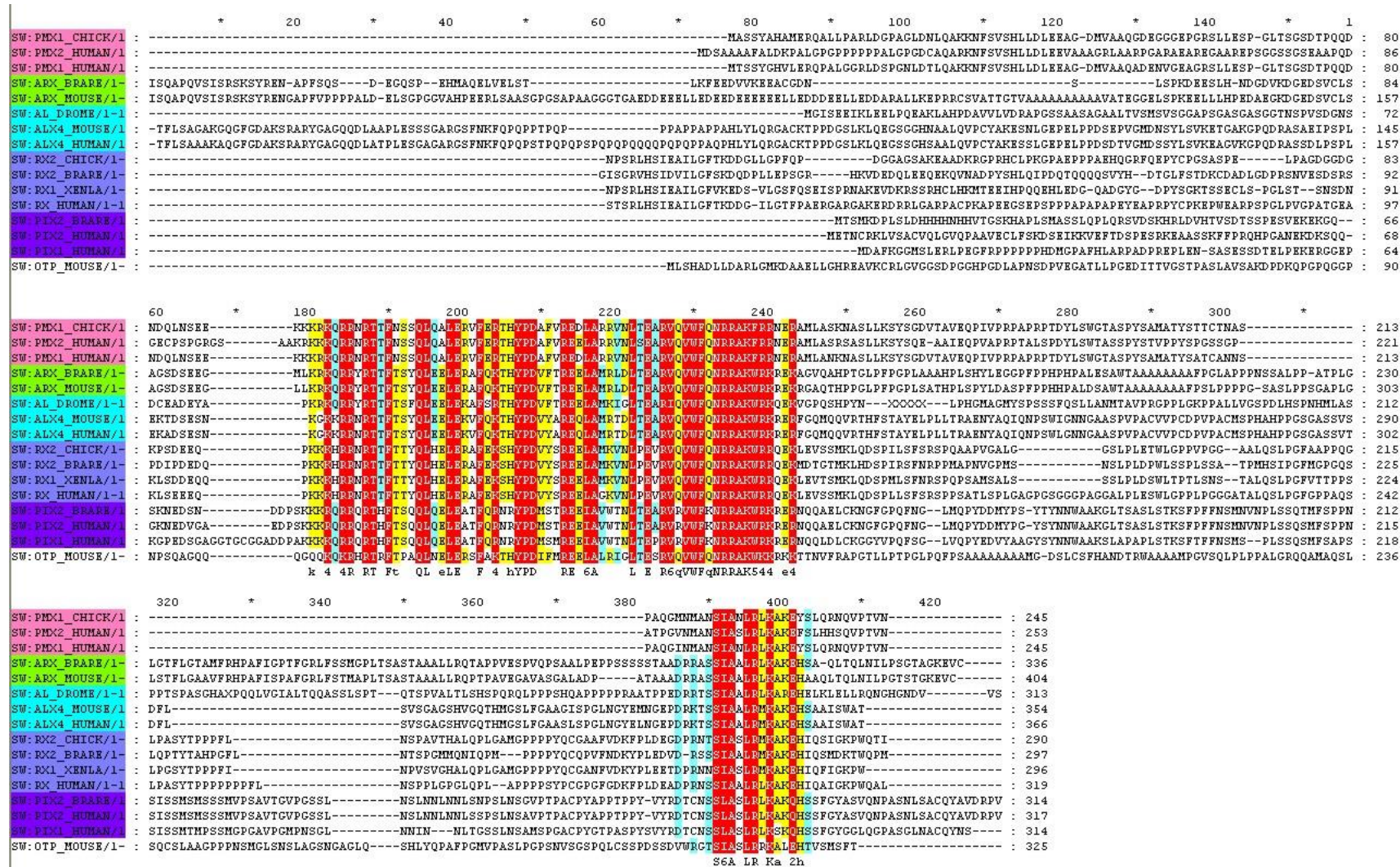
      160           *           180           *           200           *           220
THIE_LACLS : AAKHIG---IKDLRMLLENQLPIVGIGGITQDSLTELSAIGLDGLAVISLLTEAENPKKVAQMIRQKITKNG : 218
THIE_MANSM : HSPIVGMNFIRQIRQLGIDK--PCVAIGGIKEESAAILRRLGADGVAVISSAISHSVNIANTVKTLAQK----- : 220
      GG T4 6R 6 6 P V IGGI 2 S L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

```

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

Гомеобелки

Гомологичны ли эти белки?

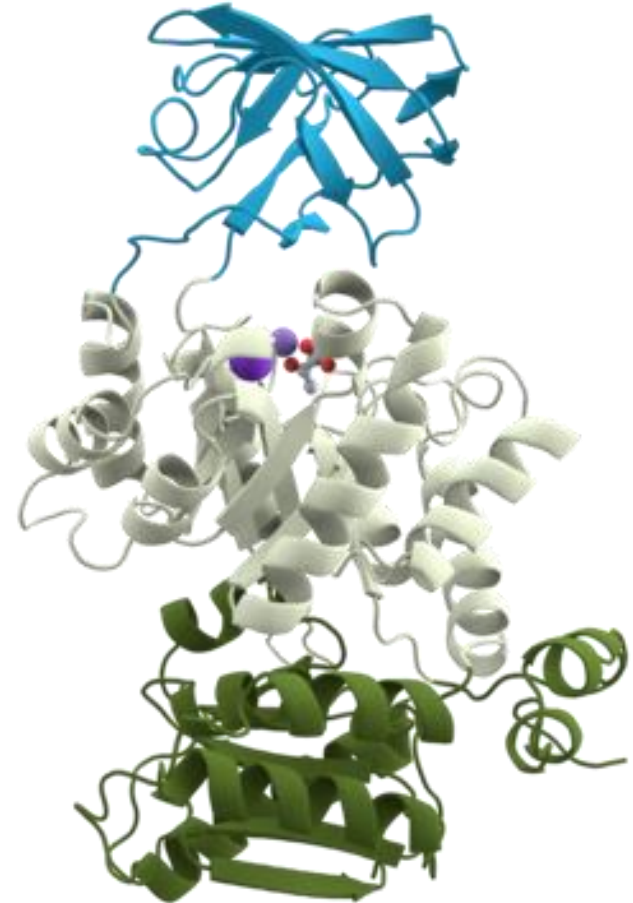


Домены белков

- ❖ **ЭВОЛЮЦИОННЫЙ ДОМЕН** – достаточно длинный (более многих десятков а.к.о.) участок предкового белка, который эволюционировал только по типу локальных мутаций.
- ❖ При этом белки-потомки могли претерпевать крупные перестройки, не затрагивающие домен.
- ❖ В белке может быть один домен, два или много.
- ❖ **Доменная архитектура** – последовательность доменов в белке

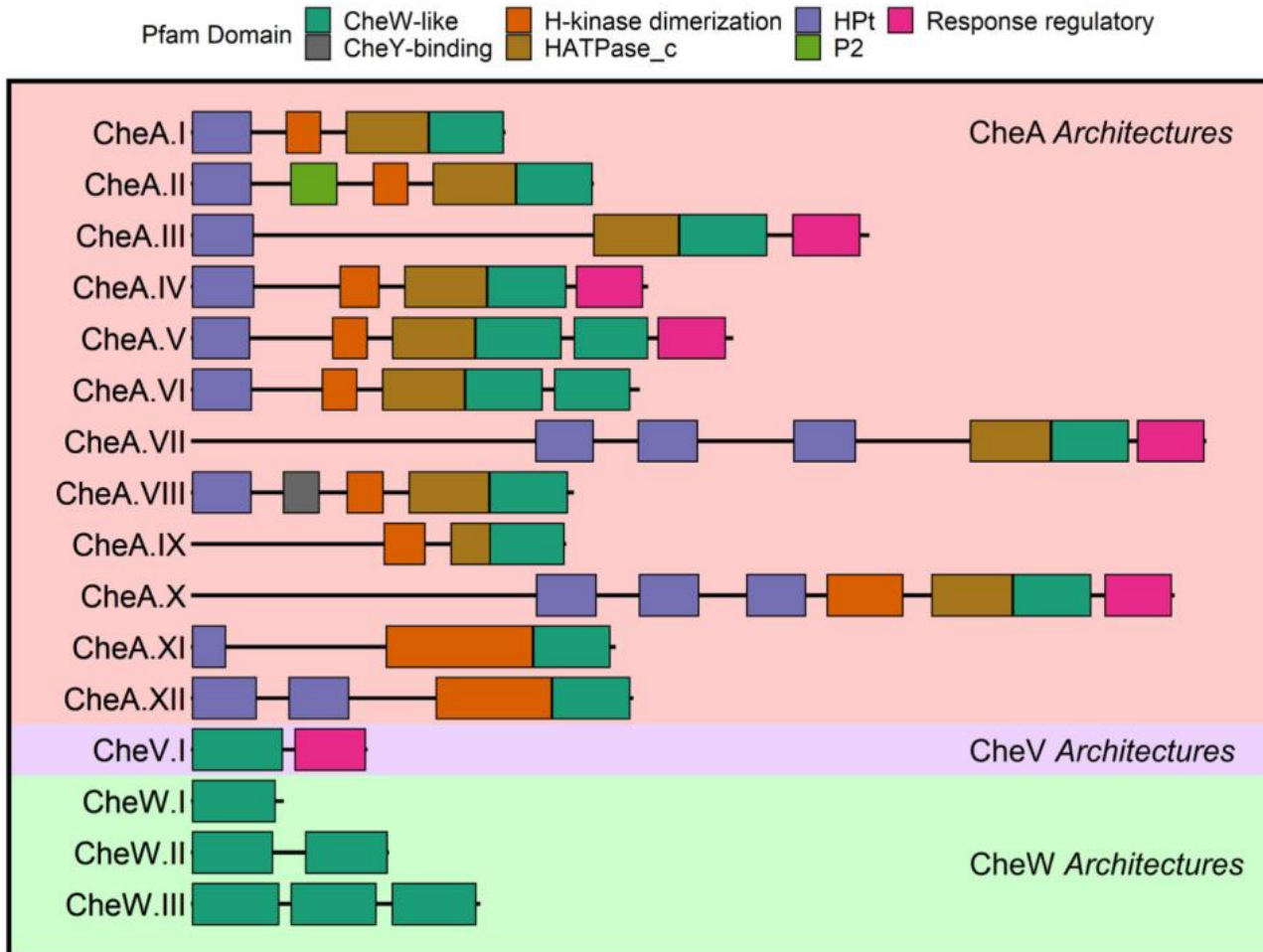
*Термин «домен белка» можно понимать по-разному

- ❖ Структурный домен – структурная единица белка, которая сворачивается независимо
- ❖ Функциональный домен – часть белка, ответственная за осуществление определенной функции

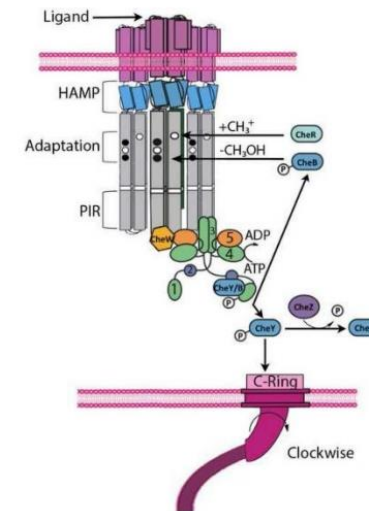


Пируваткиназа
(PDB: 1PKN)

Так изображают домены



Che белки участвуют в хемотаксисе бактерий



Доменные архитектуры белков, содержащих домен CheW-like

Многодоменные белки в эволюции образуются в результате последовательных единовременных крупных перестроек в генах белков

Домен белка – единица непрерывной эволюции белков

Семейства доменов

- ❖ ЭВОЛЮЦИОННЫЙ ДОМЕН – достаточно длинный (более многих десятков а.к.о.) участок предкового белка, который эволюционировал только по типу локальных мутаций.
- ❖ Домену дают название. Собирают представителей домена из всех белков, в которых их удаётся найти и строят выравнивание участков, соответствующих домену.
- ❖ Семейство доменов – совокупность таких участков.

Семейства доменов и базы данных

- ❖ Pfam
- ❖ PROSITE
- ❖ Prints
- ❖ SMART
- ❖ TIGERFAM



an **inte**grated **r**esource of **pro**tein families, domains and functional sites

- В занятиях работаем с базой данных Pfam (от «protein families»).
- Pfam с 2023 года доступен через интерфейс консорциума InterPro, посвященного семействам белков (до этого имел собственный интерфейс)


Pfam – protein families



- ❖ Коллекция семейств доменов (~ 24 к записей)
- ❖ Для каждого семейства имеется т.н. «НММ-профиль», позволяющий с помощью специальных программ искать представителей семейства в новых последовательностях
- ❖ Для каждого семейства хранится два выравнивания:
 - ✓ “seed” — сравнительно небольшое, использованное для построения профиля
 - ✓ “full” — из всех представителей семейства, найденных в **референсных** протеомах

JOURNAL ARTICLE

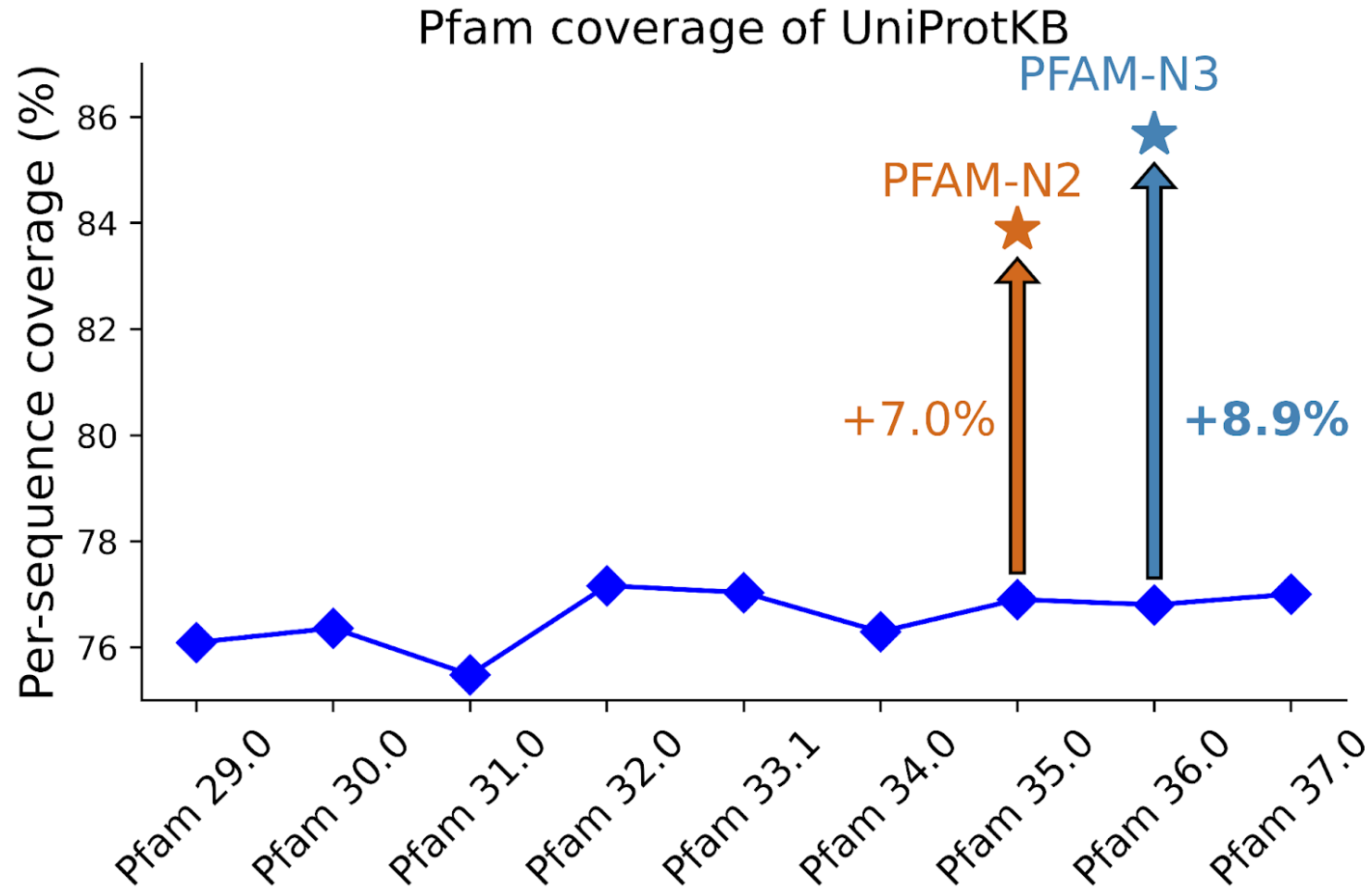
The Pfam protein families database: embracing AI/ML

Typhaine Paysan-Lafosse  , Antonina Andreeva , Matthias Blum , Sara Rocio Chuguransky , Tiago Grego , Beatriz Lazaro Pinto , Gustavo A Salazar , Maxwell L Bileschi , Felipe Llinares-López , Laetitia Meng-Papaxanthos ... [Show more](#)

Nucleic Acids Research, Volume 53, Issue D1, 6 January 2025, Pages D523–D534,
<https://doi.org/10.1093/nar/gkae997>

Published: 14 November 2024 **Article history** ▼

Pfam-N – аннотации белковых доменов, полученные с помощью нейросети

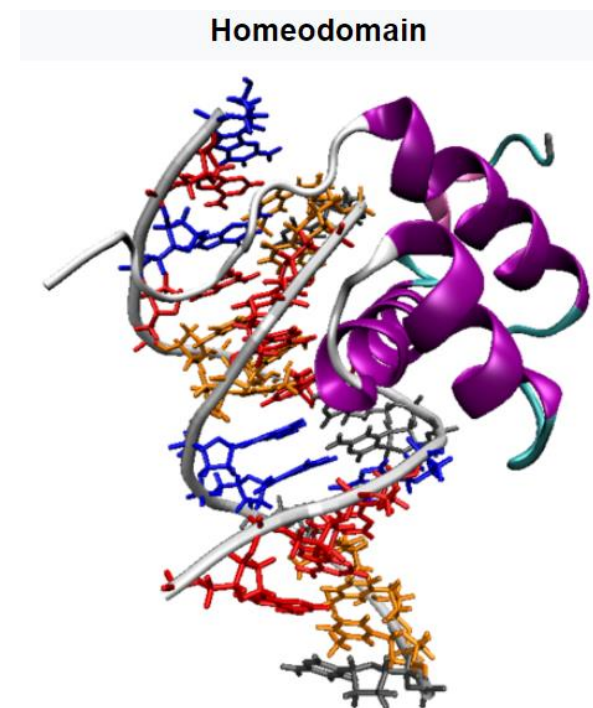


Эволюционные домены

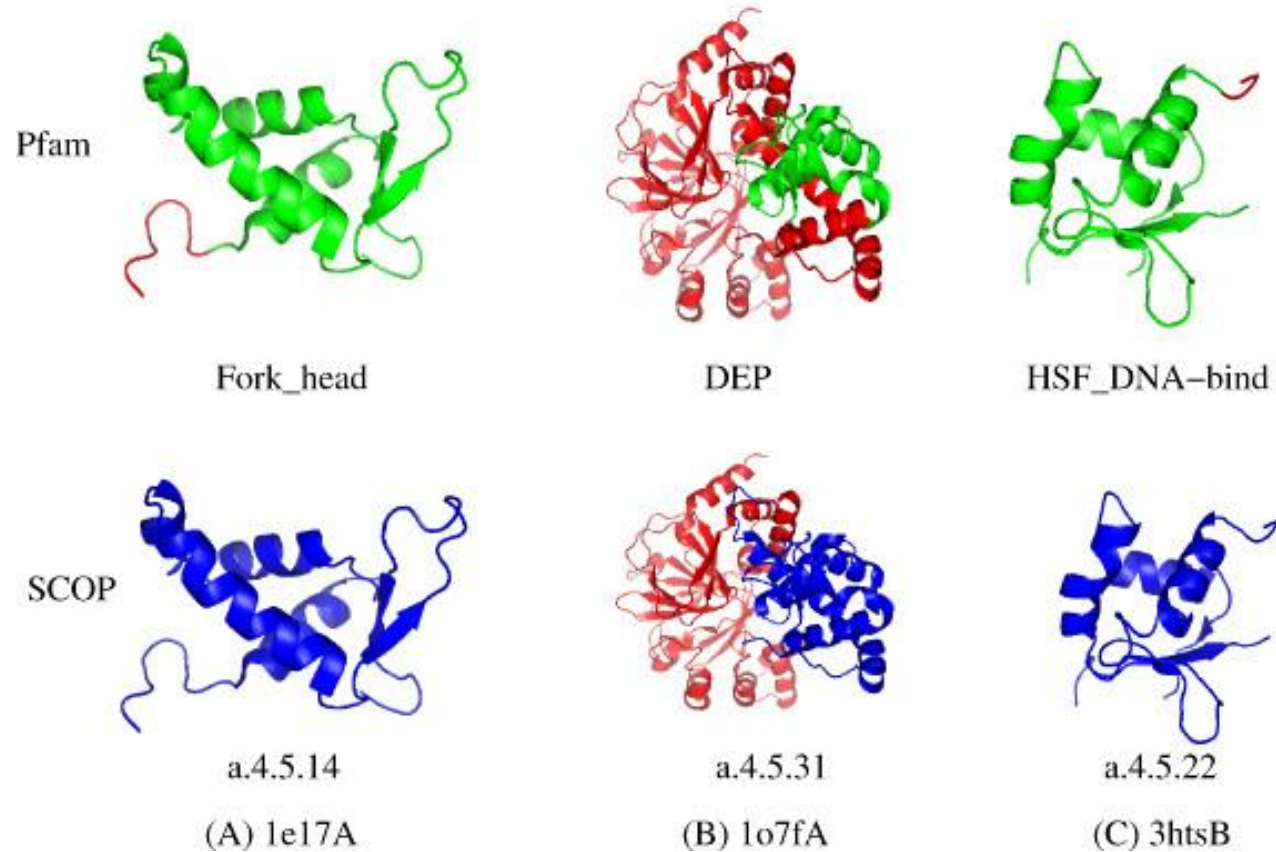
- ❖ Имеют определенную функцию (не всегда известна)
DUF – Domain of Unknown Function
- ❖ Часто совпадают со структурными доменами
(но не всегда)

Гомеодомен – ДНК СВЯЗЫВАЮЩИЙ ДОМЕН

Homeodomain proteins regulate gene expression and cell differentiation during early embryonic development, thus mutations in homeobox genes can cause developmental disorders.^[1]



Домены Pfam часто, но не всегда соответствуют структурным доменам SCOP

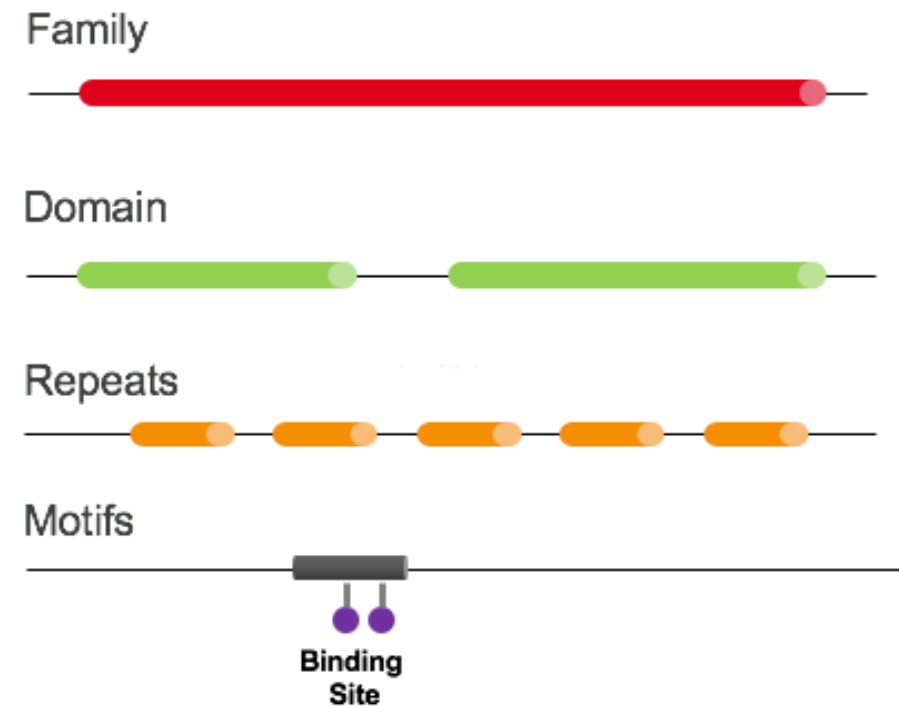


Такое соответствие
наблюдается не для
всех доменов Pfam

Pfam – protein families

❖ Типы записей в Pfam:

- ✓ Семейство – коллекция гомологичных участков белков.
- ✓ Домен – устойчивая структурная единица, или функциональный участок, встречающийся в различных белковых архитектурах.
- ✓ Повтор – короткая единица, нестабильная сама по себе, но образует стабильные структуры, если есть много копий.
- ✓ Мотив – короткий консервативный участок
- ✓ Клан – выделенная вручную группа похожих по последовательности и структуре доменов.



Pfam – protein families



- ❖ Для каждого семейства доменов есть
 - ✓ Название домена, ID AC, описание функции
 - ✓ Белки из UniProt, содержащие домен
 - ✓ Доменные архитектуры
 - ✓ Таксономическая распространенность
 - ✓ 3D структуры
 - ✓ **Множественное выравнивание**
 - **seed** - выравнивание семейства гомологичных фрагментов из разных белков, составленное экспертом для построения модели
 - **full** – выравнивание всех представителей семейства, хранящихся в Pfam (Interpro теперь позволяет скачать выравнивания PfamA целиком)
 - ✓ **НММ-профиль выравнивания**– модель семейства=математическое описание выравнивания по колонкам (Curation)

Pfam PF00046 Homeodomain

Pfam entry ⓘ

Overview

| | |
|----------------------|------|
| Proteins | 263k |
| Domain Architectures | 2k |
| Taxonomy | 14k |
| Proteomes | 3k |
| Structures | 223 |
| Profile HMM | |
| AlphaFold | 170k |
| BFVD | 8 |
| Alignment | |

| | |
|-------------------|---|
| Member database | Pfam ⓘ |
| Pfam type | domain |
| Short name | <i>Homeodomain</i> ← |
| Clan | HTH |
| Author | Eddy SR;0000-0001-6676-4706 ⓘ |
| Sequence Ontology | 0000417 |

Description ⓘ Imported from IPR001356

This entry represents the homeodomain (HD), a protein domain of approximately 60 residues that usually binds DNA. It is encoded by the homeobox sequence [7, 6, 8], which was first identified in a number of Drosophila homeotic and segmentation proteins, but is now known to be well-conserved in many other animals, including vertebrates [1, 2], as well as plants [4], fungi [5] and some species of lower eukaryotes. Many members of this group are transcriptional regulators, some of which operate differential genetic programs along the anterior-posterior axis of animal bodies [3]. This domain folds into a globular structure with three α -helices connected by two short loops that harbour a hydrophobic core. The second and third form a helix-turn-helix (HTH) motif, which make intimate contacts with the DNA: while the first helix of this motif helps to stabilise the structure, the second helix binds to DNA via a number of hydrogen bonds and hydrophobic interactions, which occur between specific side chains and the exposed bases and thymine methyl groups within the major groove of the DNA. One particularity of the HTH motif in some of these proteins arises from the stereo-chemical requirement for glycine in the turn which is needed to avoid steric interference of the β -carbon with the main chain: for *cro* and repressor proteins the glycine appears to be mandatory, while for many of the homeotic and other DNA-binding proteins the requirement is relaxed.

Accession code
домена

Короткое название домена



PF00046

Homeodomain

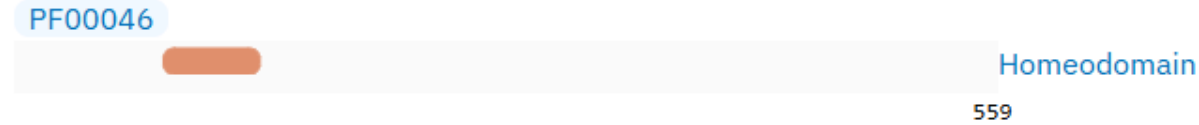
Pfam entry ⓘ

- Домены могут комбинироваться друг с другом – как модули
- Вариантов доменных архитектур может быть очень много

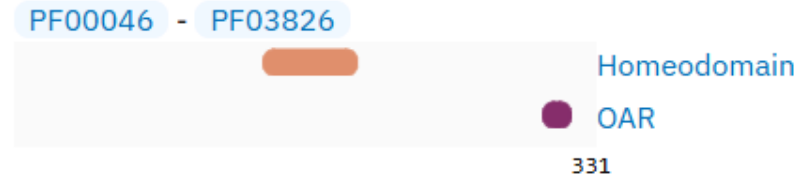
| | |
|-----------------------------|-----------|
| Overview | |
| Proteins | 263k |
| Domain Architectures | 2k |
| Taxonomy | 14k |
| Proteomes | 3k |
| Structures | 223 |
| Profile HMM | |
| AlphaFold | 170k |
| BFVD | 8 |
| Alignment | |

1881 domain architectures

There are 136209 proteins with this architecture (represented by P07269):



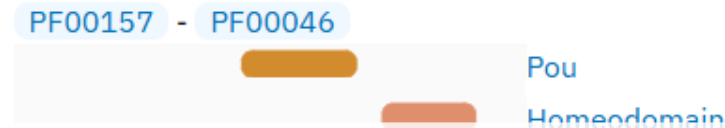
There are 14044 proteins with this architecture (represented by P70390):



There are 13163 proteins with this architecture (represented by P29673):



There are 11827 proteins with this architecture (represented by P10036):



Таксономическая представленность домена

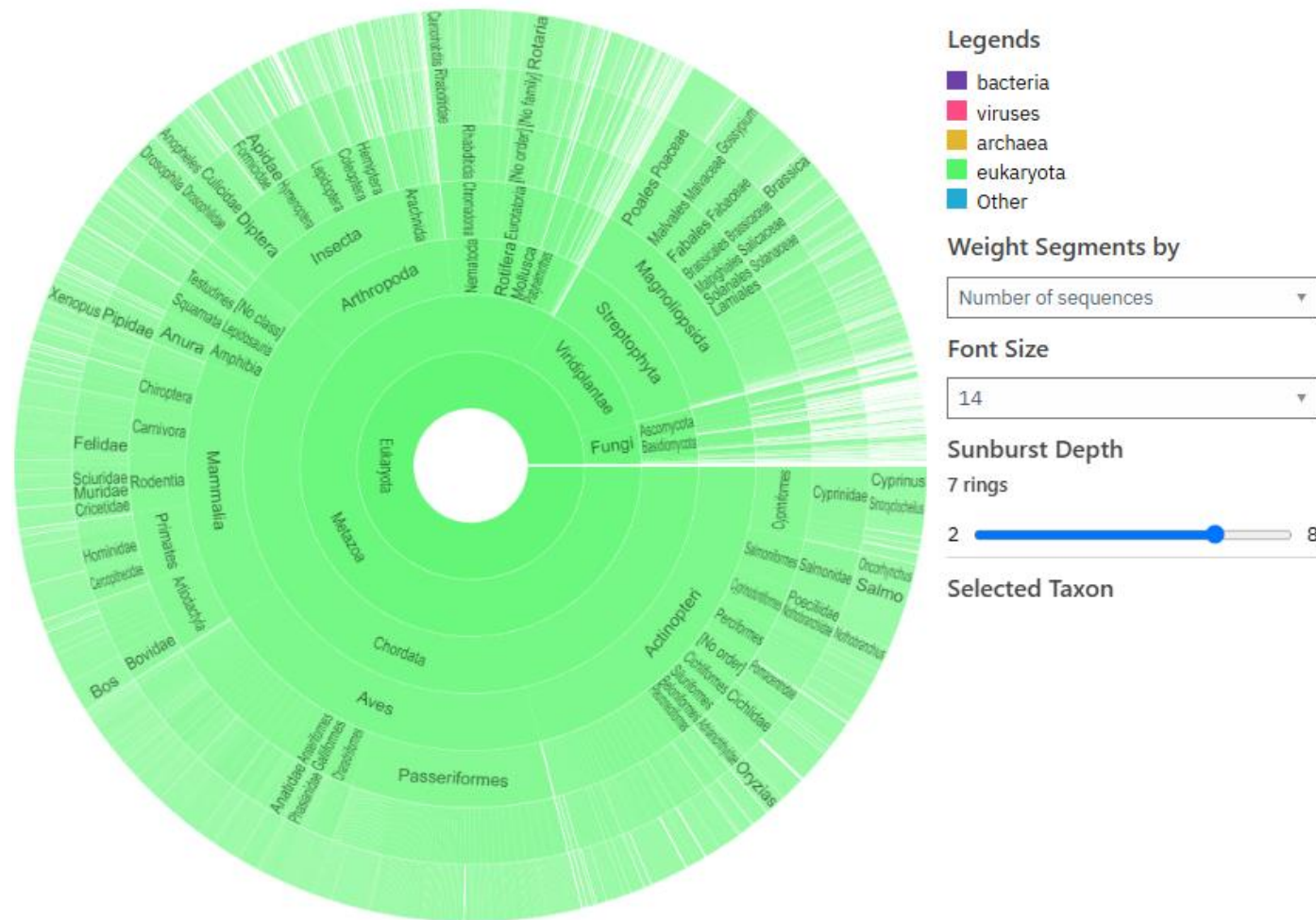
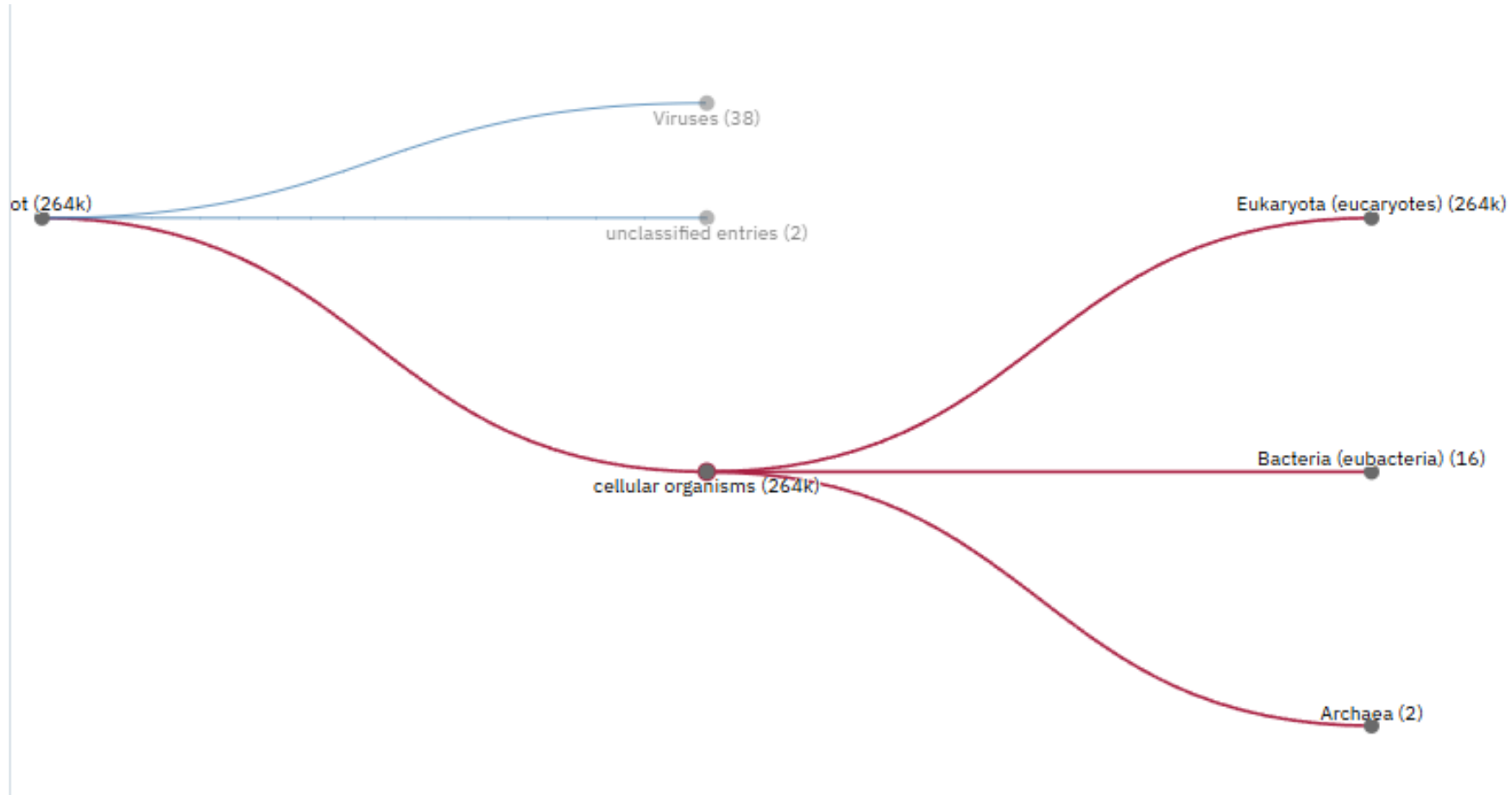


Диаграмма показывает, у **кого** встречается домен

Таксономическая представленность домена

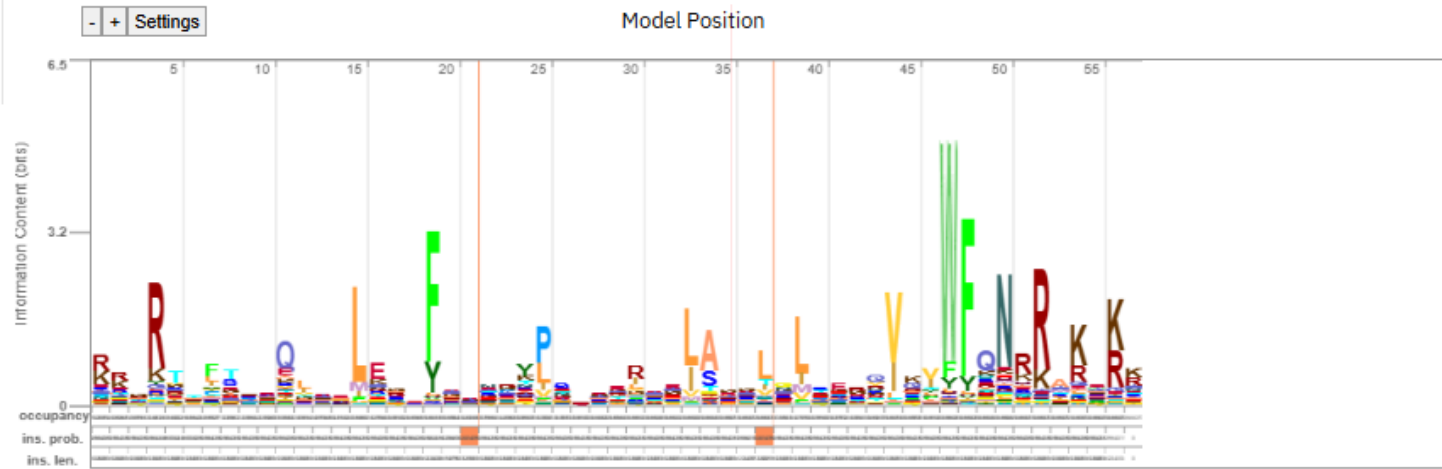


| | |
|----------------------|------|
| Overview | |
| Proteins | 263k |
| Domain Architectures | 2k |
| Taxonomy | 14k |
| Proteomes | 3k |
| Structures | 223 |
| Profile HMM | |
| AlphaFold | 170k |
| BFVD | 8 |
| Alignment | |

Profile HMM Information

| | |
|---------------------|---|
| HMM build commands | Build method: hmmbuild HMM.ann SEED.ann Search method: hmmsearch -Z 81514348 -E 1000 --cpu 4 HMM pfamseq |
| Gathering threshold | Sequence: 22.9 Domain: 22.9 |
| Download | Download the raw HMM for this family |

Profile HMM logo



- У каждого домена есть **НММ-профиль выравнивания** – модель семейства=математическое описание выравнивания по колонкам (Curation)
- НММ-профиль строится по выравниванию SEED
- Его можно использовать для поиска доменов данного семейства

- ❖ Часто белки с одинаковой доменной архитектурой имеют сходство последовательностей в границах доменов, а вне них сходство не детектируется.
- ❖ Последовательности междоменных участков обычно не имеют признаков гомологии, что происходило в эволюции – неизвестно.

[X1WJ92_ACYPI](#) [Acyrtosiphon pisum (Pea aphid)]
Uncharacterized protein (408 residues)



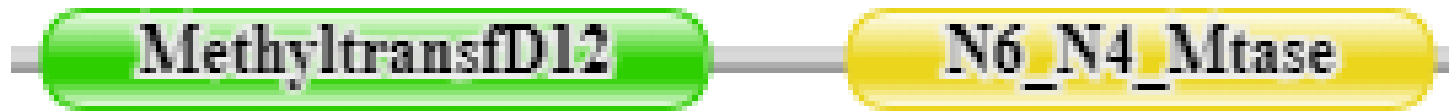
There are 14044 sequences with the following architecture: **Homeodomain, OAR**

Как выровнять эти две последовательности?

There are 9 sequences with the following architecture:

MethyltransfD12, N6_N4_Mtase

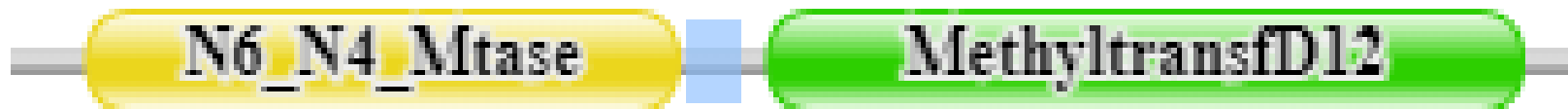
[A0A2Z5QVW5_9MICC](#) [**D12-N6_N4**]



There are 5 sequences with the following architecture:

N6_N4_Mtase, MethyltransfD12

[A0A1I7GYG0_9CLOT](#) [**N6_N4-D12**]



Как такое может возникнуть?

BLAST двух последовательностей

Nucleotide BLAST: Search nucleotide: x

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) [?](#)

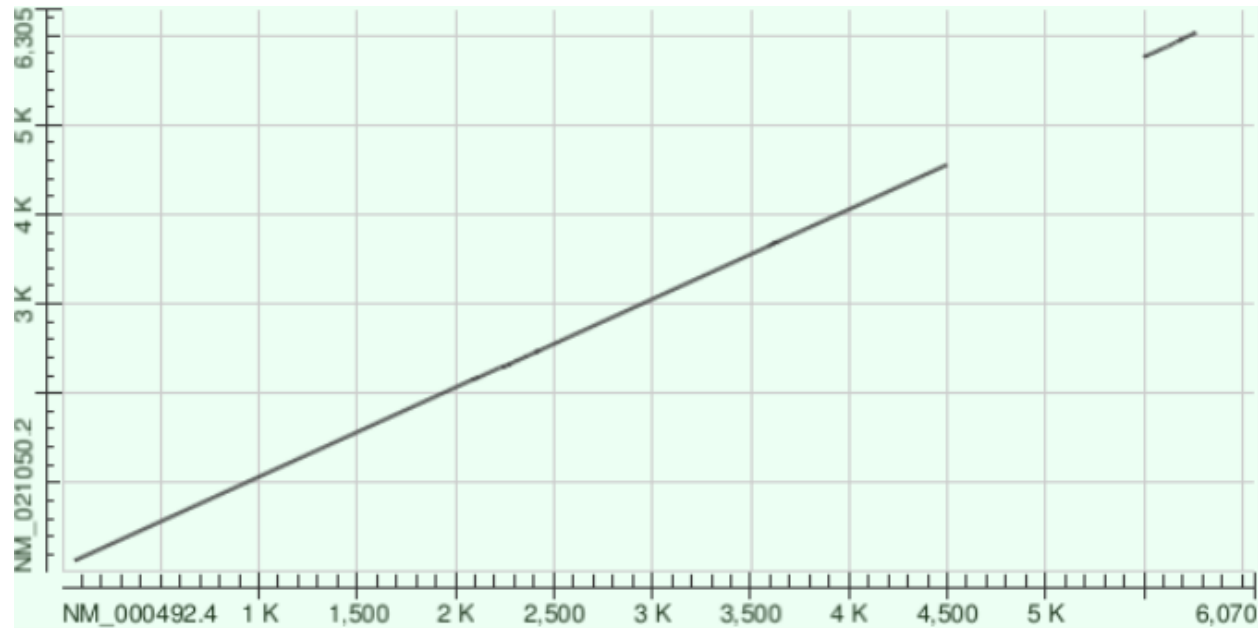
Organism Optional exclude [Add organism](#)

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

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

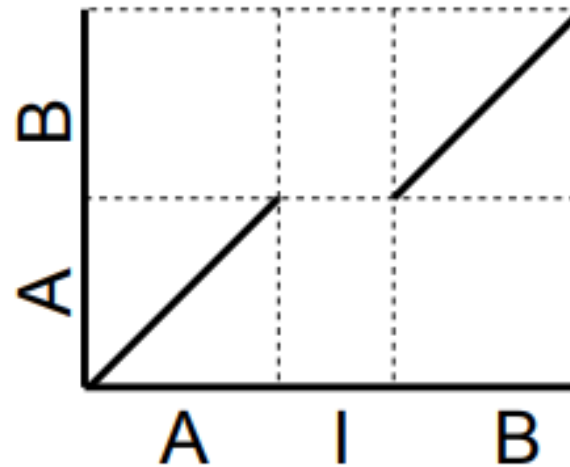
Карта локального сходства

- ❖ Визуальное представление парного выравнивания
- ❖ Отрезки отображают наличие локального выравнивания



Делеция или вставка

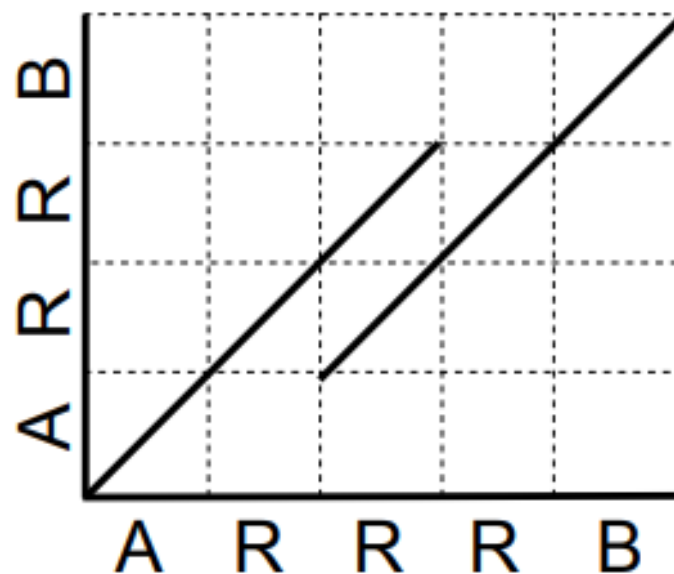
R: AIB
Q: AB



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

Дупликация

R: ARRRB
Q: ARRB

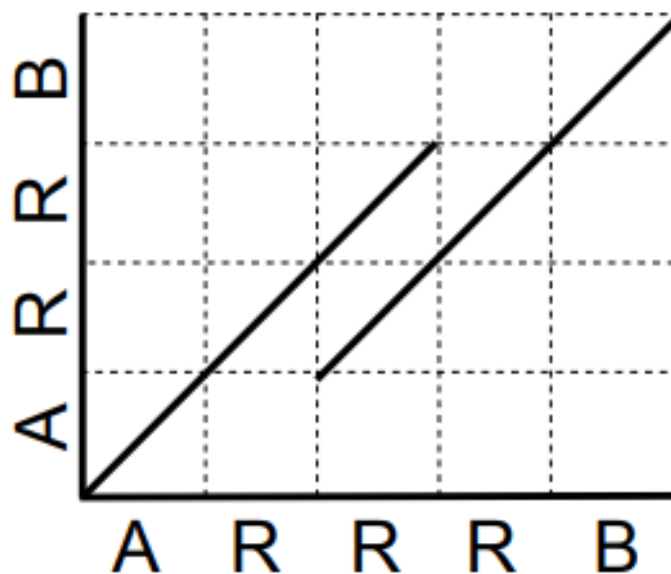


❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

Дупликация

На графике чего-то не хватает...

R: ARRRB
Q: ARRB



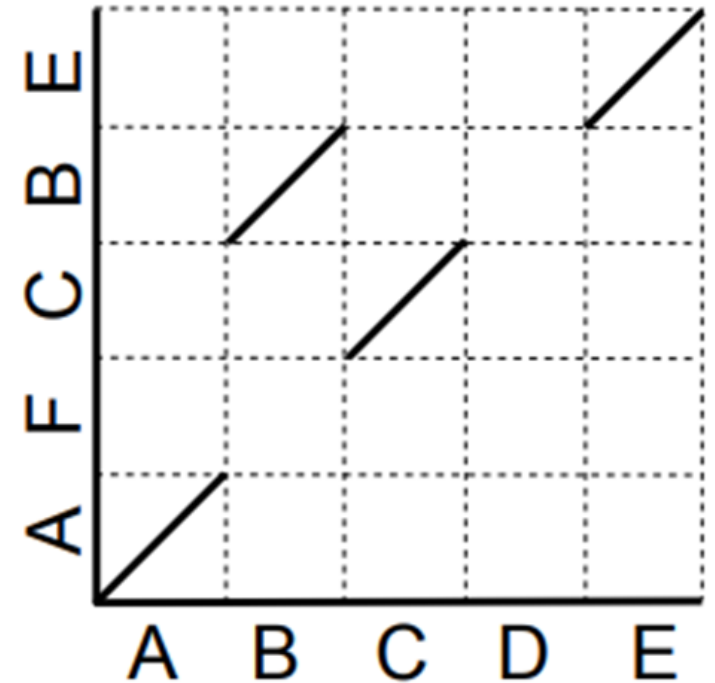
❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

Транслокация

Транслокация – перенесение фрагмента в другое место.

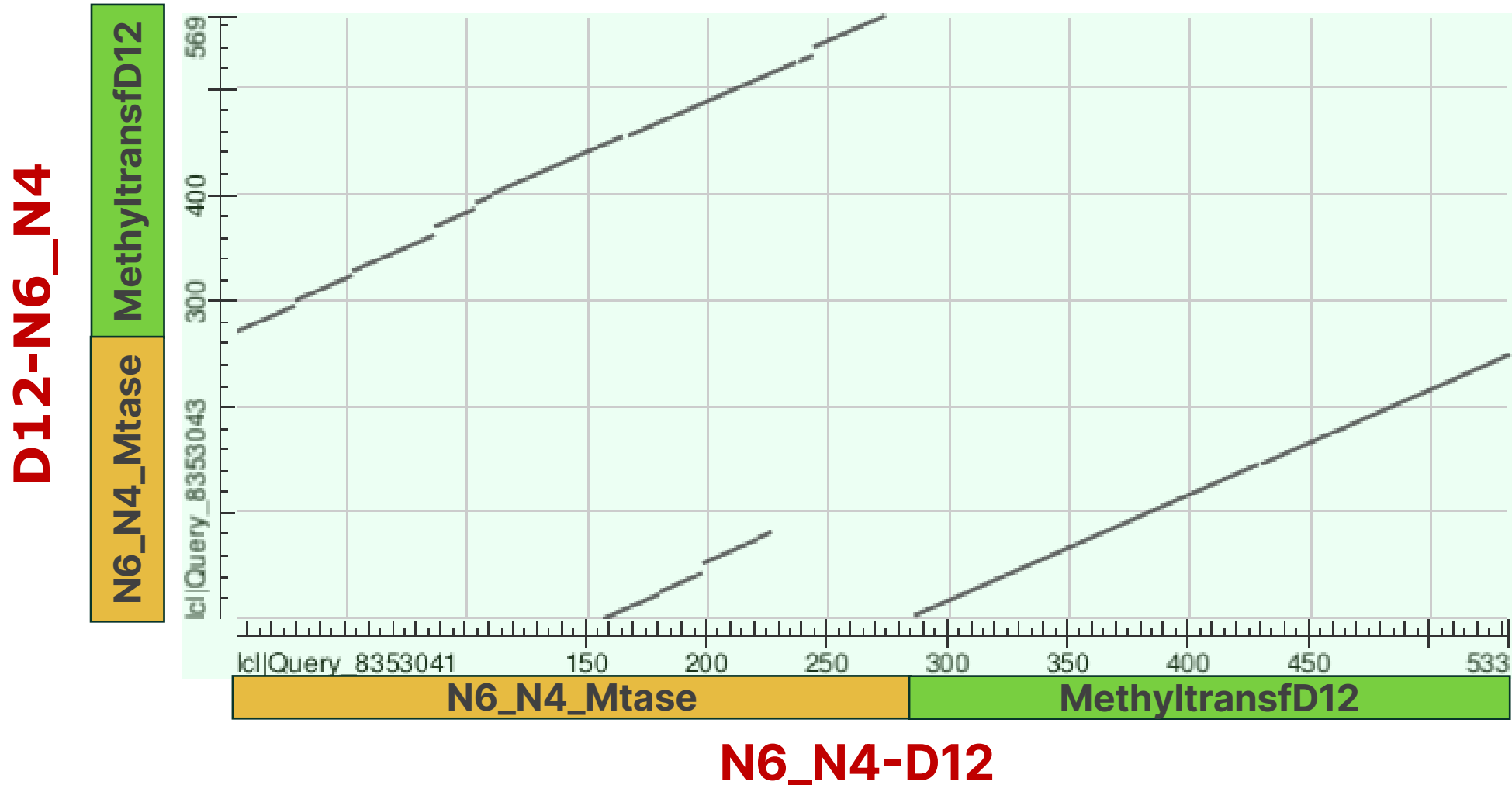
С сохранением ориентации или нет.

R: ABCDE
Q: AFCBE

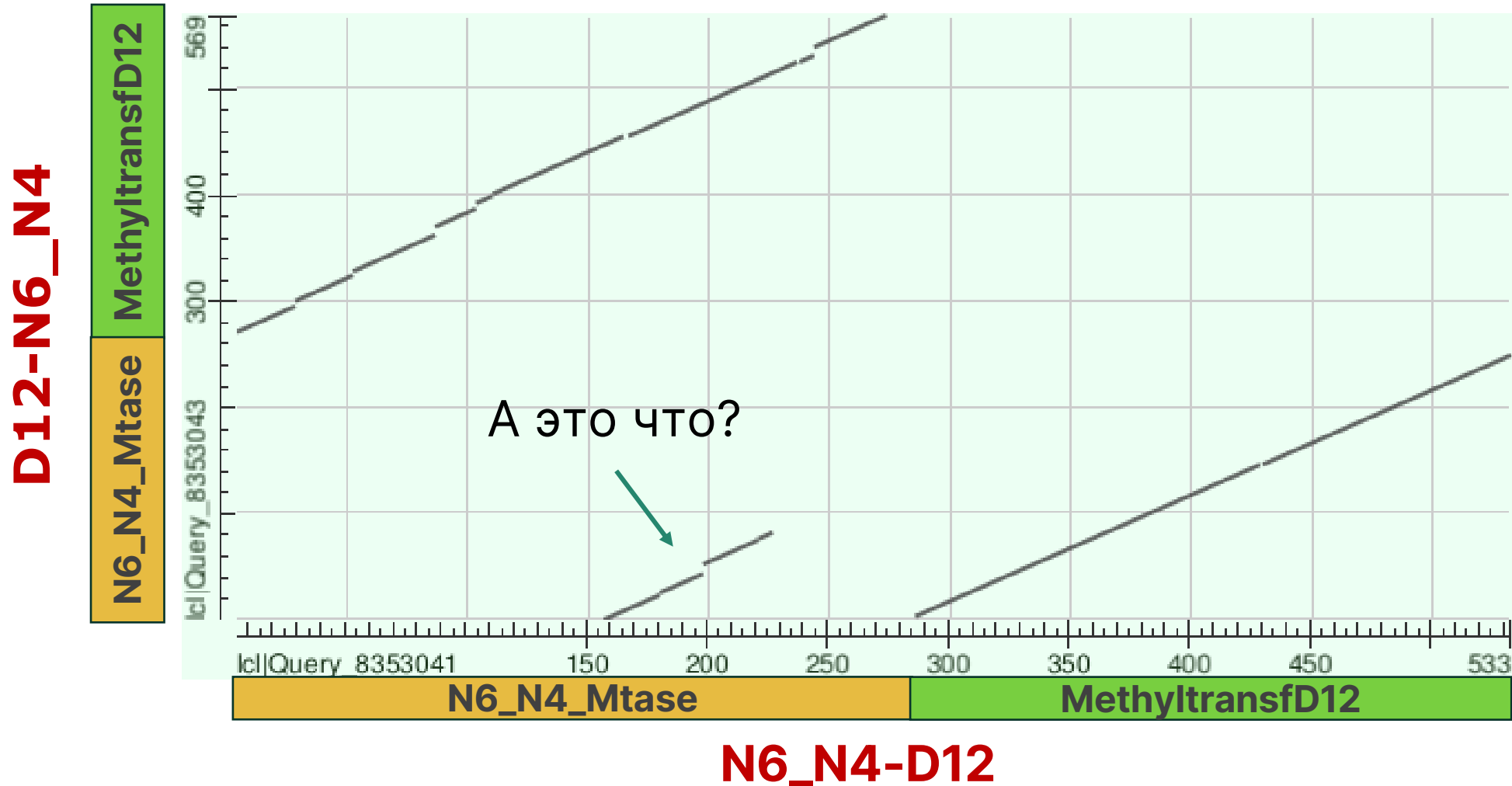


❖ <https://mummer.sourceforge.net/manual/AlignmentTypes.pdf>

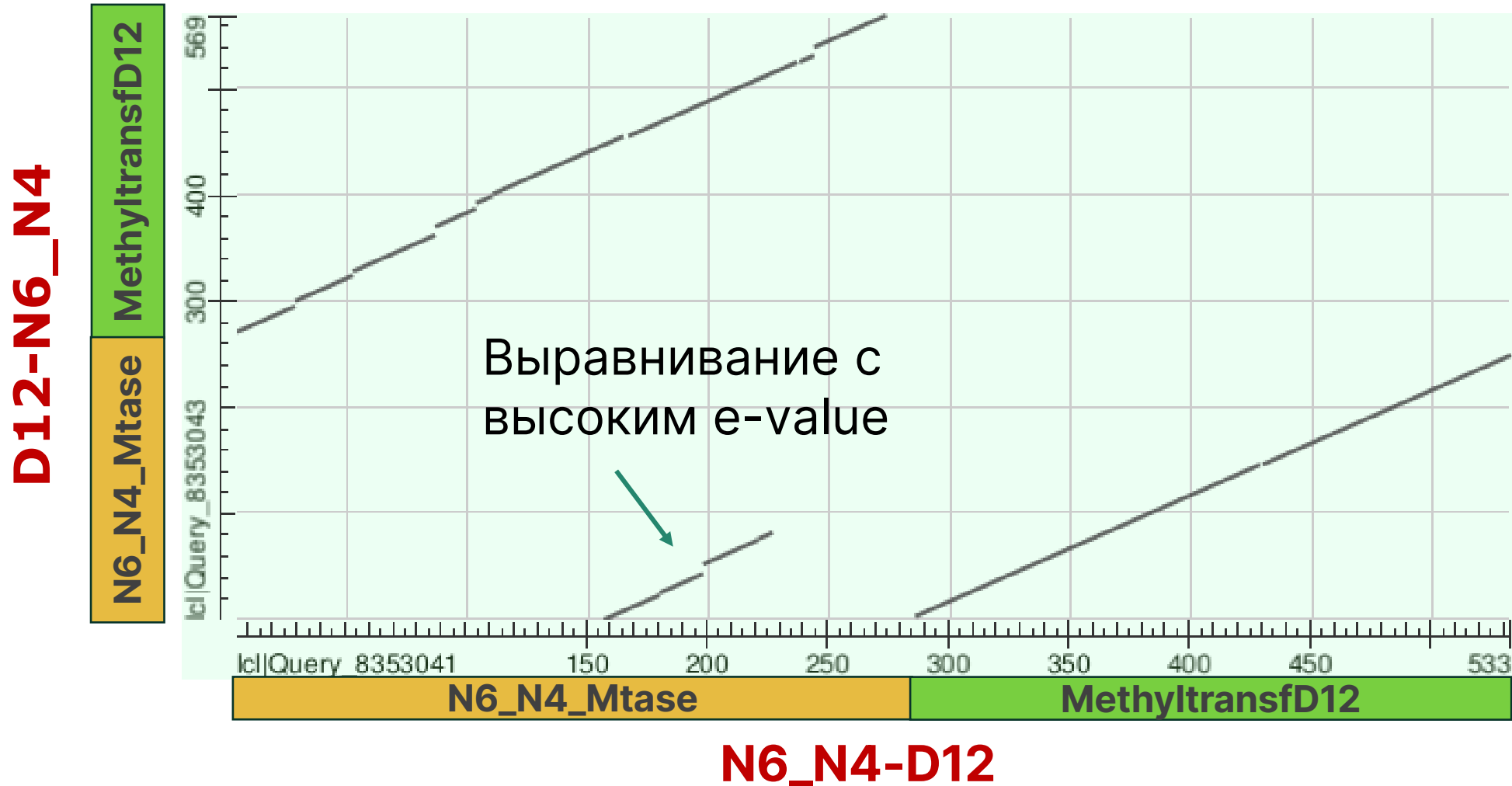
Карта локального сходства – множественные локальные выравнивания



Карта локального сходства – множественные локальные выравнивания



Карта локального сходства – множественные локальные выравнивания



JaView – редактор выравниваний

Полезные умения в Jalview

| Действие | меню | подменю | варианты | комментарии |
|---------------------------------------|---------------------------|---|---|---|
| Импорт последовательностей | File | Add sequences | <ul style="list-style-type: none"> 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 <hr/> Calculate | Make groups for selection <hr/> 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 |
| Создание новой вкладки с тем же выравниванием | View | New View | | Для создания альтернативных вариантов визуализации без дублирования окна выравнивания |
| | 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 | File | Save as | Выбрать формат выравнивания | |
| Сохранение проекта со всеми окнами | File | Save project as | | |

**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 |
|---------------------------|---|----|----|
| <i>K9Z0V0_CYAAP/42-80</i> | 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 | | |
| <i>K9SST0_9SYNE/42-80</i> | 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 | | |
| <i>K9SGD2_9CYAN/42-80</i> | 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 | | |
| <i>K9Y0S0_STAC7/42-80</i> | 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 | | |
| <i>PSBE_RIPO1/42-80</i> | 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 | | |
| <i>PSBE_SYNY3/42-80</i> | 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 | | |
| <i>PSBE_GLOC7/42-80</i> | 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 | | |
| <i>K9W859_9CYAN/42-80</i> | 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 | | |
| <i>PSBE_SYNE7/43-81</i> | 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 | | |
| <i>PSBE_ACAM1/42-80</i> | 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 | | |
| <i>C7B2G7_SELML/42-80</i> | 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 | | |
| <i>PSBE_PHATG/43-81</i> | 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 | | |
| <i>PSBE_OSTTA/42-79</i> | 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 - | | |
| <i>PSBE_CHLRE/42-79</i> | 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 - | | |
| <i>PSBE_SYNJB/42-80</i> | 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 | | |
| <i>K9TLB2_9CYAN/43-80</i> | 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 - | | |
| <i>K9S2L1_9CYAN/42-80</i> | 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 | | |
| <i>K9YE64_HALP7/42-80</i> | L A Y 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 | | |
| <i>K9R7L0_9CYAN/42-80</i> | 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 | | |
| <i>PSBE_THEVB/42-80</i> | 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 | | |
| <i>PSBE_GLOW/43-81</i> | 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 | | |
| <i>PSBE_SYNR3/43-81</i> | 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 | | |
| <i>PSBE_PROMA/43-81</i> | 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

Определения

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

| | 10 | 20 | 30 |
|---------------------------|---|-----------------------------------|----|
| <i>K9Z0V0_CYAAP/42-80</i> | 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 | |
| <i>K9SST0_9SYNE/42-80</i> | 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 | |
| <i>K9SGD2_9CYAN/42-80</i> | 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 | |
| <i>K9Y0S0_STAC7/42-80</i> | 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 | |
| <i>PSBE_RIPO1/42-80</i> | 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 | |
| <i>PSBE_SYNY3/42-80</i> | 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 | |
| <i>PSBE_GLOC7/42-80</i> | 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 | |
| <i>PSBE_OSTTA/42-79</i> | 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 | |
| <i>K9W859_9CYAN/42-80</i> | L A Y D V F G T P R P N E Y F T Q D R V E A P | P I L S D R Y E A K Q Q I Q Q F | |
| <i>PSBE_SYNE7/43-81</i> | 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 | |

Блок с гэпами

гэп

| | 10 | 20 | 30 |
|---------------------------|---|-----------------------------------|----|
| <i>K9Z0V0_CYAAP/42-80</i> | 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 | |
| <i>K9SST0_9SYNE/42-80</i> | 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 | |
| <i>K9SGD2_9CYAN/42-80</i> | 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 | |
| <i>K9Y0S0_STAC7/42-80</i> | 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 | |
| <i>PSBE_RIPO1/42-80</i> | 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 | |
| <i>PSBE_SYNY3/42-80</i> | 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 | |
| <i>PSBE_GLOC7/42-80</i> | 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 | |
| <i>PSBE_OSTTA/42-79</i> | 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 | |
| <i>K9W859_9CYAN/42-80</i> | L A Y D V F G T P R P N E Y F T Q D R V E A P | P I L S D R Y E A K Q Q I Q Q F L | |
| <i>PSBE_SYNE7/43-81</i> | 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 | |
| <i>PSBE_ACAM1/42-80</i> | 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 | |

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_GLOV/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 | 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_9NOSO/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 | LAYDVF | GTPRPDEY | FTQTRQELPIINDR |
| K9SST0_9SYNE/42-80 | LAYDIF | FGTTRPNE | YTTNDRQEAPILRDR |
| K9SGD2_9CYAN/42-80 | LAYDVF | GTPRPNE | YTLDRSDAPVLLDR |
| K9Y0S0_STAC7/42-80 | LAYDVF | GTPRPNE | YFTQDRLELPIINDR |
| PSBE_RIPO1/42-80 | LAYDVF | GTPRPD | QYFTQERQELPIISDR |
| PSBE_SYNY3/42-80 | LAYDAF | GTPRPDE | YFTQTRQELPILQER |
| PSBE_GLOC7/42-80 | LAYDVF | GTPRPDE | YTTQERLELPIILKDR |
| K9W859_9CYAN/42-80 | LAYDVF | GTPRPNE | YFTQDRVEAPILSDR |
| PSBE_SYNE7/43-81 | LAYDAF | GTPRPNE | YFTQDRTEVPIVSDR |
| PSBE_ACAM1/42-80 | LAYDVF | GTPRPNE | YFDQARQGLPLVTD |
| C7B2G7_SELML/42-80 | LAYDV | SGSPRPNE | YSTESRRDIPLITGR |
| PSBE_PHATC/43-81 | LAYDVF | GTPRPNE | YFTQDRQQIPLVND |
| PSBE_OSTTA/42-79 | LAYDVF | GTPRPNE | YFTEERQELPLISDR |
| PSBE_CHLRE/42-79 | LAYDVF | LRETPRPNE | YFTEDRQEAPLITDR |
| PSBE_SYNJB/42-80 | LAYDIF | FGTTRPNE | YFTAERQELPIVSDR |
| K9TLB2_9CYAN/43-80 | LAYDAF | GTPRPNE | YSQERLELPIVQDR |
| K9S2L1_9CYAN/42-80 | LAYDAF | GTPRPDE | YFTQQRQELPIVMDR |
| K9YE64_HALP7/42-80 | LAYDAF | GTPRPDE | YFTKQRQELPIISDR |
| K9R7L0_9CYAN/42-80 | LAYDAF | GTPRPD | TYFTQERQEVPLVND |
| PSBE_THEVB/42-80 | LAYDVF | GTPRPD | SYAQEQRSIPLVTD |
| PSBE_GLOWV/43-81 | LAYDVF | GTPRPDE | YNSDNTKKPLVNKR |
| PSBE_SYNR3/43-81 | LAYDAF | GTPRPD | SYFQAADVKA |
| PSBE_PROMA/43-81 | LAYDAF | GTPRPD | TYFQASETKA |
| K9RTX4_SYNP3/42-80 | LAYDIF | FGTTRP | DTYFSQEQQAAPVVS |
| K9ULI3_CHAP6/43-81 | LAYDAF | GTPRPDE | YFTQTRQELPIVSDR |
| K9UY34_9CYAN/42-79 | LAYDAF | GTPRPNE | YFTQTRQEVPIVNNR |
| PSBE_NOSS1/42-80 | LAYDVF | GTPRPDE | YTTQARQELPIVNNR |
| K9QJT8_9NOSQ/42-80 | LAYDVF | GTPRPNE | YTTQTRQEVPIVKNR |
| K8GLK0_9CYAN/42-80 | LAYDVF | GTPRPNE | YTTSTRQQAPIVKDR |
| K9VJB4_9CYAN/42-80 | LAYDAF | GTPRP | NQYFTPQREEVPIISDR |
| K9W2B5_9CYAN/42-80 | LAYDAF | GTPRPNE | YFTQQRQEVPIVSDR |
| K9TZN4_CHRTP/42-80 | LAYDVF | GTPRPNE | YYPQERQELPIVSDR |
| K9XG15_9CHRO/42-80 | LAYDVF | GTPRPNE | YYPQERQELPIIGDR |

Пример достоверного блока

| | | 10 | 20 | 30 | |
|--------------------|--------|----|--------|------------|---------------------|
| K9Z0V0_CYAAP/42-80 | LAYDVF | GT | PRPDEY | FTQTRQELPI | INDRYEAKNQIEQFN |
| K9SST0_9SYNE/42-80 | LAYDIF | GT | PRPNEY | YTNDRQEAPI | RDRYDANQEIPNYN |
| K9SGD2_9CYAN/42-80 | LAYDVF | GT | PRPNEY | YTLDRSDAPV | LDRFN AIDDIEDFN |
| K9Y0S0_STAC7/42-80 | LAYDVF | GT | PRPNEY | FTQDRLELPI | INDRYSASEQIKEFN |
| PSBE_RIPO1/42-80 | LAYDVF | GT | PRPDQY | FTQERQELPI | SDRYKASQQIQEFN |
| PSBE_SYNY3/42-80 | LAYDAF | GT | PRPDEY | FTQTRQELPI | QERYDINQEQEFN |
| PSBE_GLOC7/42-80 | LAYDVF | GT | PRPDEY | YTQERLELPI | KDRYNTDQQIQEFN |
| K9W859_9CYAN/42-80 | LAYDVF | GT | PRPNEY | FTQDRVEAPI | SDRYEAKQQIQQFL |
| PSBE_SYNE7/43-81 | LAYDAF | GT | PRPNEY | FTQDRTEVP | VSDRYSAKQQVDRFS |
| PSBE_ACAM1/42-80 | LAYDVF | GT | PRPNEY | FDQARQGLPL | VTDRYEGKQQIDEFT |
| C7B2G7_SELML/42-80 | LAYDVS | GS | PRPNEY | STESRRDIP | LITGRFDSSEQIDEFT |
| PSBE_PHATC/43-81 | LAYDVF | GT | PRPNEY | FTQDRQQIPL | VNDRFSAKQLEDLT |
| PSBE_OSTTA/42-79 | LAYDVF | GT | PRPNEY | FTEERQELPI | SDRYGALAQLDDL |
| PSBE_CHLRE/42-79 | LAYDVF | GT | PRPNEY | FTEDRQEAPL | ITDRFNALEQVKKL |
| PSBE_SYNJB/42-80 | LAYDIF | GT | PRPNEY | FTAERQELPI | VSDRFNALEELERLT |
| K9TLB2_9CYAN/43-80 | LAYDAF | GT | PRPNEY | FSQERLELPI | VQDRFNAKDQIDDL |
| K9S2L1_9CYAN/42-80 | LAYDAF | GT | PRPDEY | FTQQRQELPI | VMDRFGAKQQVEEF |
| K9YE64_HALP7/42-80 | LAYDAF | GT | PRPDEY | FTKQRQELPI | ISDRFQAREEVEEFN |
| K9R7L0_9CYAN/42-80 | LAYDAF | GT | PRPDTY | FTQERQEVPL | VNDRFEAKNQVDVFT |
| PSBE_THEVB/42-80 | LAYDVF | GT | PRPDSY | YAQEQRSIPL | VTDREFEAKQQVETFL |
| PSBE_GLOWV/43-81 | LAYDVF | GT | PRPDEY | YNSDNTKKPL | VNKRFEAKQQLDDEAT |
| PSBE_SYNR3/43-81 | LAYDAF | GT | PRPDSY | FQAADVKA | PVVSQRYEGKSQLDTRL |
| PSBE_PROMA/43-81 | LAYDAF | GT | PRPDTY | FQASET | KAPVVSQRFESKAQLDLRL |
| K9RTX4_SYNP3/42-80 | LAYDIF | GT | PRPDTY | FSQEQAAP | VVSDRFESKQQINEFL |
| K9ULI3_CHAP6/43-81 | LAYDAF | GT | PRPDEY | FTQTRQELPI | VSDRFESKGQLDKFV |
| K9UY34_9CYAN/42-79 | LAYDAF | GT | PRPNEY | FTQTRQEVPI | VNDRFDAKKQVETY |
| PSBE_NOSS1/42-80 | LAYDVF | GT | PRPDEY | YTQARQELPI | VNDRFEAKKQVEQLI |
| K9QJT8_9NOSQ/42-80 | LAYDVF | GT | PRPNEY | YTQTRQEVPI | VKNRYEAKQQVEKFI |
| K8GLK0_9CYAN/42-80 | LAYDVF | GT | PRPNEY | YTSTRQQAP | IVKDRFEAKQQIDQFI |
| K9VJB4_9CYAN/42-80 | LAYDAF | GT | PRPNQY | FTPQREEVPI | ISDRFEAKKQVDQFI |
| K9W2B5_9CYAN/42-80 | LAYDAF | GT | PRPNEY | FTQQRQEVPI | VSDRFESKQIDQFL |
| K9TZN4_CHRTP/42-80 | LAYDVF | GT | PRPNEY | YPQERQELPI | VSDRFESRKQIEEFI |
| K9XG15_9CHRO/42-80 | LAYDVF | GT | PRPNEY | YPQERQELPI | IGDRYQAKQEI |

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 | LAYDVF | GT | PRPDEYFTQTRQELPIINDRYEAKNQIEQFN |
| K9SST0_9SYNE/42-80 | LAYDIF | GT | PRPNEYTNDRQEAPILRDYDANQEIPNYN |
| K9SGD2_9CYAN/42-80 | LAYDVF | GT | PRPNEYTLDLRSAPVLLDRFNAIDDIEDFN |
| K9Y0S0_STAC7/42-80 | LAYDVF | GT | PRPNEYFTQDRLELPIINDRYSASEQIKEFN |
| PSBE_RIPO1/42-80 | LAYDVF | GT | PRPDQYFTQERQELPIISDRYKASQIQEFN |
| PSBE_SYNY3/42-80 | LAYDAF | GT | PRPDEYFTQTRQELPILQERYDINQEQEFN |
| PSBE_GLOC7/42-80 | LAYDVF | GT | PRPDEYFTQERLELPIILKDRYNTDQIQEFN |
| K9W859_9CYAN/42-80 | LAYDVF | GT | PRPNEYFTQDRVEAPILSDRYEAKQIQQFL |
| PSBE_SYNE7/43-81 | LAYDAF | GT | PRPNEYFTQDRTEVPVSDRYSAKQVDRFS |
| PSBE_ACAM1/42-80 | LAYDVF | GT | PRPNEYFDQARQGLPLVTDREYEGKQIQDEF |
| C7B2G7_SELML/42-80 | LAYDVSG | SP | PRPNEYSTESRRDIPITGRFDSSEQIDEFT |
| PSBE_PHATC/43-81 | LAYDVF | GT | PRPNEYFTQDRQQIPLVNDRFSAKQLEDLT |
| PSBE_OSTTA/42-79 | LAYDVF | GT | PRPNEYFTEERQELPLISDRYGALAQLDDL |
| PSBE_CHLRE/42-79 | LAYDVF | GT | PRPNEYFTEDRQEAPLITDRFNALEQVKKL |
| PSBE_SYNJB/42-80 | LAYDIF | GT | PRPNEYFTAERQELPIVSDRFNALEELERLT |
| K9TLB2_9CYAN/43-80 | LAYDAF | GT | PRPNEYFSQERLELPIVQDRFNAKDQIDDL |
| K9S2L1_9CYAN/42-80 | LAYDAF | GT | PRPDEYFTQQRQELPIVMDRFGAKQVVEEF |
| K9YE64_HALP7/42-80 | LAYDAF | GT | PRPDEYFTKQRQELPIISDRFQAREEVEEFN |
| K9R7L0_9CYAN/42-80 | LAYDAF | GT | PRPDTYFTQERQEVPLVNDRF EAKNQVDVFT |
| PSBE_THEVB/42-80 | LAYDVF | GT | PRPDSY YAQEQRSIPLVTDREAKQVETFL |
| PSBE_GLOWV/43-81 | LAYDVF | GT | PRPDEY YNSDNTKKPLVNKRFEAKQQLDEAT |
| PSBE_SYNR3/43-81 | LAYDAF | GT | PRPDSY FQAADV KAPVVSQR YEGKSQLDTRL |
| PSBE_PROMA/43-81 | LAYDAF | GT | PRPDTY FQASET KAPVVSQR FESKAQLDLRL |
| K9RTX4_SYNP3/42-80 | LAYDIF | GT | PRPDTY FSQEQQAAPVVSDR FESKQQINEFL |
| K9ULI3_CHAP6/43-81 | LAYDAF | GT | PRPDEY FTQTRQELPIVSDRFESKGQLDKFV |
| K9UY34_9CYAN/42-79 | LAYDAF | GT | PRPNEY FTQTRQEVPIVNNRFD AKKQVETY |
| PSBE_NOSS1/42-80 | LAYDVF | GT | PRPDEY YTQARQELPIVNNRFEAKKQVEQLI |
| K9QJT8_9NOSQ/42-80 | LAYDVF | GT | PRPNEY YTQTRQEVPIVKNRYEAKQVVEKFI |
| K8GLK0_9CYAN/42-80 | LAYDVF | GT | PRPNEY YTSTRQQAPIVKDRFEAKQIQDFI |
| K9VJB4_9CYAN/42-80 | LAYDAF | GT | PRPNQY FTPQREEVPIISDRFEAKKQVDQFI |
| K9W2B5_9CYAN/42-80 | LAYDAF | GT | PRPNEY FTQQRQEVPIVSDRFESKQEQIDQFL |
| K9TZN4_CHRTP/42-80 | LAYDVF | GT | PRPNEY YPQERQELPIVSDRFESRKQIEEFI |
| K9XG15_9CHRO/42-80 | LAYDVF | GT | PRPNEY YPQERQELPIIGDRYQAKQEQIKQFI |

Это максимальный достоверный блок?

| | 10 | 20 | 30 |
|--------------------|---------|----|----------------------------------|
| K9Z0V0_CYAAP/42-80 | LAYDVF | GT | PRPDEYFTQTRQELPIINDRYEAKNQIEQFN |
| K9SST0_9SYNE/42-80 | LAYDIF | GT | PRPNEYTNDRQEAPILRDYDANQEIPNYN |
| K9SGD2_9CYAN/42-80 | LAYDVF | GT | PRPNEYTLDLRSAPVLLDRFNAIDDIEDFN |
| K9Y0S0_STAC7/42-80 | LAYDVF | GT | PRPNEYFTQDRLELPIINDRYSASEQIKEFN |
| PSBE_RIPO1/42-80 | LAYDVF | GT | PRPDQYFTQERQELPIISDRYKASQIQEFN |
| PSBE_SYNY3/42-80 | LAYDAF | GT | PRPDEYFTQTRQELPILQERYDINQEQEFN |
| PSBE_GLOC7/42-80 | LAYDVF | GT | PRPDEYFTQERLELPIILKDRYNTDQIQEFN |
| K9W859_9CYAN/42-80 | LAYDVF | GT | PRPNEYFTQDRVEAPILSDRYEAKQIQQFL |
| PSBE_SYNE7/43-81 | LAYDAF | GT | PRPNEYFTQDRTEVPVSDRYSAKQVDRFS |
| PSBE_ACAM1/42-80 | LAYDVF | GT | PRPNEYFDQARQGLPLVTDRYEGKQIDEFT |
| C7B2G7_SELML/42-80 | LAYDVSG | SP | PRPNEYSTESRRDIPLITGRFDSSEQIDEFT |
| PSBE_PHATC/43-81 | LAYDVF | GT | PRPNEYFTQDRQQIPLVNDRFSAKQLEDLT |
| PSBE_OSTTA/42-79 | LAYDVF | GT | PRPNEYFTEERQELPLISDRYGALAQLDDL |
| PSBE_CHLRE/42-79 | LAYDVF | GT | PRPNEYFTEDRQEAPLITDRFNALEQVKKL |
| PSBE_SYNJB/42-80 | LAYDIF | GT | PRPNEYFTAERQELPIVSDRFNALEELERLT |
| K9TLB2_9CYAN/43-80 | LAYDAF | GT | PRPNEYFSQERLELPIVQDRFNAKDQIDDL |
| K9S2L1_9CYAN/42-80 | LAYDAF | GT | PRPDEYFTQQRQELPIVMDRFGAKQVVEFI |
| K9YE64_HALP7/42-80 | LAYDAF | GT | PRPDEYFTKQRQELPIISDRFQAREEVEEFN |
| K9R7L0_9CYAN/42-80 | LAYDAF | GT | PRPDTYFTQERQEVPLVNDRFNALEQVDFVT |
| PSBE_THEVB/42-80 | LAYDVF | GT | PRPDSYQAQEQRSIPLVTDRFNALEQVDFVT |
| PSBE_GLOWV/43-81 | LAYDVF | GT | PRPDEYNSDNTKKPLVNKRFNALEQVDFVT |
| PSBE_SYNR3/43-81 | LAYDAF | GT | PRPDSYFQAADVKAAPVVSQRVEGKSQDLTRL |
| PSBE_PROMA/43-81 | LAYDAF | GT | PRPDTYFQASETKAPVVSQRVEGKSQDLTRL |
| K9RTX4_SYNP3/42-80 | LAYDIF | GT | PRPDTYFSQEQQAAPVVSDRFNALEQVDFVT |
| K9ULI3_CHAP6/43-81 | LAYDAF | GT | PRPDEYFTQTRQELPIVSDRFNALEQVDFVT |
| K9UY34_9CYAN/42-79 | LAYDAF | GT | PRPNEYFTQTRQEVPIVNNRFDAKKQVETY |
| PSBE_NOSS1/42-80 | LAYDVF | GT | PRPDEYFTQARQELPIVNNRFNALEQVDFVT |
| K9QJT8_9NOSO/42-80 | LAYDVF | GT | PRPNEYFTQTRQEVPIVKNRYEAKQVVEKFI |
| K8GLK0_9CYAN/42-80 | LAYDVF | GT | PRPNEYTSTRQQAPIVKDRFEAKQVIDQFI |
| K9VJB4_9CYAN/42-80 | LAYDAF | GT | PRPNQYFTPQREEVPIISDRFEAKQVVDQFI |
| K9W2B5_9CYAN/42-80 | LAYDAF | GT | PRPNEYFTQQRQEVPIVSDRFNALEQVDFVT |
| K9TZN4_CHRTP/42-80 | LAYDVF | GT | PRPNEYYPQERQELPIVSDRFNALEQVDFVT |
| K9XG15_9CHRO/42-80 | LAYDVF | GT | PRPNEYYPQERQELPIIGDRYQAKQEIQKFI |

Или это?

| | 10 | 20 | 30 | | | | | |
|--------------------|------|------|-------|-----------|--------|-------|-----------|-------|
| K9Z0V0_CYAAP/42-80 | LAYD | VFGT | PRPDE | YFTQTR | QELPI | IINDR | YEAKNQI | EQFN |
| K9SST0_9SYNE/42-80 | LAYD | IFGT | TRPNE | YNTDR | QEAP | ILRDR | YDANQEI | PNYN |
| K9SGD2_9CYAN/42-80 | LAYD | VFGT | TRPNE | YTLDR | SDAP | VLLDR | FNAIDDI | EDFN |
| K9Y0S0_STAC7/42-80 | LAYD | VFGT | TRPNE | YFTQDR | LELP | IINDR | YSASEQI | KEFN |
| PSBE_RIPO1/42-80 | LAYD | VFGT | TRPDE | YFTQER | QELPI | ISDR | YKASQQI | QEFN |
| PSBE_SYNY3/42-80 | LAYD | AFGT | TRPDE | YFTQTR | QELPI | LQERY | DINQEI | QEFN |
| PSBE_GLOC7/42-80 | LAYD | VFGT | TRPDE | YTTQER | LELP | ILKDR | YNTDQQI | QEFN |
| K9W859_9CYAN/42-80 | LAYD | VFGT | TRPNE | YFTQDR | VEAP | ILSDR | YEAKQQI | QQFL |
| PSBE_SYNE7/43-81 | LAYD | AFGT | TRPNE | YFTQDR | TEVP | IVSDR | YSAKQQV | DRFS |
| PSBE_ACAM1/42-80 | LAYD | VFGT | TRPNE | YFDQAR | QGLP | LVTD | RYEGKQQI | DEFT |
| C7B2G7_SELML/42-80 | LAYD | VSGS | TRPNE | YSTES | RRDI | PLITG | RFDSSQI | DEFT |
| PSBE_PHATC/43-81 | LAYD | VFGT | TRPNE | YFTQDR | QQIP | LVNDR | FSAKQEL | EDLT |
| PSBE_OSTTA/42-79 | LAYD | VFGT | TRPNE | YFTEER | QELP | LISDR | YGALAQL | DDL |
| PSBE_CHLRE/42-79 | LAYD | VFGT | TRPNE | YFTEDR | QEAP | LITDR | FNALEQV | KKL |
| PSBE_SYNJB/42-80 | LAYD | IFGT | TRPNE | YFTAER | QELPI | IVSDR | FNALEEL | ERLT |
| K9TLB2_9CYAN/43-80 | LAYD | AFGT | TRPNE | YSQER | LELP | IVQDR | FNAKDQI | DDL |
| K9S2L1_9CYAN/42-80 | LAYD | AFGT | TRPDE | YFTQQR | QELPI | VMDR | FGAKQQV | EEFI |
| K9YE64_HALP7/42-80 | LAYD | AFGT | TRPDE | YFTKQR | QELPI | ISDR | FQAREEV | EEFN |
| K9R7L0_9CYAN/42-80 | LAYD | AFGT | TRPDT | YFTQER | QEVPL | VNDR | F EAKNQV | DVFT |
| PSBE_THEVB/42-80 | LAYD | VFGT | TRPDS | Y YAQEQRS | IPL | VTD | RFEAKQQV | ETFL |
| PSBE_GLOWV/43-81 | LAYD | VFGT | TRPDE | YNSDNT | KKPL | VNKR | RFEAKQQL | DEAT |
| PSBE_SYNR3/43-81 | LAYD | AFGT | TRPDS | YFQAADV | KAPV | VVSQR | YEGKSQ | LDTRL |
| PSBE_PROMA/43-81 | LAYD | AFGT | TRPDT | YFQASET | KAPV | VVSQR | RFEASKAQL | DLRL |
| K9RTX4_SYNP3/42-80 | LAYD | IFGT | TRPDT | YFSQEQQA | APV | VSDR | RFEASKQQ | INEFL |
| K9ULI3_CHAP6/43-81 | LAYD | AFGT | TRPDE | YFTQTR | QELPI | IVSDR | RFEASKGQL | DKFV |
| K9UY34_9CYAN/42-79 | LAYD | AFGT | TRPNE | YFTQTR | QEVPI | VNDR | RFDKAKQV | ETY |
| PSBE_NOSS1/42-80 | LAYD | VFGT | TRPDE | YTTQAR | QELPI | VNDR | RFEAKKQV | EQLI |
| K9QJT8_9NOSQ/42-80 | LAYD | VFGT | TRPNE | YTTQTR | QEVPI | VKNR | RYEAKQQV | EKFI |
| K8GLK0_9CYAN/42-80 | LAYD | VFGT | TRPNE | YTTSTR | QQAP | IVKDR | RFEAKQQI | DQFI |
| K9VJB4_9CYAN/42-80 | LAYD | AFGT | TRPN | QYFTP | QREEVP | IISDR | RFEAKKQV | DQFI |
| K9W2B5_9CYAN/42-80 | LAYD | AFGT | TRPNE | YFTQQR | QEVPI | VSDR | RFEASKQE | IDQFL |
| K9TZN4_CHRTP/42-80 | LAYD | VFGT | TRPNE | YYPQER | QELPI | VSDR | RFEARKQI | EEFI |
| K9XG15_9CHRO/42-80 | LAYD | VFGT | TRPNE | YYPQER | QELPI | IGDR | RYQAKQE | IKQFI |

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

А, может, это максимально достоверный блок?

| | 10 | 20 | 30 | | | | | | | |
|--------------------|------|------|--------|------------|--------------|--------------|----------|---------|-------|-------|
| K9Z0V0_CYAAP/42-80 | LAYD | VFGT | PRPDEY | FTQTRQELPI | IINDRYEAKNQI | EQFN | | | | |
| K9SST0_9SYNE/42-80 | LAYD | IFGT | TRPNE | YNTDRQEAP | ILRD | RYDANQEI | PNYN | | | |
| K9SGD2_9CYAN/42-80 | LAYD | VFGT | TRPNE | YTLDRSDAP | VLLDR | FNAIDDI | EDFN | | | |
| K9Y0S0_STAC7/42-80 | LAYD | VFGT | TRPNE | YFTQDR | LELPI | IINDRYSASEQ | IKEFN | | | |
| PSBE_RIPO1/42-80 | LAYD | VFGT | TRP | DQYFTQER | QELPI | ISDRYKASQQ | IQEFN | | | |
| PSBE_SYNY3/42-80 | LAYD | AFGT | TRPDE | YFTQTR | QELPI | LQERYDINQE | IQEFN | | | |
| PSBE_GLOC7/42-80 | LAYD | VFGT | TRPDE | YTTQER | LELPI | LKDRYNTDQQ | IQEFN | | | |
| K9W859_9CYAN/42-80 | LAYD | VFGT | TRPNE | YFTQDR | VEAP | ILSDRYEAKQQ | IQQFL | | | |
| PSBE_SYNE7/43-81 | LAYD | AFGT | TRPNE | YFTQDR | TEVP | IVSDRYSAKQQV | DRFS | | | |
| PSBE_ACAM1/42-80 | LAYD | VFGT | TRPNE | YFDQAR | QGLPL | VTDREYEGKQQ | IDEF | | | |
| C7B2G7_SELML/42-80 | LAYD | VSGS | TRPNE | YSTES | RRDI | PLITGR | FDSSEQ | IDEF | | |
| PSBE_PHATC/43-81 | LAYD | VFGT | TRPNE | YFTQDR | QQI | PLVND | RFSAKQEL | EDLT | | |
| PSBE_OSTTA/42-79 | LAYD | VFGT | TRPNE | YFTEER | QELPI | ISDRYGALAQL | DDL | | | |
| PSBE_CHLRE/42-79 | LAYD | VFLG | TRPNE | YFTEDR | QEAP | LITDR | FNALEQV | KKL | | |
| PSBE_SYNJB/42-80 | LAYD | IFGT | TRPNE | YFTAER | QELPI | IVSDRF | NALEEL | ERLT | | |
| K9TLB2_9CYAN/43-80 | LAYD | AFGT | TRPNE | YSQER | LELPI | VQDR | FNAKDQ | IDL | | |
| K9S2L1_9CYAN/42-80 | LAYD | AFGT | TRPDE | YFTQQR | QELPI | VMDR | FGAKQQV | EEFI | | |
| K9YE64_HALP7/42-80 | LAYD | AFGT | TRPDE | YFTKQR | QELPI | ISDR | FQAREEV | EEFN | | |
| K9R7L0_9CYAN/42-80 | LAYD | AFGT | TRP | DTYFTQER | QEVPL | VND | FEAKNQV | DVFT | | |
| PSBE_THEVB/42-80 | LAYD | VFGT | TRP | DSY | YAQE | QRSI | PLVTD | FEAKQQV | ETFL | |
| PSBE_GLOWV/43-81 | LAYD | VFGT | TRPDE | YNSD | NTKK | PLVN | KRFEAK | QQLD | DEAT | |
| PSBE_SYNR3/43-81 | LAYD | AFGT | TRP | DSY | FQAAD | VKAP | VVSQR | YEGK | SQLD | TRL |
| PSBE_PROMA/43-81 | LAYD | AFGT | TRP | DTY | FQAS | EKAP | VVSQR | RFES | KAQL | DLRL |
| K9RTX4_SYNP3/42-80 | LAYD | IFGT | TRP | DTY | FSQE | QQAAP | VVSDR | RFES | KQQ | INEFL |
| K9ULI3_CHAP6/43-81 | LAYD | AFGT | TRPDE | YFTQTR | QELPI | IVSDR | RFES | KGQL | DKFV | |
| K9UY34_9CYAN/42-79 | LAYD | AFGT | TRPNE | YFTQTR | QEVPI | VNNR | FD | AKKQ | VETY | |
| PSBE_NOSS1/42-80 | LAYD | VFGT | TRPDE | YTTQAR | QELPI | VNNR | FEAK | KQVE | QLI | |
| K9QJT8_9NOSQ/42-80 | LAYD | VFGT | TRPNE | YTTQTR | QEVPI | VKNR | RYEAK | QQVE | KFI | |
| K8GLK0_9CYAN/42-80 | LAYD | VFGT | TRPNE | YTTSTR | QQAP | IVKDR | FEAK | QQ | IDQFI | |
| K9VJB4_9CYAN/42-80 | LAYD | AFGT | TRP | NQY | FTPQ | REEVP | ISDR | FEAK | KQVD | QFI |
| K9W2B5_9CYAN/42-80 | LAYD | AFGT | TRPNE | YFTQQR | QEVPI | VSDR | RFES | KQE | IDQFL | |
| K9TZN4_CHRTP/42-80 | LAYD | VFGT | TRPNE | YYPQER | QELPI | VSDR | RFES | RKQ | IEEFI | |
| K9XG15_9CHRO/42-80 | LAYD | VFGT | TRPNE | YYPQER | QELPI | IGDR | RYQAK | QE | IKQFI | |

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

А, может, это максимально достоверный блок?

| | 10 | 20 | 30 |
|--------------------|------|------|-----------------------------------|
| K9Z0V0_CYAAP/42-80 | LAYD | VFGT | PRPDEYFTQTRQELPIINDRYEAKNQIQFN |
| K9SST0_9SYNE/42-80 | LAYD | IFGT | PRPNEYYTNDRQEAPILDRYDANQEIPNYN |
| K9SGD2_9CYAN/42-80 | LAYD | VFGT | PRPNEYYTLD RSDAPVLLDRFNAIDDIEDFN |
| K9Y0S0_STAC7/42-80 | LAYD | VFGT | PRPNEYFTQDRLELPIINDRYSASEQIKEFN |
| PSBE_RIPO1/42-80 | LAYD | VFGT | PRPDQYFTQERQELPIISDRYKASQQIQDEFN |
| PSBE_SYNY3/42-80 | LAYD | AFGT | PRPDEYFTQTRQELPILQERYDINQEIQDEFN |
| PSBE_GLOC7/42-80 | LAYD | VFGT | PRPDEYFTQERLELPIILKDRYNTDQQIQDEFN |
| K9W859_9CYAN/42-80 | LAYD | VFGT | PRPNEYFTQDRVEAPILSDRYEAKQQIQQFL |
| PSBE_SYNE7/43-81 | LAYD | AFGT | PRPNEYFTQDRTEVPIVSDRYSAKQQVDRFS |
| PSBE_ACAM1/42-80 | LAYD | VFGT | PRPNEYFDQARQGLPLVTD RYEGKQQIQDEFN |
| C7B2G7_SELML/42-80 | LAYD | VSGS | PRPNEYSTESRRDIP LITGRFDSSEQIQDEFN |
| PSBE_PHATC/43-81 | LAYD | VFGT | PRPNEYFTQDRQQIPLVND RFSAKQELLEDLT |
| PSBE_OSTTA/42-79 | LAYD | VFGT | PRPNEYFTEERQELPLISDRYGALAQLDDL |
| PSBE_CHLRE/42-79 | LAYD | VFGT | PRPNEYFTEDRQEAPLITDRFNALQVKKL |
| PSBE_SYNJB/42-80 | LAYD | IFGT | PRPNEYFTAERQELPIVSDRFNALEELERLT |
| K9TLB2_9CYAN/43-80 | LAYD | AFGT | PRPNEYFSQERLELPIVQDRFNAKDQIQDDL |
| K9S2L1_9CYAN/42-80 | LAYD | AFGT | PRPDEYFTQQRQELPIVMDRFQAKQQVEEFI |
| K9YE64_HALP7/42-80 | LAYD | AFGT | PRPDEYFTKQRQELPIISDRFQAREEVVEEFN |
| K9R7L0_9CYAN/42-80 | LAYD | AFGT | PRPDTYFTQERQEVPLVND RFEAKNQVQVFT |
| PSBE_THEVB/42-80 | LAYD | VFGT | PRPDSYQAQEQRSIPLVTD RFEAKQQVETFL |
| PSBE_GLOW/43-81 | LAYD | VFGT | PRPDEYNSDNTKKPLVNKRFEAKQQLDDEAT |
| PSBE_SYNR3/43-81 | LAYD | AFGT | PRPDSYFQAADV KAPVVSQR YEGKSQLDTRL |
| PSBE_PROMA/43-81 | LAYD | AFGT | PRPDTYFQASET KAPVVSQR FESKAQLDLRL |
| K9RTX4_SYNP3/42-80 | LAYD | IFGT | PRPDTYFSQEQQAAPVVS DRFESKQQINEFL |
| K9ULI3_CHAP6/43-81 | LAYD | AFGT | PRPDEYFTQTRQELPIVSD RFEESKQQLDKFV |
| K9UY34_9CYAN/42-79 | LAYD | AFGT | PRPNEYFTQTRQEVPIVNN RFDKQVETY |
| PSBE_NOSS1/42-80 | LAYD | VFGT | PRPDEYTTQARQELPIVNN RFEAKKQVEQLI |
| K9QJTB_9NOSO/42-80 | LAYD | VFGT | PRPNEYTTQTRQEVPIVKN RYEAKQQVEKFI |
| K8GLK0_9CYAN/42-80 | LAYD | VFGT | PRPNEYTTSTRQQA PIVKDRFEAKQQIDQFI |
| K9VJB4_9CYAN/42-80 | LAYD | AFGT | PRPNQYFTPQREEVPIISDR FFEAKKQVDQFI |
| K9W2B5_9CYAN/42-80 | LAYD | AFGT | PRPNEYFTQQRQEVPIVSD RFEESKQIDQFL |
| K9TZN4_CHRTP/42-80 | LAYD | VFGT | PRPNEYYPQERQELPIVSD RFEESRKQIEEFI |
| K9XG15_9CHRO/42-80 | LAYD | VFGT | PRPNEYYPQERQELPIIGDR YQAKQEI 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 | LAYDVF | GT | PRPDEYFTQTRQELPIINDRYEAKNQIQEFN |
| K9SST0_9SYNE/42-80 | LAYDIF | GT | PRPNEYTNDRQEAPILRDRYDANQEIPNYN |
| K9SGD2_9CYAN/42-80 | LAYDVF | GT | PRPNEYTLDLRSAPVLLDRFNAIDDIEDFN |
| K9Y0S0_STAC7/42-80 | LAYDVF | GT | PRPNEYFTQDRLELPIINDRYSASEQIKEFN |
| PSBE_RIPO1/42-80 | LAYDVF | GT | PRPDQYFTQERQELPIISDRYKASQIQEFN |
| PSBE_SYNY3/42-80 | LAYDAF | GT | PRPDEYFTQTRQELPILQERYDINQEIQEFN |
| PSBE_GLOC7/42-80 | LAYDVF | GT | PRPDEYFTQERLELPIILKDRYNTDQIQEFN |
| K9W859_9CYAN/42-80 | LAYDVF | GT | PRPNEYFTQDRVEAPILSDRYEAKQIQQFL |
| PSBE_SYNE7/43-81 | LAYDAF | GT | PRPNEYFTQDRTEVPVSDRYSKQVDRFS |
| PSBE_ACAM1/42-80 | LAYDVF | GT | PRPNEYFDQARQGLPLVTDRYEGKQIQDEF |
| C7B2G7_SELML/42-80 | LAYDV | SG | SPRPNSEYSTERDIPLITGRFDSSEIQDEF |
| PSBE_PHATC/43-81 | LAYDVF | GT | PRPNEYFTQDRQQIPLVNDRFSAKQELDLT |
| PSBE_OSTTA/42-79 | LAYDVF | GT | PRPNEYFTEERQELPLISDRYGALAQLDDL |
| PSBE_CHLRE/42-79 | LAYDVF | GT | PRPNEYFTEDRQEAPLITDRFNALEQVKKL |
| PSBE_SYNJB/42-80 | LAYDIF | GT | PRPNEYFTAERQELPIVSDRFNALEELERLT |
| K9TLB2_9CYAN/43-80 | LAYDAF | GT | PRPNEYFSQERLELPIVQDRFNAKDQIDL |
| K9S2L1_9CYAN/42-80 | LAYDAF | GT | PRPDEYFTQQRQELPIVMDRFGAKQVVEFI |
| K9YE64_HALP7/42-80 | LAYDAF | GT | PRPDEYFTKQRQELPIISDRFQAREEVVEFN |
| K9R7L0_9CYAN/42-80 | LAYDAF | GT | PRPDTYFTQERQEVPLVNDRF EAKNQVDVFT |
| PSBE_THEVB/42-80 | LAYDVF | GT | PRPDSYQAQEQRSIPLVTDRF EAKQVETFL |
| PSBE_GLOWV/43-81 | LAYDVF | GT | PRPDEYNSDNTKKPLVNKRFEAKQQLDEAT |
| PSBE_SYNR3/43-81 | LAYDAF | GT | PRPDSYFQAADVKA PVVSQRFEYEGKSQDLTRL |
| PSBE_PROMA/43-81 | LAYDAF | GT | PRPDTYFQASETKA PVVSQRFEESKAQLDLRL |
| K9RTX4_SYNP3/42-80 | LAYDIF | GT | PRPDTYFSQEQQAAPVVSDRFEKQQINEFL |
| K9ULI3_CHAP6/43-81 | LAYDAF | GT | PRPDEYFTQTRQELPIVSDRFESKGQLDKFV |
| K9UY34_9CYAN/42-79 | LAYDAF | GT | PRPNEYFTQTRQEVPIVNNRFDKQVETY |
| PSBE_NOSS1/42-80 | LAYDVF | GT | PRPDEYFTQARQELPIVNNRFEAKKQVEQLI |
| K9QJT8_9NOSQ/42-80 | LAYDVF | GT | PRPNEYFTQTRQEVPIVKNRYEAKQVVEKFI |
| K8GLK0_9CYAN/42-80 | LAYDVF | GT | PRPNEYTSTRQQAPIVKDRFEAKQVIDQFI |
| K9VJB4_9CYAN/42-80 | LAYDAF | GT | PRPNQYFTPQREEVPIISDRFEAKKQVDQFI |
| K9W2B5_9CYAN/42-80 | LAYDAF | GT | PRPNEYFTQQRQEVPIVSDRFESKQVIDQFL |
| K9TZN4_CHRTP/42-80 | LAYDVF | GT | PRPNEYYPQERQELPIVSDRFESRKQIEEFI |
| K9XG15_9CHRO/42-80 | LAYDVF | GT | PRPNEYYPQERQELPIIGDRYQAKQEI KQFI |

- Для выделения блоков в выравнивании очень полезен JalView

Практическое задание

- a. Заполните таблицу с информацией о домене PF00284 :
 - a. Сколько всего белков с доменом
 - b. Сколько из них из SwissProt (reviewed)
 - c. Сколько из бактерий
 - d. Для скольких определена пространственная структура
 - e. Сколько в выравнивании seed, которое использовалось для поиска всех доменов в белках

- b. Сохраните выравнивание последовательностей домена C-5 cytosine-specific DNA methylase (PF00145) из SwissProt (revised)
 - a. Pfam => proteins => Generate Fasta => download (имя файла начните с вашей фамилии)
 - b. Откройте в Jalview
 - c. Выровняйте последовательности программой mafft. Ужаснитесь!!!
 - d. Раскрасьте Clustal и by conservation
 - e. Найдите консервативные участки понижая порог conservation пока их не увидите.
 - f. Сделайте группы по консервативным позициям 2го участка
 - g. Сортировка по группам
 - h. Сколько последовательностей не имеют канонических консервативных а.к...