

Гомология

и

выравнивание

Множественное выравнивание последовательностей
гомологических белков

План

1. Задание
2. Гомология (3 слайда)
 - a. Термин
 - b. Эволюция геномов бактерий
 - c. Гомологию белков и фрагментов геномов определяют по сходству
3. Эволюционное выравнивание (2)
 1. Отражение эволюции с помощью выравнивания
 2. Смысл выравнивания последовательностей гомологичных белков
4. Множественное выравнивание (10 + | Jalview demo)
 1. Анализ выравнивания
 2. Консервативное значит важное
 3. Множественное выравнивание против парного
 4. Алгоритмы.
5. Эволюционные домены
 1. Определение
 2. Примеры из Pfam
 3. Как происходят изменения доменной архитектуры
6. Крупные перестройки в геномах
 1. Следствия для белков. Карты локального сходства белков
 2. Эволюционные домены
 3. БД Pfam
7. Jalview
 - a. Эволюция белков
 - b. Эволюционные события
 - c. Эволюционные домены

1. Задание по теме "гомология и выравнивание"

Результат:

- **Тривиальная часть** - описание одного семейства доменов по информации из Pfam и Uniprot
- **Нетривиальная** (самому или самой надо думать и принимать решения) - одно выравнивание двух подгрупп доменов семейства с обоснованием их различий и отличий от всех остальных доменов семейства
Отнеситесь к заданию как к игре с выравниванием:.) Очень надеюсь на ваш интерес. ААл

Методы:

- Сервисы базы данных Pfam
- Редактор выравниваний Jalview
- Blast выравнивание 2х последовательностей, в формате Dot Plot
- Uniprot поиск и скачивание результата в табличном формате

1. Выберите семейство доменов из Pfam для анализа

От выбора зависит всё дальнейшее

Ограничения, направлены на то, чтобы обезопасить вас от больших технических трудностей, они не являются абсолютными

2. Опишите семейство доменов

Укажите число доменных архитектур с этим доменом

Выберите две достаточно представленные доменные архитектуры и укажите какие именно выбрали, их названия и число белков с каждой из них

Укажите число разных белков с доменом семейства, для которых известна 3D структура. *Разные структуры одного и того же белка (по Uniprot ID) считать за одну.*

Укажите число белков с доменом по таксонам самого высокого ранга. Типично - по суперцарствам(они же домены жизни) - бактерии, археи, эукариоты.

3. Постройте карту локального сходства (Dot Plot) двух белков с доменом семейства, но с разной доменной архитектурой

Придумайте эволюционный сценарий наблюдаемого

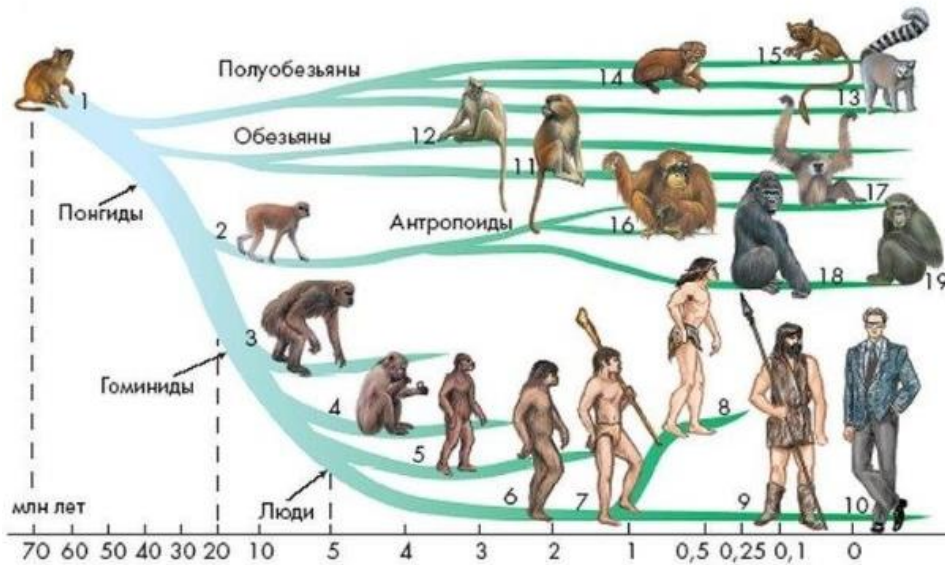
4. В выравнивании доменов семейства выделите на основании сходства две подгруппы доменов Pfam

В ответе - выравнивание, содержащее обе подгруппы и обоснование различий подгрупп и отличий от всех остальных доменов семейства.

5. Сохраните таблицу со всеми белками из Uniprot с доменом семейства Pfam

В таблице сохраните колонки:

2. Гомология – наличие общего предка



В словарик:

* Последний общий предок (LCA)

* Гомология

Последний общий предок ныне живущих обезьян.

Гомоло́гия ([др.-греч. ὁμοιος](#) «подобный, похожий» + [λογος](#) «слово, закон») в биологии — сопоставимость частей сравниваемых биологических объектов, обусловленная общностью происхождения

wiki

Для целых организмов термин «гомология» не употребляют

Эволюция геномов бактерий

Половое размножение отсутствует

Небольшие, локальные изменения от поколения к поколению.

Замены нуклеотидов, короткие делеции и вставки

Изменения локальные, но их может быть много в эволюции

Происходят случайно. С разной частотой*)

Контролируются отбором – носители вредных и слабо вредных мутации удаляются из популяции.

Накапливаются от поколения к поколению. Пытаются по их числу измерять время от потомков до последнего общего предка

Крупные единовременные изменения генома

тем более – под отбором! Сохраняются только удачные.

Но из-за огромного количества организмов, мы их видим

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

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

Как говорить можно, и как нельзя

- Высокая ~~ГОМОЛОГИЯ~~ последовательностей

+ Высокое СХОДСТВО
последовательностей

*) почему почти?

3. Эволюционное выравнивание

Выравнивание последовательностей потомков относительно предка

предок	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	
предок	TATGCGAATGCCCTGAA	
сын	TATG A GAATGCCCTGAA	замена
внук	TATG C GAATG C TCTGAA	замена
правнук	TATG C GAAT C G C TCTGAA	вставка 1 п.н.
праправнуку	TATG A GA A A C G C TCTGAA	замена
прапраправнук	T G A GA A A C G C TCTGAA	делеция 2 п.н.
потомок	1 4 5 6 7 8 9 9a 10 11 12 13 14 15 16 17	

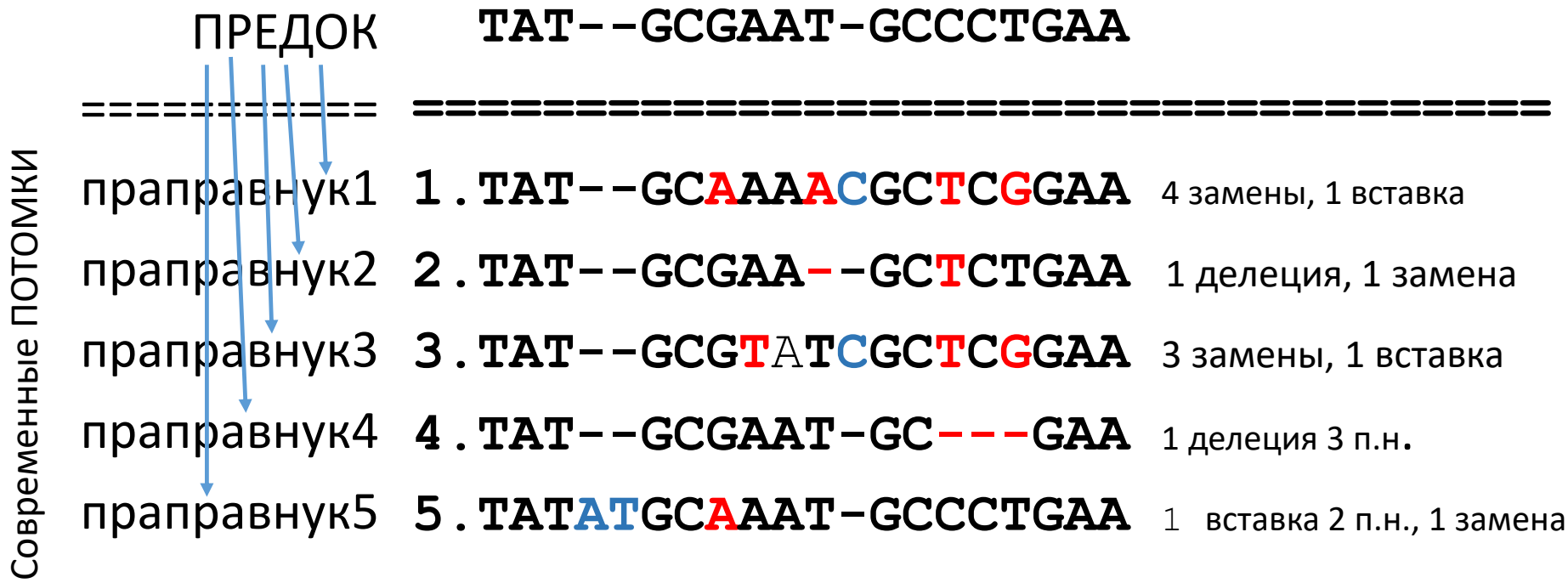
Нукл-ы потомка с номерами как у предка являются гомологами нукл-в предка

предок	TATGCGAAT-GCCCTGAA
сын	TATG A GAAT-GCCCTGAA
внук	TATG C GAAT-G C TCTGAA
правнук	TATG C GAAT C G C TCTGAA
праправнуку	TATG A GA A A C G C TCTGAA
прапраправнук	T--G A GA A A C G C TCTGAA

Выравнивание: гомологичные нуклеотиды - друг под другом

Идеальное выравнивание потомков относительно общего предка

Синий – вставка
 Красный – замена
 - Делеция
 относительно ПРЕДКА



Такое выравнивание бывает только в экспериментах по изучению эволюции. E.coli (Ленский), шизофилум (А.Кондрашов), др.

Сокращения

* а.к.о. – аминокислотный остаток

* aa – amino acid residue

4. Множественное выравнивание последовательностей гомологичных белков (анализ результата выравнивания)

Эволюционное выравнивание (редко достижимый идеал) :
в каждой колонке стоят гомологичные аминокислотные
остатки (или “-” – символы гэпа)

Смысл выравнивания последовательностей гомологичных белков

- Некоторые кодоны а.к.о. гомологичных белков потомков *произошли из одного кодона последнего общего предка* этих белков, или были делятированы 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!

Для этого нужно набираться опыта на примерах.

Биологический смысл выравнивания

Обсуждаем смысл колонок и гэпов

```
          10          20          30          40          50          60
EJL77459.1 GVDLVF GGPPCQGFSQIGMRR- LDDER- NEL YQQYTR I VAKLKPRVFLMENVPNLALMNKGH
RXK67093.1 DLDVVF GGPPCQGYSSQIGTRR- LDDER- NEL YLQYAR I VEKQRPRMFLMENVPNMVL LNKGH
OJY44288.1 NVDLVF GGPPCQGYSSQIGTRD- LHDPR- NRL FEEFARVVATLKP KLFMENVPNL LLLLNKGH
TRU90449.1 NPEMIV GSPPCQDFSSAGKRNEGLGR- - ANL TLTFAE I VTRVSPQWFVMENVD- - - RIEKSK
OXI46696.1 GTDLVF GGPPCQGFSQIGMRR- LDDER- NEL YKQYTRV VSTLRPRVFLMENVPNLALMNKGH
AVZ30243.1 EIDVVF GGPPCQGFSLIGKRS- FEDPR- NSL VFH Y IRLVLELSPKFFV IENVKGMTAGNHQA
AFZ12381.1 DIIGF IGGAPCPDFSVGGKNRGSEGDK- GKLSASY IEL I CQQKPDFFLFENVKGLYKTKKHR
HCQ21462.1 HIIGF IGGPPCPDFSVGGKNKGHLGDN- GKLSASY IEL I CQNLPDFFLFENVKGLWRTTKHR
EDN77159.1 SLIGF IGGPPCPDFSIAGKNKGKGDGN- GKLSLSYTNL I IEMKPDFFLFENVKGLWRTARHR
SOD91684.1 EVSLVVGGAPCQPF SNIGKKLGKNDERNGDL FLEFVRMVKGIQPEAF I FENVVGI TQNKHSD
QCS48280.1 NVVGF IGGPPCPDFSI GGKNRGRQGDH- GKLSesy IIDL I IQHQPDFFI FENVKGLYRTKKHR
SMB95934.1 GLFG IGGPPCPDFSVGGKNRGENGEQ- GRLSKVFVDK I LDLQP VFFLYENVPGL I RTAKHR
RUO38876.1 SPVGF IGGPPCPDFSVGGKNRGHEGEN- GRLTRTYVDG I IKYAPDFFI FENVKGLWRTKRHR
OIP70538.1 TIDL ICGPPCQGFSTIGTND- KKDHR- NFL FFEFLRMVETFKPNFI I LENVTGLLAKKNES
AFY60915.1 NLVGFVGGPPCPDFSI GGKNKGQYGDN- GKLTKVYVD I I IENQPDFFV FENVKGLWRTRSRHR
CUR30340.1 DLIGF IAGPPCPDFSVGGKNRGKNGDQ- GKLTACYVEL I CQQRPDFFV FENVKGLWSTKKHR
TAK03971.1 QAALVVGGAPCQPF SNL GSKRGTADSR- GTL FQDF I RIVKGV RPKGFI FENVEGLTQDKHKG
AEE51071.1 KVALVVGGAPCQPF SNIGKKEGENDAKNGDL FLEFVRMVKGIQPEAF I FENVAGI I QSKHSK
RTR31666.1 RLVGFVGGPPCPDFSVGGKNKGSEGEN- GKLTRTY IIDL I VKDNP DYFI FENVKGLWRTTRHR
PTU64472.1 NIDLVF GGPPCQGFSQIGTRR- LDDER- NEL YKQYTR I VKTLKPRVFLMENVPNLAMMNKGH
```

DNA (cytosine-5-)-methyltransferase

Не всегда так хорошо как на предыдущем слайде

Продолжение того же выравнивания

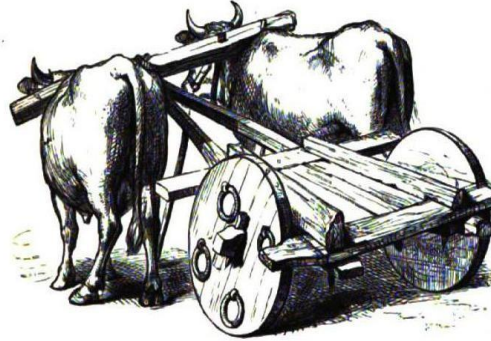
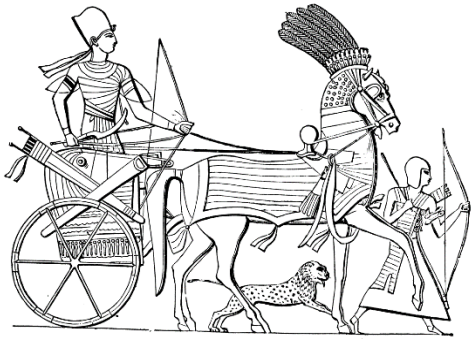
```
      210      220      230      240      250      260      270      280
QDRKRVFIVGYREDLNLK-----FEFPKPLNKKVTLRD-----AIGDLPE-F
QDRERVFYVGF RKDLNLSN-----FE-FYP ISEKERKYLKD-----SIWDLKDN/
QERKRVFYIGFRKDLEIKF-----SFPKGSTVEDKDKITLKD-----VIWDLQDT/
QDRKRVFYIGFRKELNIN-----YLPP IPHL I KPTFKD-----VIWDLKDNF
QQRDRLVLF AAKQG-----VIKI I PPTHTPENYR-----TVRDVIGSLATN\
QSRQRVFFIGLKSDRPLNQQ-----ILTP-----PSKVI ESEYTSLEEAI SDLPVIE-----AGEGGEVQDYPVAF
QLRKRTFVIGHRHGS IAD-----LANVLQQRLAKQSL-----TVRDYFG--
QSRTRFSLIGKLNSEHNF-----LIPTLSRKLSDKPM-----TVRDYLG--
QRRHRI I I VGI RKDQD-----VAFRVPEPTHKEKYR-----TASEALADIPED/
HQRHRWFCLAIRKDYEP EE-----I I VSVNATKFDWENNEPPCQVDNK-----SYENSTLVRLAGYS
QNRERIYIVGFRN-----IEHYKNFNFPMPQP-----LTLTIKDMINLS
QNRERLYIIGIREDLIKNEE-----WSLDFKRKDI LQKGKQRLVELDI KSFNFRWTAQ-----SAATKRLKDLL EE\
QNRERVFCISILN-----PNEDFTFPQKQ-----NLTL SMNDLLEEN
QARRRVFMISTLNEF-----VELPKGDKKPKS-----IKKVLNKIVSE\
QNRERIYLVGF-----LNHDVDFRFPQP-----IGQATAVGDILEA\
QNRERIYIVGFDRKS-----ISNYSDFQMPTP-----LQEKTRVGNILES\
QNRERIYIVGFNKEK-----VRNHEHFTFPTP-----LKT KTRVGDILEK\
QNRERIYIVGFHKS-----TGVNSFSYPEP-----LDKIVTFADIREEF
QNRRLRVYIVGLDQSQPELT-----ITSHIGATDSHKFKQLSNQASLFD-----TNKIMLVRDILED\
QNRVRIYILGILGSKPKLT-----LTSNVGAADSHKYK--NEQISLFD-----ES-YATVKDILED\
QKRKR FYLVAF LNQN-----IHFEF PKP-----PMISKDIGEVLES\
QRRERIVIVGFHPDLG-----INDFSFPKGN-----PDNKVPINAIL EH\
QKRERIYMICFRNDLN-----IQNFQFPKP-----FELNTFVKDLLLP
QRRRRVFI FG YKQDLNYSKAME-----ESPLDKI IYHNGLFAEAFP I EDYANKNR-----VNRTHITHDIVDISDN
QRRKRAI IRLNKKGT IWN-----LPLKQNI VSVEQ-----AIGNLPSIESG\
QVRERVF I TATLVPERMRDER I PRTETGE I DAEA I GPKPVATMNDRFP I KKGTEL FHPGDRKSGWNLL TSGI I REGDPEF
QNRDRVFI IGIQQKLGVPD-----FSFPEYSESEQRLYD I LDNLQTPS I I-----PESLPIQRNLFGEF
QNRDRVFI VGIQQKLDLNG-----FSFPEYAESDQRLYH I LDNLEAPETK-----LES IPIQRNLFGEF
QKRERIV I IGIREDLVK-----EQKYPFRFPLAQ-----VYKPV LKDV LKD\
QLRPRVLFVALKNEYTN-----FFKWPEPNSEQPK-----TVGELLFDLMSE\
```

Смотрим в Jalview

- Файл PF00145_seed.fasta

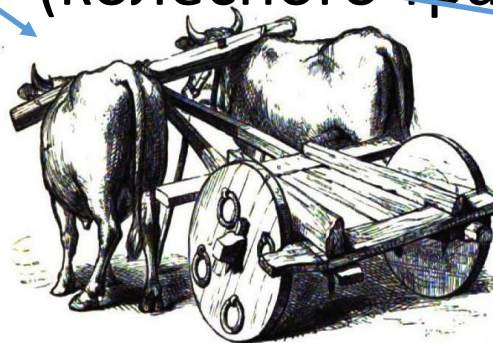
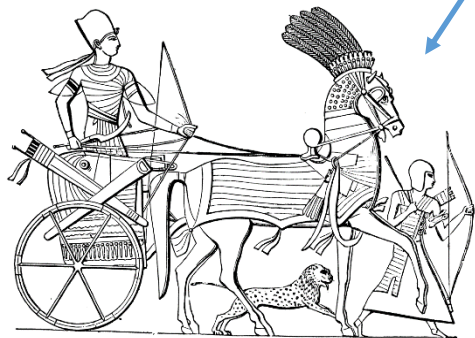
Консервативное значит важное

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



LUCA

(колесного транспорта)

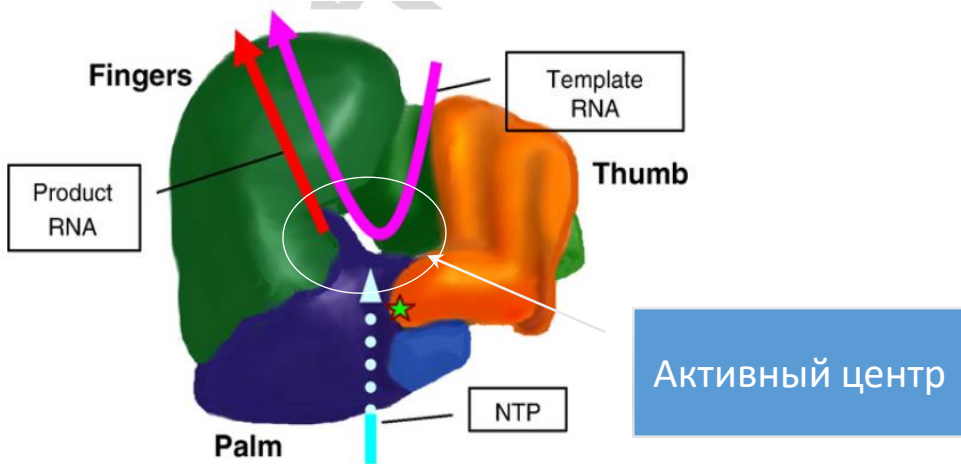


В выравниваниях белков – то же самое:
Сохраняющееся в эволюции (консервативное) – важно

РНК зависимаая РНК полимераза (RdRP), консервативные участки

```

*      320      *      340      *      360      *      380      *      400      *      420      *      440      *      460
FKTMIRFGDVGLDLDFSAFDASLSPPFMIREA..GRIMSELS...GTPSHFGTALINTIIYSKHLIYNCCY.....HVCGSMPSGSPCTALNSTINNVIYVFSKIFGKSPVFF.....CQALKILC.YGDDVIVFVSRD
EVAMQG.FERVYDVVYSNEDSTHSVAMFRLL..A...EEFF.TPENGFDPLTREYLESIAISTHAFEKRF.....LTGGLPSCCAATSMINTIMNNIIIRAGLYLTYNFEFDD.....VKVLS.YGDDLIVATNYQL
ETHFAQ.YKNVWVDVLYSADANHCSDAMNIMFEEVFERTEFG.....FHPNAEWILKTLVNTTEHAYENKRI.....VVEGCMPSGCSATSIINTILNNIYVLYALRRHYEGVELDT.....YTMIS.YGDDIVVASYDYD
.....WSLCVATIVSDHDTFWPGWLRDLICDELINMGYA.PWVVKLFETSLKLPVYVGAFAPEQGHLLGDPSNPDLVGLSSGQGATDLMGTLIMSTIYLVMLQDHTAPHLNSRIKMPSACRFLDSYWQGHEETROIS.KSDDAILGWTKGR
LRLRLE.NWVYCADGSOEDSSSLTPYLINAV..LTLRSTYMEDWDVGLQMLRNLYTEIVYTPISTPDGTIV.....KKFRGNNSGQPSIVDNLMLVVIAMHYALIKECFEVEEID.....STCVFFV.NGDDLIAVNPEK
HDKLNRPGWLHGSGDGRDSSIDPFFFDVV..KTKRKHEL..PSEHHRaidLIYDEILNTTICLANGMVI.....KKNVGTQR.QPSTVDNTLVMITAFLYAYIHKTGDRELAL.....LNERFIFVC.NGDDNKFAISPQF
AISLASFSPYGFNCFANEDGMFHPSSFSMV..SEIANIFY...GNFLSTERDNLTRMLTNRFSLMKGAIL.....RVPGGSPSGFFMTVFNSEINLFYLQSAWIMLARFNGRQDISH.....PCNFPKYVRACV.YGDDNIVAIMEV
AARMKEKGNVDVLCODYSSEFDGLLSKQVMDVI..ASVINELC.GGEDQLKNARRNLLMACCSRIAICKNTVW.....RVECGIPSGFFMTVFNSEINLFYLRHYHKIMREQQAPELMV.....QSFDKLIGLVT.YGDDNLSVNAV
YAEHAK.YKNHFDADYIADSTQNRQIMTES..FSIMSRLT...ASPELAEVVAQDLAPSEMVDVGYVI.....RVKEGLPSGFFPCTSQVNSINHWITLCALEATGLSPDVV.....QMSYFYSFYGDDIVSTDIDF
NNLTSKASDFLCLDYSKFDSTMSPCVVRIA..IDLADCC...EQTELTKSVVLTILKSHFMTILAMIV.....QTKRGLPSGMPFTSVINSHCHWLLWSAAVYKSCAEIGLHCS.....NLYEDAPFYT.YGDDGVYAMTEMM
IQRIKS.AAKVYAVDYSKFDSTQSPRVSAAAS..IDLRYFS...DRSPIVDSAANTLKSPPIAIFNGVAV.....KVSSGLPSGMPFTSVINSHCHWLLWSAAVYKSCAEIGLHCS.....NLYEDAPFYT.YGDDGVYAMTEMM
TKRLERPKHdrycVLYSKFDSTQPPKVSQS..IDLRHFT...DKSPIVDSACATLKSNIPIGIFNGVAF.....KVAGGLPSGMPFTSVINSHCHWLLWSAAVYKSCAEIGLHCS.....NLYEDAPFYT.YGDDGVYIVPPLI
D      D      g      sg      T      n3      gDD
    
```



На каких участках выравнивание правильное – совпадает с эволюционным?

Fig. 1. Schematic architecture of “small” RdRP. The hairpin between the palm and thumb domains is in light blue. The predicted approximate location of MV RdRP Trp460 is marked by star.

Рисунок мой :) хвастаюсь

Немного об алгоритмах

Множественное правильнее парного!

Множественное даёт аргументы, опровергающие оптимальное парное выравнивание. Пример.

```

      *           100           *           120           *           140           *           160
THEIE_LACLS : AGVSEFIVNDDVELARELNADGHIHGQTDSEVSKVREKVGQEMWLGLSVTKADELKTAQ-SSGADYLGIGPIYPTNSKND : 14
THEIE_MANSM : FQVBEFIVNDDVELALSIQADGHIHVGQKDTAVETILRNTRNKPIIIGLSINTLAQALANKDRQDIDYFVGVPPIFPTNSKADH : 15
THEIE_STRA3 : YQVBEFIIDDDIDLVELIDADGHIHGQNDLPVDEARRRLPDKI-IGLSVSTMAEYQKSQ-LSVVDYIIGIGPFNPQSKADA : 14:
THEIE_LISIN : YQVBEFIINDDDVALALEIGADGHIHVGQNDDEEIRQVIASCAGKMKIIGLSVHVSVEAEEAERLGSVDYIIGVGPPIFPTISKADA : 14:
THEIE_ANOFW : YNIBEFIVNDDVDLALALQADGVHVGQDEVEAERVDRDRIGDKY-LGVSVHNLNEVKKAL-AACADYVGLGPIFPTVSKEDA : 14:
THEIE_GEOTN : YGVBEFIVNDDVELAIAIDADGVHVGQDDEADARRVREKIGDKI-LGVSAHNVEEARAAL-EAGADYIIGVGPPIYPTRSKDDA : 14
THEIE_BACSU : AGVBEFIVNDDVELALNLKADGHIHGQEDANAEVRAAIGDMI-LGVSAMTSEVKQAE-EDGADYVGLGPIYPTETTKDD : 14
THEIE_BACA2 : AGIBEFIVNDDVELALRLEADGVHIGQDDADAEEETRAAIGDMI-LGVSAMTSEVKQAE-AAGADYVGMGPVYPTETTKDD : 14
THEIE_OCEIH : FQIBEFIVNDDVDLAKQLDADGHIHGQDDQPVVVRKQFENKI-IGLSISTNNELNQSP-LDLVDYIIGVGPPIFDTNTKEDA : 14
THEIE_STAAB : YNVBEFIVNDDVSLAKEINADGHIHVGQDDAKVKEIAQYFTDKI-IGLSISDLGEYAKSD-LTHVDYIIGVGPPIYPTPSKHDA : 14:
THEIE_STACT : YNVBEFIVNDDVALAEEIDADGHIHVGQDDEAVDDFNRRFEGKI-IGLSIGNLEELNASD-LTYVDYIIGVGPPIFATPSKDDA : 14:
      6pFI61DD6 La 6 ADG6H6GQ D 6G68 2 DY G6GP pT 3K Da

```

```

      *           180           *           200           *           220           *           240
THEIE_LACLS : AKETIGIKDLR-LMLLENQLPIVIGGGITQDSLTELSAIGLDGLAVISLLTEAENPKKVAQMIRQKITKNG~~~~~ : 21
THEIE_MANSM : SPIVGMNFIRQIRQLGIDKPCVAIGGITEESAAILRRLGADGVAVISAIHSHSVNIANTVKTLAQK~~~~~ : 22
THEIE_STRA3 : KPAVGNRTTKAVREINQDIPVVAIGGGITSDVFVDIIESGADGLAVISAIISKANHIVDATRQLRYEVEKALVNRQKRSDVI : 22:
THEIE_LISIN : EPVSGTAILEEIRRAGIKLPIVIGGGITNETNSAEVLTAGADGVSVISAITRSEDCQSVIKQLKNPGSPS~~~~~ : 21
THEIE_ANOFW : KQACGLTMEHIRAEKRVPLVAIGGITEQTAQVIEAGADGLAVISAIKRAEHIYEQTKRLYEMVMRAKQKQKDR~~~~~ : 21
THEIE_GEOTN : NEAQQPGILRHLRREQGITIPIVVAIGGITADNTRAVIEAGADGVSVISAIASAPEPKAAAAALATAVREANL---R~~~~~ : 22
THEIE_BACSU : RAVQGVSLIEAVRRQGISIPVIGGGITIDNAAPVIEAGADGVSMISAISQAEDPESAARKFREEIQTYKTG--R~~~~~ : 22:
THEIE_BACA2 : EAVQGVTLIEEVRQGITIPVIGGGITADNAAPVIEAGADGVSMISAISQAEDPKAAARKFSEEIRRSKAGLSR~~~~~ : 22
THEIE_OCEIH : KTAVGLEWIIQSLKKQHPSLPIVVAIGGITNTNAQEIIEAGADGVSVISAITETDHIILQAVQRL~~~~~ : 20
THEIE_STAAB : HTPVGPMEIATFKEMNPQLPIVVAIGGITSNVAPIVEAGANGISVISAIKXSENIKTVNRFKDFFN~~~~~ : 21:
THEIE_STACT : SEPVGPKMIETLRKEVGDLEIPIVVAIGGITSLDNVQEVAKTSADGVSVISAIARSPHVTETVHKFLQYFK~~~~~ : 21:

```

```

      80           *           100           *           120           *           140           *
THEIE_LACLS : VSEFIVNDDVELARELNADGHIHGQTDSEVSKVREKVGQEMWLGLSV-TKADELKTAQSSGADYLGIGPIYPTNSKND :
THEIE_MANSM : VEFIVNDDVELALSIQADGHIHVGQKDTAVETILRNTRNKPIIIGLSINTLAQALANKDRQDIDYFVGVPPIFPTNSKAD :
      V FIVNDDVELA 6 ADGIH6GQ D V 6 6GLS6 T A L DY G6GPI5PTNSK D

```

```

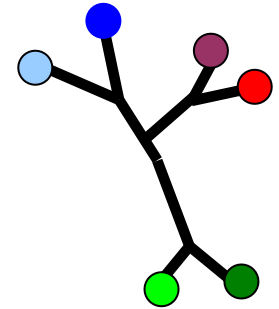
      160           *           180           *           200           *           220
THEIE_LACLS : AAKPTG---TKDLRLMLLENQLPIVIGGGITQDSLTELSAIGLDGLAVISLLTEAENPKKVAQMIRQKITKNG : 218
THEIE_MANSM : HSPVGMNFIRQIRQLGIDK--PCVAIGGITEESAAILRRLGADGVAVISAIHSHSVNIANTVKTLAQK----- : 220
      6G T4 6R 6 6 P V IGGI 2 S L 6G DG6AVIS 63 N 6 QK

```

В красном овале во множественном выравнивании – одна делеция между консервативными позициями.

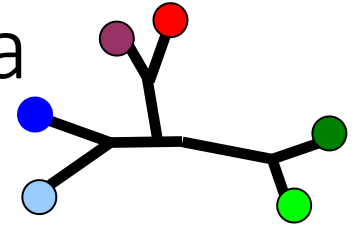
В оптимальном парном выравнивании первых двух последовательностей в красном овале – четыре делеции. Участки те же.

Иерархический алгоритм выравнивания многих последовательностей



- Основная идея: выравнивание двух выравниваний с помощью динамического программирования
- Этапы алгоритма
 - Построение направляющего дерева
 - Итерация выравнивания выравниваний
 - “Рафинирование” (refinement) выравнивания
- Результат – ГЛОБАЛЬНОЕ множественное выравнивание

Построение направляющего дерева



- Для ВСЕХ ПАР последовательностей строится парное выравнивание.
- Вес парного выравнивания пересчитывается в расстояние между последовательностями:
 - чем больше вес, тем меньше расстояние;
 - расстояние между совпадающими последовательностями равно 0.
- Получается матрица расстояний между послед-ми
- Есть алгоритмы, превращающие матрицу попарных расстояний в дерево.
 - Расстояния между листьями по дереву отражают сходство последовательностей

5. Наблюдаемый результат крупных перестроек генома: домены белков

Семейства доменов в БД Pfam (Protein families)

Домены белков

[Длинные] гомологичные участки из разных белков, которые эволюционируют только по типу локальных мутаций, **и максимальной длины, с сохранением этого свойства,** называются

ЭВОЛЮЦИОННЫМИ ДОМЕНАМИ

Терминологическая проблема.

ДОМЕН – набор фрагментов последовательностей и их выравнивание. Имеет короткое название. Например, RdRP_1

ДОМЕН белка – фрагмент последовательности, входящий в определенный домен, например, в RdRP_1

ДОМЕННАЯ АРХИТЕКТУРА – последовательность доменов в белке

ДВА ДОМЕНА гомеобелков: гомеодомен и OAR домен

SW: PMX1_CHICK/1	-----MASSYAHAMERQALLPARLDGPACLDNLQAKNFSVSHLLDLEEAG-DMVAAQDCEGGGPRGRSLLLESP-GLTSGSDTPQQD	: 80
SW: PMX2_HUMAN/1	-----MDSAAAAFALDKPALGCPGPPPPALGCPGDCQAQRNFSVSHLLDLEEVAAAGRLAARPGARABAREGAAREPSCGSSGSEAAAPQD	: 86
SW: PMX1_HUMAN/1	-----MTSSYGHVLEQPALGCRRLDSPGLMDTLQAKNFSVSHLLDLEEAG-DMVAAQADENVGACRSLLLESP-GLTSGSDTPQQD	: 80
SW: ARX_BRARE/1	ISQAPQVVISRSKSYREN-APFSQS---D-EQSP--EHAQELVELST-----LKFEEDEVVKEEACQDN-----S-----LSPKDEESLH-NDGVDKCDSDSVCLS	: 84
SW: ARX_MOUSE/1	ISQAPQVVISRSKSYRENGCAPVPPPPALD-ELSGPCGVHPEERLSAASGPGSAPAAGCGTCAEDDEEELLEDEEEDEREELLEDDDEELLEDDARALLKEPERRCVATTCTVAAAAAATAVATEGGELSPKRELLLHPEDARCKDGEDSVCLS	: 157
SW: AL_DROME/1-1	-----MGISEEIKLEELPQAKLHAHDPAVVLVDRAPGSSAASAGAAALTVSMVSYSGGAPSCASGASGCTNSPVSDGNS	: 72
SW: ALX4_MOUSE/1	-TFLSAGAKQCPCDAKSRARYGACQDLAAPLESSSGARGSPNKFQPPQPTQP-----PPAPPAPPAHLYLQRCACKTPDQCSLKLQEGSSGCHNAALQVPCYAKRESNLCEPELPPDSFVPCVMDNSYLSVKRTGARCPQDASAEIIPSL	: 145
SW: ALX4_HUMAN/1	-TFLSAAAQAQPCDAKSRARYGACQDLATPLESSGARGSPNKFQPPQPTQPQPPQPPQPPQPPQPPQPPHLYLQRCACKTPDQCSLKLQEGSSGCHNAALQVPCYAKRESNLCEPELPPDSFVPCVMDNSYLSVKREAGVRCQDRASSDLPSPL	: 157
SW: RX2_CHICK/1	-----NPSRLHSIEAILGFTKDDGLLGFQFP-----DGGAGSAAKAAADKRGPRHCLPKGPAEPPPAEHQGRFQEPYPCGASAPF-----LPAGCGDG	: 83
SW: RX2_BRARE/1	-----GISCRVHSIDVILGFSKDDPPLLEPSGR-----HKVDLEDQLEEQEKQVADPYSHLQIPDQIQQQQSVYH---DTGLFSTDKCADLGDPRSINVEDSRS	: 92
SW: RX1_XENLA/1	-----NPSRLHSIEAILGFKEDS-VLGSFQSEIISPRNAKEVDKRSRHLCHMTKEIHPQEHLEDG-QADCYG--DPYSGRTSSECLP-PCLS--SNSDN	: 91
SW: RX_HUMAN/1-1	-----STSRLHSIEAILGFTKDDG-ILGTFPAERGARGAKEDRRLGARCPACPKAPEGSEEPSPPAPAPAPYEAPRPPYCPKPEWRAPSPGLPVGPAATGEA	: 97
SW: PIX2_BRARE/1	-----MTSHKPLSLDHHHHHHVTCGSHAPLSMASSLQPLQRSVDSKHLRDLVHTVSDTSSPEVKEKRCQ--	: 66
SW: PIX2_HUMAN/1	-----METNCRKLVSAVGLQVPAAEVCLFSKDSSEIKKVFETDPSRKRKAASAKFPHPHQCANFKDQSKQ--	: 68
SW: PIX1_HUMAN/1	-----MDAFKCGMSLERLPEGFPPPPPHDMGPAFHRLAPDPRPLEN-SASESSDTLPEKREKGGEP	: 64
SW: OTP_MOUSE/1	-----MLSHADLLDARLCHKDAEALLGHREAVKRLGVGCSDPGCHPCDLAPNSDPVRCATLLPREDITTVGSTPASLAVSAKDPDKQPGPQCGP	: 90
SW: PMX1_CHICK/1	NDQLNSEE-----KKKRRRRRRTFTFNSSQLQALERWERETHYDPAFVRBDLARRVNDTEARVQVVFQNRRAKFRNNEAMLSKMASLLKSYSGDVTAVEQIIVPRPAPRPTDYLWSGTASPYSAMATYSTTCTMAS-----	: 213
SW: PMX2_HUMAN/1	GCPCSPGRCG-----AAKRRKRRRRTFTFNSSQLQALERWERETHYDPAFVREELARRVNDTEARVQVVFQNRRAKFRNNEAMLSRSASLLKSYSGE-AAEQPVAPRPTALSPTDYLWSWTASSPYSTVPPYSPGSSCP-----	: 221
SW: PMX1_HUMAN/1	NDQLNSEE-----KKKRRRRRRTFTFNSSQLQALERWERETHYDPAFVRBDLARRVNDTEARVQVVFQNRRAKFRNNEAMLANKNASLLKSYSGDVTAVEQIIVPRPAPRPTDYLWSGTASPYSAMATYSATCANNS-----	: 213
SW: ARX_BRARE/1	AGSDSEEG-----MLKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRLDLDEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 230
SW: ARX_MOUSE/1	AGSDSEEG-----LLKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRLDLDEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 303
SW: AL_DROME/1-1	EKADSEY-----PKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRLDLDEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 212
SW: ALX4_MOUSE/1	EKTDSESN-----KCKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRLDLDEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 290
SW: ALX4_HUMAN/1	EKADSESN-----KCKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRLDLDEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 302
SW: RX2_CHICK/1	KPSDEEQ-----PKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 215
SW: RX2_BRARE/1	PDIPDEDQ-----PKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 225
SW: RX1_XENLA/1	KLSDEEQ-----PKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 224
SW: RX_HUMAN/1-1	KLSEEQ-----PKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 242
SW: PIX2_BRARE/1	SKNEDSW-----DDPSKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 212
SW: PIX2_HUMAN/1	GRNEDVGA-----EDPSKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 215
SW: PIX1_HUMAN/1	KCPEDSCAGCTGCCGADDPAKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 218
SW: OTP_MOUSE/1	NPSQACQQ-----CQQQKRRRRTFTFTSYQLEELERAFQTHYDPAFVREELARRVNDTEARVQVVFQNRRAKWRREERCAQTHPPGLPFPFPCPLSATHPLSPYLDASFPFPHHPALDSAWTAAAAAAPPSPPLPPPG-SASLPSPCAPLG	: 236
	k 4 4R RT Ft QL EL E R F 4 HYPD RE 6A L E R6qVWFQNRRAK54 e4	
SW: PMX1_CHICK/1	-----PAQGMNMANSLALPLAKRHSYSLQRNQVPTVN-----	: 245
SW: PMX2_HUMAN/1	-----ATPCVMNMANSLALPLAKRHSYSLQRNQVPTVN-----	: 253
SW: PMX1_HUMAN/1	-----PAQGINMANSLALPLAKRHSYSLQRNQVPTVN-----	: 245
SW: ARX_BRARE/1	LGTFLCTAMFRHFAFIPTFCRLFFSSMCLPTSASTAAALLRQTAPPVSPVQSAALPEPPSSSSSTAADRRASSIAALPLAKRHSYSLQRNQVPTVN-----	: 336
SW: ARX_MOUSE/1	LSTFLGAAVFRHFAFISPAFCRLFFSTMAPLTSASTAAALLRQTAPPVGAASGALADP-----ATAAADRRASSIAALPLAKRHSYSLQRNQVPTVN-----	: 404
SW: AL_DROME/1-1	PPTSPASGHAXPQVLVGIATLQQAASSLSPT---QTSFVALTSHSPQRQLPSPHQAPPPPPAAATPPEDRRTSSIAALPLAKRHSYSLQRNQVPTVN-----	: 313
SW: ALX4_MOUSE/1	DFL-----SVSGACSHVQTHMCSLFGAACISPLNGCYELNCGPDRKTSIAALPLAKRHSYSLQRNQVPTVN-----	: 354
SW: ALX4_HUMAN/1	DFL-----SVSGACSHVQTHMCSLFGAACISPLNGCYELNCGPDRKTSIAALPLAKRHSYSLQRNQVPTVN-----	: 366
SW: RX2_CHICK/1	LPASYTPPPFL-----NSPVTGHALQPLGAMGPPPPYQCGAFAVDFKFLDEGDPNRTSSIAALPLAKRHSYSLQRNQVPTVN-----	: 290
SW: RX2_BRARE/1	LQPTYTAHPCFL-----NTSPGMHNQIQPM---PPPPYQVPPVFNKDYPLEEDVD---RSSIAALPLAKRHSYSLQRNQVPTVN-----	: 297
SW: RX1_XENLA/1	LPASYTPPPFI-----NPVSVGHALQPLGAMGPPPPYQCGAFAVDFKFLDEEDPDMSSIAALPLAKRHSYSLQRNQVPTVN-----	: 296
SW: RX_HUMAN/1-1	LPASYTPPPPPFL-----NSPPLGCPQLPL---APPPSYPCGCFGDKFLDEADPNSSIAALPLAKRHSYSLQRNQVPTVN-----	: 319
SW: PIX2_BRARE/1	SISMSMSSSMVPSAVTCVPGSSL-----NSLNNLNLNSPNSLNSVPTPACPYAPPTPPY-VYRDTCNSSLASLPLAKRHSYSLQRNQVPTVN-----	: 314
SW: PIX2_HUMAN/1	SISMSMSSSMVPSAVTCVPGSSL-----NSLNNLNLNSPNSLNSVPTPACPYAPPTPPY-VYRDTCNSSLASLPLAKRHSYSLQRNQVPTVN-----	: 317
SW: PIX1_HUMAN/1	SISMTMPSSMCPGAVPGMNSCL-----MNIN---MLTGSSLNSAMSFGACPYCTPASPYVYRDTCNSSLASLPLAKRHSYSLQRNQVPTVN-----	: 314
SW: OTP_MOUSE/1	SQCSLAAGPPPNMCLNSLNSLNSGACGLQ---SHLYQPAFFGMVPSALPGSPNSVSGSLQCLSSPDSVDRGTSIAALPLAKRHSYSLQRNQVPTVN-----	: 325

Домены принято изображать так

[X1WJ92 ACYPI](#) [Acyrtosiphon pisum (Pea aphid)]
Uncharacterized protein (408 residues)



There are 1836 sequences with the following architecture:
Homeodomain, OAR

Гомеодомен является ДОМЕНОМ

Доказательство

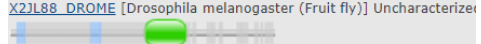
- Выравнивание, со сходимостью, свидетельствующим о гомологичности последовательностей
- Представленность домена в негомологичных белках (обычно, но не обязательно)

```

G4VRE6_SCHMA/203-259
HME2B_DANRE/173-229
HM1N_BOVW2/373-429
TIEHES_HELRO/41-97
HM05_CAEL/36-92
G3IBX4_CRIGR/25-81
DLX3B_DANRE/136-192
F1QFR3_DANRE/136-192
B8A5N9_DANRE/135-191
Q91967_CHICK/77-133
DLX3B_DANRE/126-182
HM04_CAEL/193-159
HM23_CAEL/212-268
MSX3_MOUSE/88-144
HM30_CAEL/96-152
BARH2_RAT/230-286
BARH1_DROME/300-356
BARX2_MOUSE/138-194
DSH1_DROME/275-331
H2XU06_CIOIN/470-524
HM19_CAEL/95-151
SLOU_DROME/546-602
F6VWQ6_XENTR/112-168
TIN_DROME/302-358
NKX25_RAT/138-194
H0XK12_OTOGA/100-156
HM09_CAEL/71-127
H2VEX2_TAKRU/135-69
U3K517_FICAL/59-115
TLX3_CHICK/173-229
U3ZQ6_FICAL/136-188
LBX1_MOUSE/126-182
G4VGG4_SCHMA/38-94
BCD_DROME/98-153
BCD_DROME/98-153 (55)
VENTX_HUMAN/92-148
VENT1_XENTR/128-184
Q804C9_XENTR/190-246
K48B21_SOLLCL/24-79
PHO2_YEAST/78-134
WOX9_ARATH/52-113
WOX9_ORYSJ/111-72
WOX2_ORYSJ/24-85
WOX4_ARATH/87-148
WOX1_ARATH/73-134
WOX2_ARATH/11-72
WOX5_ORYSJ/41-102
WUS_SOLLCL/25-85
WOX6_ARATH/58-119
YHP1_YEAST/174-230
YOX1_YEAST/177-233
HARA_DICDI/163-219
PHX1_SCHPO/169-223
CUT_DROME/1746-1802
CUX2_MOUSE/1114-1170
CUX1_MOUSE/1240-1296
Q22810_CAEL/212-268
HBX2_DICDI/486-542

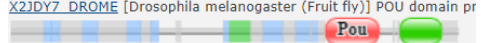
```

There are 25976 sequences with the following architect



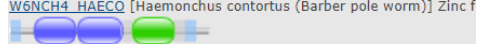
[Show all sequences with this architecture.](#)

There are 2311 sequences with the following architectu



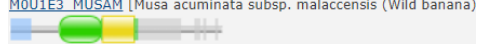
[Show all sequences with this architecture.](#)

There are 2108 sequences with the following architectu



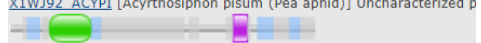
[Show all sequences with this architecture.](#)

There are 1903 sequences with the following architectu



[Show all sequences with this architecture.](#)

There are 1836 sequences with the following architectu



Эволюционные домены

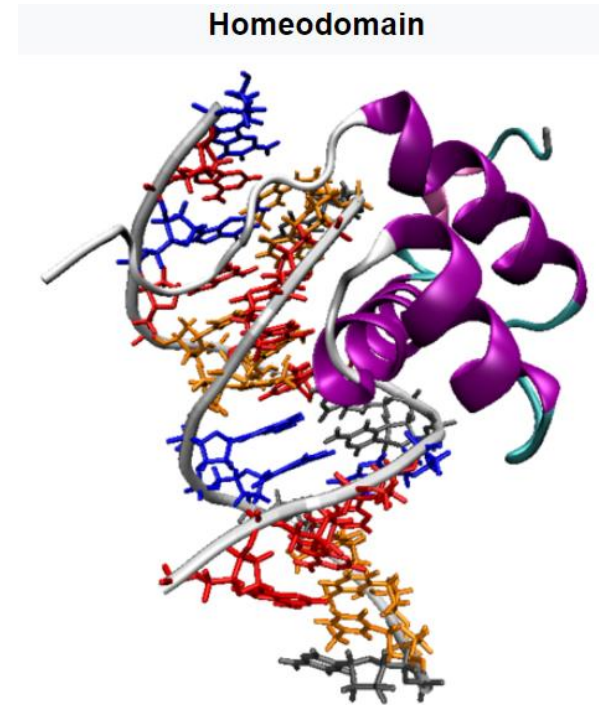
- Имеют определенную функцию (не всегда известна)

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]

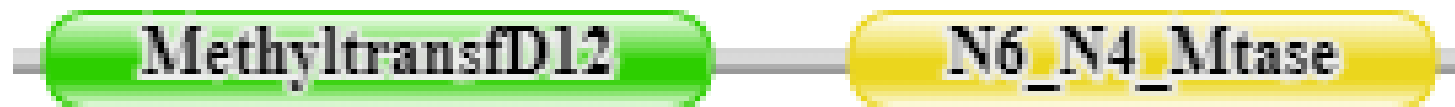


Как выровнять эти две
последовательности?

There are 9 sequences with the following architecture:

MethyltransfD12, N6_N4_Mtase

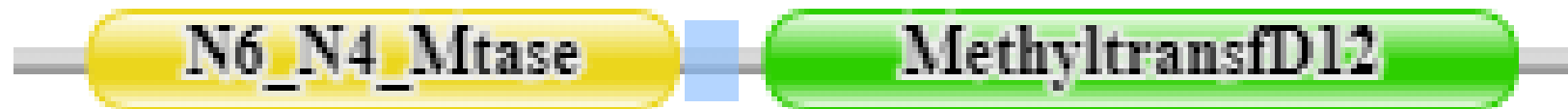
[A0A2Z5QVW5](#) [9MICC](#) [**D12-N6_N4**]



There are 5 sequences with the following architecture:

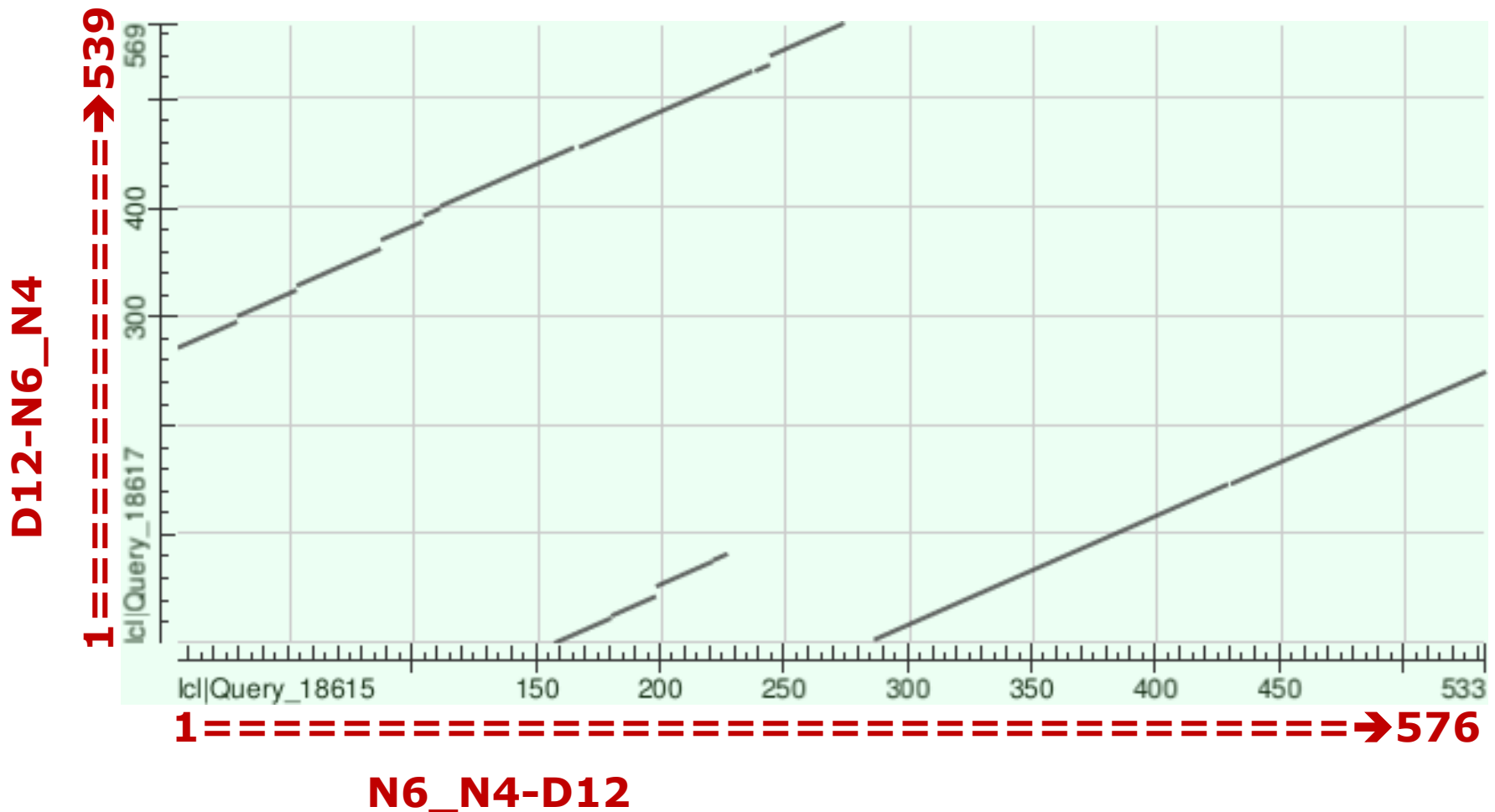
N6_N4_Mtase, MethyltransfD12

[A0A1I7GYG0](#) [9CLOT](#) [**N6_N4-D12**]



Как такое может возникнуть?

Лучшее парное выравнивание:
алгоритм множественных локальных
выравниваний.



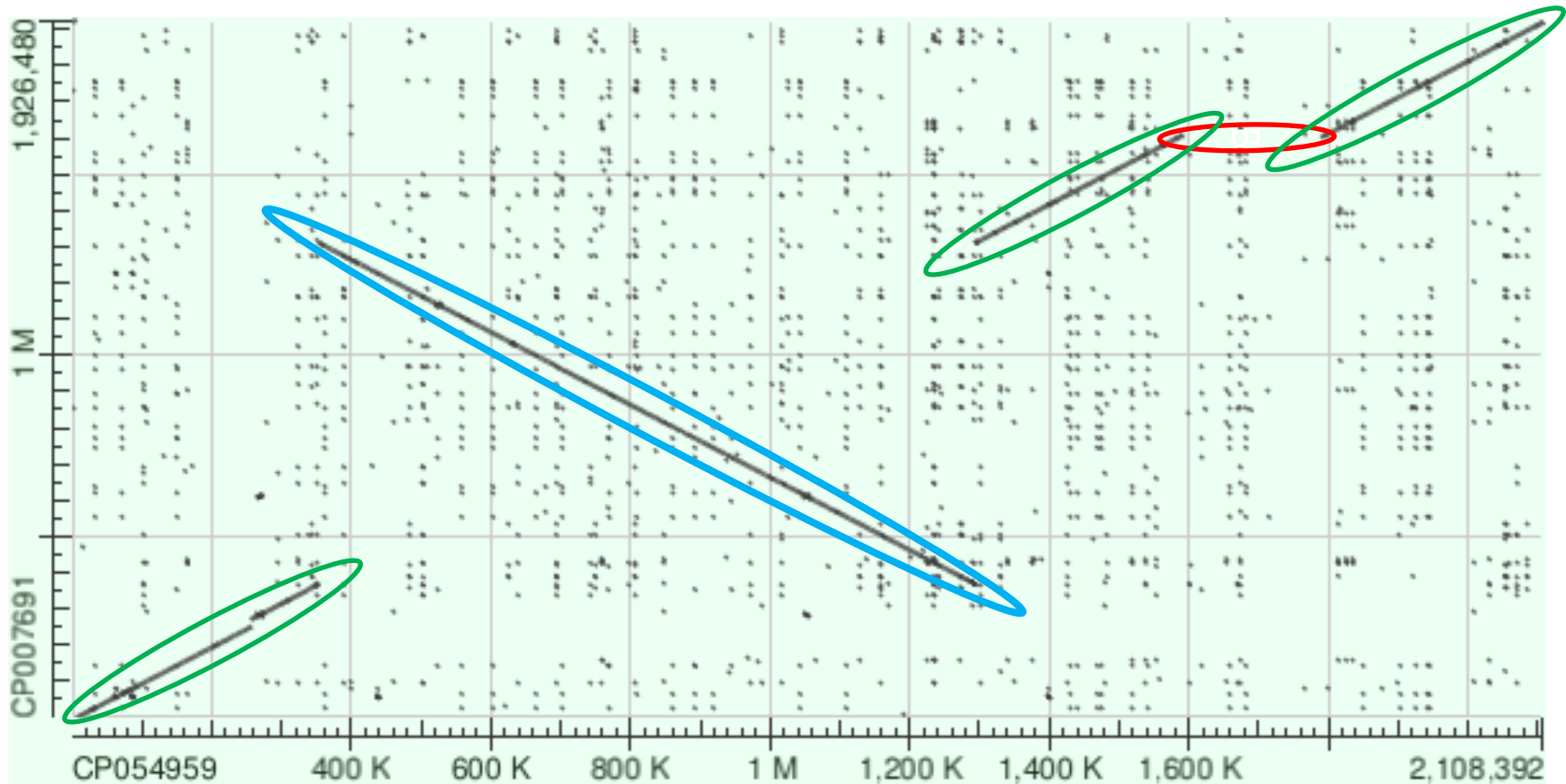
Программа BLASTp. Визуализация Dot Plot

Эволюция геномов бактерий

Крупные - единовременные изменения в геномах:

- Делеция большого участка (многие сотни, тысячи и миллионы пар нуклеотидов)
- Дупликация большого участка
- Горизонтальный перенос, т.е. вставка большого участка из чужого генома
- Инверсия большого участка
- Транслокация большого участка

Карта локального сходства двух геномов родственных штаммов бактерий

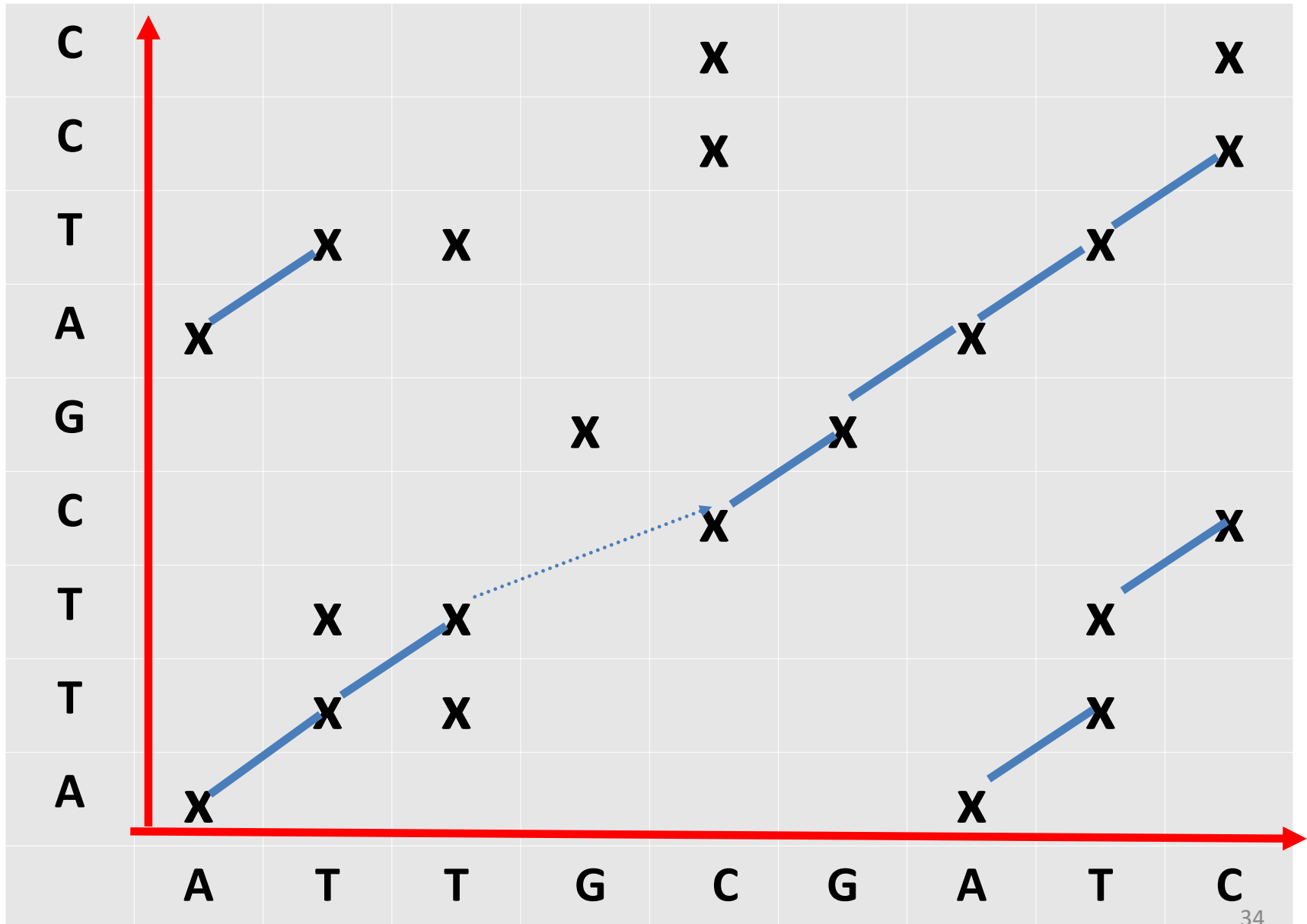


Сходство прямых цепочек

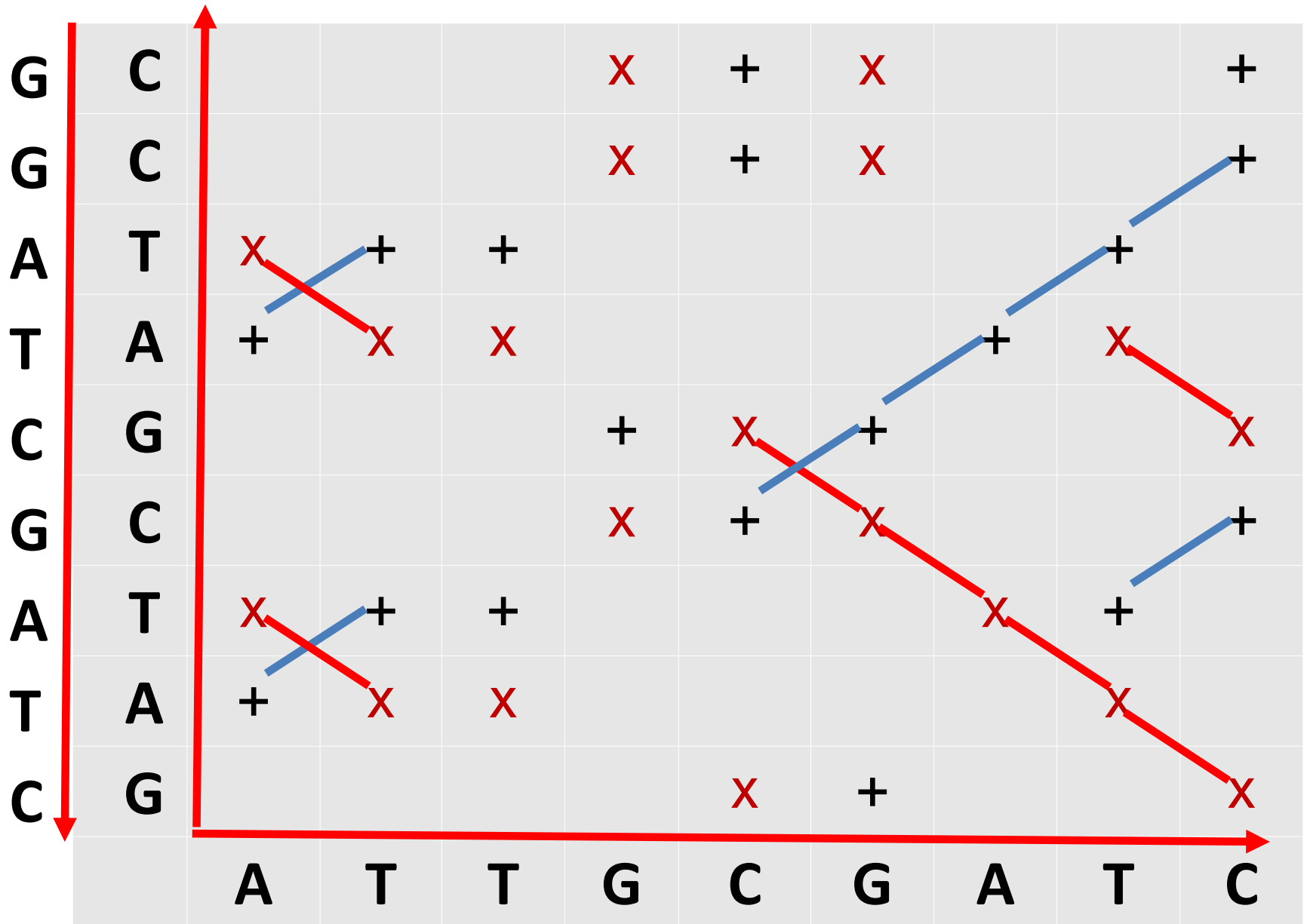
Инверсия - сходство прямой цепочки
с комплементарной второго генома

Делеция / вставка

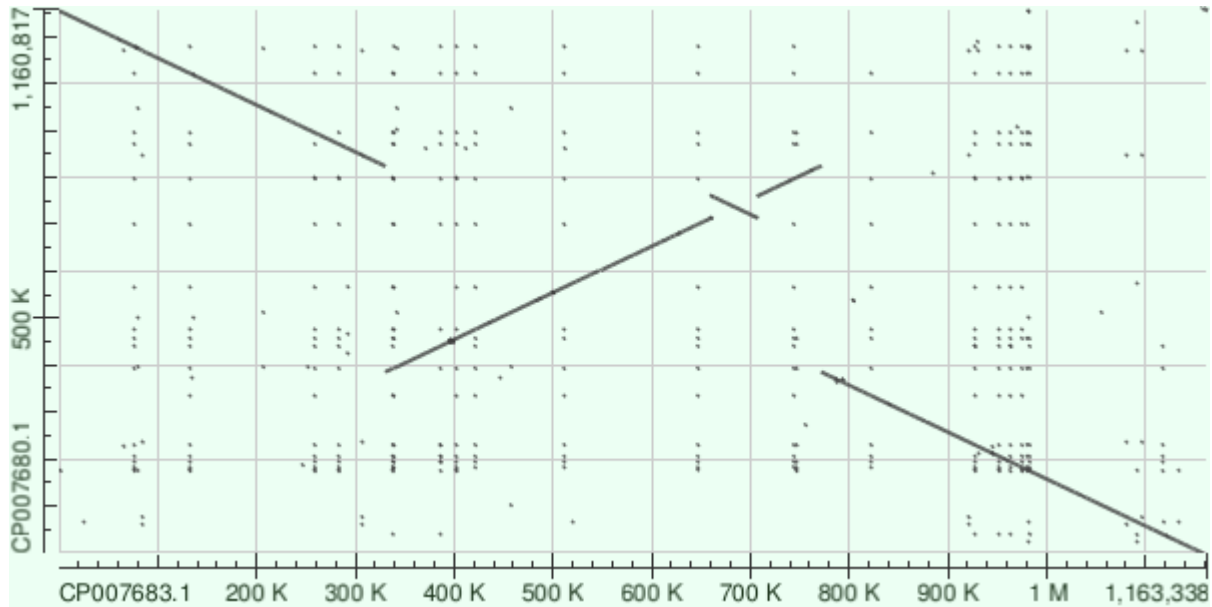
Объяснение Карты локального сходства



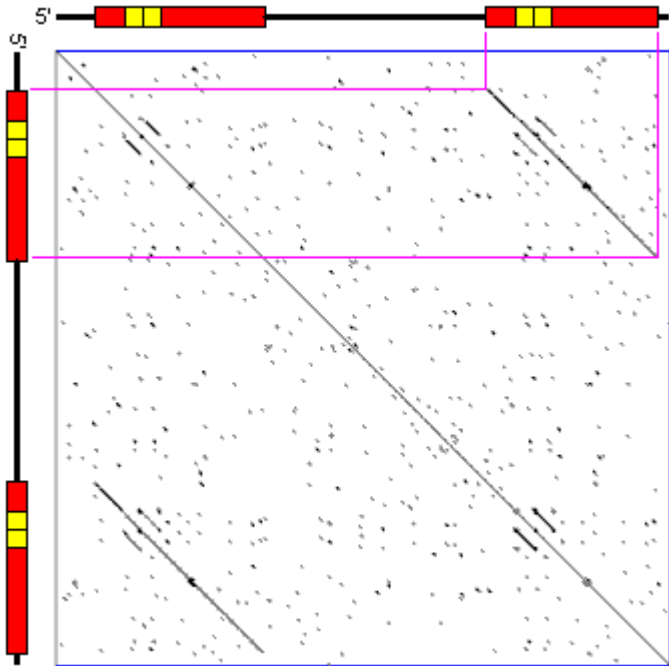
Карта сходства с учетом комплементарной цепочки



Что видим на этой карте?



Дупликация



Пример выполнения задания
Семейство доменов

КОНЕЦ ПРЕЗЕНТАЦИИ

Какие выравнивания тех же последовательностей совпадают?

	1	2	3	4	5	6	7	8	9	10	11	12	13
Seq1	M	K	F	R	-	S	S	H	Y	A	-	S	
Seq2	M	K	Y	R	R	R	-	H	Y	A	-	S	
Seq3	M	E	F	R	R	R	R	S	H	Y	A	-	R

	1	2	3	4	5	6	7	8	9	10	11	12
Seq1	M	K	F	-	R	S	S	H	Y	A	S	
Seq2	M	K	Y	R	R	R	-	H	Y	A	S	
Seq3	M	E	F	R	R	R	R	S	H	Y	A	R

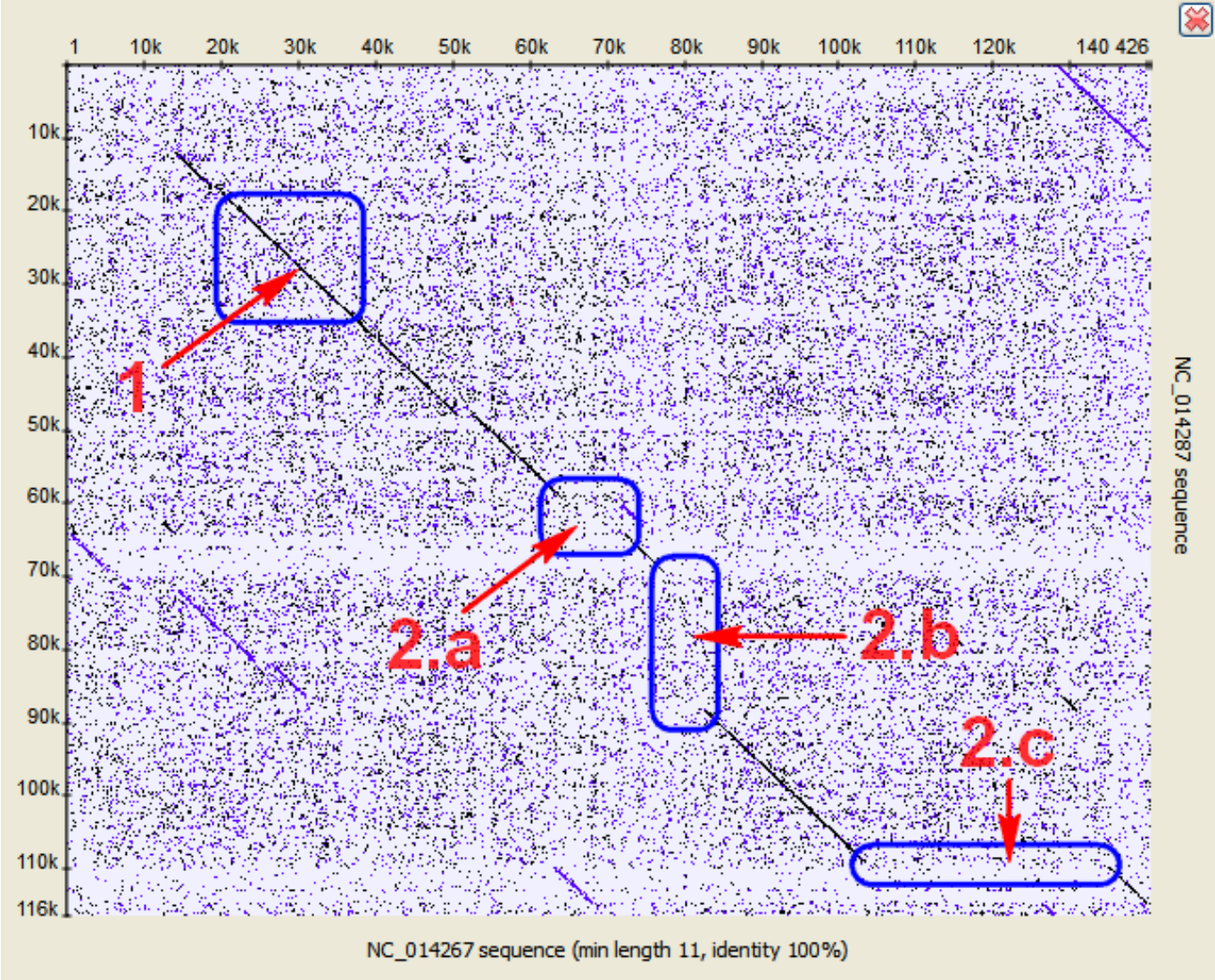
	1	2	3	4	5	6	7	8	9	10	11	12	13
Seq1	M	K	F	R	-	S	S	H	Y	A	-	S	
Seq2	M	K	Y	R	-	R	S	H	Y	A	-	S	
Seq3	M	E	F	R	R	R	R	S	H	Y	A	-	R

	1	2	3	4	5	6	7	8	9	10	11	12	
Seq1	-	M	K	F	R	-	S	S	H	Y	A	S	
Seq2	-	M	K	Y	R	-	R	S	H	Y	A	S	
Seq3	-	M	E	F	R	R	R	R	S	H	Y	A	R

Колонка i выравнивания X совпадает с колонкой j выравнивания Y если в них – те же самые остатки; те же самые значит – с теми же номерами, а не с теми же буквами!

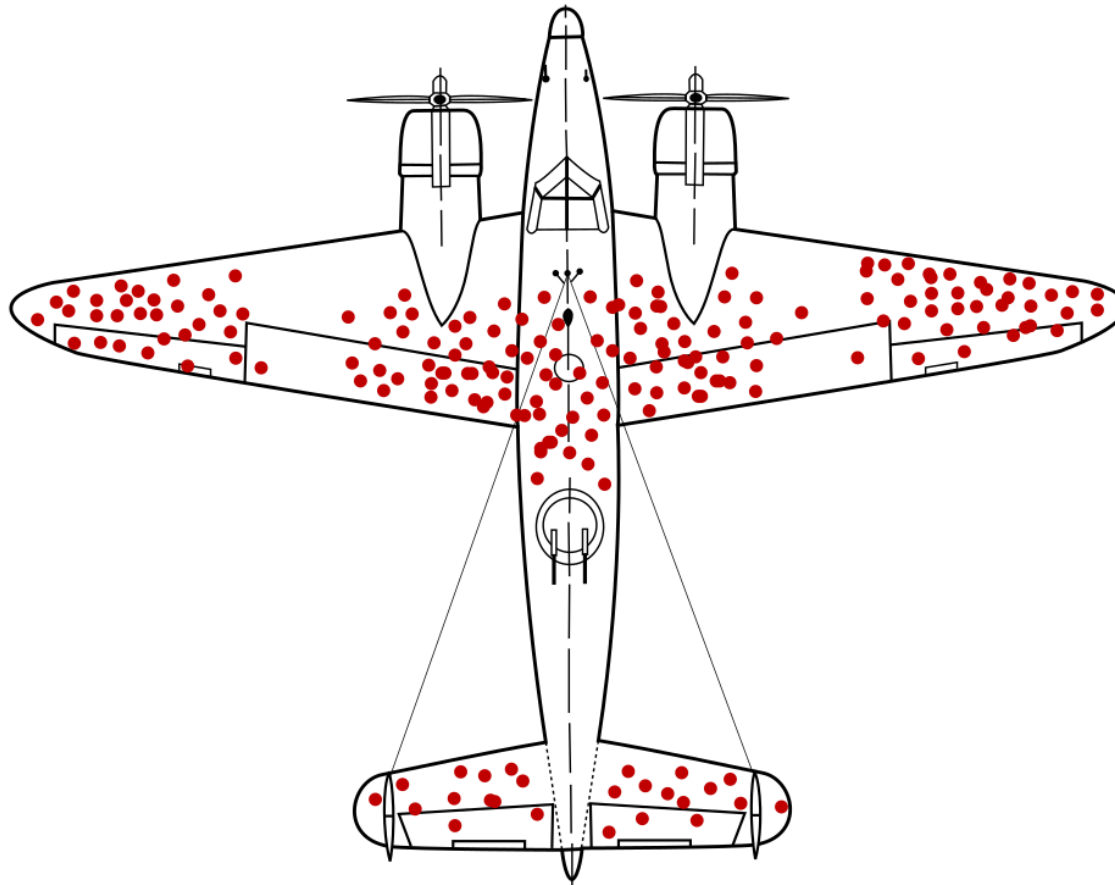
Interpreting Dotplot: Identifying Matches, Mutations, Inversions, etc

Создатель Yuliya Algaer, 2014



Пробоины на вернувшихся американских самолётах во время второй мировой войны.

Какие части укреплять броней?



Эволюция белков

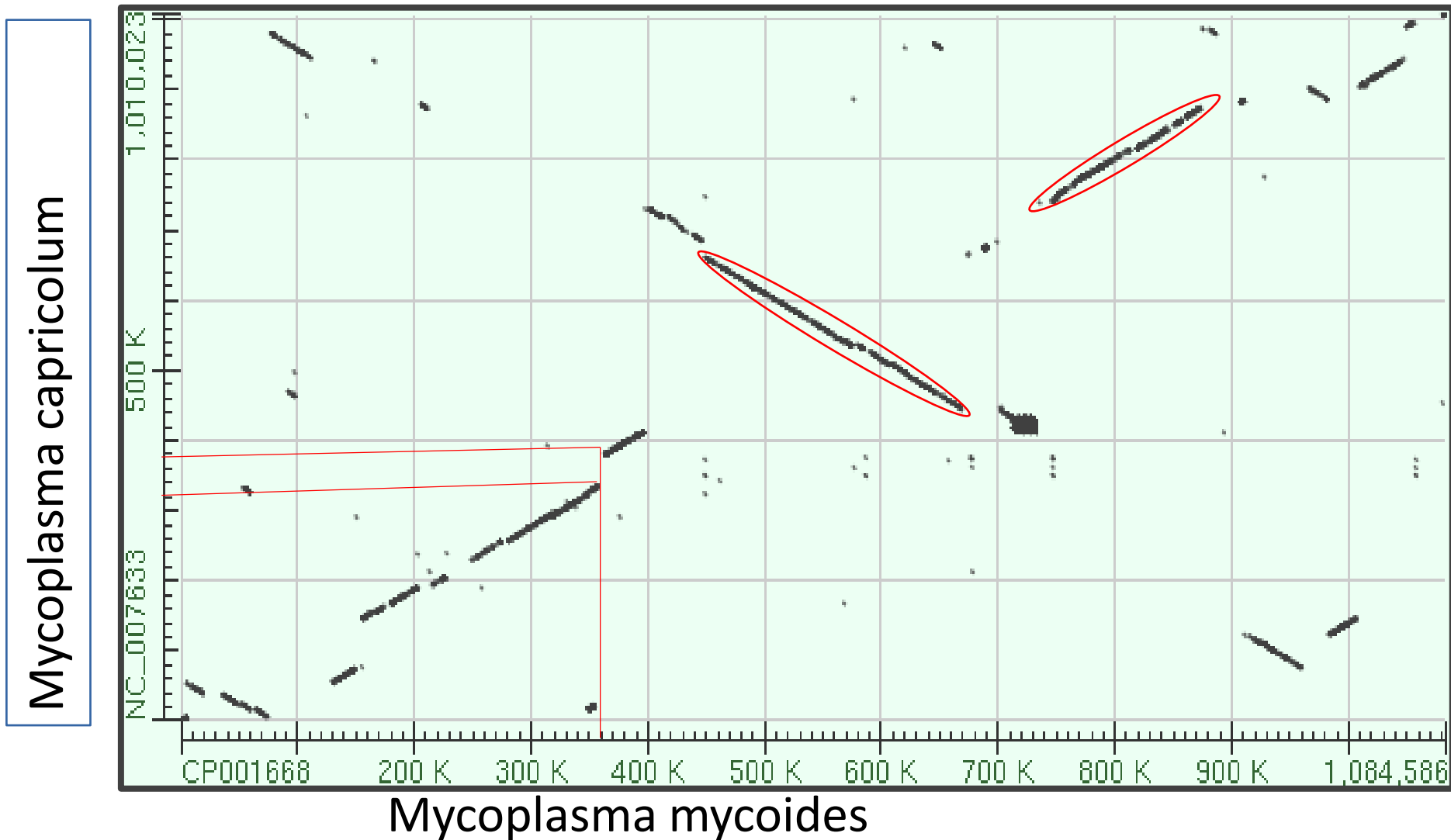
Локальная - небольшие изменения в гене
(Замены а.к. Делеции Вставки)

Большие изменения:

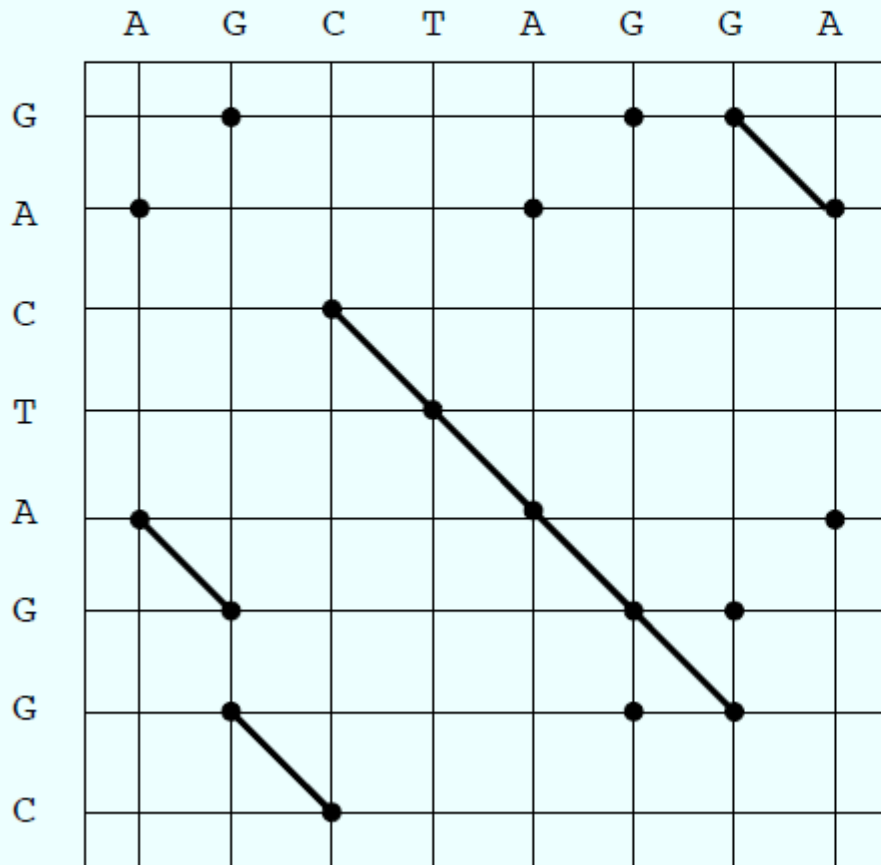
- 1) Накопленные небольшие изменения
- 2) Небольшие изменения гена ~~ведущие к большим~~ изменениям белка
 - 1) Мутация стоп кодона => удлинение последовательности белка
 - 2) Мутация кодона на стоп кодон
 - 1) гибель белка = псевдогенизация или
 - 2) Укорочение последовательности белка
 - 3) Программируемый сдвиг рамки считывания
 - 3) Мутация в сайте инициации (начала) трансляции
- 3) Крупные перестройки генома, затрагивающие гены!

2. Парное выравнивание геномов

Карта локального сходства геномов M. capricolum и M. mycoides



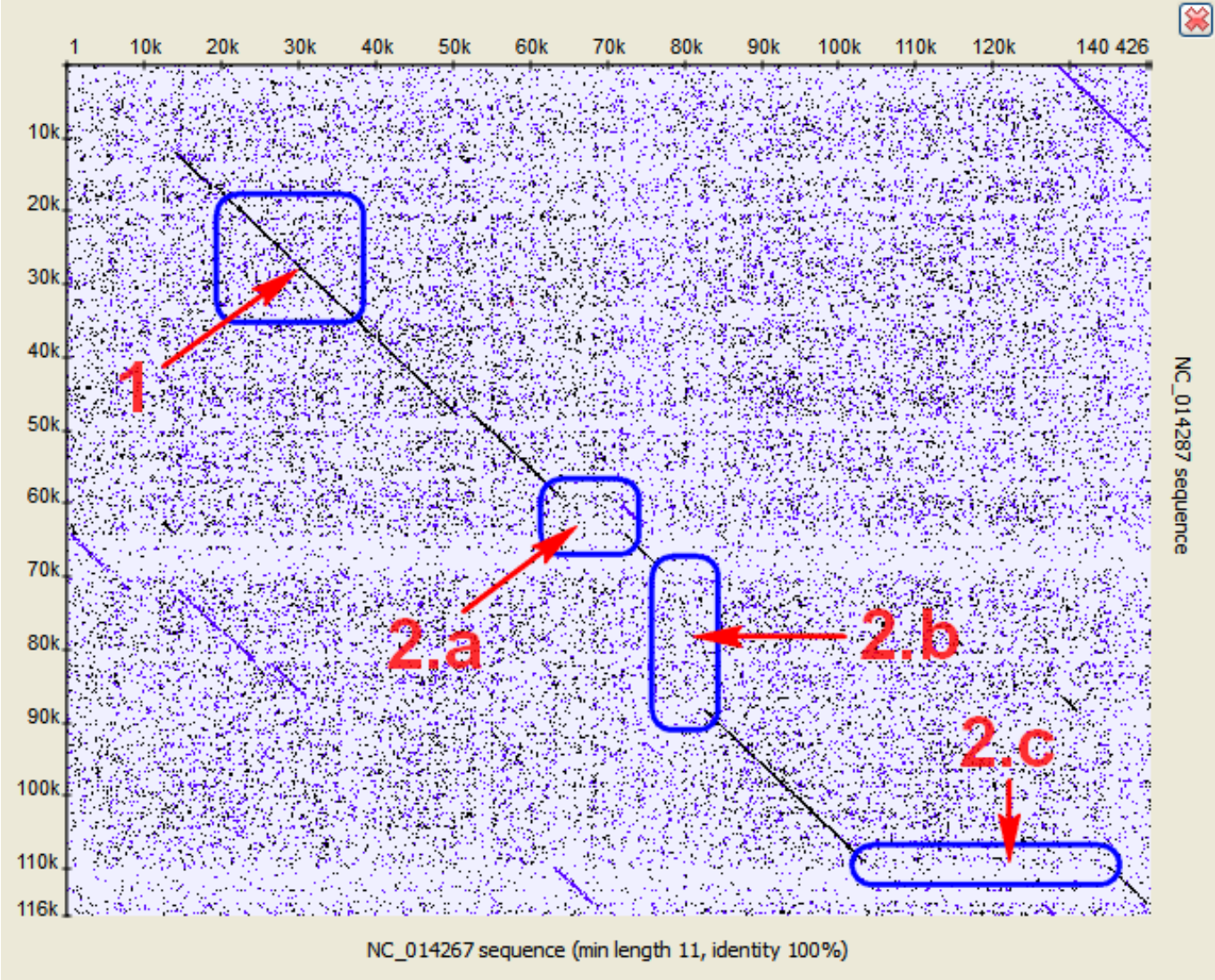
- established in 1970 by A.J. Gibbs and G.A.McIntyre
- method for comparing two amino acid or nucleotide sequences



- each sequence builds one axis of the grid
- one puts a dot, at the intersection of same letters appearing in both sequences
- scan the graph for a series of dots
 - reveals similarity
 - or a string of same characters
- longer sequences can also be compared on a single page, by using smaller dots

Interpreting Dotplot: Identifying Matches, Mutations, Inversions, etc

Создатель Yuliya Algaer, 2014



ВЫРАВНИВАНИЕ

DNAK_THEAC 82 KFKVFDKEFTPQQISAFILQKIKKDA-EAFLGEPVNEAVITVPAYFNDNQRR 131
 DNAK_PICTO 82 KYKIFGKEYTPQQISAFILQKIKRDA-EAFLGEPVTDAVITVPAYFNDNQRR 131
 HSCA_ACIF2 116 RLRTVAGEKSPVEVSAEILRVLKERAVETLGGEPEGAVITVPAYFDEAQR 165
 HSCA_ACIF5 116 RLRTVAGEKSPVEVSAEILRVLKERAVETLGGEPEGAVITVPAYFDEAQR 165

DNAK_THEAC 132 QATKDAGTIAGFDVKRIINEPTAAALAYGVDKSGKSEKILVFDLGGGTLDV 182
 DNAK_PICTO 132 QATKDAGAIAGLNVRRRIINEPTAACLAYGIDKLNQTLKIVIYDLGGGTLDV 182
 HSCA_ACIF2 166 QATKDAARLAGLNVLLAEPTAAAVAYGLDKGSEGI-FAIYDLGGGTFDI 215
 HSCA_ACIF5 166 QATKDAARLAGLNVLLAEPTAAAVAYGLDKGSEGI-FAIYDLGGGTFDI 215

DNAK_THEAC 183 TIMDFGDGVFQVLSSTSGDTRLGGTDMDEAIVNYIADDFQKKEGIDLKDRS 233
 DNAK_PICTO 183 TIMDFGQGVFQVLSSTSGDTHLGGTDMDEAIVNFLADNFQRENGIDLKDHHS 233
 HSCA_ACIF2 216 SILRLQAGVFEVLATAGDSALGGDDMDHALAEWLMQE- - -EGGDASDPLW 262
 HSCA_ACIF5 216 SILRLQAGVFEVLATAGDSALGGDDMDHALAEWLMQE- - -EGGDASDPLW 262

DNAK_THEAC 234 AYIRLRDAAEKAKKIELSTTLSTDIDLPIYITVTNSGPKHKIKMTLTRAKLEEL 284
 DNAK_PICTO 234 AYIRLRDAAEKAKKIELSTVLETEINLPIYITATQDGPKHLQYTLTRAKFEEL 284
 HSCA_ACIF2 263 RRQVLQQ-ARTAKEALSVAEET- - -MIVLTPSGRAAREIKLSRGRLES 307
 HSCA_ACIF5 263 RRQVLQQ-ARTAKEALSVAEET- - -MIVLTPSGRAAREIKLSRGRLES 307

DNAK_THEAC 285 ISPIVERVKGPIDKALEGAKLKKTEITKLLFVGGPTRIPYVRKYVEDYLG 335
 DNAK_PICTO 285 IAPIVDRSKVPLDTALEGAKLKKGDIDKIILIGGPTRIPYVRKYVEDYFGR 335
 HSCA_ACIF2 308 IQPVIQRSLPACRRALRDAGLKLDEIEGVVLVGGATRVPVAVRAMVEEFFRQ 358
 HSCA_ACIF5 308 IQPVIQRSLPACRRALRDAGLKLDEIEGVVLVGGATRVPVAVRAMVEEFFRQ 358