

Выравнивание ...



... последовательностей белков
и его биологический смысл

Мутации белка – следствия мутаций кодирующей последовательности его ген

- Последовательность одного и того же белка у потомка может отличаться от последовательности предка
- Отличия происходят из-за мутаций в кодирующей последовательности гена

CDS, coding sequence –
кодирующая
последовательность гена

atgсссаааgсс.....tga

Начало CDS: **atg**

Стоп-кодоны: **taa, tag, tga**

Геномная ДНК

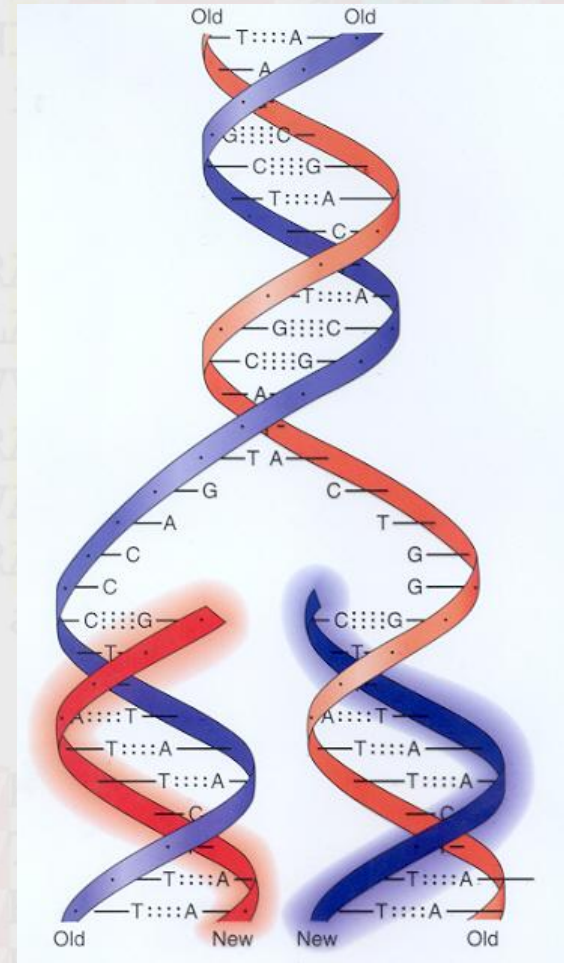
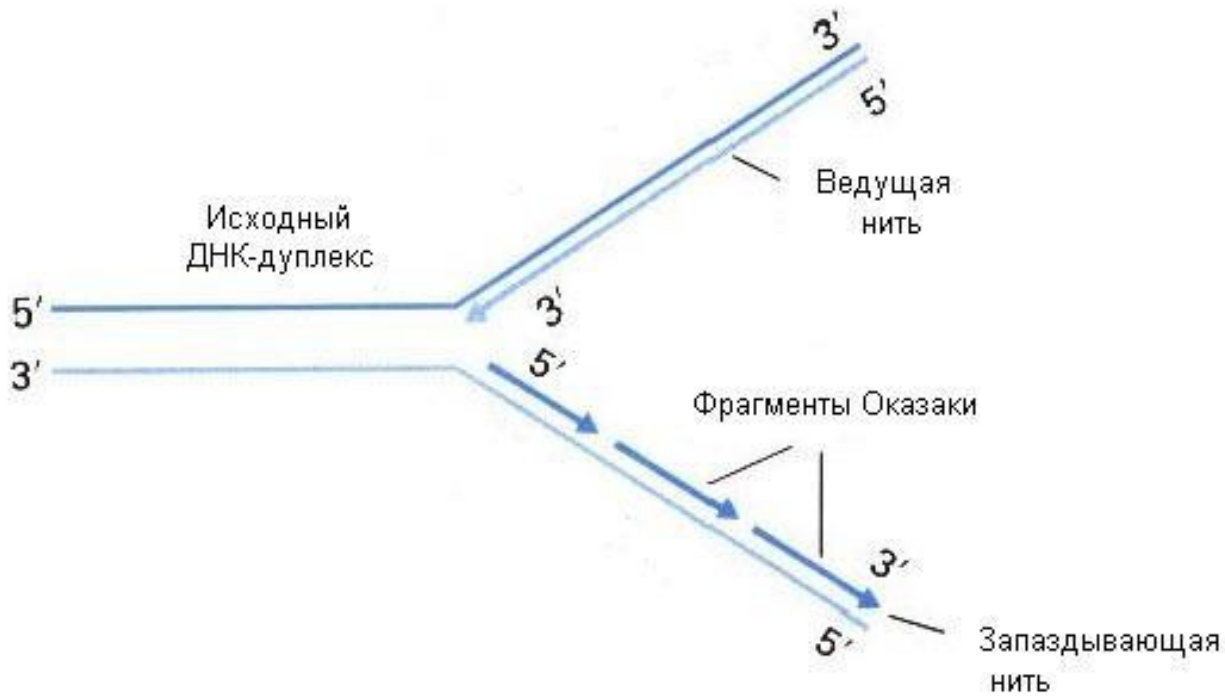
- Большая молекула
- Полимер
- Линейный
- Ориентированный
- Гетерополимер
- Двухцепочечный
- Цепочки ориентированы
противоположно

Каждая комплементарная пара нуклеотидов ДНК потомка происходит из комплементарной пары ДНК предка

- ДНК потомка всегда получается из предковой ДНК репликацией
- Поэтому ДНК-потомок имеет ту же последовательность, что и ДНК-предок
- *Закон биологии: из каждого закона биологии существуют исключения*
- *Вопрос: существуют ли исключения из этого закона?*

Схема репликации

РЕПЛИКАТИВНАЯ ВИЛКА



Виды мутаций

- Замена нуклеотида на другой
- Делеция одного или нескольких подряд нуклеотидов
- Вставка одного или нескольких подряд нуклеотидов
- Дупликация участка
- Перестановка участка - рекомбинация

Причины мутаций разнообразны

- Ошибки ДНК-зависимой ДНК-полимеразы
- Мутации из-за мутагенов – химических соединений, радиации, ультрафиолета и др.
- Ошибки репарации
- Работа разнообразных систем рекомбинации
- Специально запланированные механизмы вариации ДНК

• ...

Мутации CDS и отбор

- Мутации белка происходят из-за мутаций кодирующей последовательности ДНК
- Последовательность белка находится под прессом отбора:
 - каждый белок имеет свои биологические функции
 - функция белка зависит от структуры белка и от остатков, напрямую участвующих в выполнении функции
 - свойства и структура белка зависят от последовательности
- Малое число мутаций CDS имеют шанс закрепиться в эволюции
- Мутации последовательности белка происходят редко; *но если подождать 10 млн лет, то можно обнаружить довольно много!*

Как отобразить отношение предок – потомок для нуклеотидов?

ДНК предка:**AAASTGATGCAACGTGA**.....



ДНК потомка:**AAtCttTGATAcсTGA**.....

Выравнивание – общепринятый способ отражения “родства” нуклеотидов или аминокислотных остатков

seq1:.....**AAAC--TGATGCAACGTGA**.....

seq2:.....**AAtCttTGAT---AcсTGA**.....

“Наследование”

аминокислотного остатка белка

- Если все три нуклеотида **кодона** потомка унаследованы от трех нуклеотидов **кодона** предка, то можно говорить о том, что остаток потомка произошел от остатка предка
- Такие остатки следует помещать в одну колонку выравнивания аминокислотных последовательностей
- Это еще не вся правда про выравнивание последовательностей белков!

Проблема выравнивания

- Время наблюдения за последовательностью много меньше 10 млн лет 😞
- Мы наблюдаем только потомков общего предка, а самого предка не знаем
- Про родство аминокислотных остатков приходится догадываться
- Как догадываться – об этом на следующих занятиях

Аминокислотные остатки помещают в одну колонку выравнивания если они

- происходят от одного предкового остатка последовательности белка – общего предка (**эволюция**)
- их C_{alpha} атомы находятся в участках полипептидной цепи сходной конформации (**структура**)
- играют сходную роль в белке (**функция**)

Пример выравнивания

```

                *           240           *           260           *           280
1zjh:A : FG-VEQDVDMVFASFI RKASDVHEVRKVLGE-KGKNIKIISKIENHEGVRRFDEIL : 260
1a3x:B : FG-VKNGVHMVFASFI RTANDVLTIREVLGE-QGKDVKIIIVKIENQQGVNMFDEIL : 254
3eoe:B : NFgIPMGCNFIAASFVQSADDVRYIRGLLGP-RGRHIRIIPKIENVEGLVNFDEIL : 249
1aqf:B : FG-VEQDVDMVFASFI RKAADVHEVRKILGE-KGKNIKIISKIENHEGVRRFDEIL : 272
1pk1:G : FG-VEQGVDMIFASFI RSAEQVGDVRKALGP-KGRDIMIICKIENHQGVQNIDSI : 252
1e0t:A : FG-CEQGVDFVAASFIRKRSDVIEIREHLKAhGGENIHIIISKIENQEGLNMFDEIL : 230
fg      v      ASfir a dV      R      Lg      G      i      II      KIEN      G      fDeil

```

```

                *           300           *           320           *
1zjh:A : EASDGI MVARGLGIEI pA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmikkpR : 315
1a3x:B : KVT DGV MVARGLGIEI pA-PEVLAVQK KLI AKSNLAGKPVICATQmlesmtynpR : 309
3eoe:B : AEADGI MIARGDLGMEI pP-EKVFLAQKMMI AKCNVVGKPVITATQmlesmiknpR : 304
1aqf:B : EASDGI MVARGLGIEI pA-EKVFLAQKMI I GRCNRAGKPVICATQmlesmikkpR : 327
1pk1:G : EESDGI MVARGLGVEI -PaEKVVVAQKILI SKCNVAGKPVICATQmlesmtynpR : 307
1e0t:A : EASDGI MVARGLGVEI pV-EEVIFAQKMMI EK CIRARKVVI TATM-----R : 276
DGiMvARGDLG EI p e V aQK I cn agKpVI ATqmlesm pr

```

```

                340           *           360           *           380           *
1zjh:A : PTRAE GSDVANAVLDGAD C IMLSGETakgdy PLEAVRMQH LIAREAEAaiyhlqlf : 371
1a3x:B : PTRAE VSDVGNAILD GAD CVMLSGETakgny PINAVTTMAETAVIAEQaiaylpny : 365
3eoe:B : PTRAE AADVANAVLDG TDCVMLSGETangef PVITVETMARICYEAE Tcvdypaly : 360
1aqf:B : PTRAE GSDVANAVLDGAD C IMLSGETakgdy PLEAVRMQH LIAREAEAamfhrklf : 383
1pk1:G : PTRAE VSDVANAVFNGAD CVMLSGETakgky PNEVVQYMARICLEAQSalneyvff : 363
1e0t:A : PTDAE AGDVANAILDGT DAVMLSGET----- PLEAVSIMATICERTDR----- : 318
PTrAE  DVaNA ldG Dc MLSGETa g P V i a

```

Что делать с остатками, которые не должны выравниваться?

1zjh:A : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL
1a3x:B : FG-VKNGVHMVFASFIKASDVHEVRKVIIGE-QCKDVKLIIVKIENQOQGVNDFDEIL
3eoe:B : NFgIPMGCNFIAASEVQSADDVRYIRGLIGP-RGRHIRIIPKTIENVEGLVNFDEIL
1aqf:B : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL
1pk1:G : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL
1e0t:A : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL
fg v ASfir a dV R Lg G i II KIEN G fDeIl

300 320

1zjh:A : EASDGI MVARGDLGIEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmikkpR
1a3x:B : KVTIDGVMVARGDLGIEIpA-PEVLAVQKKLIAKSNLAGKPVICATQmlesmtynpR
3eoe:B : AEADGIMVARGDLGMEIpP-EKVFLAQKMMIAKCNVVGKPVITATQmlesmiknpR
1aqf:B : EASDGI MVARGDLGIEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmikkpR
1pk1:G : EESDGI MVARGDLGVEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmtynpR
1e0t:A : EASDGI MVARGDLGVEIpV-EEVIFAQKMMIEKCI RARKVVITATM-----R
DGIMvARGDLG EIp e V aQK I cn agKpVI ATqmlesm pR

340 * 360 * 380 *

1zjh:A : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAAaiyhqlf
1a3x:B : PTRAEVSDVGNAILDGCDCVMSLGETakgnypINAVTTMAETAVIAEQaiaylpny
3eoe:B : PTRAEAAADVANAVLDGTDCCVMSLGETangefDVITVETMARICYEAETcvdypaly
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAAamfhrklf
1pk1:G : PTRAEVSDVANAVFNGADCDCVMSLGETakgkypNEVVQYMARICLEAQSaIneyvff
1e0t:A : PTDAAEAGDVANAILDGTDAVMSLGE-----PLEAVSIMATICERTDR-----
PTrAE DVaNA ldG Dc MLSGETa q P V i a

Выравнивание и эволюция

```
          *          20          *          4
POLG_CXB4J : GAQVSTQKTGAHETSLASGNSIIHYTNINYYKDAASNS : 39
POLG_CXB4E : GAQVSTQKTGAHETSLSATGNSIIHYTNINYYKDAASNS : 39

          0          *          60
POLG_CXB4J : ANRQDFTQDPSKFTEPVKDVMIKSLPALN : 68
POLG_CXB4E : ANRQDFTQDPSKFTEPVKDVMIKSLPALN : 68
```

Последовательности белка оболочки из двух штаммов
вируса Коксаки

| | | | | | | | | | |
|--------------|------------------|--------------|-------------|---------|--------|-------------|---|----|--|
| | | * | | 20 | | * | | 4 | |
| POLG_CXB4J : | GAQVSTQKTGAHETSL | SASGNSIIHYTN | INYYKDAASNS | : | 39 | | | | |
| POLG_CXB4E : | GAQVSTQKTGAHETSL | SATGNSIIHYTN | INYYKDAASNS | : | 39 | | | | |
| POLG_HE71B : | GSQVSTQRS | GS | HENSNS | SATEGST | INYYTT | INYYKDSYAAT | : | 39 | |

| | | | | | | |
|--------------|--------------|-----------|------------|---|----|--|
| | 0 | | * | | 60 | |
| POLG_CXB4J : | ANRQDFTQDPSK | FTEPVKD | VMIKSLPALN | : | 68 | |
| POLG_CXB4E : | ANRQDFTQDPSK | FTEPVKD | VMIKSLPALN | : | 68 | |
| POLG_HE71B : | AGKQSLKQDP | DFEANPVKD | IFTEMAAPLK | : | 68 | |

Последовательности белка оболочки из двух штаммов вируса Коксаки и энтеровируса человека

Что есть что (редактор GeneDoc)

Номер столбца выравнивания

| | | | | | | | | | | | | | | | | |
|------------|---|------|-------|-----|------|----|------|--------|-------|-----|----|----|----|-----|----|----|
| | | | * | | 20 | | * | | | | | | | | | |
| MTA1_YEAST | : | --- | KSSIS | SPQ | ARAF | LE | QV | RRK--- | QSLNS | : | 24 | | | | | |
| MAT2_YEAST | : | KPYR | GH | RFT | KE | NV | RI | LE | SW | EAK | NI | EN | PY | LDT | : | 31 |
| | | | | 3 | 2 | | LE | F | 4 | | | | | L13 | | |
| | | | | | | | | | | | | | | | | |
| | | | | 40 | | * | | 60 | | | | | | | | |
| MTA1_YEAST | : | KEKE | EE | VAK | KCG | IT | PL | QVR | VWF | INK | MR | SK | - | : | 53 | |
| MAT2_YEAST | : | KGLE | NI | MKN | TS | LS | RI | QIK | NW | VS | NR | RR | KE | KT | : | 61 |
| | | K | E | 6 | K | 63 | 6Q64 | W | N4R | 4 | K | | | | | |

Название последовательности

Консервативный остаток

Функционально консервативная позиция

Номер последнего в строке остатка ИЗ ЭТОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Сегодня все

1zjh:A : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL
1a3x:B : FG-VKNGVHMVFASFIKASDVHEVRKVIIGE-QCKDVKLIIVKIENQOQGVNDFDEIL
3eoe:B : NFgIPMGCNFIAASEVQSADDVRYIRGLIGP-RGRHIRIIPKTIENVEGLVNFDEIL
1aqf:B : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL
1pk1:G : FG-VEQGVDMIFAASFIKASDVHEVRKVIIGE-RDIMLICKIENHQGVQNDSDIIL
1e0t:A : FG-CEQGVDFVAASFIKASDVHEVRKVIIGE-KAHGSENIHLSKIENQOGLNDFDEIL
fg v ASfir a dV R Lg G i II KIEN G fDeIl

* 300 * 320 *

1zjh:A : EASDGI MVARGDLGIEIpA-EKVFLAQKMMIGRCNRAGKPVICATQmlesmikkpR
1a3x:B : KVTGDMVARGDLGIEIpA-PEVLAVQKLIIAKSNLAGKPVICATQmlesmtynpR
3eoe:B : AEADGIMVARGDLGMEIpP-EKVFLAQKMMIAKCNVVGKPVITATQmlesmiknpR
1aqf:B : EASDGI MVARGDLGIEIpA-EKVFLAQKMMIGRCNRAGKPVICATQmlesmikkpR
1pk1:G : EESDGI MVARGDLGVEIpA-EKVVVAQKILISKCNVAGKPVICATQmlesmtynpR
1e0t:A : EASDGI MVARGDLGVEIpV-EEVIFAQKMMIEKCIRARKVVITATM-----R
DGIMvARGDLG EIp e V aQK I cn agKpVI ATqmlesm pR

340 * 360 * 380 *

1zjh:A : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAaiyhqlf
1a3x:B : PTRAEVSDVGNAILDGCDCVMSLGETakgnypINAVTTMAETAVIAEQaiaylpny
3eoe:B : PTRAEAADVANAVLDGCTDCVMSLGETangefDVITVETMARICYEAETcvdypaly
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAamfhrklf
1pk1:G : PTRAEVSDVANAVFNGADCDCVMSLGETakgkypNEVVQYMARICLEAQSalneyvff
1e0t:A : PTRAEAGDVANAILDGTDAVMSLGE-----PLEAVSIMATICERTDR-----
PTRAE DVaNA ldG Dc MLSGETa q P V i a