

# Выравнивание ...



... последовательностей белков  
и его биологический смысл

1zjh:A : FG-VEQDVMVFASFIKASDVHEVRKVIKGE-KCKNIKLIISKIENHEGVRRFDEIL  
1a3x:B : FG-VKNGVHMVFASFIKRTANDVLTIREVIGE-QCKDVKLIIVKIENQQGVNDFDEIL  
3eoe:B : NFgIPMGCNFIAASEVQSADDVRYIRGLIGP-RGRHIRIIPKTIENVEGLVNFDEIL  
1aqf:B : FG-VEQDVMVFASFIKKAADVHEVRKIIGE-KCKNIKLIISKIENHEGVRRFDEIL  
1pk1:G : FG-VEQGVDMI FASEIRSAEQVGDVRKAI GP-KGRDIMLICKIENHQGVQNLDSIL  
1e0t:A : FG-CEQGVDFVAASEIRKRSDVIEIREHKAhGSENHLSKIENQOGLNDFDEIL  
fg v ASfir a dV R Lg G i II KIEN G fDeIl

\* 300 \* 320 \*

# Мутации и отбор

Сегодня основное внимание  
уделим мутациям

1zjh:A : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAaiyhqlf  
1a3x:B : PTRAEVSDVGNAILDGCDCVMLSGETakgnypINAVTTMAETA VIAEQaiaylpny  
3eoe:B : PTRAEAADVANAVLDGTDCCVMLSGETangefDVITVETMARICYEAETcvdypaly  
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAamfhrklf  
1pk1:G : PTRAEVSDVANAVFNGADCDCVMLSGETakgkypNEVVQYMARICLEAQSaIneyvff  
1e0t:A : PTDAEAGDVANAAILDGTDAVMLSGET-----PLEAVSIMATICERTDR-----  
PTAE DVaNA ldG Dc MLSGETa q P V i a

# 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

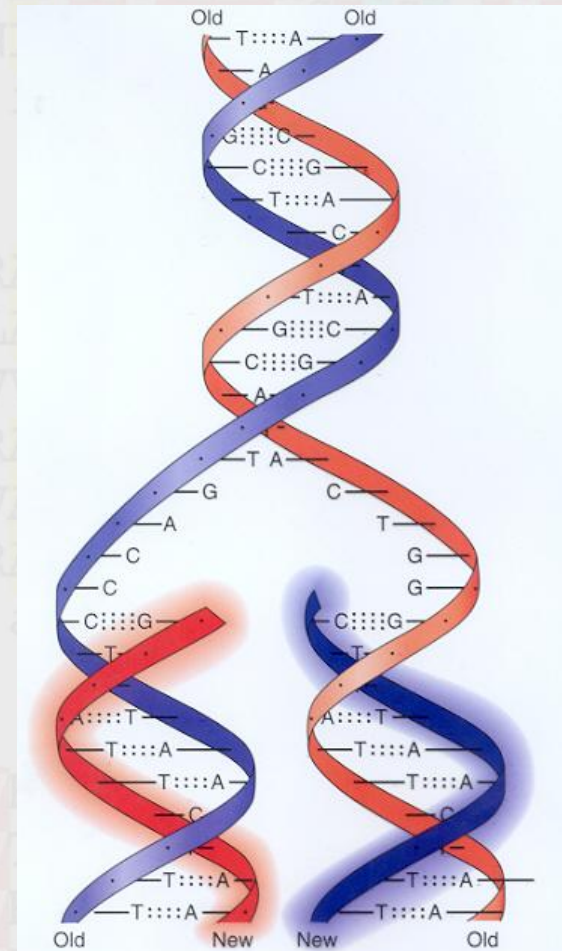
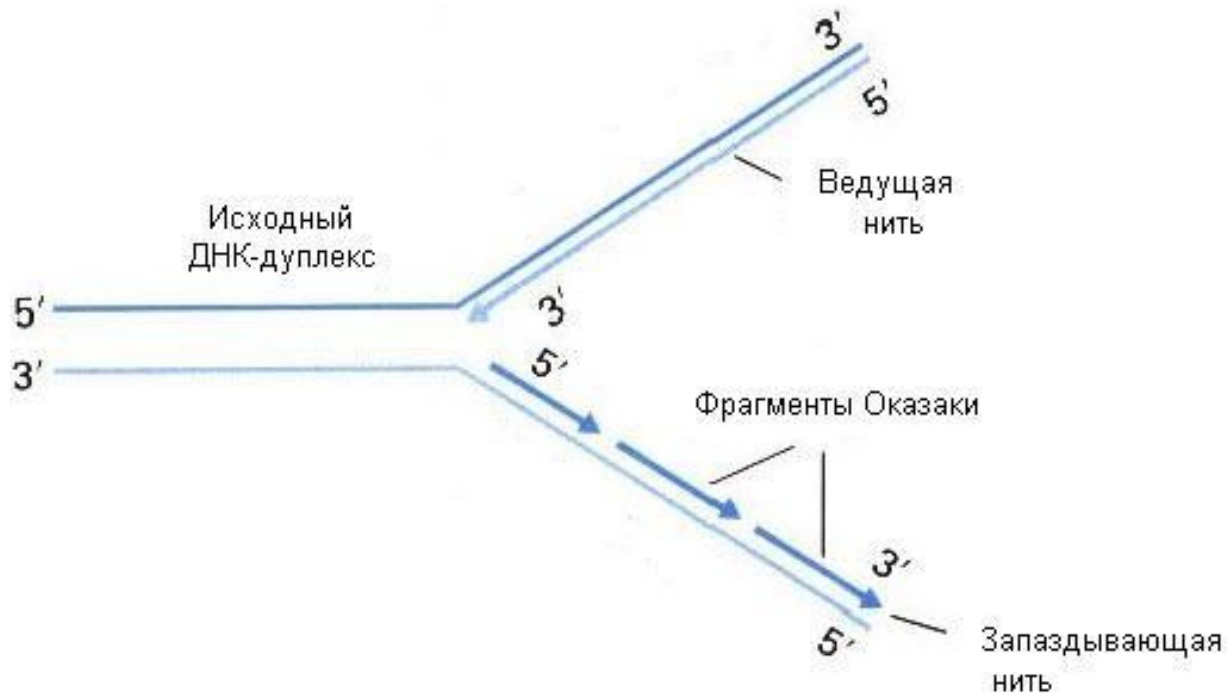
- Большая молекула
- Полимер
- Линейный
- Ориентированный
- Гетерополимер
- Двухцепочечный
- Цепочки ориентированы  
противоположно

# 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<sub>alpha</sub> атомы находятся в участках полипептидной цепи сходной конформации (**структура**)
- играют сходную роль в белке (**функция**)

# Пример выравнивания

```

                *           240           *           260           *           280
1zjh:A : FG-VEQDVDMVFASFIRKASDVHEVRKVLGE-KGKNIKIISKIENHEGVRRFDEIL : 260
1a3x:B : FG-VKNGVHMVFASFIRTANDVLTIREVLGE-QGKDVKIIIVKIENQQGVNMFDEIL : 254
3eoe:B : NFgIPMGCNFIAASFVQSADDVRYIRGLLGP-RGRHIRIIPKIENVEGLVNFDEIL : 249
1aqf:B : FG-VEQDVDMVFASFIRKAADVHEVRKILGE-KGKNIKIISKIENHEGVRRFDEIL : 272
1pkl:G : FG-VEQGVDMIFASFIRSAEQVGDVRKALGP-KGRDIMIICKIENHQGVQNIDSI : 252
1e0t:A : FG-CEQGVDFVAASFIRKRSDVIEIREHLKAhGGENIHIIISKIENQEGLNMFDEIL : 230
fg      v      ASfir a dV      R      Lg      G      i      II      KIEN      G      fDeil
    
```

```

                *           300           *           320           *
1zjh:A : EASDGIMVARGDLGIEIpa-EKVFLAQKMMIGRCNRAGKPVICATQmlesmikkpR : 315
1a3x:B : KVTDGVMVARGDLGIEIpa-PEVLAVQKKLIAKSNLAGKPVICATQmlesmtynpR : 309
3eoe:B : AEADGIMIARGDLGMEIpp-EKVFLAQKMMIAKCNVVGKPVITATQmlesmiknpR : 304
1aqf:B : EASDGIMVARGDLGIEIpa-EKVFLAQKMMIGRCNRAGKPVICATQmlesmikkpR : 327
1pkl:G : EESDGIMVARGDLGVEI-PaEKVVVAQKILISKCNVAGKPVICATQmlesmtynpR : 307
1e0t:A : EASDGIMVARGDLGVEIpp-EEVIFAQKMMIEKCIRARKVVI TATM-----R : 276
DGIMvARGDLG EIp      e V      aQK      I      cn      agKpVI      ATqmlesm      pR
    
```

```

                340           *           360           *           380           *
1zjh:A : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMQH LIAREAEAaiyhlqlf : 371
1a3x:B : PTRAEVSDVGNAILD GADCVMLSGETakgnyPINAVTTMAETAVIAEQaiaylpny : 365
3eoe:B : PTRAEAADVANAVLDGTD CVMLSGETangefPVITVETMARICYEAE Tcvdypaly : 360
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMQH LIAREAEAamfhrklf : 383
1pkl:G : PTRAEVSDVANAVFNGADC VMLSGETakgkyPNEVVQYMARICLEAQSalneyvff : 363
1e0t:A : PTDAEAGDVANAILDGTDAVMLSGE-----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-PEVLAVQKLLIAKSNLAGKPVICATQmlesmtynpR  
3eoe:B : AEADGIMVARGDLGMEIpP-EKVFLAQKMMIAKCNVVGKPVITATQmlesmiknpR  
1aqf:B : EASDGI MVARGDLGIEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmikkpR  
1pk1:G : EESDGI MVARGDLGVEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmtynpR  
1e0t:A : EASDGI MVARGDLGVEIpV-EEVIFAQKMMIEKCI RARKVVITATM-----R  
DGIMvARGDLG EI p e V aQK I cn agKpVI ATqmlesm pR  
340 \* 360 \* 380 \*

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



# Выравнивание и эволюция

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

          0          *          60
POLG_CXB4J : ANRQDFTQDPSKFTPEPVKDVMIKSLPALN : 68
POLG_CXB4E : ANRQDFTQDPSKFTPEPVKDVMIKSLPALN : 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        | GSHENSNS     | SATEGSTIN   | YTTINYYKDSYAAT | :  | 39 |  |   |  |

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

Последовательности белка оболочки из двух штаммов вируса Коксаки и энтеровируса человека

# Что есть что (редактор GeneDoc)

Номер столбца выравнивания

|            |            |   |   |    |    |    |    |     |       |      |
|------------|------------|---|---|----|----|----|----|-----|-------|------|
| MTA1_YEAST | : ---KSSIS | 3 | 2 | LE | 20 | F  | 4  | --- | QSLNS | : 24 |
| MAT2_YEAST | : KPYRGHRE |   |   | LE |    | F  |    |     | LDT   | : 31 |
|            |            |   |   |    |    |    |    |     |       |      |
| MTA1_YEAST | : KEKEEVA  | 6 | 6 | K  | 40 | 63 | 64 | W   | N4R   | : 53 |
| MAT2_YEAST | : KGLENI   |   |   | M  |    |    |    |     |       | : 61 |
|            |            |   |   |    |    |    |    |     |       |      |

Название последовательности

Консервативный остаток

Функционально консервативная позиция

Номер последнего в строке остатка ИЗ ЭТОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

# Сегодня все

1zjh:A : FG-VEQDVDMVFASEIRKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL  
1a3x:B : FG-VKNGVHMVFASEIRTANDVLTIREVIIGE-QCKDVKLIIVKIENQOQGVNNFDEIL  
3eoe:B : NFgIPMGCNFIAASEVQSADDVRYIRGLIGP-RGRHIRIIPKTIENVEGLVNFDEIL  
1aqf:B : FG-VEQDVDMVFASEIRKAADVHEVRKIIGE-KCKNIKLIISKIENHEGVRRFDEIL  
1pk1:G : FG-VEQGVDMIFASEIRKASDVHEVRKVIIGE-KCKNIKLIISKIENHQGVQNLDSIL  
1e0t:A : FG-CEQGVDFVAASEIRKRSDVLEIREHKAHGSENIHLSKIENQOGLNNFDEIL  
fg v ASfir a dV R Lg G i II KIEN G fDeIl

\* 300 \* 320 \*

1zjh:A : EASDGI MVARGDLGIEIpA-EKVFLAQKMMIGRCNRAGKPVICATQmlesmikkpR  
1a3x:B : KVTGGMVARGDLGIEIpA-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 : PTRAEVSDVGNAILDGCDCVMSGETakgnypINAVTTMAETA VIAEQaiaylpny  
3eoe:B : PTRAEAADVANAVLDGCTDCVMSGETangefDVITVETMARICYEAETcvdypaly  
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAamfhrklf  
1pk1:G : PTRAEVSDVANAVFNGADCDCVMSGETakgkypNEVVQYMARICLEAQSaIneyvff  
1e0t:A : PTRAEAGDVANAILDGTDAVMSGET-----PLEAVSIMATICERTDR-----  
PTRAE DVaNA ldG Dc MLSGETa q P V i a