

# Выравнивание ...



... последовательностей белков  
и его биологический смысл

1zjh:A : FG-VEQDVMVFASFIKASDVHEVRKVI GE-KCKNIKLI SKIENHEGVRRFDEIL  
1a3x:B : FG-VKNGVHMVFASFIKRTANDVLTIREVIG E-QCKDVKLI VKIENQQGVNNFDEIL  
3eoe:B : NFgIPMGCNFIAASEVQSADDVRYIRGLIGP-RGRHIRIIPKTIENVEGLVNFDEIL  
1aqf:B : FG-VEQDVMVFASFIKKAADVHEVRKIIG E-KCKNIKLI SKIENHEGVRRFDEIL  
1pk1:G : FG-VEQGVDMI FASFIKRSAEQVGDVIRKAI GP-KGRDIMLICKTIENHQGVQNLDSIL  
1e0t:A : FG-CEQGVDFVAASEFIKRSADVIEIREHKAhGSENIHLSKIENQOGLNNFDEIL  
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 GSDVANAVLDGADCIMLSG E t a k g d y P L E A V R M Q H L I A R E A E A a i y h l q l f  
1a3x:B : PTRAE VSDVGNAILD G A D C V M L S G E t a k g n y P I N A V T T M A E T A V I A E Q a i a y l p n y  
3eoe:B : PTRAE AADVANAVLDG T D C V M L S G E t a n g e f D V I T V E T M A R I C Y E A E T c v d y p a l y  
1aqf:B : PTRAE GSDVANAVLDGADCIMLSG E t a k g d y P L E A V R M Q H L I A R E A E A a m f h r k l f  
1pk1:G : PTRAE VSDVANAVFNGADCVM L S G E t a k g k y P N E V V Q Y M A R I C L E A Q S a l n e y v f f  
1e0t:A : PTD A E A G D V A N A I L D G T D A V M L S G E ----- P L E A V S I M A T I C E R T D R -----  
PT r A E D V a N A l d G D c M L S G E t a g 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

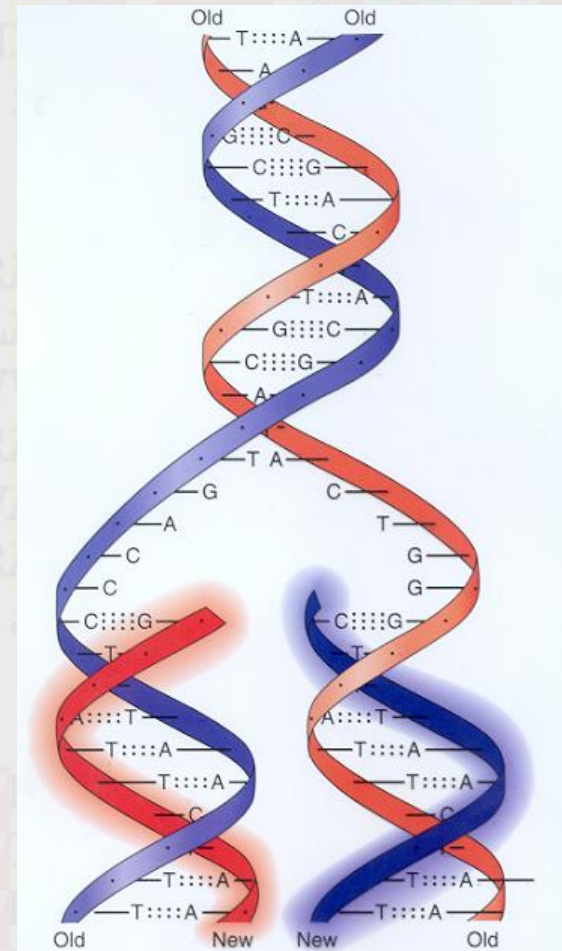
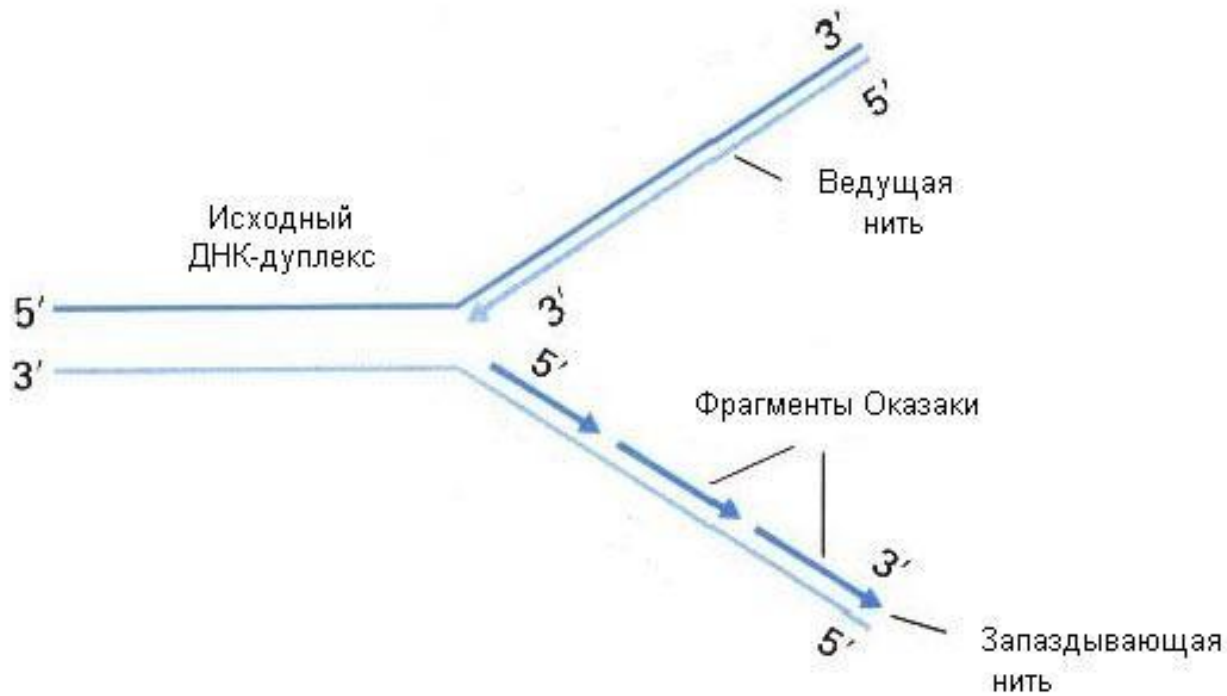
- Большая молекула
- Полимер
- Линейный
- Ориентированный
- Гетерополимер
- Двухцепочечный
- Цепочки ориентированы  
противоположно

# 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-QGKDVKIIVKIENQQGVNMFDEIL : 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 : PTRAEVSDVANAVFNGADCVMLSGETakgkyPNEVVQYMARICLEAQSalneyvff : 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-PEVLAVQKLI AKSNLAGKPVICATQmlesmtynpR  
3eoe:B : AEADGIMVARGDLGMEIpP-EKVFLAQKMMI AKCNVVGKPVITATQmlesmiknpR  
1aqf:B : EASDGI MVARGDLGIEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmikkpR  
1pk1:G : EESDGI MVARGDLGVEIpA-EKVFLAQKMMI SKCNVAGKPVICATQmlesmtynpR  
1e0t:A : EASDGI MVARGDLGVEIpV-EEVIFAQKMMI EKCI RARKVVITATM-----R  
DGIMvARGDLG EI p 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 :	GAQVSTQKTGAHETSL	SAS	SGNSIIHYTNINYYKDAASNS	:	39			
POLG_CXB4E :	GAQVSTQKTGAHETSL	SAT	SGNSIIHYTNINYYKDAASNS	:	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	GSHENSNSATEG	STINYYT	INYYKDSYAAT	:	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

			*		20		*																								
MTA1_YEAST	:	----	KSSIS	SP	QARAF	LE	QV	RRK---	QS	LNS	:	24																			
MAT2_YEAST	:	KPYR	GH	RF	TK	EN	VR	IL	LE	SW	EAK	NI	EN	PY	LDT	:	31														
					3	2			LE	F	4				L13																
					40		*		60																						
MTA1_YEAST	:	KE	KE	EV	AK	KCG	IT	PL	QVR	VWF	IN	KR	MR	SK	-	:	53														
MAT2_YEAST	:	K	G	LE	N	M	K	N	T	S	L	S	R	I	Q	I	K	N	W	V	S	N	R	R	R	K	E	K	T	:	61
		K	E	6	K				63	6Q	64	W		N	4	R	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-RGRHIRLIPKIENVEGLVNFDEIL  
1aqf:B : FG-VEQDVMVFASFIKASDVHEVRKVIIGE-KCKNIKLIISKIENHEGVRRFDEIL  
1pk1:G : FG-VEQGVDMIFAKASDVHEVRKVIIGE-RDIMLICKIENHQGVQNLDSIL  
1e0t:A : FG-CEQGVDFVAASEFIKASDVHEVRKVIIGE-KCKNIKLIISKIENQOGLNDFDEIL  
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 : PTRAEVSDVGNAILDGCDCVMSGETakgnypINAVTTMAETAVIAEQaiaylpny  
3eoe:B : PTRAEAADVANAVLDGCTDCVMSGETangefDVITVETMARICYEAETcvdypaly  
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMOHLIAREAEAamfhrklf  
1pk1:G : PTRAEVSDVANAVFNGADCDCVMSGETakgkypNEVVQYMARICLEAQSaIneyvff  
1e0t:A : PTRAEAGDVANAILDGTDAVMSGET-----PLEAVSIMATICERTDR-----  
PTRAE DVaNA ldG Dc MLSGETa q P V i a