

# 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



# Геномная ДНК

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

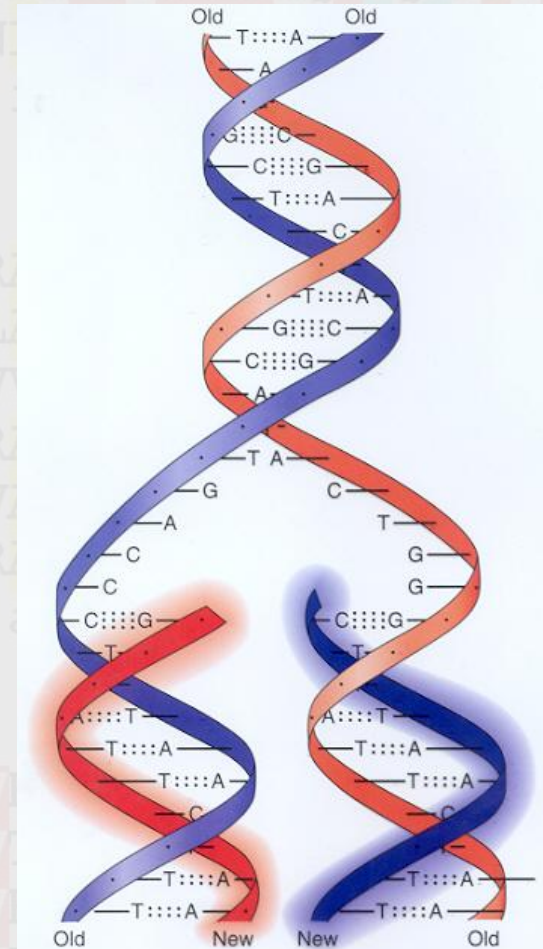
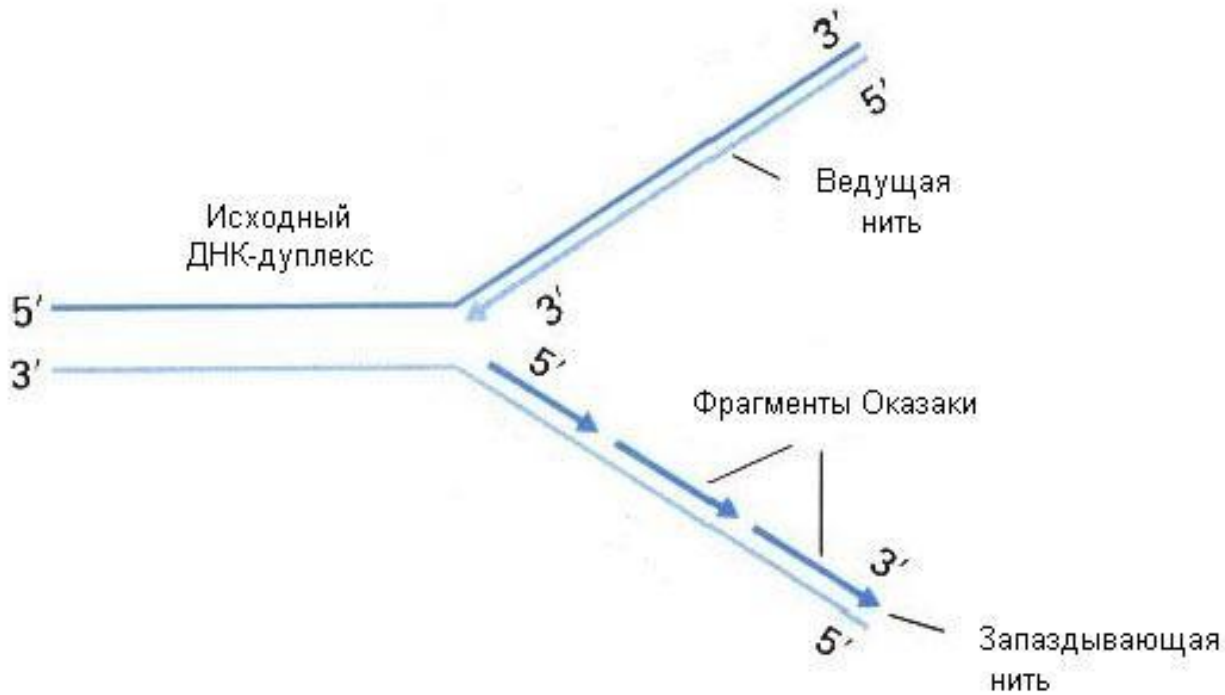


# Каждая комплементарная пара нуклеотидов ДНК потомка происходит из комплементарной пары ДНК предка

- 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-VEQD	VDMVFASFI	RKASDVHEV	RKVLGE-KG	KNIKIIS	SKIENHEG	VRRFDEIL	:	260		
1a3x:B	:	FG-VKNG	VHMVFASFI	RTANDVLT	IREVLGE-Q	GKDVKIIV	KIENQQG	VNNFDEIL	:	254		
3eoe:B	:	NFgIPMG	CNFIAASF	VQSADDV	RYIRGLL	GPR-RGR	HIRIIPK	IENVEGL	VNFDEIL	:	249	
1aqf:B	:	FG-VEQD	VDMVFASFI	RKAADVHEV	RKILGE-KG	KNIKIIS	SKIENHEG	VRRFDEIL	:	272		
1pkl:G	:	FG-VEQG	VDMIFASFI	RSAEQVGD	VRKALGP-K	GRDIMII	CKIENHQ	GVQNIDSI	:	252		
1e0t:A	:	FG-CEQG	VDFVAASF	IRKRSDV	IEIREHL	KAhGGEN	IHIISKI	ENQEG	LNNFDEIL	:	230	
		fg	v	ASfir	a dV	R	Lg	G	i II	KIEN	G	fDeil

	*	300	*	320	*							
1zjh:A	:	EASDGI	MVARGDLG	IEIpa-	EKVFLA	QKMMI	GRCNRAG	KPVICAT	Qmlesmik	kpR	:	315
1a3x:B	:	KVTDG	VMVARGDLG	IEIpa-	PEVLAV	QKCLI	AKSNLAG	KPVICAT	Qmlesmtyn	pr	:	309
3eoe:B	:	AEADGI	MVARGDLG	MEIpp-	EKVFLA	QKMMI	AKCNVVG	KPVITAT	Qmlesmik	knpr	:	304
1aqf:B	:	EASDGI	MVARGDLG	IEIpa-	EKVFLA	QKMMI	GRCNRAG	KPVICAT	Qmlesmik	kpR	:	327
1pkl:G	:	EESDGI	MVARGDLG	VEI-Pa	EKVVVA	QKILI	SKCNVAG	KPVICAT	Qmlesmtyn	pr	:	307
1e0t:A	:	EASDGI	MVARGDLG	VEIpp-	EEVIFA	QKMMI	EKCIRAR	KVVITAT	M-----	R	:	276
		DGiMv	ARGDLG	EIp	e V	aQK	I	cn	agKpVI	ATqmlesm	pr	

	340	*	360	*	380	*					
1zjh:A	:	PTRAEG	SDVANAVLD	GADCIM	LSGETak	gdyPLEA	VRMQHLI	AREAEA	Aaiyhlqlf	:	371
1a3x:B	:	PTRAEV	SDVGNAILD	GADCVM	LSGETak	gnyPIN	AVTTMA	ETAVIAE	Qaiaylpny	:	365
3eoe:B	:	PTRAE	AADVANAVLD	GTD	CVMLSGE	tangef	PVITVET	MARICYE	AEtcvdypaly	:	360
1aqf:B	:	PTRAEG	SDVANAVLD	GADCIM	LSGETak	gdyPLEA	VRMQHLI	AREAEA	Aamfhrklf	:	383
1pkl:G	:	PTRAEV	SDVANAVFNG	ADCVM	LSGETak	gkyPNE	VVQYMAR	ICLEAQ	Salnevvff	:	363
1e0t:A	:	PTDAE	AGDVANAILD	GTD	AVMLSGE	-----	PLEAV	SIMATIC	ERTDR-----	:	318
		PtrAE	DVaNA	ldG	Dc	MLSGETa	g	P	V	i	a





# Выравнивание и эволюция

		*		20	*		4		
POLG_CXB4J	:	GAQVSTQKTGAHETSLSASGNSIIHYTNINYYKDAASNS						:	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	GSHENSNSATEG	STINYYTTINYYKDSYAAT	:	39				

	0		*		60	
POLG_CXB4J :	ANRQDFTQDPSK	FTEPVKDVMIKSLPALN	:	68		
POLG_CXB4E :	ANRQDFTQDPSK	FTEPVKDVMIKSLPALN	:	68		
POLG_HE71B :	AGKQSLKQDPD	KEANPVKDI	FTEMAAPLK	:	68	

Последовательности белка оболочки из двух штаммов вируса Коксаки и энтеровируса человека



# Что есть что (редактор GeneDoc)

Номер столбца выравнивания

```

                *                20                *
MTA1_YEAST : ----KSSIS P Q A R A F L E Q V E R R K ---- Q S L N S : 24
MAT2_YEAST : K P Y R G H R F T K E N V R I L E S W E A K N I E N P Y L D T : 31
                3 2                L E F 4                L 1 3
    
```

```

                40                *                60
MTA1_YEAST : K E K E E V A K K C G I T P L Q V R V W F I N K R M R S K - : 53
MAT2_YEAST : K G L E N I 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 6Q64 W N 4R 4 K
    
```

Название последовательности

Консервативный остаток

Функционально консервативная позиция

Номер последнего в строке остатка ИЗ ЭТОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

# Сегодня все

1zjh:A : FG-VEQDVMVFASFIRKASDVHEVRKVLGE-KGKNIKLIISKIENHEGVRRFDEIL  
1a3x:B : FG-VKNGVHMVFASFIRTANDVLTIREVLGE-QGKDVKLIIVKIENQQGVNNFDEIL  
3eoe:B : NFgIPMGCNFIAASFVQSADDVRYIRGLLGP-RGRHIRLIPKIENVEGLVNFDEIL  
1aqf:B : FG-VEQDVMVFASFIRKAADVHEVRKILGE-KGKNIKLIISKIENHEGVRRFDEIL  
1pkl:G : FG-VEQGVDMI FASFIRKASDVHEVRKILGE-KGKNIKLIISKIENHQGVQNIIDSTI  
1e0t:A : FG-CEQGVDFVAASFIRKRSDVIEIREHUKAhGGENIHIIISKIENQEGLNNFDEIL  
fg v ASfir a dV R Lg G i II KIEN G fDeIl

\* 300 \* 320 \*

1zjh:A : EASDGI MVARGDLGIEIpA-EKVFLAQKMMI GRCNRAGKPVICATQmlesmikkpR  
1a3x:B : KVTDGVMVARGDLGIEIpA-PEVLAVQKKLI AKSNLAGKPVICATQmlesmtynpR  
3eoe:B : AEADGIMVARGDLGMEIpP-EKVFLAQKMMI AKCNVVGKPVITATQmlesmiknpR  
1aqf:B : EASDGI MVARGDLGIEIpA-EKVFLAQKMI I GRCNRAGKPVICATQmlesmikkpR  
1pkl:G : EESDGI MVARGDLGVEI-PaEKVVVAQKIL I SKCNVAGKPVICATQmlesmtynpR  
1e0t:A : EASDGI MVARGDLGVEIpV-EEVIFAQKMMI EK CIRARKVVITATM-----R  
DGiMvARGDLG EI p e V aQK I cn agKpVI ATqmlesm pR

340 \* 360 \* 380 \*

1zjh:A : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMQH LIAREAEAAaiyhqlf  
1a3x:B : PTRAEVSDVGNAILD GADCVMLSGETakgnypINAVTTMAETAVIAEQaiaylpny  
3eoe:B : PTRAEAADVANAVLDGTDCVMSLSEtange fPVITVETMARICYEAETcvdypaly  
1aqf:B : PTRAE GSDVANAVLDGADCIMLSGETakgdyPLEAVRMQH LIAREAEAAamfhrklf  
1pkl:G : PTRAEVSDVANAVFNGADCVMLSGETakgkypNEVVQYMARICLEAQSalneyvff  
1e0t:A : PTDAEAGDVANA I LDGTD AVMSLSE-----PLEAVSIMATICERTDR-----  
PTrAE DVaNA ldG Dc MLSGETa g P V i a