

Сравнение строк

(продолжение)

Николай Вяххи

vyahhi@bioinformaticsinstitute.ru

Computer Science Club
Санкт-Петербург, 2013



**Институт
Биоинформатики**



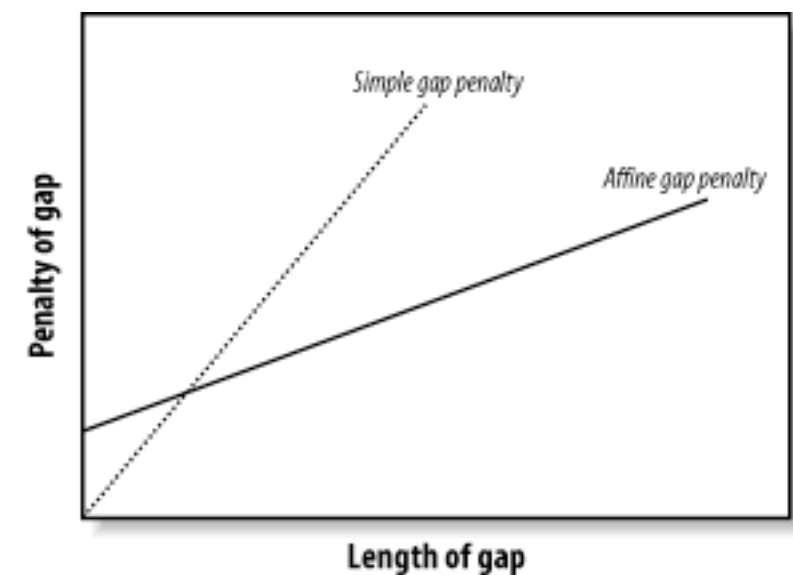
Аффинная функция штрафов

Удаление X букв **подряд** случается чаще, чем удаление X букв **по отдельности**.

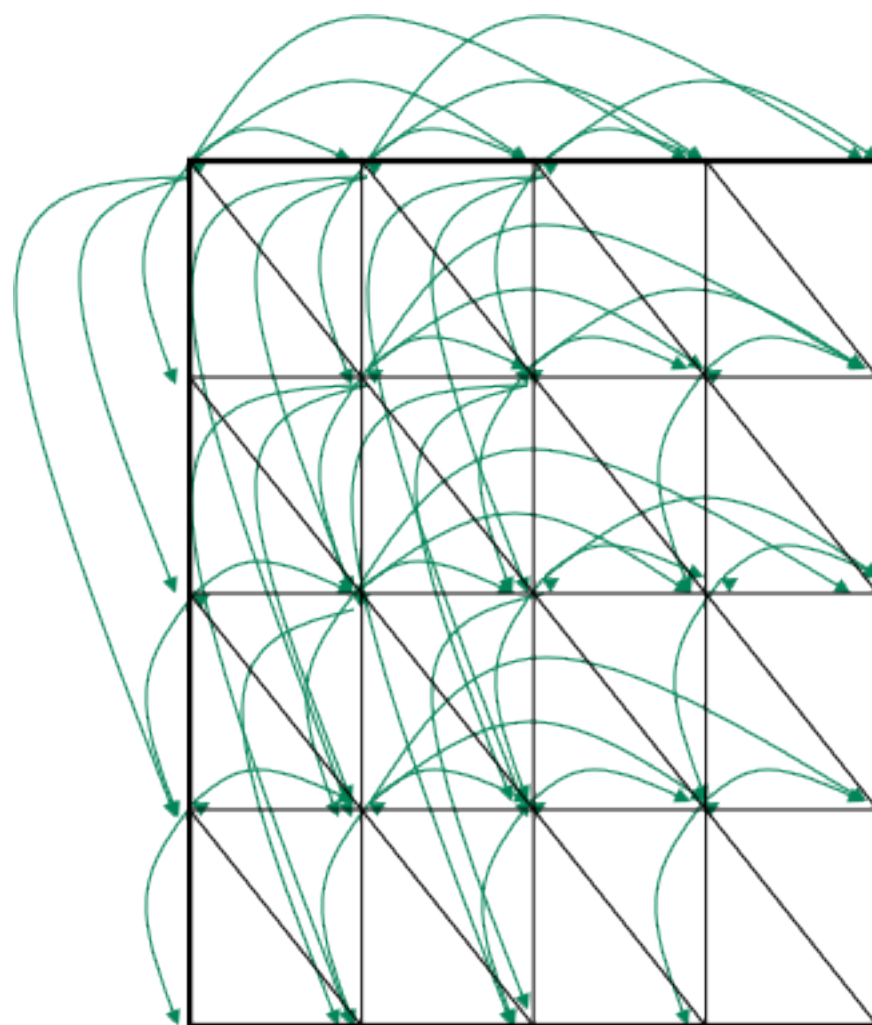
$$\text{Gap}(L) = A + B \cdot (L-1)$$

A — gap opening penalty

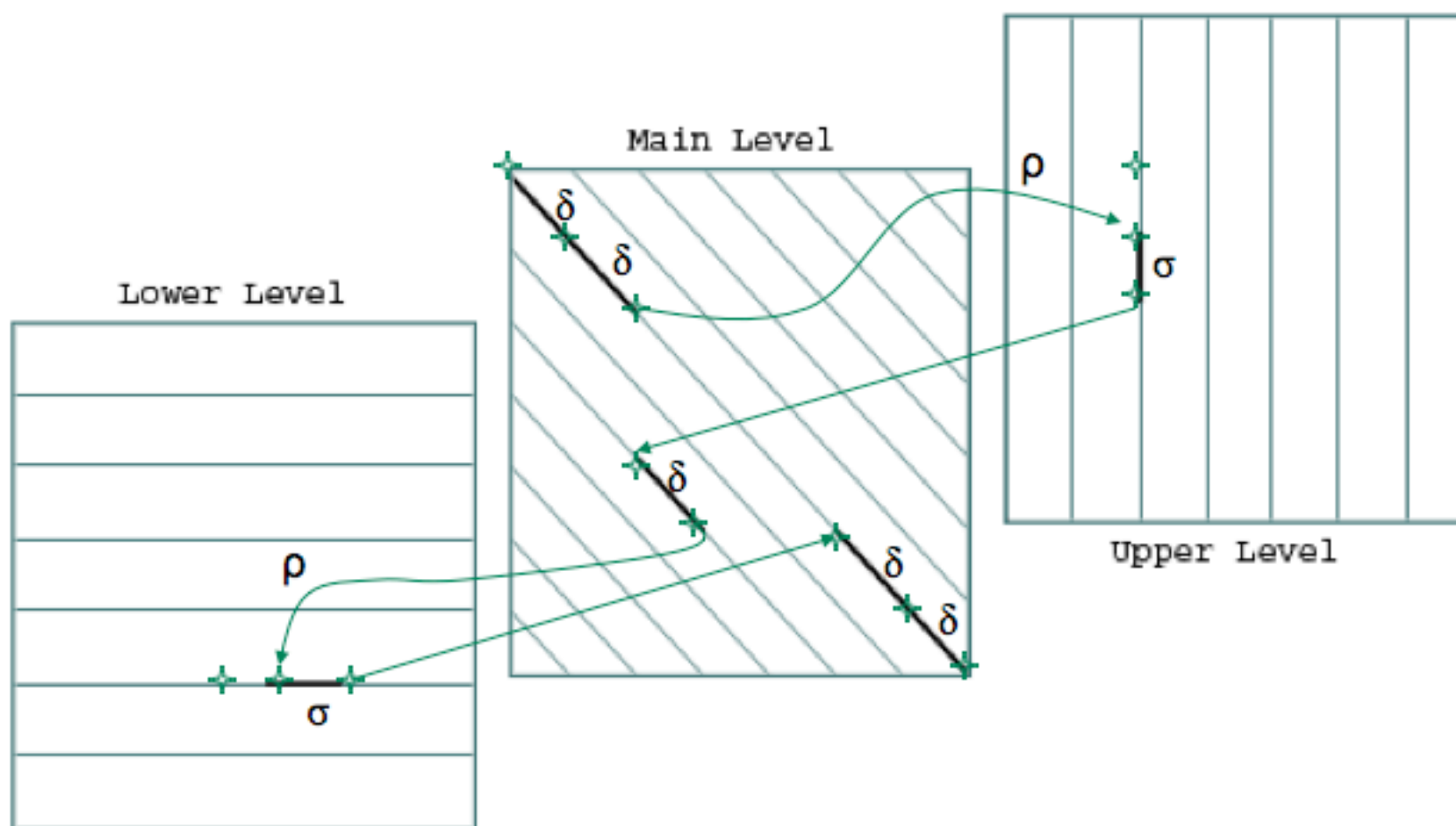
B — gap extension penalty



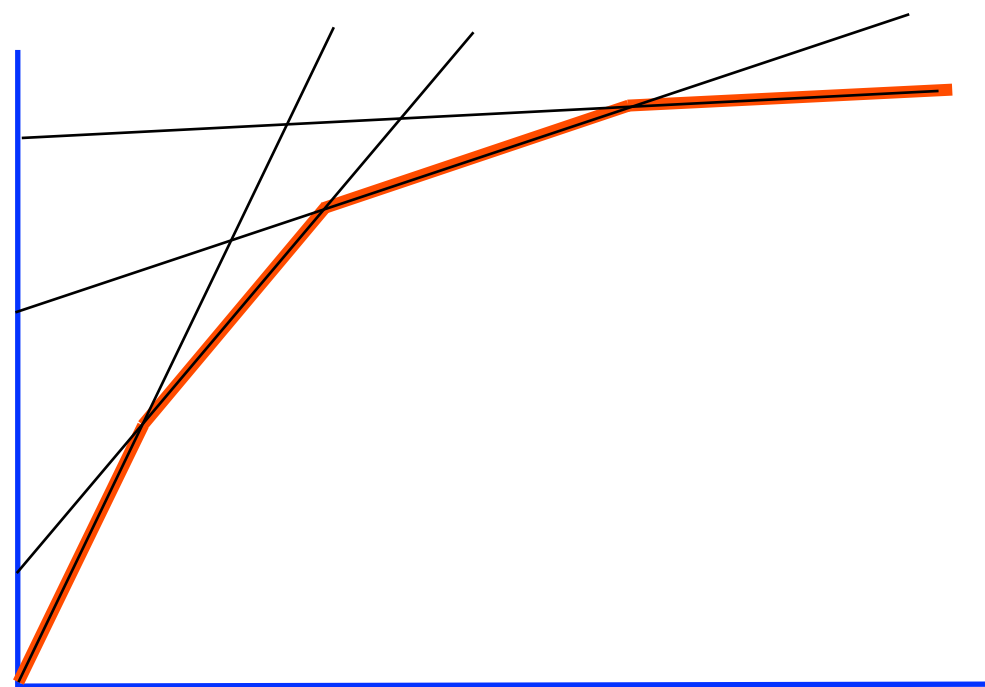
Аффинная функция штрафов



Аффинная функция штрафов



Сложная функция штрафов



Алгоритм Хиршберга

		A	C	G	T	A	C	G	T	A	C	G	T
A													
G													
T													
A													
C													
C													
T													
A													
C													
C													
G													
T													

Алгоритм Хиршберга

		A	C	G	T	A		C		G	T	A	C	G	T
	0														
A	1														
G	2														
T	3														
A	4														
C	5														
C	6														
T	7														
A	8														
C	9														
C	10														
G	11														
T	12														

Алгоритм Хиршберга

		A	C	G	T	A		C
	0	1						
A	1	0						
G	2	1						
T	3	2						
A	4	3						
C	5	4						
C	6	5						
T	7	6						
A	8	7						
C	9	8						
C	10							
G	11							
T	12							

Алгоритм Хиршберга

		A	C	G	T	A		C		G	T	A	C	G	T
	0	1	2	3	4	5	6	6	7	8	9	10	11	12	
A	1	0	1	2	3	4	5	5	6	7	8	9	10	11	
G	2	1	1	1	2	3	4	5	5	6	7	8	9	10	
T	3	2	2	2	1	2	3	4	4	5	6	7	8	9	
A	4	3	3	3	2	1	2	3	3	4	5	6	7	8	
C	5	4	3	4	3	2	1	2	2	3	4	5	6	7	
C	6	5	4	4	4	3	2	2	1	2	3	4	5	6	
T	7	6	5	5	4	4	3	3	2	1	2	3	4	5	
A	8	7	6	6	5	4	4	3	2	1	1	2	3	4	
C	9	8	7	7	6	5	4	3	2	1	0	1	2	3	
C	10	9	8	8	7	6	5	4	3	2	1	0	1	2	
G	11	10	9	8	8	7	6	5	4	3	2	1	0	1	
T	12	11	10	9	8	8	7	6	5	4	3	2	1	0	

Алгоритм Хиршберга

		A	C	G	T	A		C		G	T	A	C	G	T
	0	1	2	3	4	5	6	12	6	7	8	9	10	11	12
A	1	0	1	2	3	4	5	10	5	6	7	8	9	10	11
G	2	1	1	1	2	3	4	9	5	5	6	7	8	9	10
T	3	2	2	2	1	2	3	7	4	4	5	6	7	8	9
A	4	3	3	3	2	1	2	5	3	3	4	5	6	7	8
C	5	4	3	4	3	2	1	3	2	2	3	4	5	6	7
C	6	5	4	4	4	3	2	4	2	1	2	3	4	5	6
T	7	6	5	5	4	4	3	6	3	2	1	2	3	4	5
A	8	7	6	6	5	4	4	7	3	2	1	1	2	3	4
C	9	8	7	7	6	5	4	7	3	2	1	0	1	2	3
C	10	9	8	8	7	6	5	9	4	3	2	1	0	1	2
G	11	10	9	8	8	7	6	11	5	4	3	2	1	0	1
T	12	11	10	9	9	8	7	13	6	5	4	3	2	1	0

Алгоритм Хиршберга

		A	C		G		T	A	C	G
		0	1	2	3	5	0	3	4	5
A	1	0	1	2	3	3	1	2	3	4
G	2	1	1	1	1	1	0	1	2	3
T	3	2	0	2	3	3	0	1	2	
A	4	3	3	3	5	5	2	1	0	1
C	5	4	3	4	7	7	3	2	1	0
C										

Алгоритм Хиршберга

		A	C	G	T	A	C	G	T	A	C		G		T	
A		○	○													
G				○												
T					○											
A						○										
C							○									
C								○								
T									○							
A										○						
C												0	1	3	2	3
C												1	1	2	1	2
G												2	1	1	0	1
T												1	2	3	1	0

Алгоритм Хиршберга

		A	C	G	T	A	C	G	T	A	C	G	T
A		○	○										
G				○									
T					○								
A						○							
C							○						
C								○					
T									○				
A										○			
C											○		
C												○	
G												○	
T													○

Метод 4-х русских

V. Arlazarov, E. Dinic, M. Kronrod, I. Faradzev

Разделить матрицу на блоки размером $t \times t$.

Использовать таблицу поиска для быстрого подсчёта значений в блоке.

Метод 4-х русских

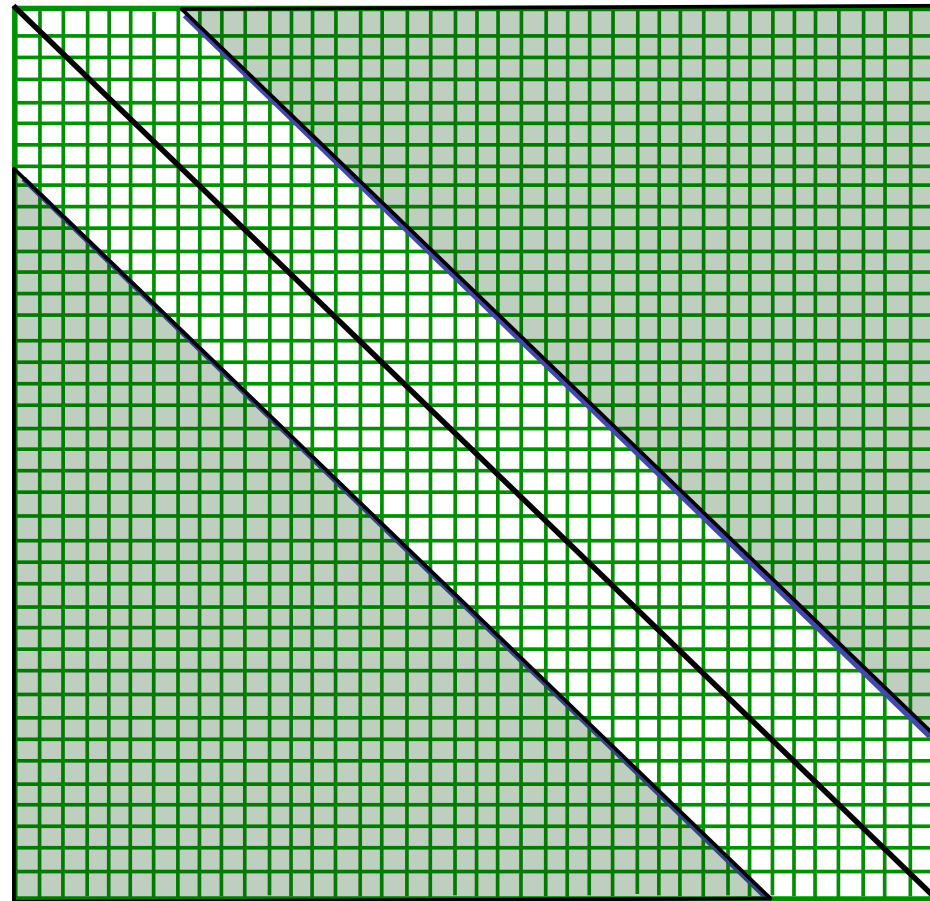
V. Arlazarov, E. Dinic, M. Kronrod, I. Faradzev

Разделить матрицу на блоки размером $t \times t$.

Использовать таблицу поиска для быстрого подсчёта значений в блоке.

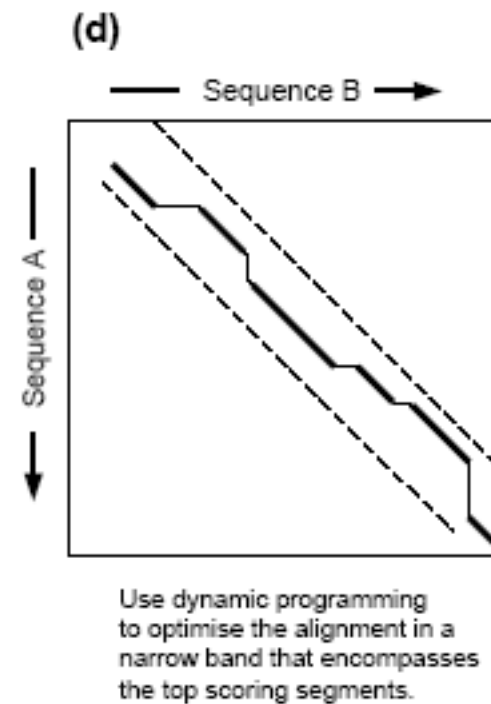
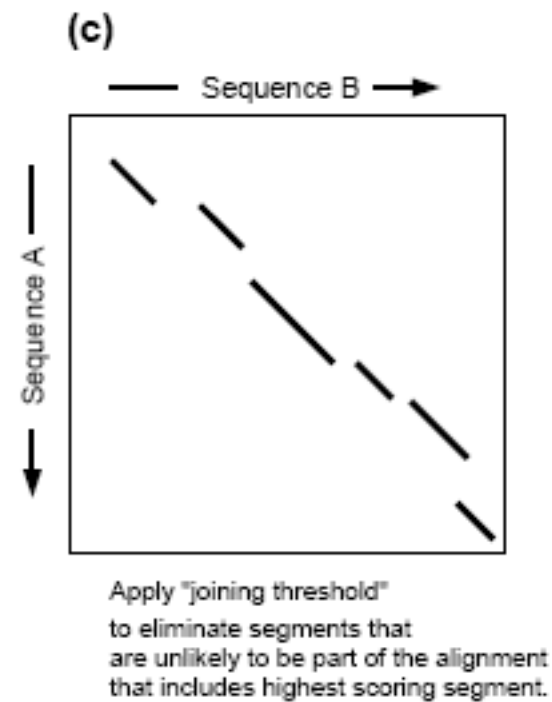
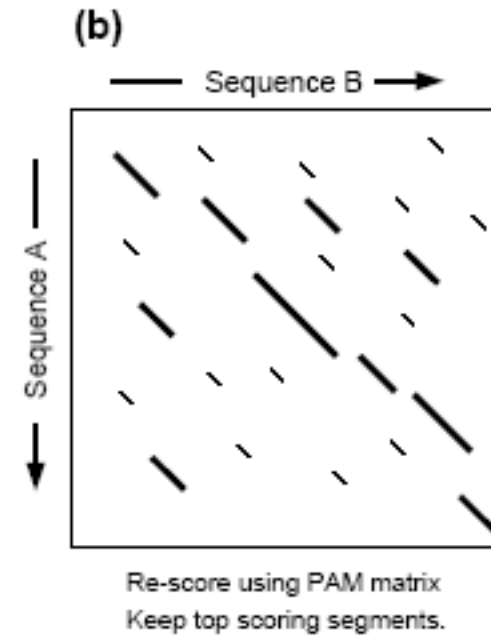
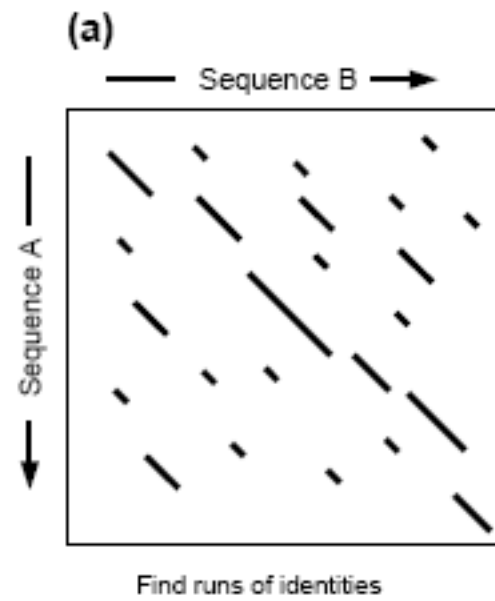
$$t \approx \log(N)$$

Ограниченное выравнивание



FASTA

FASTA Algorithm



D. Lipman
W. Pearson

FASTA формат

```
>gi|48994873|gb|U00096.2| Escherichia coli str. K-12 substr. MG1655  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAGAGTGTCTGATAGCAGC  
TTCTGAACTGGTTACCTGCCGTGAGTAAATTTAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAA  
TATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC  
ATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAG  
CCCGCACCTGACAGTGCGGGCTTTTTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTTGAA  
GTTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCGATATTCTGGAAAGCAATGCC  
AGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTGGCGATGATTG  
AAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTTGCCGAACTTTT  
GACGGGACTCGCCGCCGCCAGCCGGGGTTCCTCGCTGGCGCAATTGAAAACCTTCGTCGATCAGGAATTT
```

BLAST

«Basic Local Alignment Search Tool»

Altschul, Gish, Miller, Myers, Lipman

(самая цитируемая статья 1990х)

<http://blast.ncbi.nlm.nih.gov>

BLAST

Query sequence: PQGEFG

Word 1: PQG

Word 2: QGE

Word 3: GEF

Word 4: EFG

Query sequence: R P P Q G L F

Database sequence: D P P E G V V

↳ Exact match is scanned.

Score: -2 7 7 2 6 1 -1

↳ HSP

Optimal accumulated score = $7+7+2+6+1 = 23$

Query sequence

Newly joined region,
then it is extended to be
an HSP region.

Distance < A

Database sequence

BLAST

Sequences producing significant alignments:	Score	E
	(bits)	Value
gi 18858329 ref NP_571095.1 ba1 globin [Danio rerio] >gi 147757...	171	3e-44
gi 18858331 ref NP_571096.1 ba2 globin; SI:dZ118J2.3 [Danio rer...	170	7e-44
gi 37606100 emb CAE48992.1 SI:bY187G17.6 (novel beta globin) [D...	170	7e-44
gi 31419195 gb AAH53176.1 Ba1 protein [Danio rerio]	168	3e-43

ALIGNMENTS

>gi|18858329|ref|NP_571095.1| ba1 globin [Danio rerio]

Length = 148

Score = 171 bits (434), Expect = 3e-44

Identities = 76/148 (51%), Positives = 106/148 (71%), Gaps = 1/148 (0%)

Query: 1 MVHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK 60

MV T E++A+ LWGK+N+DE+G +AL R L+VYPWTQR+F +FG+LS+P A+MGNPK

Sbjct: 1 MVEWTDAERTAILGLWGKLNIDEIGPQALSRCCLIVYPWTQRYFATFGNLSSPAAIMGPNK 60

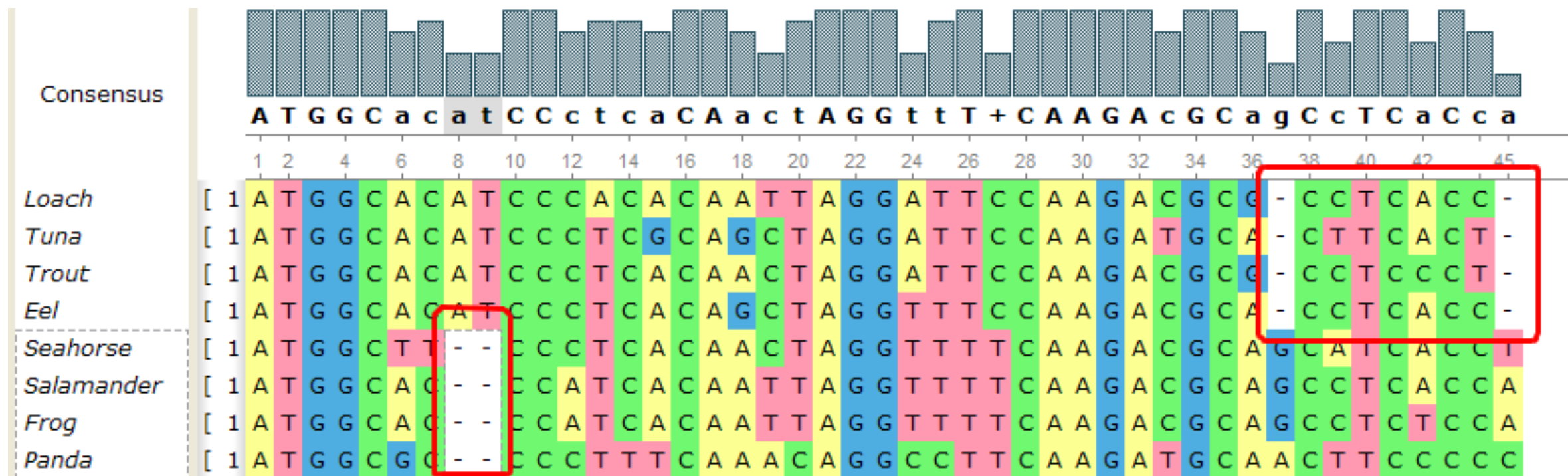
Query: 61 VKAHGKKVLGAFSDGLAHLNLDKGTFFATLSELHCDKLHVDPENFRLLGNVLCVLAHHFG 120

V AHG+ V+G + ++DN+K T+A L.S +H +K.LHVDP+NFRT.L + + A FG

Множественное выравнивание

Расширение попарного выравнивания
на случай большого количества
сравниваемых строк.

Множественное выравнивание



Множественное выравнивание

			*	.	:	.	*	:	:	:	.																																																																																	
Q5E940_BOVIN	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	Q	I	R	M	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76												
RLA0_HUMAN	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	Q	I	R	M	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76												
RLA0_MOUSE	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	Q	I	R	M	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76												
RLA0_RAT	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	Q	I	R	M	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76												
RLA0_CHICK	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	M	K	I	I	Q	L	L	D	D	Y	P	K	C	F	V	G	A	D	N	V	G	S	K	O	M	Q	Q	I	R	M	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76													
RLA0_RANSY	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	Q	I	R	M	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	S	A	L	E	76												
Q7ZUG3_BRARE	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	T	I	R	L	S	L	R	G	K	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76												
RLA0 ICTPU	-----	M	P	R	E	D	R	A	T	W	K	S	N	Y	F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	T	I	R	L	S	L	R	G	K	-	A	I	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	A	L	E	76												
RLA0_DROME	-----	M	V	R	E	N	K	A	A	W	K	A	Q	Y	F	I	K	V	V	E	L	F	D	E	F	P	K	C	F	I	V	G	A	D	N	V	G	S	K	O	M	Q	N	I	R	T	S	L	R	G	L	-	A	V	V	L	M	G	K	N	T	M	M	R	K	A	I	R	G	H	L	E	N	N	--	P	Q	L	E	76												
RLA0_DICDI	-----	M	S	G	A	G	-	S	K	R	K	L	F	I	E	K	A	T	K	L	F	T	T	Y	D	K	M	I	V	A	E	A	D	F	V	G	S	S	Q	L	Q	K	I	R	K	S	I	R	G	I	-	G	A	V	L	M	G	K	K	T	M	I	R	K	V	I	R	D	L	A	D	S	K	--	P	E	L	D	75													
Q54LP0_DICDI	-----	M	S	G	A	G	-	S	K	R	K	N	V	F	I	E	K	A	T	K	L	F	T	T	Y	D	K	M	I	V	A	E	A	D	F	V	G	S	S	Q	L	Q	K	I	R	K	S	I	R	G	I	-	G	A	V	L	M	G	K	K	T	M	I	R	K	V	I	R	D	L	A	D	S	K	--	P	E	L	D	75												
RLA0_PLAF8	-----	M	A	K	L	S	K	Q	Q	K	Q	M	Y	I	E	K	L	S	S	L	I	Q	Q	Y	S	K	I	L	I	V	H	D	N	V	G	S	N	O	M	A	S	V	R	K	S	L	R	G	K	-	A	T	I	L	M	G	K	N	T	R	I	R	T	A	L	K	N	L	O	A	V	--	P	Q	I	E	76															
RLA0_SULAC	----	M	I	G	L	A	V	T	T	T	K	K	I	A	K	W	K	V	D	E	V	A	E	L	T	E	K	L	K	T	H	K	T	I	I	I	A	N	I	E	G	F	P	A	D	K	L	H	E	I	R	K	K	L	R	G	K	-	A	D	I	K	V	T	K	N	N	L	F	N	I	A	L	K	N	A	G	----	Y	D	T	K	79									
RLA0_SULTO	----	M	R	I	M	A	V	I	T	Q	E	R	K	I	A	K	W	K	I	E	E	V	K	E	L	E	Q	K	L	R	E	Y	H	T	I	I	I	A	N	I	E	G	F	P	A	D	K	L	H	D	I	R	K	K	M	R	G	M	-	A	E	I	K	V	T	K	N	T	L	F	G	I	A	A	K	N	A	G	----	L	D	V	S	80								
RLA0_SULSO	----	M	K	R	L	A	L	A	L	K	Q	R	K	V	A	S	W	K	L	E	E	V	K	E	L	T	E	L	I	K	N	S	N	T	I	L	I	G	N	L	E	G	F	P	A	D	K	L	H	E	I	R	K	K	L	R	G	K	-	A	T	I	K	V	T	K	N	T	L	F	K	I	A	A	K	N	A	G	----	I	D	I	E	80								
RLA0_AERPE	MSVVS	S	L	V	G	Q	M	Y	K	R	E	K	P	I	P	E	W	K	T	L	M	L	R	E	L	E	L	F	S	K	H	R	V	V	L	F	A	D	L	T	G	T	P	T	F	V	V	Q	R	V	R	K	K	L	W	K	K	-	Y	P	M	M	V	A	K	K	R	I	L	R	A	M	K	A	A	G	L	E	----	L	D	D	N	86								
RLA0_PYRAE	-MMLA	I	G	K	R	R	Y	V	R	T	R	Q	Y	P	A	R	K	V	K	I	V	S	E	A	T	E	L	L	Q	K	Y	P	Y	V	F	L	F	D	L	H	G	L	S	S	R	I	L	H	E	Y	R	Y	R	L	R	R	Y	-	G	V	I	K	I	K	P	T	L	F	K	I	A	F	T	K	V	Y	G	G	----	I	P	A	E	85								
RLA0_METAC	-----	M	A	E	R	H	H	T	E	H	I	P	Q	W	K	K	D	E	I	E	N	I	K	E	L	I	Q	S	H	K	V	F	G	M	V	G	I	E	G	I	L	A	T	K	M	Q	K	I	R	R	D	L	K	D	V	-	A	V	L	K	V	S	R	N	T	L	I	E	R	A	L	N	Q	L	G	----	E	T	I	P	78											
RLA0_METMA	-----	M	A	E	R	H	H	T	E	H	I	P	Q	W	K	K	D	E	I	E	N	I	K	E	L	I	Q	S	H	K	V	F	G	M	V	R	I	E	G	I	L	A	T	K	I	Q	K	I	R	R	D	L	K	D	V	-	A	V	L	K	V	S	R	N	T	L	I	E	R	A	L	N	Q	L	G	----	E	S	I	P	78											
RLA0_ARCFU	-----	M	A	A	V	R	G	S	----	P	P	E	Y	K	V	R	A	V	E	E	I	K	R	M	I	S	S	K	P	V	V	A	I	V	S	F	R	N	V	P	A	G	O	M	Q	K	I	R	R	E	F	R	G	K	-	A	E	I	K	V	V	K	N	T	L	L	E	R	A	L	D	A	L	G	----	G	D	Y	L	75												
RLA0_METKA	MAVKAK	G	Q	P	P	S	G	Y	E	P	K	V	A	E	W	K	R	R	E	V	K	E	L	M	D	E	Y	E	N	V	G	L	V	D	L	E	G	I	P	A	P	Q	L	O	E	I	R	A	K	L	R	E	R	D	T	I	I	R	M	S	R	N	T	L	M	R	I	A	L	E	E	K	I	D	E	R	----	P	E	L	E	88										
RLA0_METTH	-----	M	A	H	V	A	E	W	K	K	K	E	V	Q	E	L	H	D	L	I	K	G	Y	E	V	G	I	A	N	L	A	D	I	P	A	R	Q	L	Q	K	M	R	Q	T	L	R	D	S	-	A	L	I	R	M	S	K	K	T	L	I	S	L	A	L	E	K	A	G	R	E	L	----	E	N	V	D	74															
RLA0_METTL	-----	M	I	T	A	E	S	E	H	K	I	A	P	W	K	I	E	E	V	N	K	L	K	E	L	L	K	N	G	Q	I	V	A	L	V	D	M	M	E	V	P	A	R	Q	L	O	E	I	R	D	K	I	R	-	G	T	M	T	L	K	M	S	R	N	T	L	I	E	R	A	I	K	E	V	A	E	E	T	G	N	P	E	F	A	----	G	N	P	E	F	A	82
RLA0_METVA	-----	M	I	D	A	K	S	E	H	K	I	A	P	W	K	I	E	E	V	N	A	L	K	E	L	L	K	S	A	N	V	I	A	L	I	D	M	M	E	V	P	A	V	Q	L	O	E	I	R	D	K	I	R	-	D	Q	M	T	L	K	M	S	R	N	T	L	I	K	R	A	V	E	E	V	A	E	E	T	G	N	P	E	F	A	----	G	N	P	E	F	A	82
RLA0_METJA	-----	M	E	T	K	V	K	A	H	V	A	P	W	K	I	E	E	V	K	T	L	K	G	L	I	K	S	K	P	V	V	A	I	V	D	M	M	D	V	P	A	P	Q	L	O	E	I	R	D	K	I	R	-	D	K	V	K	L	R	M	S	R	N	T	L	I	I	R	A	L	K	E	A	A	E	L	N	N	P	K	L	A	----	K	L	A	81					
RLA0_PYRAB	-----	M	A	H	V	A	E	W	K	K	K	E	V	E	L	A	N	L	I	K	S	Y	P	V	I	A	L	V	D	V	S	S	M	P	A	Y	P	L	S	Q	M	R	R	L	I	R	E	N	G	G	L	L	R	V	S	R	N	T	L	I	E	L	A	I	K	K	A	A	Q	E	L	G	K	P	E	L	E	----	P	E	L	77										
RLA0_PYRHO	-----	M	A	H	V	A	E	W	K	K	K	E	V	E	L	A	K	L	I	K	S	Y	P	V	I	A	L	V	D	V	S	S	M	P	A	Y	P	L	S	Q	M	R	R	L	I	R	E	N	G	G	L	L	R	V	S	R	N	T	L	I	E	L	A	I	K	K	A	A	K	E	L	G	K	P	E	L	E	----	P	E	L	77										
RLA0_PYRFU	-----	M	A	H	V	A	E	W	K	K	K	E	V	E	L	A	N	L	I	K	S	Y	P	V	V	A	L	V	D	V	S	S	M	P	A	Y	P	L	S	Q	M	R	R	L	I	R	E	N	N	G	L	L	R	V	S	R	N	T	L	I	E	L	A																													

Множественное выравнивание

```
                220      230      240      250      260      270      280      290      300      310      320      330      340      350
oth OP3D SHEEP/1-355 L I V I F F C Y G Q L V F T U K E A A A Q Q Q E S A T T Q K A E K E V T R M V I I M V I A F L I C W L P Y A G U A F Y I F T H Q G S D F G P I F M T I P A F F A K S S S V Y N P V I Y I M M N K Q F R N C M L T T L C C G K N P L G D D E - - - - A S T - - - T V S K T E T S Q V A P A
oth OP3D BOVIN/1-355 L I V I F F C Y G Q L V F T U K E A A A Q Q Q E S A T T Q K A E K E V T R M V I I M V I A F L I C W L P Y A G U A F Y I F T H Q G S D F G P I F M T I P A F F A K T S A V Y N P V I Y I M M N K Q F R N C M V T T L C C G K N P L G D D E - - - - A S T - - - T V S K T E T S Q V A P A
oth OP3D MOUSE/1-355 M I V I F F C Y G Q L V F T U K E A A A Q Q Q E S A T T Q K A E K E V T R M V I I M V I F F L I C W L P Y A S V A F Y I F T H Q G S N F G P I F M T L P A F F A K S S S I Y N P V I Y I M L N K Q F R N C M L T T L C C G K N P L G D D D - - - - A S A - - - T A S K T E T S Q V A P A
oth OP3D HUMAN/1-355 M I I I F F C Y G Q L V F T U K E A A A Q Q Q E S A T T Q K A E K E V T R M V I I M V I A F L I C W P Y A S V A F Y I F T H Q G S N F G P I F M T I P A F F A K S A A I Y N P V I Y I M M N K Q F R N C M L T T I C C G K N P L G D D E - - - - A S A - - - T V S K T E T S Q V A P A
oth OP3D CHICK/1-355 L A V I F F C Y G N L V C T U K E A A A Q Q Q E S A T T Q K A E K E V T R M V I I M V I A F L I C W P Y A S V A F Y I F T H Q G S D F G P I F M T I P A F F A K S S A I Y N P V I Y I M M N K Q F R N C M I T T L C C G K N P L G D E D - - - - T S A G K T E T S S V S T S Q V S P A
oth OPSU BRARE/1-355 L I V I F F C Y G R L V C T U K E A A R Q Q Q E S E T T Q R A E R E V T R M V I I M V I A F L I C W L P Y A G U A W Y I F T H Q G S E F G P V F M T L P A F F A K T S A V Y N P C I Y I C M N K Q F R H C M I T T L C C G K N P F E E E G A S - T T A S K T E A S S V S S S S V S P A
oth OP3D XENLA/1-355 L I V I F F C Y G R L L C T U K E A A A Q Q Q E S L T T Q K A E K E V T R M V I M V I F F L I C W P Y A Y A F Y I F T H Q G S N F G P V F M T V P A F F A K S S A I Y N P V I Y I V L N K Q F R N C L I T T L C C G K N P F G D E D G S S - A A T S K T E A S S V S S S Q V S P A
oth OP3D LAMJA/1-355 F V I I F F C Y G R L L C T U K E A A A A Q Q Q E S A S T Q K A E K E V T R M V U L M V I G F L V C W P Y A S V A F Y I F T H Q G S D F G A T F M T L P A F F A K S S A L Y N P V I Y I L M N K Q F R N C M I T T L C C G K N P L G D D E - S G - A S T S K T E V S S V S T S P V S P A
oth OP3H CARAU/1-355 V T I I F F T Y G R L V C T U K A A A A Q Q Q D S A S T Q K A E R E V T K M V I L M V L G F L V A W T P Y A T V A A W I F F N K G A A F S A Q F M A I P A F F S K T S A L Y N P V I Y V L L N K Q F R S C M L T T L F C G K N P L G D - E E S T V S T S K T E V S S V S - - - - - P A
oth OP3G CARAU/1-355 V A V I F F T Y G R L V C T U K A A A A Q Q Q D S A S T Q K A E R E V T K M V I L M V F G F L I A W T P Y A T V A A W I F F N K G A D F S A K F M A I P A F F S K S S A L Y N P V I Y V L L N K Q F R N C M L T T I F C G K N P L G D - D E S T V S T S K T E V S S V S - - - - - P A
oth OP3G CHICK/1-355 W W V I F F S Y G R L I C K V R E A A A Q Q Q E S A T T Q K A E K E V T R M V I L M V L G F M L A W T P Y A V V A F W I F T N K G A D I T A T L M A V P A F F S K S S S L Y N P I I Y V L M N K Q F R N C M I T T I C C G K N P F G D E D V S S T V S Q S K T E V S S V S S S Q V S P A
oth_OP3B_GECGE/1-355 M W V I F F S Y G R L V C K V R E A A A Q Q Q E S A T T Q K A E K E V T R M V I L M V L G F L L A W T P Y A A T A I W I F T N R G A A F S V T F M T I P A F F S K S S S I Y N P I I Y V L L N K Q F R N C M V T T I C C G K N P F G D E D V S S S V S Q S K T E V S S V S S S Q V A P A
```

Quality/1-355



Профиль выравнивания

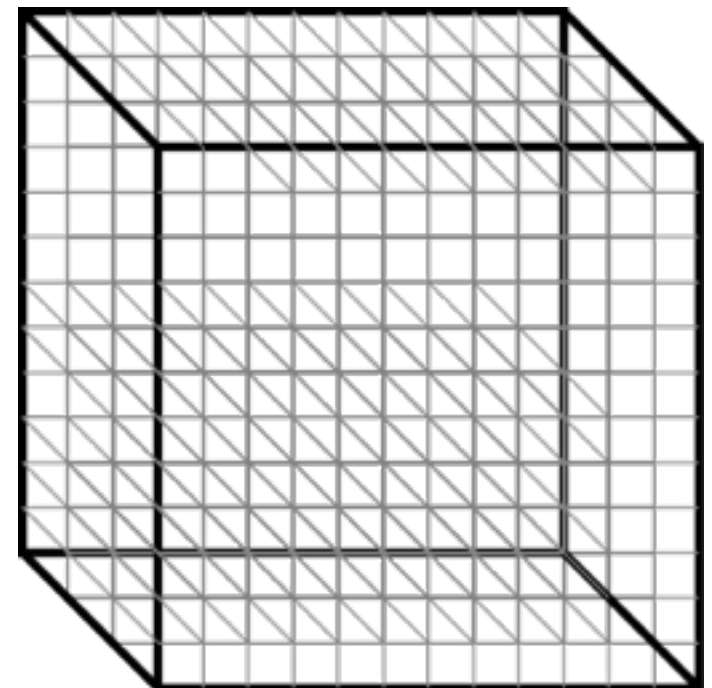
		A	T	C	C	A	G	C	T
		G	G	G	C	A	A	C	T
		A	T	G	G	A	T	C	T
Alignment		A	A	G	C	A	A	C	C
		T	T	G	G	A	A	C	T
		A	T	G	C	C	A	T	T
		A	T	G	G	C	A	C	T
	A	5	1	0	0	5	5	0	0
Profile	T	1	5	0	0	0	1	1	6
	G	1	1	6	3	0	1	0	0
	C	0	0	1	4	2	0	6	1
Consensus		A	T	G	C	A	A	C	T

Множественное выравнивание

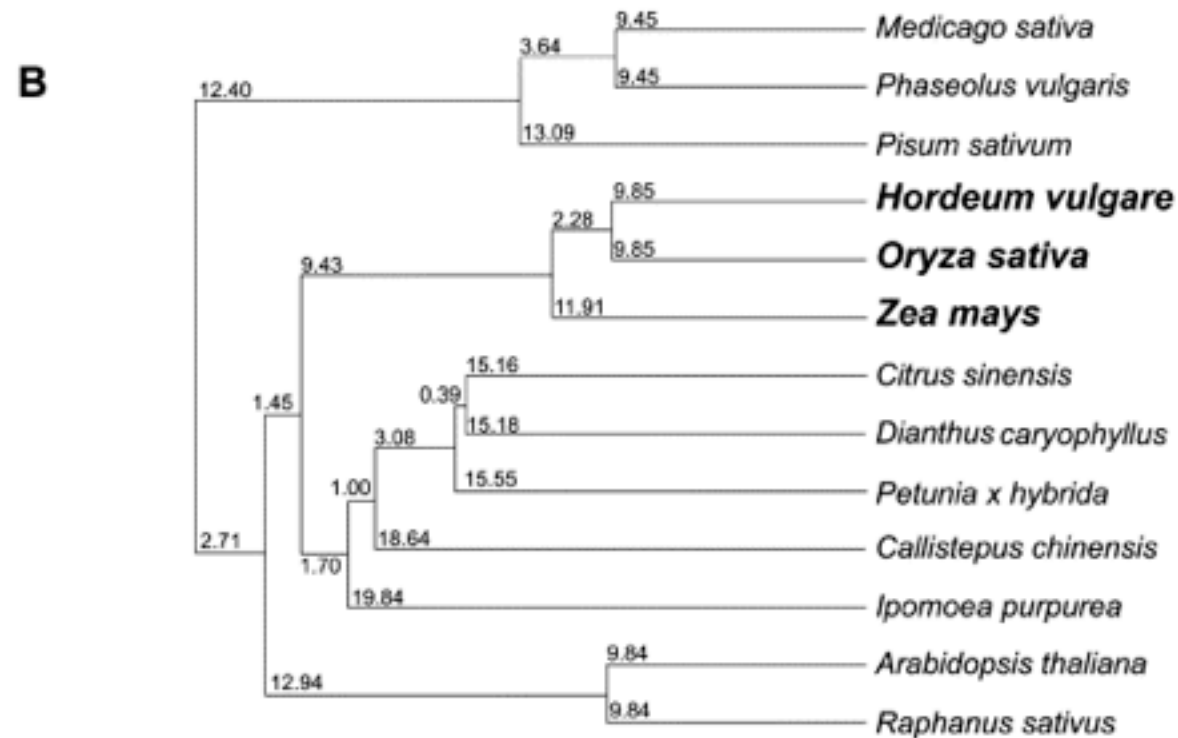
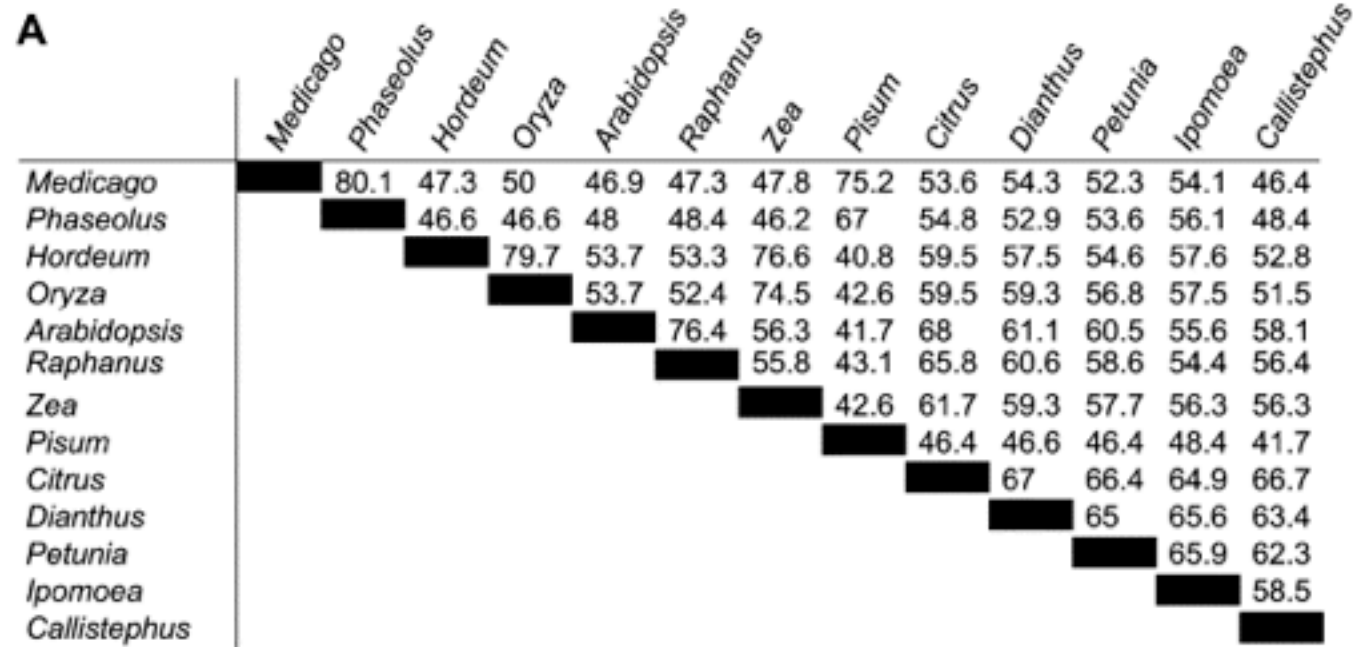
Динамическое программирование

Прогрессивные (иерархические) методы

Итеративные методы



ClustalW



Геномные перестройки

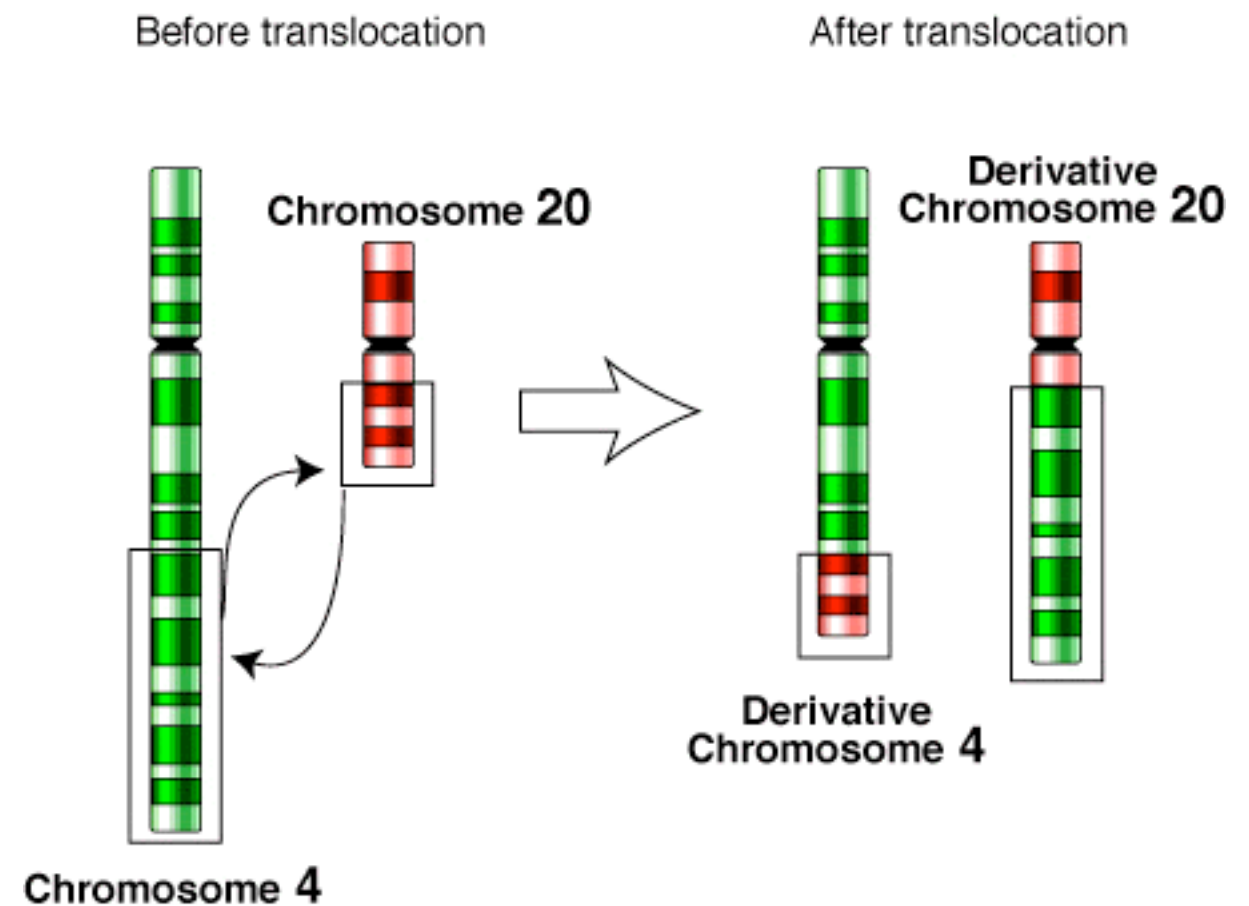
Крупные мутации:

реверсии

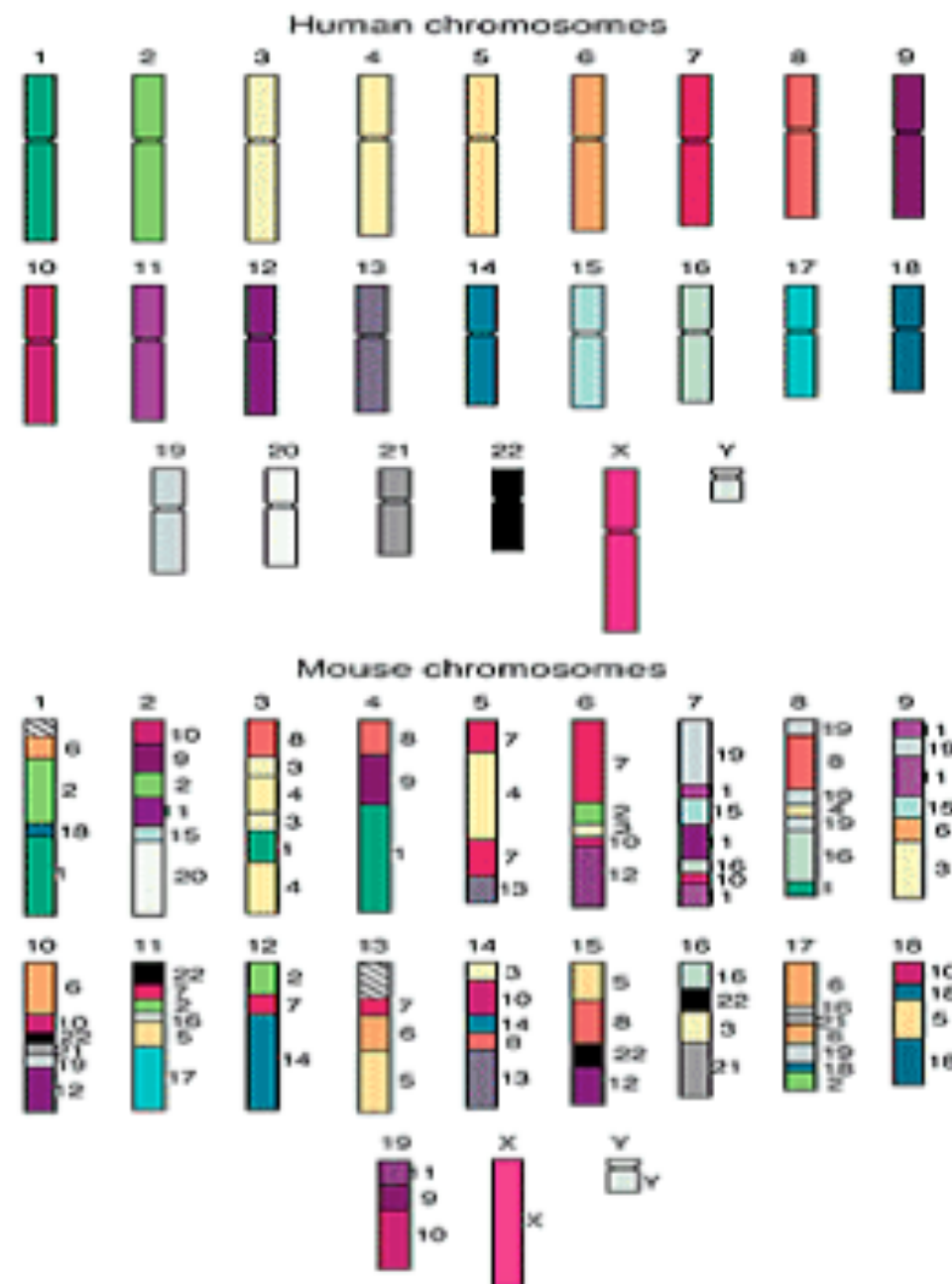
транслокации

слияния

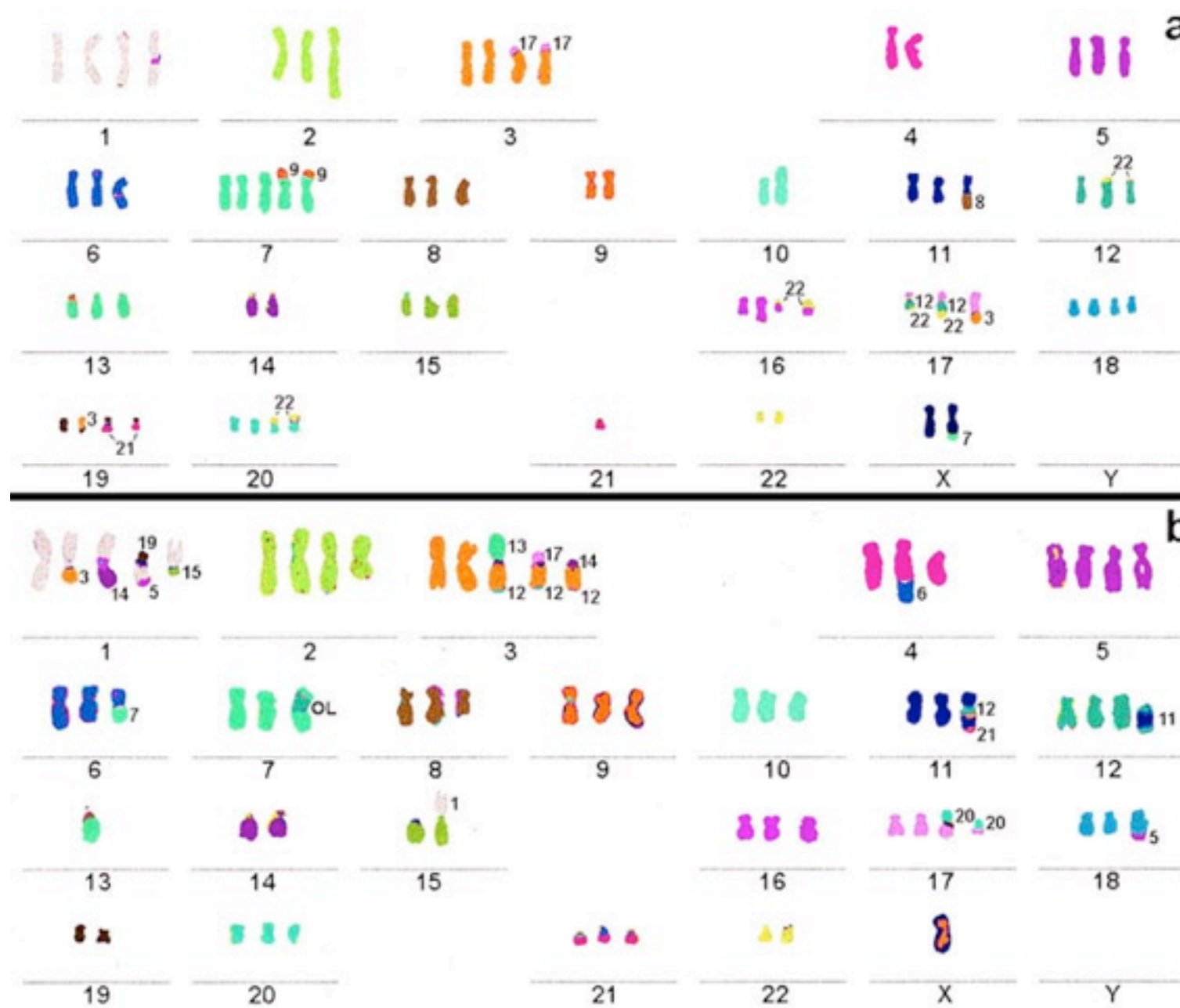
расщепления



Геномные перестройки



Геномные перестройки



Сортировка реверсиями

1 2 3 4 5 6 7 8 9

1 2 3 9 8 7 6 5 4

1 8 9 3 2 7 6 5 4

Что мы узнали

Редакционное расстояние

Глобальное и локальное выравнивание

Аффинная модель вставки

Всё за $O(N^2)$

Линейная память и 4 русских

FASTA, BLAST

Множественное выравнивание

Геномные перестройки

Формат обучения

12 лекций по воскресеньям

Квизы для самопроверки

Домашние задания и вопросы онлайн

<http://rosalind.info/classes/enroll/ff45302de4/>

