



Алгоритмы и программы множественного выравнивания Сравнение MSA

Алешина Юлия Александровна
Спирин Сергей Александрович

Цель множественного выравнивания последовательностей гомологичных доменов или белков

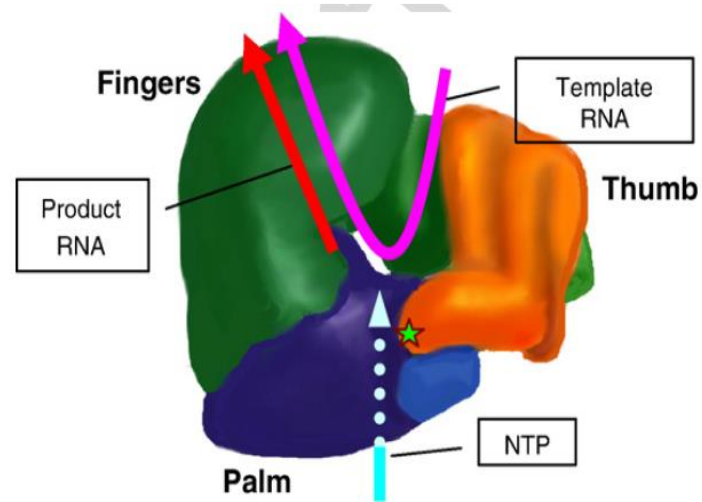
- ❖ Реконструкция эволюции белков от общего предка. А именно, в колонке выравнивания стоят аминокислотные остатки (ако), унаследованные от ако общего предка всех этих белков. Наследуются ДНК гена, т.е. кодоны, это подразумевается.
- ❖ Основа для установления гомологии – сходство последовательностей. Оно неравномерно на разных участках (слайд +1)
- ❖ Достигнута ли цель - практически не проверяемо (не считая парочки долговременных экспериментов (Ленски – E.coli, Кондрашов- *Schizophyllum commune*)
- ❖ На практике стремятся к этой цели, но следует понимать, что она недостижима. Разные программы множественного выравнивания используют разные алгоритмы и разные эвристики.

Выравнивание РНК-зависимых РНК полимераз (RdRP)

```

*      320      *      340      *      360      *      380      *      400      *      420      *      440      *      460
'KTMIRFGDVGLDLDFSAFASLSPFMIREA..GRIMSELS...GTPSHFGTALINTIISKHLLYNCCY.....HVCGSMPSGSPTALINSIINNVLYYVFSKIFGKSPVFF.....CQALKILC.YGDDVLI VFSRDV
SVAMQG.FERVYDVLYSNEDSTHSVAMFRLI..A...EEFF.TPENGFDPLTREYLESIAISTHAFEEKRF.....LITGGLPSGCAATSMINTIMNNIIRAGLYLTYNFEFDD.....VKVLS.YGDDLIVATNYQL
STHFAQ.YKNVWDVDYSAEDANHCSTAMNIMFEEVTEFG.....FHPNAEWILKTLVNTTEHAYENKRI.....VVEGCMPSGCSATSIINTILNNIYVLYALRRHYEGVELDT.....YTMIS.YGDDIVASDYDL
.....WSLCVATIVSDHDTFWPGWLRDLICDELINMGYA.PWVVKLFETSLKLPVYVGAFAPEQGHITLLGDPSNPDLEVGLSSGQGATDLMGTILMSITYLVMQLDHTAPHLNSRIKMPACSACRFLDSYWQGHEEIRQIS.KSDAILGWTGKR
RRRIPE.NWVYCDAGSQEDSSLTPYLINAV..LTIRSTYMEDWDVGLQMLRNLYTEIVYTPISTPDGTIV.....KKFRGNNSGQSTVDNSIMVVIAMHYALIKECFEVEEID.....STCVFFV.NGDDLIAVNPEK
IDKLNRPGLWHGSGDGRDSSIDPFFFDVV..KTRKHFL..PSEHHKAIDLIYDEILNTTICLANGMVI.....KKNVGTQR.QPSTVVDNTLVLMTAFLYAYIHKTGDRELAL.....LNERFIFVC.NGDDNKFAISPQF
VLSIASFSPYGFNGDFANFDGMFHPSSFSMV..SEIANIFY...GNFLSTERDNLRLTNRFSLMKGAIL.....RVPGGPGSGFEMTVIENSFINLFYI QSAWIMLARFNQRQDISH.....PCNFPKYVRACV.YGDDNTVAIKMEV
VARMKEKGNDVLCDDYSEDGLLSKQVMDVI..ASINELC.GGEDQLKNARRNLLMACCSRLAICKNTW.....RVECGIPSGGFEMTVIVNSIFNETILIRYHYKKLMREQQAPELMV.....QSFDKLI GLVT.YGDDNTISVNAVY
AEHAK.YKNHFDADYIANDSTQNRQIMTES..FSIMSRLT...ASPELAEVVAQDLLAPSEMDVGDYVI.....RVKEGLPSGFPCTSQVNSINHWTITLCAISEATGLSPDVV.....QSMSYFSFYGDDEIVSTDIDF
NNLTSKASDFLCDDYSKNDSTMSPCVRIA..IDLADCC...EQTELTKSVVLTIKSHFMTILEAMIV.....QTKRGLPSGMPPTSVINSICHWLWSAAVYKSCAEIGLHCS.....NLYELAPFYT.YGDDGVYAMTFMM
IQRIKS.AAKVYAVDYSKNDSTQSPRVSAA..IDLRYFS...DRSPIVDSAANTLKSPPIAIFNGVAV.....KVSSGLPSGMPITSVINSINHCLYVGCAILQSLEARGVPVTW.....NLFSTFCMMT.YGDDGVYMFMMF
'KRLEPKHDRVYCDYSKNDSTQPPKVTQS..IDLRHET....DKSPIVDSACATLKSNPIGIFNGVAF.....KVAGGLPSGMPITSVINSINHCLMGSAAVKALED SGVRVTW.....NIFDSMDLFT.YGDDGVYIVPPLI
D      D                                     g  sg  T  n3                                     gDD

```



Динамическое программирование

Алгоритм Нидлмана-Вунша
сложность $O(n*n)$

match = 1 mismatch = -1 gap = -1

		G	C	A	T	G	C	G	
		0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5	
A	-2	0	0	1	0	-1	-2	-3	
T	-3	-1	-1	0	2	1	0	-1	
T	-4	-2	-2	-1	1	1	0	-1	
A	-5	-3	-3	-1	0	0	0	-1	
C	-6	-4	-2	-2	-1	-1	1	0	
A	-7	-5	-3	-1	-2	-2	0	0	

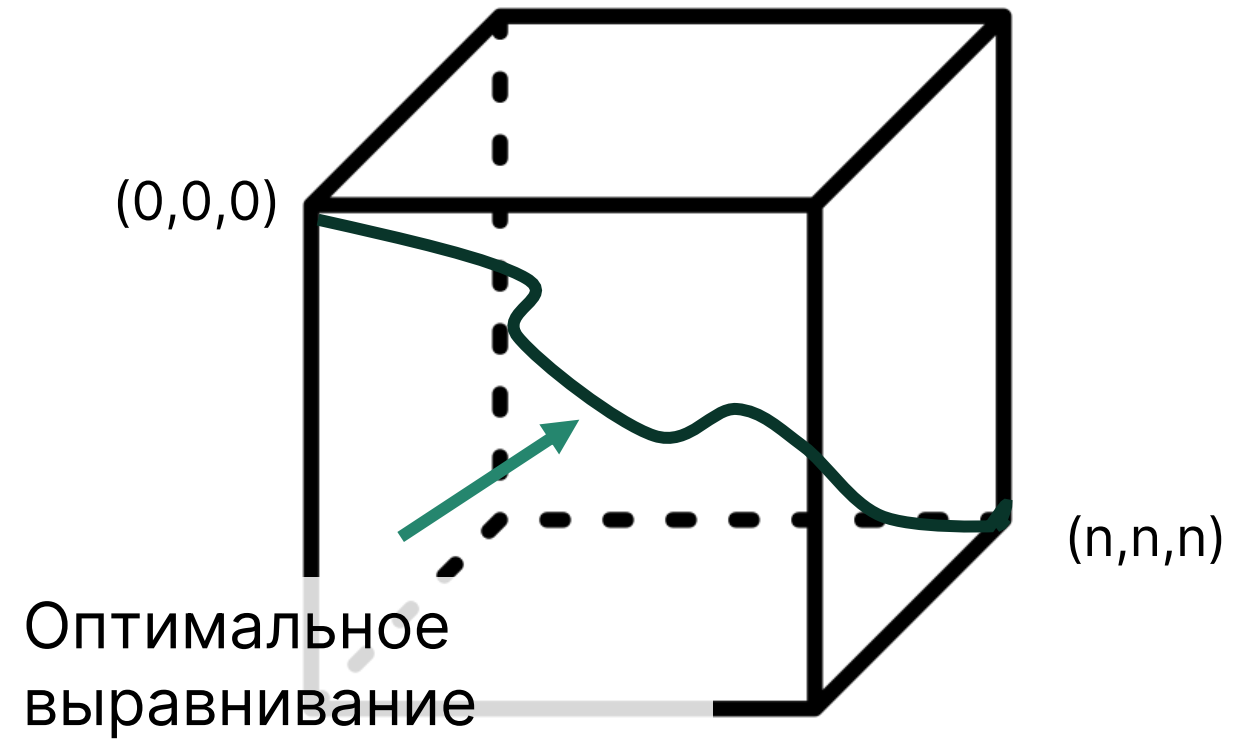
Динамическое программирование

Алгоритм Нидлмана-Вунша
сложность $O(n*n)$

match = 1 mismatch = -1 gap = -1

		G	C	A	T	G	C	G	
		0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5	
A	-2	0	0	1	0	-1	-2	-3	
T	-3	-1	-1	0	2	1	0	-1	
T	-4	-2	-2	-1	1	1	0	-1	
A	-5	-3	-3	-1	0	0	0	-1	
C	-6	-4	-2	-2	-1	-1	1	0	
A	-7	-5	-3	-1	-2	-2	0	0	

Множественное для трех
последовательностей



Динамическое программирование

Алгоритм Нидлмана-Вунша
сложность $O(n*n)$

match = 1 mismatch = -1 gap = -1

		G	C	A	T	G	C	G	
		0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5	
A	-2	0	0	1	0	-1	-2	-3	
T	-3	-1	-1	0	2	1	0	-1	
T	-4	-2	-2	-1	1	1	0	-1	
A	-5	-3	-3	-1	0	0	0	-1	
C	-6	-4	-2	-2	-1	-1	1	0	
A	-7	-5	-3	-1	-2	-2	0	0	

Множественное выравнивание
Сложность $O(n^k)$

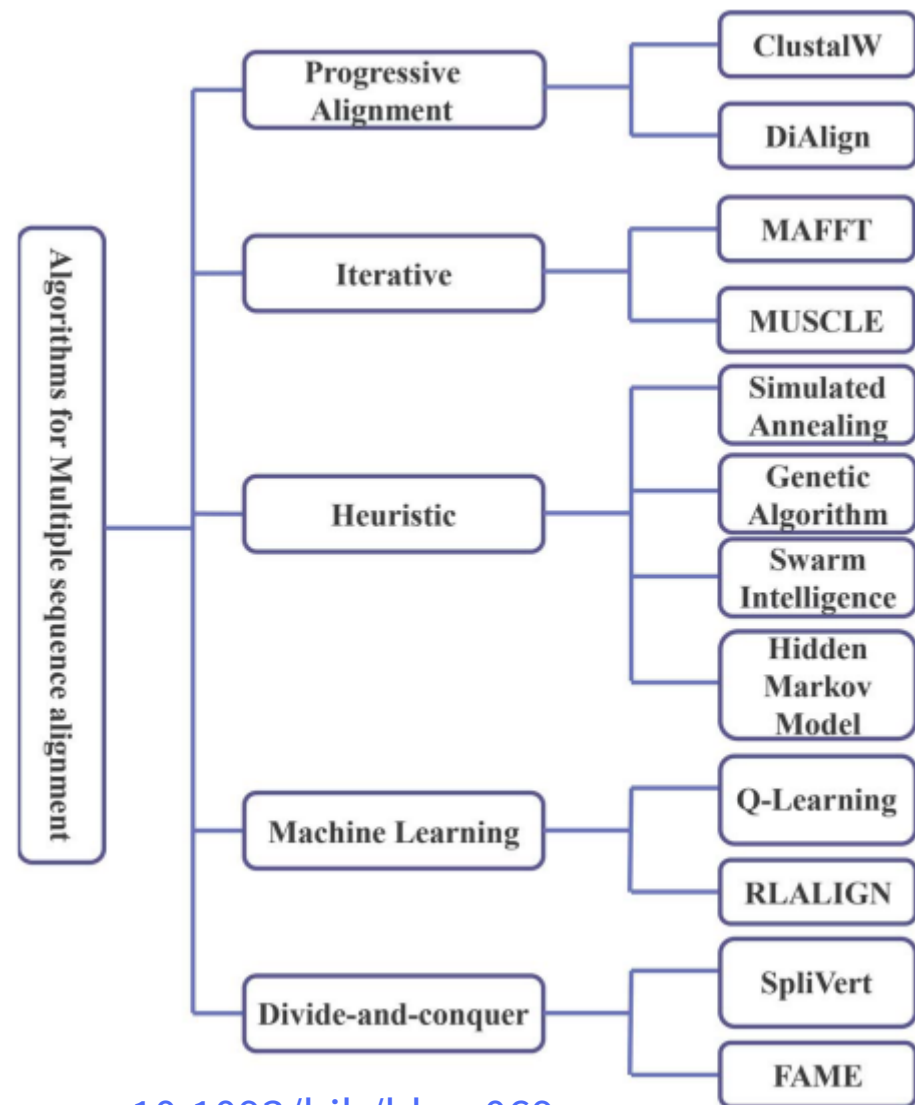
СЛОЖНОСТЬ $O(n^2)$

match = 1 mismatch = -1 gap = -1

		G	C	A	T	G	C	G	
		0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5	
A	-2	0	0	1	0	-1	-2	-3	
T	-3	-1	-1	0	2	1	0	-1	
T	-4	-2	-2	-1	1	1	0	-1	
A	-5	-3	-3	-1	0	0	0	-1	
C	-6	-4	-2	-2	-1	-1	1	0	
A	-7	-5	-3	-1	-2	-2	0	0	

Программы множественного выравнивания

- ❖ Программ много, и все они используют разные **эвристические** алгоритмы
- ❖ * и классификации в литературе встречаются разные (например, https://link.springer.com/protocol/10.1007/978-1-0716-1036-7_17)

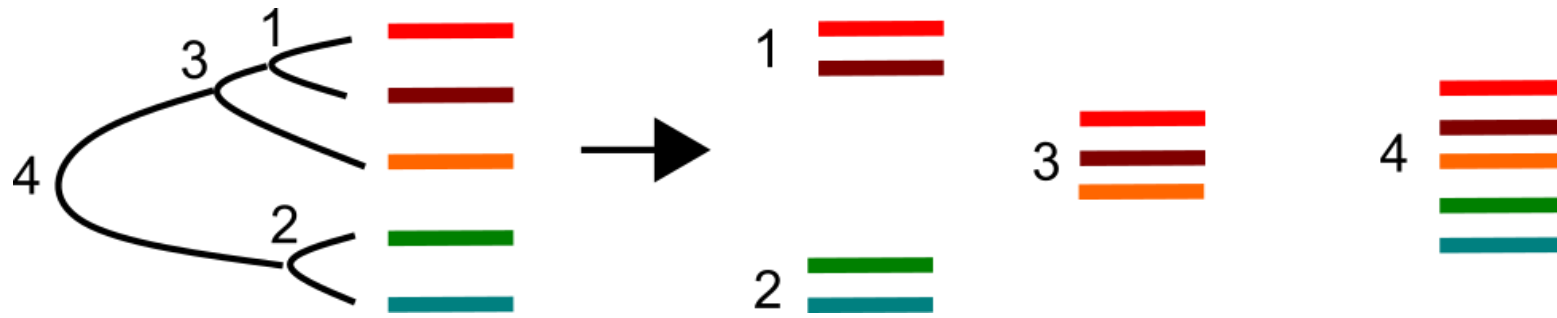


Список популярных программ MSA

Program	Год	в Jalview	на kodomo	Website
ClustalO	2011	yes		https://www.ebi.ac.uk/jdispatcher/msa/clustalo
ClustalW	1994	yes	yes	http://www.clustal.org/
MAFFT	2002	yes	yes	https://mafft.cbrc.jp/alignment/software/
MSAProbs	2010	yes		http://msaprobs.sourceforge.net/
MUSCLE	2004	yes	yes	http://www.drive5.com/muscle/
ProbCons	2005	yes		http://probcons.stanford.edu/
T-Coffee	2000	yes		http://tcoffee.crg.cat/
Glprobs	2015	yes		
DiAlign	1998		yes	http://dialign.gobics.de/
FAME	2020			http://github.com/naznoosh/msa
Kalign	2005			https://www.ebi.ac.uk/Tools/msa/kalign/
NX4	2019			https://www.nx.io
PRANK	2008		yes	http://wasabiapp.org/software/prank/
Probalign	2006			http://probalign.njit.edu/standalone.html

Прогрессивное выравнивание

- ❖ Основная идея: выравнивание двух выравниваний с помощью динамического программирования
- ❖ Этапы:
 - ✓ Построить дерево родственности всех последовательностей – направляющее дерево
 - ✓ Выравниваем стопки выровненных последовательностей от листьев до корня

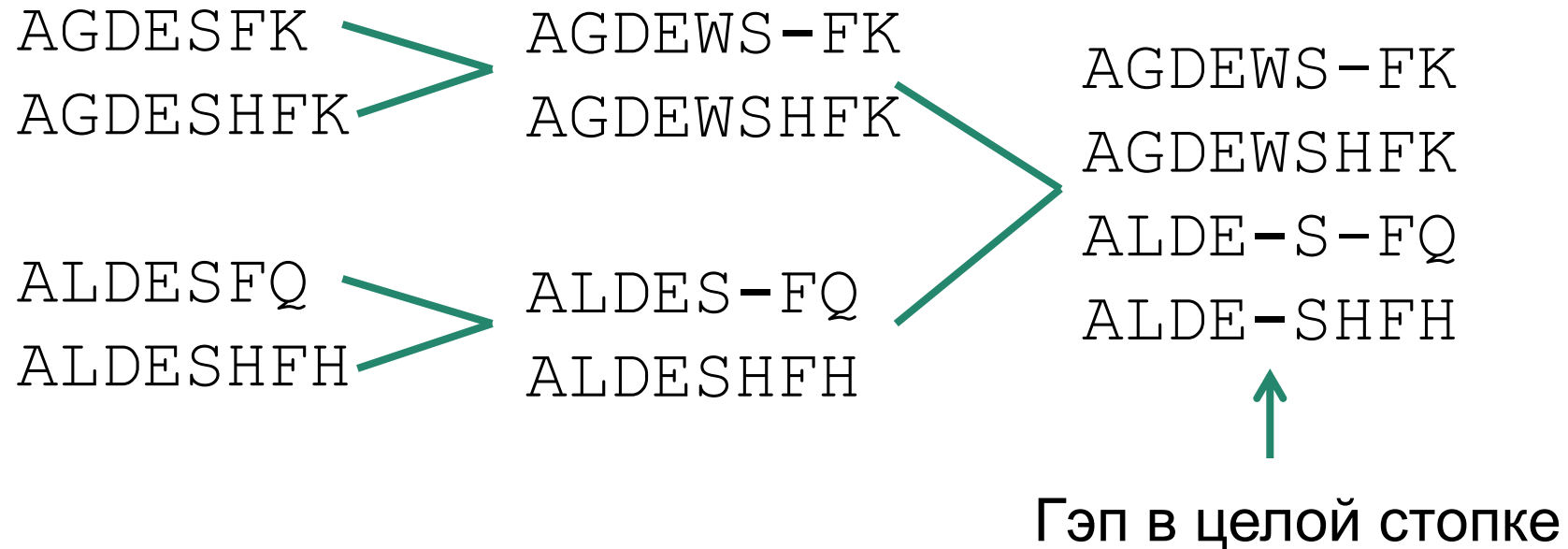


Построение направляющего дерева

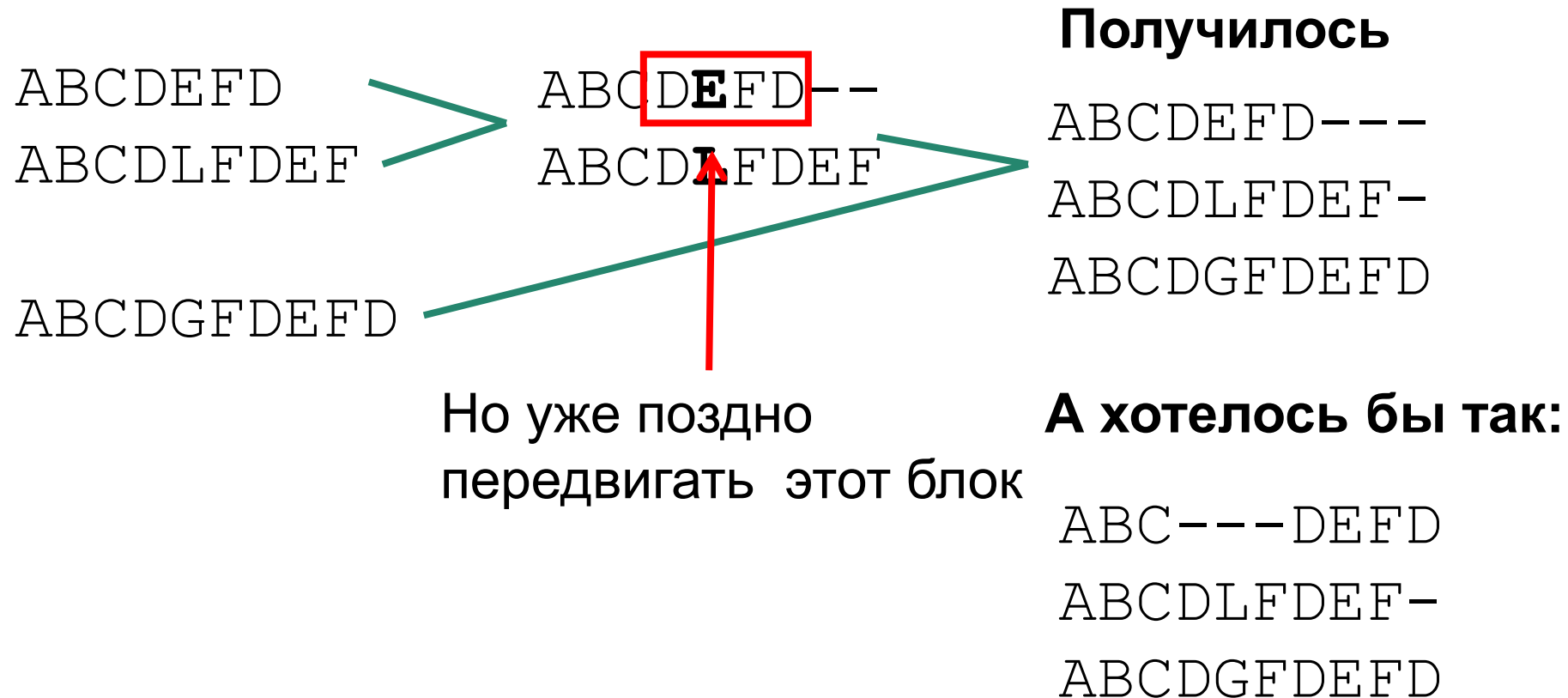
- ❖ Для ВСЕХ ПАР последовательностей строится парное выравнивание.
- ❖ Вес парного выравнивания пересчитывается в расстояние между последовательностями:
 - ✓ чем больше вес, тем меньше расстояние;
 - ✓ расстояние между совпадающими последовательностями равно 0.
- ❖ Получается матрица расстояний между послед-ми
- ❖ Есть алгоритмы, превращающие матрицу попарных расстояний в дерево.
 - ✓ Расстояния между листьями по дереву отражают сходство последовательностей



Пример прогрессивного выравнивания

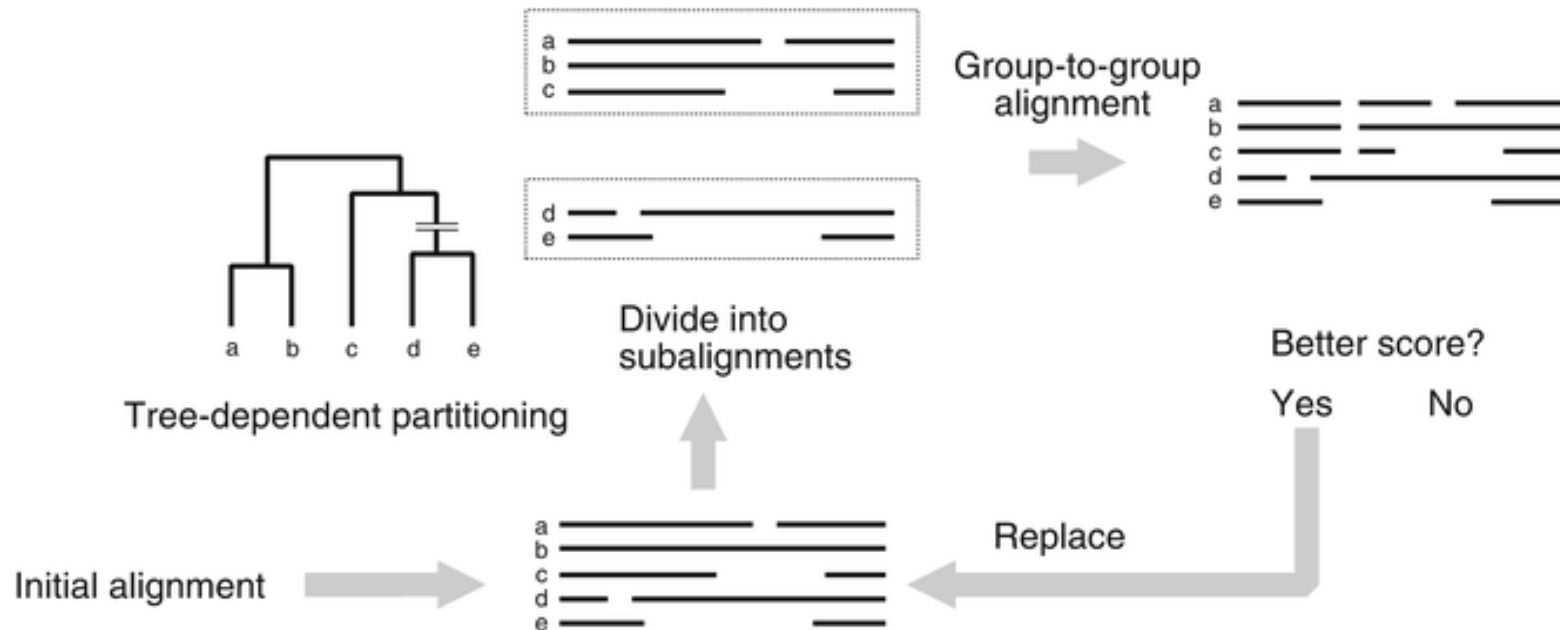


Проблема: если ошиблись, то ошибка никуда не денется



Итеративное рафинирование выравнивания

- i. Построить множественное выравнивание
- ii. Разделить его на две группы
- iii. Перевыровнять две группы. Повторить ii



Иерархический алгоритм выравнивания многих последовательностей

- ❖ Основная идея: выравнивание двух выравниваний с помощью динамического программирования
- ❖ Этапы алгоритма
 - ✓ Построение направляющего дерева
 - ✓ Итерация выравнивания выравниваний
 - ✓ “Рафинирование” (refinement) выравнивания
- ❖ Результат – ГЛОБАЛЬНОЕ множественное выравнивание

Список популярных программ множественного выравнивания

- ❖ Прогрессивное выравнивание (ClustalW)
- ❖ Итеративное рафинирование (MUSCLE, PRALINE, IterAlign, MAFFT)
- ❖ Методы, основанные на согласованности (PROBCONS)
- ❖ Методы, использующие структурные данные (3D-Coffee)

А какая программа лучше?

- ❖ Все алгоритмы множественного выравнивания эвристические

Как сравнить программы?

- ❖ Базы данных идеальных выравниваний! Обычно 3D-наложение структур + ручная проверка экспертами
- ❖ BAliBASE (Benchmark Alignment dataBASE) – наиболее признанная
 - ✓ 10 референсных выравниваний, покрывающих разные особенности (высокая вариабельность, повторы и т д)
- ❖ Есть и другие
 - ✓ SABMark(SequenceAlignmentBenchMark)
 - ✓ OXBench(OXfordBenchmark)
 - ✓ SMART(SimpleModularArchitectureResearchTool)
 - ✓ PREFAB(ProteinREferenceAlignmentBenchmark) и др

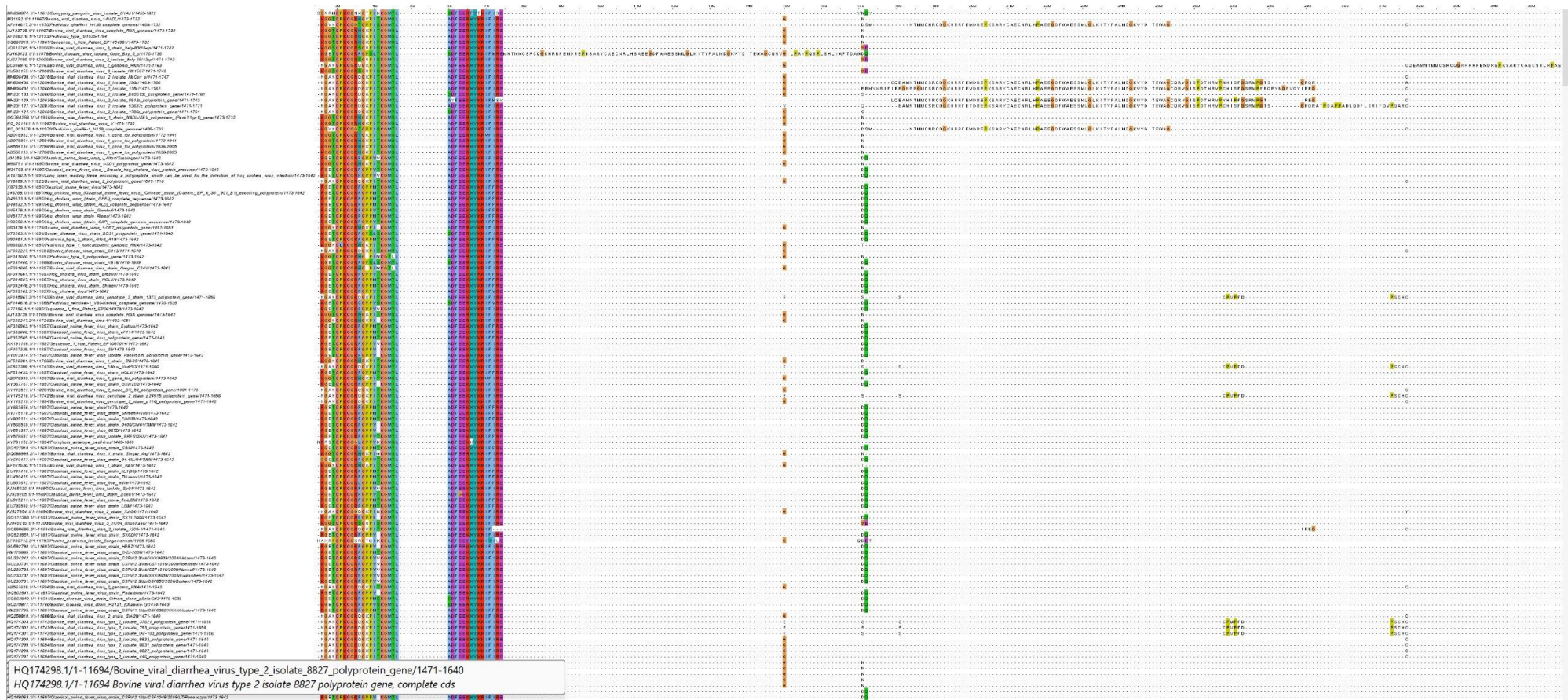
Результаты сравнения программ на BaliBase

Program	R1-1	R1-2	R2	R3	R4	R5	Avg Score	Tot time(s)
MSAProbs [122]	0.441	0.865	0.464	0.607	0.622	0.608	0.607	12382.00
Probalign [84]	0.453	0.862	0.439	0.566	0.603	0.549	0.589	10095.20
MAFFT [14]	0.439	0.831	0.450	0.581	0.605	0.591	0.588	1475.40
Probcons [17]	0.417	0.855	0.406	0.544	0.532	0.573	0.558	13086.30
Clustal Omega [123]	0.358	0.789	0.450	0.575	0.579	0.533	0.554	539.91
T-Coffee [71]	0.410	0.848	0.402	0.491	0.545	0.587	0.551	81041.50
Kalign [72]	0.365	0.790	0.360	0.476	0.504	0.435	0.501	21.88
MUSCLE [15]	0.318	0.804	0.350	0.409	0.450	0.460	0.475	789.57
FSA [136]	0.270	0.818	0.187	0.259	0.474	0.398	0.419	53648.10
DiAlign [12]	0.265	0.696	0.292	0.312	0.441	0.425	0.415	3977.44
PRANK [137]	0.223	0.680	0.257	0.321	0.360	0.356	0.376	128355.00
ClustalW [11]	0.227	0.712	0.220	0.272	0.396	0.308	0.374	766.47

- Колонки 2-7 – разные наборы тестовых выравниваний.
- Указан TC (total column) вес сравнения с референсным выравниванием.
- Значения от 0 до 1. Чем больше – тем более похоже выравнивание на референсное.
- Последняя колонка – время работы программы

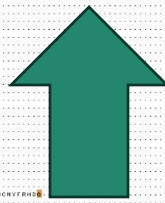
Результаты выравнивания одних и тех же последовательностей разными программами могут существенно отличаться!

Пример из жизни. Выравнивание участка с вставкой Jiv90 полипротеинов пестивирусов.



Выравнивание тех же последовательностей с помощью MUSCLE

q_muscle.fast



Сравнение разных выравниваний одних и тех же последовательностей

Построенных разными алгоритмами множественного выравнивания

Какие выравнивания тех же последовательностей совпадают?

1

```
Seq1 MKFR-SSHYA-S  
Seq2 MKYRRS-HYA-S  
Seq3 MEFRRRSHYA-R
```

2

```
Seq1 MKF-RSSHYAS  
Seq2 MKYRRS-HYAS  
Seq3 MEFRRRSHYAR
```

4

```
Seq1 MKFR-SSHYA-S  
Seq2 MKYR-RSHYA-S  
Seq3 MEFRRRSHYA-R
```

3

```
Seq1 -MKFR-SSHYAS  
Seq2 -MKYR-RSHYAS  
Seq3 -MEFRRRSHYAR
```

	1	2	3	4	5	6	7	8	9	10	11	12
Seq1	M	K	F	R	-	S	S	H	Y	A	-	S
Seq2	M	K	Y	R	R	S	-	H	Y	A	-	S
Seq3	M	E	F	R	R	R	S	H	Y	A	-	R

	1	2	3	4	5	6	7	8	9	10	11
Seq1	M	K	F	-	R	S	S	H	Y	A	S
Seq2	M	K	Y	R	R	S	-	H	Y	A	S
Seq3	M	E	F	R	R	R	S	H	Y	A	R

	1	2	3	4	5	6	7	8	9	10	11	12
Seq1	M	K	F	R	-	S	S	H	Y	A	-	S
Seq2	M	K	Y	R	-	R	S	H	Y	A	-	S
Seq3	M	E	F	R	R	R	S	H	Y	A	-	R

	1	2	3	4	5	6	7	8	9	10	11	12
Seq1	-	M	K	F	R	-	S	S	H	Y	A	S
Seq2	-	M	K	Y	R	-	R	S	H	Y	A	S
Seq3	-	M	E	F	R	R	R	S	H	Y	A	R

Колонка i выравнивания X совпадает с колонкой j выравнивания Y если в них – те же самые остатки;
те же самые значит – с теми же номерами, а не с теми же буквами!


```
Seq1  MKFR-SSHYA-S  
Seq2  MKYRRS-HYA-S  
Seq3  MEFRRRSHYA-R
```

```
Seq1  MKF-RSSHYPAS  
Seq2  MKYRRS-HYPAS  
Seq3  MEFRRRSHYPAR
```

Вместо аминокислоты укажем номер остатка в последовательности:

```
Seq1  1234-5678 9 -10  
Seq2  123456-78 9 -10  
Seq3  123456789 10-11
```

```
Seq1  123-45678 9 10  
Seq2  123456-78 9 10  
Seq3  123456789 10 11
```

Seq1 MKFR-SSHYA-S
Seq2 MKYRRS-HYA-S
Seq3 MEFRRRSHYA-R

Seq1 MKF-RSSHYPAS
Seq2 MKYRRS-HYPAS
Seq3 MEFRRRSHYPAR

Вместо аминокислоты укажем номер остатка в последовательности:

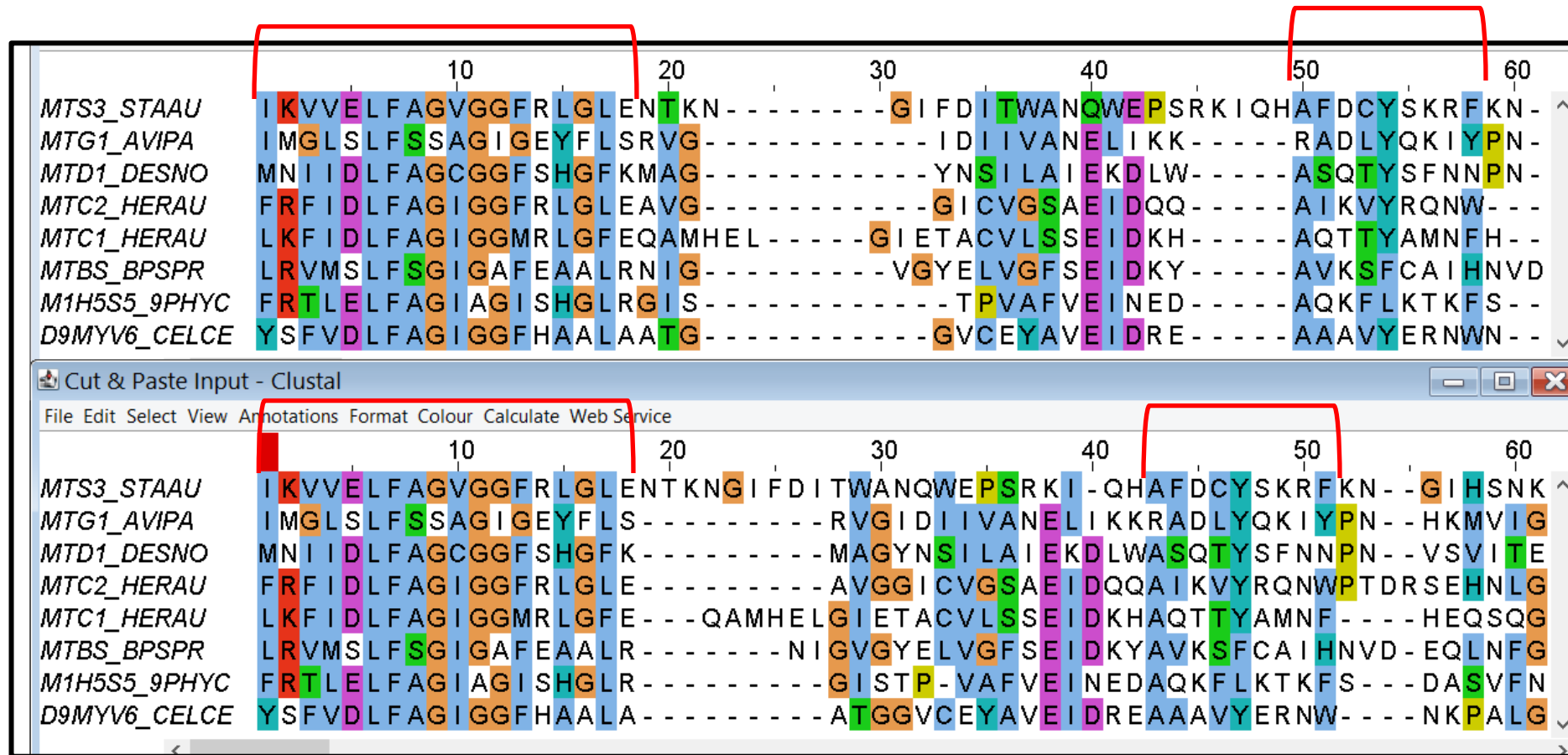
Seq1 123**4**-5678 9 -10
Seq2 123**45**6-78 9 -10
Seq3 123**45**6789 10-11

Seq1 123-**4**5678 9 10
Seq2 123**45**6-78 9 10
Seq3 123**45**6789 10 11

Точное сравнение двух выравниваний

- ❖ Два выравнивания I и II тех же последовательностей совпадают если в каждой колонке i ($i = 1, 2, \dots, N$) выравниваний I и II стоят те же самые буквы.
- ❖ Буквы не в смысле а.к.о., а в смысле номера буквы в последовательности
- ❖ Различие выравниваний I и II определяется числом колонок, для которых это не так и их расположением

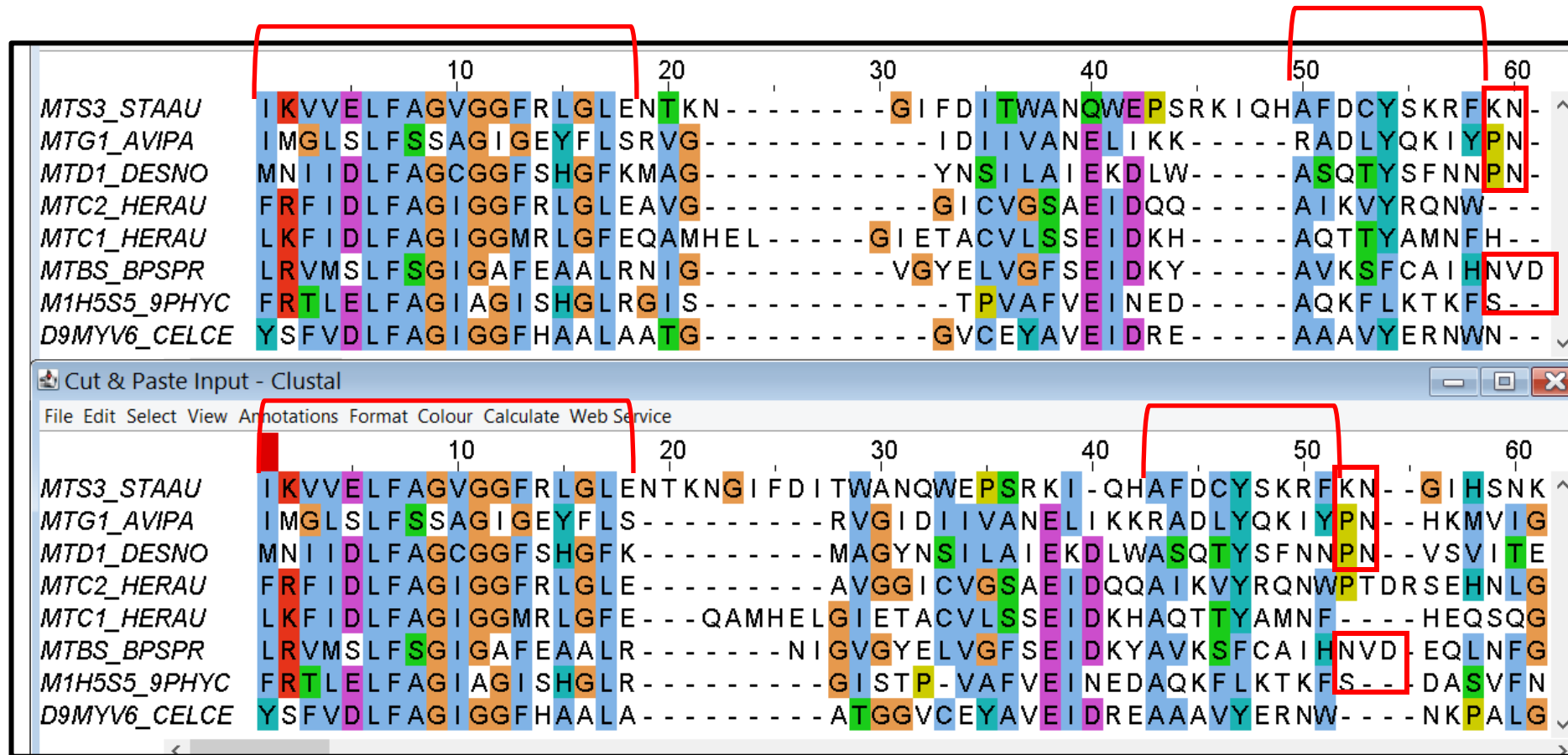
Выравнивания совпадают на двух выделенных участках
Ещё есть совпадения не на всех, а на подмножествах последовательностей.
НАЙДИТЕ!



Выравнивания совпадают на двух выделенных участках
Ещё есть совпадения не на всех, а на подмножествах последовательностей.
НАЙДИТЕ!

Ещё есть совпадения не на всех, а на подмножествах последовательностей.

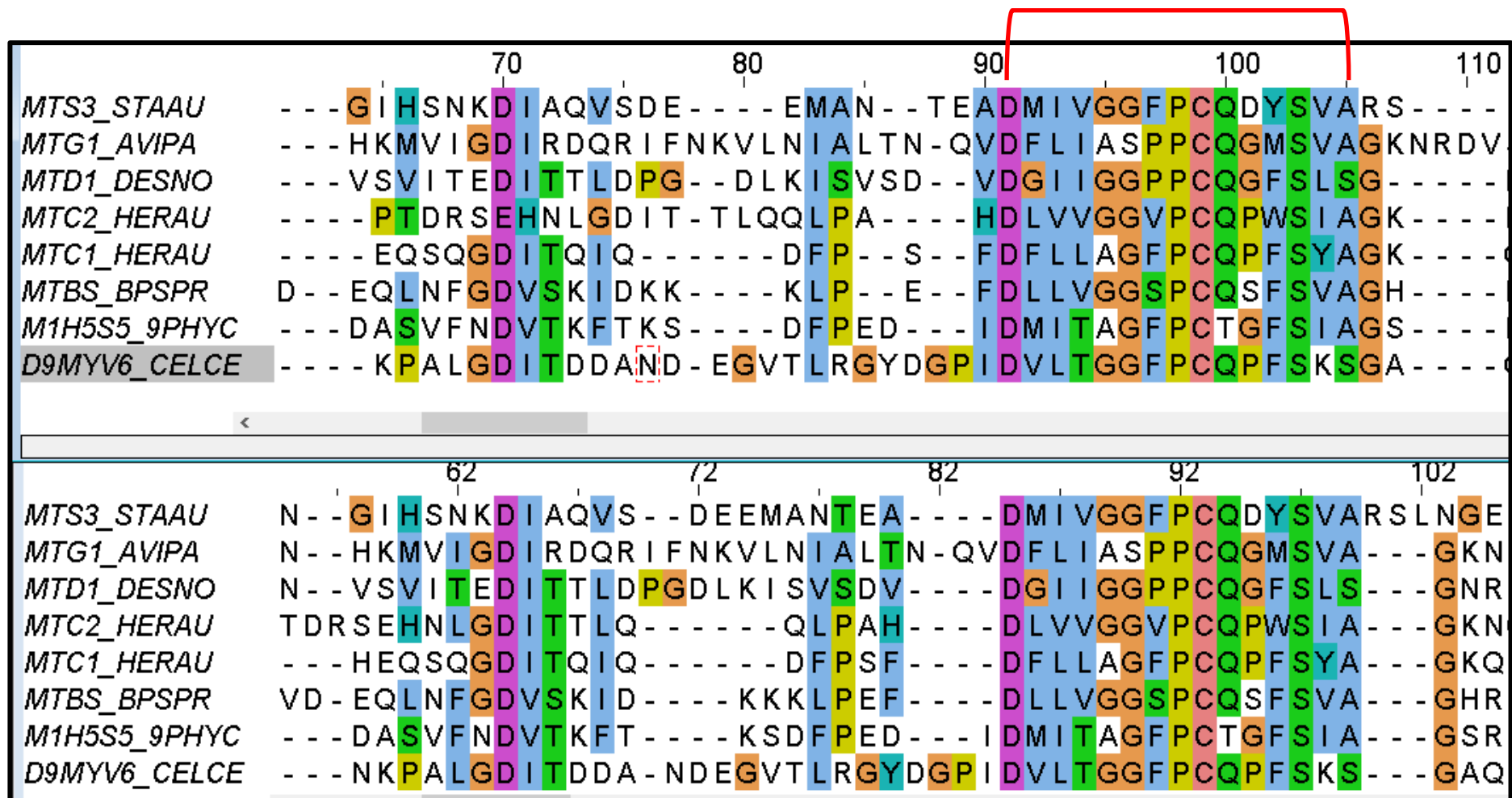
НАЙДИТЕ!



Продолжение. Разметьте самостоятельно!

		70		80		90		100		110																																							
MTS3_STAAU	- - -	G	I	H	S	N	K	D	I	A	Q	V	S	D	E	- - -	E	M	A	N	- -	T	E	A	D	M	I	V	G	G	F	P	C	Q	D	Y	S	V	A	R	S	- - -							
MTG1_AVIPA	- - -	H	K	M	V	I	G	D	I	R	D	Q	R	I	F	N	K	V	L	N	I	A	L	T	N	-	Q	V	D	F	L	I	A	S	P	P	C	Q	G	M	S	V	A	G	K	N	R	D	V
MTD1_DESNO	- - -	V	S	V	I	T	E	D	I	T	T	L	D	P	G	- -	D	L	K	I	S	V	S	D	- -	V	D	G	I	I	G	G	P	P	C	Q	G	F	S	L	S	G	- - -						
MTC2_HERAU	- - -	P	T	D	R	S	E	H	N	L	G	D	I	T	-	T	L	Q	Q	L	P	A	- - -	H	D	L	V	V	G	G	V	P	C	Q	P	W	S	I	A	G	K	- - -							
MTC1_HERAU	- - -	E	Q	S	Q	G	D	I	T	Q	I	Q	- - -	- -	D	F	P	- -	S	- -	F	D	F	L	L	A	G	F	P	C	Q	P	F	S	Y	A	G	K	- - -										
MTBS_BPSPR	D - -	E	Q	L	N	F	G	D	V	S	K	I	D	K	K	- - -	K	L	P	- -	E	- -	F	D	L	L	V	G	G	S	P	C	Q	S	F	S	V	A	G	H	- - -								
M1H5S5_9PHYC	- - -	D	A	S	V	F	N	D	V	T	K	F	T	K	S	- - -	D	F	P	E	D	- -	I	D	M	I	T	A	G	F	P	C	T	G	F	S	I	A	G	S	- - -								
D9MYV6_CELCE	- - -	K	P	A	L	G	D	I	T	D	D	A	N	D	-	E	G	V	T	L	R	G	Y	D	G	P	I	D	V	L	T	G	G	F	P	C	Q	P	F	S	K	S	G	A	- - -				
		62		72		82		92		102																																							
MTS3_STAAU	N - -	G	I	H	S	N	K	D	I	A	Q	V	S	- -	D	E	E	M	A	N	T	E	A	- - -	D	M	I	V	G	G	F	P	C	Q	D	Y	S	V	A	R	S	L	N	G	E				
MTG1_AVIPA	N - -	H	K	M	V	I	G	D	I	R	D	Q	R	I	F	N	K	V	L	N	I	A	L	T	N	-	Q	V	D	F	L	I	A	S	P	P	C	Q	G	M	S	V	A	- - -	G	K	N		
MTD1_DESNO	N - -	V	S	V	I	T	E	D	I	T	T	L	D	P	G	D	L	K	I	S	V	S	D	V	- - -	D	G	I	I	G	G	P	P	C	Q	G	F	S	L	S	- - -	G	N	R					
MTC2_HERAU	T	D	R	S	E	H	N	L	G	D	I	T	T	L	Q	- - - - -	Q	L	P	A	H	- - -	D	L	V	V	G	G	V	P	C	Q	P	W	S	I	A	- - -	G	K	N								
MTC1_HERAU	- - -	H	E	Q	S	Q	G	D	I	T	Q	I	Q	- - - - -	D	F	P	S	F	- - -	D	F	L	L	A	G	F	P	C	Q	P	F	S	Y	A	- - -	G	K	Q										
MTBS_BPSPR	V	D	-	E	Q	L	N	F	G	D	V	S	K	I	D	- - -	K	K	K	L	P	E	F	- - -	D	L	L	V	G	G	S	P	C	Q	S	F	S	V	A	- - -	G	H	R						
M1H5S5_9PHYC	- - -	D	A	S	V	F	N	D	V	T	K	F	T	- - -	K	S	D	F	P	E	D	- -	I	D	M	I	T	A	G	F	P	C	T	G	F	S	I	A	- - -	G	S	R							
D9MYV6_CELCE	- - -	N	K	P	A	L	G	D	I	T	D	D	A	-	N	D	E	G	V	T	L	R	G	Y	D	G	P	I	D	V	L	T	G	G	F	P	C	Q	P	F	S	K	S	- - -	G	A	Q		

Если в двух блоках без гэпов из разных выравниваний есть хотя бы одна одинаково выровненная позиция, то блоки выровнены одинаково



Как сравнить два множественных выравнивания

1. Отсортировать последовательности по ID в обоих выравниваниях.
2. Если колонка (номеров) букв в выравнивании A отличается хотя бы одной буквой от колонки букв в выравнивании B, то они выровнены не одинаково
3. Алгоритм сравнения

1) Для каждого выравнивания идём по колонкам $N = 1, 2, 3, \dots$ и составляем вектор S:

$N: SI(N) = s_1, s_2, \dots, s_n$

s_1 – номер буквы в первой последовательности

s_2 – номер буквы во второй последовательности

.....

Если в последовательности i стоит гэп, то $s_i = \text{"-"}$

2) Если $SI(N) = SII(N')$, то колонка N выравнивания I выровнена одинаково с колонкой N' выравнивания II

3) Последовательно идущие в обоих выравниваниях одинаково выровненные колонки объявляются блоком одинаково выровненных фрагментов

Готовые программы для сравнения выравниваний

- ❖ VerAlign: a multiple sequence alignment assessment tool
(<https://www.ibi.vu.nl/programs/veralignwww/>)
- ❖ AlignStat: A tool for the statistical comparison of alternative multiple sequence alignments
(<https://github.com/TS404/AlignStat?tab=readme-ov-file>)

VerAlign – программа для ручного сравнения – раскрашивает колонки второго выравнивание цветами первого

```
GVLAGLUNDPS-----CRYAYGVNKR---TKFLEKSSISEVDGKELNALYFNNQ--HKILVSCAFPCQDESSQYTK-
EIVAAVDNWRP-----AINTYQGNF-----THPIHELDLAQIDAAVSLIKTHS--PELILGGFPCCQDESSAG--
NSILALEKDLW-----ASQTYSFNNPN--VSVITEDITTLDPG--DLKISVSD--VDGILGGFPCCQGFSSLSG--
ICVGSAEIDQQ-----AIKVYRQNW-----PTDRSEHNLGDIIT-TLQQLFA---HDLVVGGVPCQPFWSIAGK-
ACVLSSEIDKH--AQT---TYAMNHFH-----EQSQGDITQIQ-----DFP--S--FDFELLAGFPCQPFSSYAGK-
ELVGFSEIDKY-----AVKSFCATHNVD--EQLNFGDVSKIDKK-----KLP--E--FDLLVGGSPCQSFSSVAGH-
KCVFSSEWDKY--AAQ---TYEANYG-----EKPHGDITKINEN-----DIP--D--QDVLLAGFPCQPFSSNIGK-
ECVLSSEIDKK--ACE---TYALNFK-----EEPQGDITHEIT-----SFP--E--FDFELLAGFPCQPFSSYAGK-
ETVWANEYDKN-----AAITYQSNFKN-----KLIIDDIRNIKVE-----DVP--D--FDVLVSGFPCCTSFSSVAGY-
VCVASAEIDQQ-----AIKVYRQNW-----PTDGVVDHNLGDIIT-AIQQLFA---HDVLVGGVPCQPFWSIAGK-
TPVAFVEINED-----AQKFLTKKFS---DASVFNDVTKFTKS---DFPED--IDMITAGFPCCTGFSSIAGS-
ECVYSNEWDKY--AQE---VYEMNFG-----EKPEGDITQVNEK-----TIF--D--HDILCAGFPCQAFSSISGK-
VCEYAVEIDRE-----AAAVYERNWN-----KPALGDITDDAND--EGVTLRGYDGPIDVLVGGFPCQPFSSKSGA-
NVVFSSEWDKF--AQK---TYHANYG-----DFPDGDITKIDEK-----DIP--D--HEILVGGFPCVAFSSQAGL-
EHAWANDIDEW-----ACETFTTNICPDRPD SVVCGDVRELDIKSLGEEKFG---EIDAFTEGFPCCNDYSSIVGE-
```



```
FEI--PAANEYD---K-----TIWATFANHPKT--HLIEGDIRKI-----KE-----E-D-----F---
PVSNGYWKRRKKD---D-----ELKIIYNAIKLSEK--EGNIFDIRDL-----YK-----R-T-----L---
-----FENRADKLGGQ-----KLKDMYLANKLNK---NFGDIRSI-----DP-----K-K-----L---
VKS--VFSSEID---K-----FAIKTYANFGD---E-PHGDITKI-----DE-----K-D-----I---
LST--YGAVEID---K-----NAAETLNNRPKW--KVIENDIEFI-----AD-----NLDDEFI-----D---
FDI--TWNQWE---PSRKIQHAFDCY--KRFKNGI--H-SNKDIAQV-----SD-----E-E-----M---
VRC--VFSSEID---K-----YAVQTYQANHGG--T-VCGDITQT-----DV-----A-D-----I---
FSH--VALIEIE---P-----SACQTLNNRPDW--NVIEGDVRIFF-----QG-----E-G-----Y---
GKC--VFSSEID---P-----FAKFTYTNFEGV--V-PFGDITKV-----EA-----T-T-----I---
FNI--VFANDNW---K-----GCWKTFKNHGI--KINKKPIEWL-----KP-----S-E-----I---
FEI--CAAFENW---E-----KAIEIYNNFESH--PIYNIDLRNE-----KE-----AV-E-K---IKK---
FRI--ICANEYD---K-----SIWKTYSNHSA--KLIKGDISKI-----SS-----D-E-----F---
```

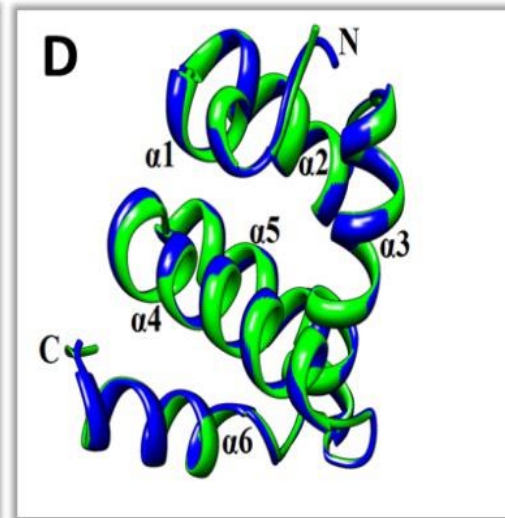
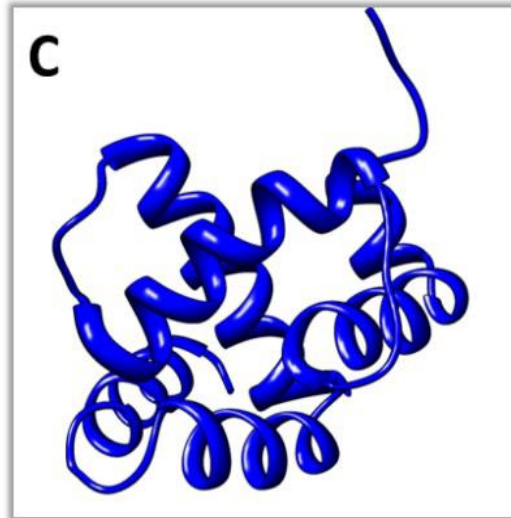
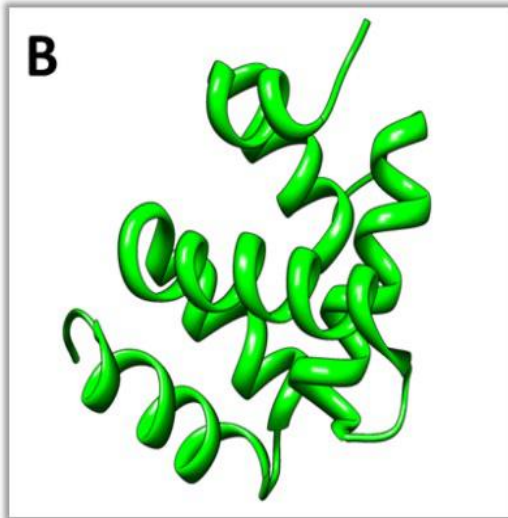
**Как верифицировать выравнивание
белков?**

Выравнивание, построенное по совмещению полипептидных цепей

A

		*	20		*	40		*	60		*				
MsepCSP5_	:	MNSFTVLCIFALVALAVARPDG	KYTD	RYDS	VNLD	QILSNRRLIV	VPYIK	CMLDQ	GKCTPD	GKELK	THIREA	:	70		
MbraCSFA6	:	-----	ED	KYTD	KYDN	NINLDEIL	ANKRLIV	AVVNC	VMERG	KCSPE	GKELKE	HLQDA	:	50	
			KYTD	YD	NLD	IL	N	RLIV	Y	C	GKC	P	GKELK	H	A

		80		*	100		*	120		*								
MsepCSP5_	:	LEQDC	AKCT	KAQR	DGTR	QVMG	HLIN	HEVD	YNEL	KAKY	DEK	NLYST	THHE	QELR	KLKQ	-----	:	127
MbraCSFA6	:	IENG	CKKCT	ENG	EKG	AYRV	IEHL	IKNE	IEIWR	ELTAK	YDPT	GNWR	KKYED	RAKA	AGIV	IPPEE	:	112
		E	C	KCT	Q	G	V	HLI	E	W	EL	AKYDP	K	E				



- ✓ С-alpha атомы а.к.о. из одной колонки при совмещении 3D-структур этих белков оказываются в одном положении

Множественное совмещение структур белков

- ❖ PDBeFold – сайт работает, но множественное совмещение не удалось запустить

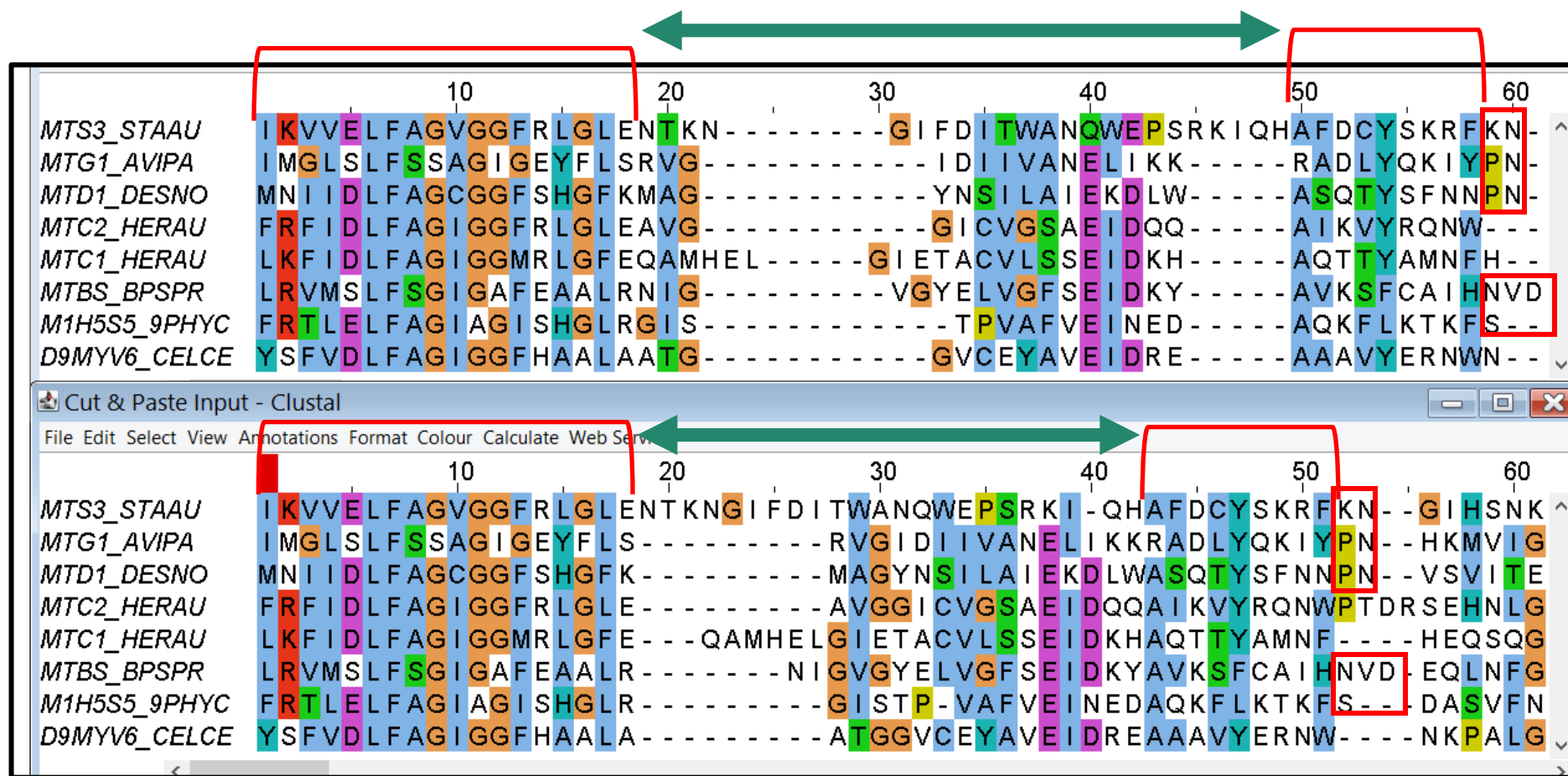
<https://www.ebi.ac.uk/msd-srv/ssm>

- ❖ Программа FATCAT множественного совмещения структур белков **сервис недоступен**
- ❖ Реализация FATCAT в PDB для парных сравнений работает
- ❖ Caretta

Практическое задание в классе

- ❖ Дано: два выравнивания pf00145_seed-reduced.fasta и pf00145_seed-tcoffee-reduced.fasta одних и тех же последовательностей
- ❖ Найти
 - ✓ блок одинаково выровненных колонок в двух выравниваниях
 - ✓ участок, на котором выравнивания полностью различны
 - ✓ результаты сохранить в гугл-форме

Участок несовпадающих выравниваний – между двумя совпадающими блоками





Спасибо за внимание!