

Ejercicios2_algoritmos

Tania Gonzalo Daniel Parra

WORKSHEET

Ejercicio 1 (4 puntos)

En este ejercicio probarás el algoritmo Needleman-Wunsch en una secuencia corta de partes de hemoglobina (código PDB 1AOW) y mioglobina 1 (código PDB 1AZI). Aquí alinearé la secuencia HGSAQVKGHG con la secuencia KTEAEMKASEDLKKHGT.

Las dos secuencias están dispuestas en una matriz en la Tabla 1. Las secuencias comienzan en la esquina superior derecha, y las penalizaciones por desfase inicial se enumeran en cada posición inicial de desfase. La penalización por desfase se considera -8. Las puntuaciones de similitud $S_{i,j}$ procedentes de la búsqueda de coincidencias proceden de la tabla BLOSUM40.

Algoritmos

Ejercicio 1

	H	G	S	A	Q	V	K	G	H	G	
Q	0	-8	-16	-24	-32	-40	-48	-56	-64	-72	-80
K	-8	-1	-9	-16	-24	-31	-39	-42	-50	-58	-66
T	-16	-9	-3	-7	-15	-23	-30	-38	-44	-52	-60
E	-24	-16	-11	-3	-8	-13	-21	-29	-37	-44	-52
A	-32	-24	-15	-10	-2	-6	-13	-21	-28	-36	-43
E	-40	-32	-23	-15	-6	4	-4	-12	-20	-28	-36
M	-48	-39	-31	-23	-14	-4	5	-3	-11	-19	-27
K	-56	-47	-39	-31	-22	-12	-3	11	3	-5	-13
A	-64	-55	-46	-38	-26	-20	-11	3	12	4	-4
S	-72	-63	-54	-41	-34	-25	-19	-5	4	11	4
E	-80	-71	-62	-49	-42	-32	-27	-13	-4	4	8
D	-88	-79	-70	-57	-50	-40	-35	-21	-12	-4	2
L	-96	-87	-78	-65	-58	-48	-38	-29	-20	-12	-6
K	-104	-95	-86	-73	-66	-56	-46	-32	-28	-20	-14
K	-112	-103	-94	-81	-74	-64	-54	-40	-34	-28	-22
H	-120	-111	-102	-89	-82	-72	-62	-48	-42	-21	-29
G	-128	-119	-110	-97	-88	-80	-70	-56	-40	-29	-13
T	-136	-127	-119	-109	-96	-88	-78	-64	-48	-37	-21

Obtenemos dos alineamientos con puntuación -21

* Alineamiento 1

-	-	H	G	S	-	-	A	-	Q	-	V	K	G	H	G	-
K	T	E	A	E	M	K	A	S	E	D	L	K	K	H	G	T

* Alineamiento 2

-	-	H	G	-	-	S	A	-	Q	-	V	K	G	H	G	-
K	T	E	A	E	M	K	A	S	E	D	L	K	K	H	G	T

Figura 1: Matriz y alineamientos obtenidos.

Como se puede observar, La puntuación obtenida es -21 y se han encontrado dos alineamientos globales posibles.

Ejercicio 2 (6 puntos)

Dado el conjunto de secuencias múltiples:

- S1: PPGVKSDCAS
- S2: PADGVKDCAS
- S3: PPDGKSDS
- S4: GADGKDCCS
- S5: GADGKDCAS

Utilice el popular método de alineación progresiva para alinear globalmente el conjunto anterior de secuencias. Genere el árbol guía por unión de vecinos. Compare su resultado (alineamiento) con el de Clustal-Omega.

Con el alineamiento final representa el logo. Para este proposito los caracteres nulos o gap son ignorados y no cuentan para el número de observaciones de una columna.

Para este ejercicio decidimos utilizar EMBOSS Needle, Pairwise Sequence Alignment (PSA) para realizar los distintos alineamientos dos a dos; y Clustal Omega (1.2.4), Multiple Sequence Alignment (MSA) para el alineamiento múltiple.

Alineamientos dos a dos

Los parámetros que utilizamos para realizar este proceso son:

OUTPUT FORMAT ⓘ	MATRIX ⓘ	GAP OPEN ⓘ
pair ▼	BLOSUM40 ▼	10 ▼
GAP EXTEND ⓘ	END GAP ⓘ	END GAP OPEN ⓘ
10.0 ▼	false ▼	10 ▼
END GAP EXTEND ⓘ		
0.5 ▼		

Figura 2: Parámetros para el alineamiento dos a dos.

Decidimos intentar mantener unos parámetros próximos a los utilizados en el ejercicio 1: una tabla BLOSUM40 y restar 10 por cada gap (lo más próximo a 8 que estaba disponible).

Obtuvimos los siguientes alineamientos con sus scores correspondientes, calculados dividiendo el número de coincidencias entre el número de residuos sin tener en cuenta los gaps:

S1	1 P-PGVKSDCAS	10
	.	
S2	1 PADGVK-DCAS	10

Figura 3: Alineamiento dos a dos de S1 vs S2, con 10 aa cada secuencia y 8 coincidencias, tiene un score 8/10.

S1	1 PPGVKSDCAS	10
	.. .	
S3	1 PPDGKSDS--	8

Figura 4: Alineamiento dos a dos de S1 vs S3, con 10 aa en S1 y 8 aa en S3; y 5 coincidencias, tiene un score 5/9.

S1	1 PPGVK-SDCAS	10
	.. .	
S4	1 --GADGKDCCS	9

Figura 5: Alineamiento dos a dos de S1 vs S4, con 10 aa en S1 y 9 aa en S4; y 4 coincidencias, tiene un score 8/19.

S1	1 PPGVK-SDCAS	10
	.. .	
S5	1 --GADGKDCAS	9

Figura 6: Alineamiento dos a dos de S1 vs S5, con 10 aa en S1 y 9 aa en S5; y 5 coincidencias, tiene un score 10/19.

S2	1 PADGVKDCAS	10
	
S3	1 PPDGKSDS--	8

Figura 7: Alineamiento dos a dos de S2 vs S3, con 10 aa en S2 y 8 aa en S3; y 4 coincidencias, tiene un score 4/9.

S2	1 PADGVKDCAS	10
	. .	
S4	1 GADG-KDCCS	9

Figura 8: Alineamiento dos a dos de S2 vs S4, con 10 aa en S2 y 9 aa en S4; y 7 coincidencias, tiene un score 14/19.

S2	1 PADGVKDCAS	10
	.	
S5	1 GADG-KDCAS	9

Figura 9: Alineamiento dos a dos de S2 vs S5, con 10 aa en S2 y 9 aa en S5; y 8 coincidencias, tiene un score 16/19.

S3	1 PPDGKSDS--	8
	.. .	
S4	1 GADGK-DCCS	9

Figura 10: Alineamiento dos a dos de S3 vs S4, con 8 aa en S3 y 9 aa en S4; y 4 coincidencias, tiene un score 8/17.

S3	1 PPDGKSDS-	8
:	
S5	1 GADGKDCAS	9

Figura 11: Alineamiento dos a dos de S3 vs S5, con 8 aa en S3 y 9 aa en S5; y 3 coincidencias, tiene un score 6/17.

S4	1	GADGKDCCS	9
		.	
S5	1	GADGKDCAS	9

Figura 12: Alineamiento dos a dos de S4 vs S5, con 9 aa cada secuencia y 8 coincidencias, tiene un score 8/9.

Y construimos la matriz de distancias, para la cual hacemos $1 - \text{el score de similitud del alineaiento}$:

Ejercicio 2

- Matriz de distancias tras los alineamientos Pairwise

S1	0				
S2	0,2	0			
S3	0,44	0,56	0		
S4	0,58	0,26	0,53	0	
S5	0,47	0,16	0,65	0,11	0
	S1	S2	S3	S4	S5

$$S2-S5 = 1 - \frac{16}{19} = 0,16$$

$$S3-S5 = 1 - \frac{6}{17} = 0,65$$

$$\left[\begin{matrix} S4 \\ S5 \end{matrix} \right] A$$

$$A-S1 = \frac{0,58 + 0,47}{2} = 0,525$$

$$A-S2 = \frac{0,26 + 0,16}{2} = 0,21$$

$$A-S3 = \frac{0,53 + 0,65}{2} = 0,59$$

B	0		
S3	0,5	0	
A	0,3675	0,59	0
	B	S3	A

$$\left[\begin{matrix} A \\ B \end{matrix} \right] S3$$

$$S1-S2 = 1 - \frac{8}{10} = 0,2$$

$$S1-S3 = 1 - \frac{5}{9} = 0,44$$

$$S1-S4 = 1 - \frac{8}{14} = 0,58$$

$$S1-S5 = 1 - \frac{10}{19} = 0,47$$

$$S2-S3 = 1 - \frac{4}{9} = 0,56$$

$$S2-S4 = 1 - \frac{14}{19} = 0,26$$

$$S3-S4 = 1 - \frac{8}{17} = 0,53$$

$$S4-S5 = 1 - \frac{8}{9} = 0,11$$

S1	0			
S2	0,2	0		
S3	0,44	0,56	0	
A	0,525	0,21	0,59	0
	S1	S2	S3	A

$$\left[\begin{matrix} S1 \\ S2 \end{matrix} \right] B$$

$$B-S3 = \frac{0,44 + 0,56}{2} = 0,5$$

$$B-A = \frac{0,525 + 0,21}{2} = 0,3675$$

Figura 13: Construcción de matrices de distancias.

Finalmente, construimos el guide tree:

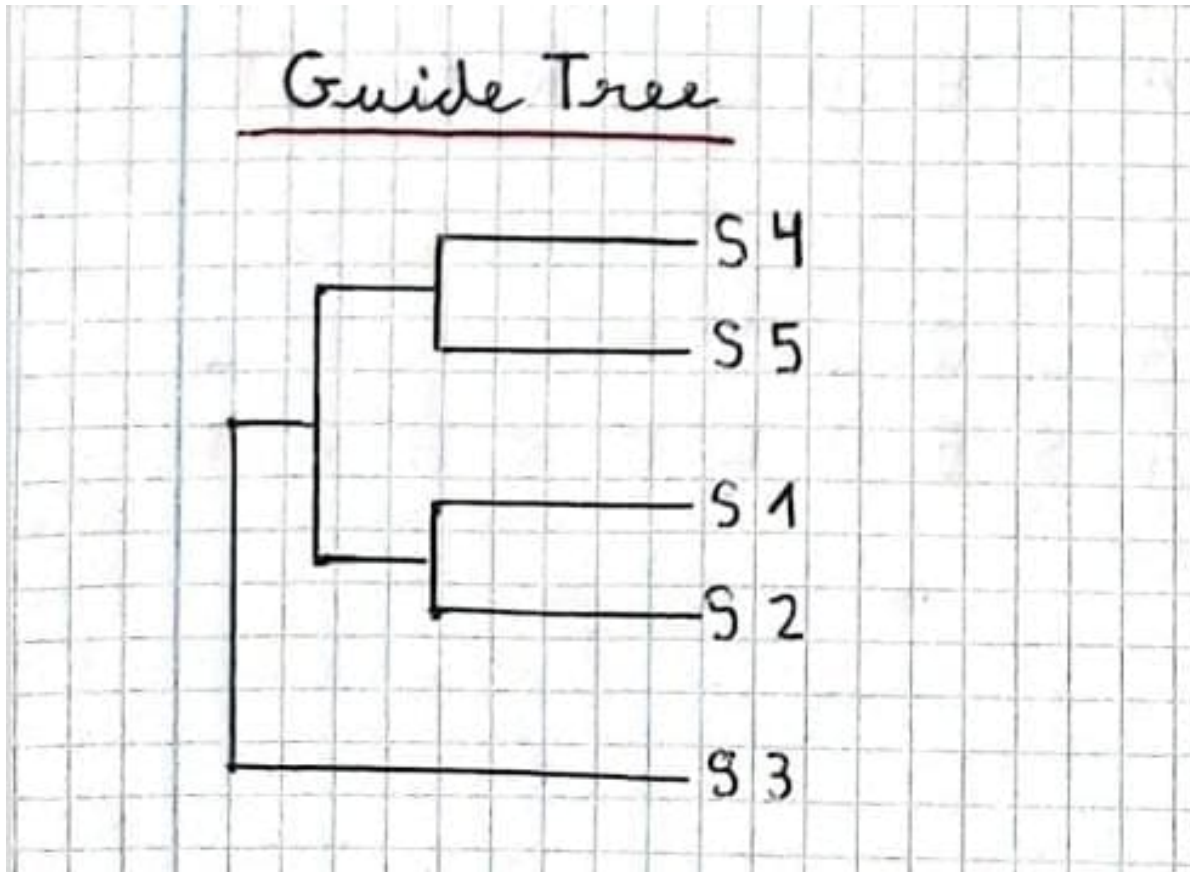


Figura 14: Guide Tree obtenido partir de alineamientos 2 a 2 con parámetros similares al ejercicio 1.

Y el alineamiento ¿forward?: ¿Lo hacemos también?

Con parámetros por defecto

También, debido a que Clustal Omega utiliza una tabla BLOSUM62, decidimos repetir este proceso pero dejando los parámetros que vienen por defecto:

OUTPUT FORMAT ⓘ pair ▼	MATRIX ⓘ BLOSUM62 ▼	GAP OPEN ⓘ 10 ▼
GAP EXTEND ⓘ 0.5 ▼	END GAP ⓘ false ▼	END GAP OPEN ⓘ 10 ▼
END GAP EXTEND ⓘ 0.5 ▼		

Figura 15: Parámetros para el alineamiento dos a dos por defecto.

Los alineamientos y scores obtenidos con estos parámetros fueron:

S1	1 -PPGVKSDCAS	10
	..	
S2	1 PADGVK-DCAS	10

Figura 16: Alineamiento dos a dos con parámetros por defecto, de S1 vs S2, con 10 aa cada secuencia y 7 coincidencias, tiene un score 7/10.

S1	1 PPGVKSDCAS	10
	.. .	
S3	1 PPDGKSDS--	8

Figura 17: Alineamiento dos a dos con parámetros por defecto, de S1 vs S3, con 10 aa en S1 y 8 aa en S3; y 5 coincidencias, tiene un score 5/9.

S1	1 PPGVK-SDCAS	10
	
S4	1 --GADGKDCCS	9

Figura 18: Alineamiento dos a dos con parámetros por defecto, de S1 vs S4, con 10 aa en S1 y 9 aa en S4; y 4 coincidencias, tiene un score 8/19.

S1	1 PPGVK-SDCAS	10
	
S5	1 --GADGKDCAS	9

Figura 19: Alineamiento dos a dos con parámetros por defecto, de S1 vs S5, con 10 aa en S1 y 9 aa en S5; y 5 coincidencias, tiene un score 10/19.

S2	1 PADGVKDCAS	10
	
S3	1 PPDGKSDS--	8

Figura 20: Alineamiento dos a dos con parámetros por defecto, de S2 vs S3, con 10 aa en S2 y 8 aa en S3; y 4 coincidencias, tiene un score 4/9.

S2	1 PADGVKDCAS	10
	. .	
S4	1 GADG-KDCCS	9

Figura 21: Alineamiento dos a dos con parámetros por defecto, de S2 vs S4, con 10 aa en S2 y 9 aa en S4; y 7 coincidencias, tiene un score 14/19.

S2	1 PADGVKDCAS	10
	.	
S5	1 GADG-KDCAS	9

Figura 22: Alineamiento dos a dos con parámetros por defecto, de S2 vs S5, con 10 aa en S2 y 9 aa en S5; y 8 coincidencias, tiene un score 16/19.

S3	1 PPDGKSDS-	8
	
S4	1 GADGKDCCS	9

Figura 23: Alineamiento dos a dos con parámetros por defecto, de S3 vs S4, con 8 aa en S3 y 9 aa en S4; y 3 coincidencias, tiene un score 6/17.

S3	1 PPDGKSDS-	8
:	
S5	1 GADGKDCAS	9

Figura 24: Alineamiento dos a dos con parámetros por defecto, de S3 vs S5, con 8 aa en S3 y 9 aa en S5; y 3 coincidencias, tiene un score 6/17.

S4	1 GADGKDCCS	9
	.	
S5	1 GADGKDCAS	9

Figura 25: Alineamiento dos a dos con parámetros por defecto, de S4 vs S5, con 9 aa cada secuencia y 8 coincidencias, tiene un score 8/9.

La matriz de distancias correspondiente a estos alineamientos sería:

Finalmente, construimos el guide tree:

Y el alineamiento ¿forward?: ¿Lo hacemos también?

Alineamiento múltiple

Los parámetros utilizados para realizar este proceso fueron:

OUTPUT FORMAT ⓘ		DEALIGN INPUT ⓘ	
ClustalW with character counts ▼		no ▼	
MBED-LIKE CLUSTERING GUIDE-TREE ⓘ	MBED-LIKE CLUSTERING ITERATION ⓘ	COMBINED ITERATIONS ⓘ	
no ▼	yes ▼	default(0) ▼	
MAX GUIDE TREE ⓘ	MAX HMM ITERATIONS ⓘ	ORDER ⓘ	DISTANCE MATRIX ⓘ
default ▼	default ▼	aligned ▼	yes ▼
OUTPUT GUIDE TREE ⓘ			
yes ▼			

Figura 26: Parámetros para el alineamiento múltiple.

Obtuvimos el siguiente guide tree:

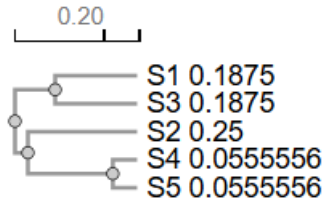


Figura 27: Guide tree del alineamiento múltiple por Clustal Omega.

Y este alineamiento múltiple:

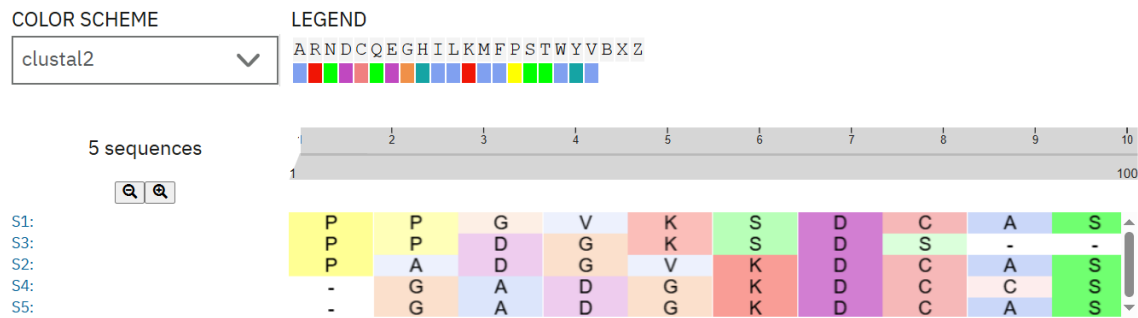


Figura 28: Alineamiento múltiple por Clustal Omega.

Comparación y logo

Observamos que los guide tree obtenidos por ambos procesos son mmmmmmmmmmm

El logo que se obtiene utilizando el alineamiento múltiple mediante Clustal Omega es:

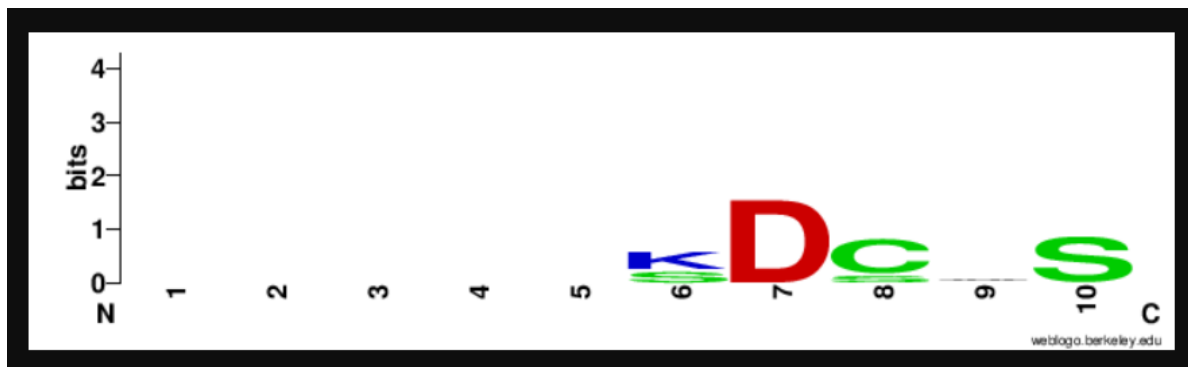


Figura 29: Logo del MSA.

Utilizamos la herramienta proporcionada en este enlace: <https://weblogo.berkeley.edu/logo.cgi>