

1. پژوهشگری قصد دارد دو توالی آمینو اسیدی زیر را از طریق الگوریتم Align، Needleman–Wunsch کند. او جدول زیر را کشیده است و می خواسته از ماتریس PAM250 استفاده کند. جدول را کامل کرده و هم ردیفی یا هم ردیفی های بهینه را بنویسید.

(15 نمره مثبت برای جدول درست و 15 نمره برای همردیفی درست)

**نمره دهی: برای جدول هر فلش یا عدد غلط 0.25 نمره کسر می گردد. حداقل نمره برابر صفر می باشد.**

**نمره همردیفی تنها به همردیفی کاملاً درست داده خواهد شد.**

توالی 1: AWNNTCS

توالی 2: WNSTSS

Gap Open: 1

Gap Extend: 0.5

End Gap Open: 0.25

End Gap Extend: 0.25

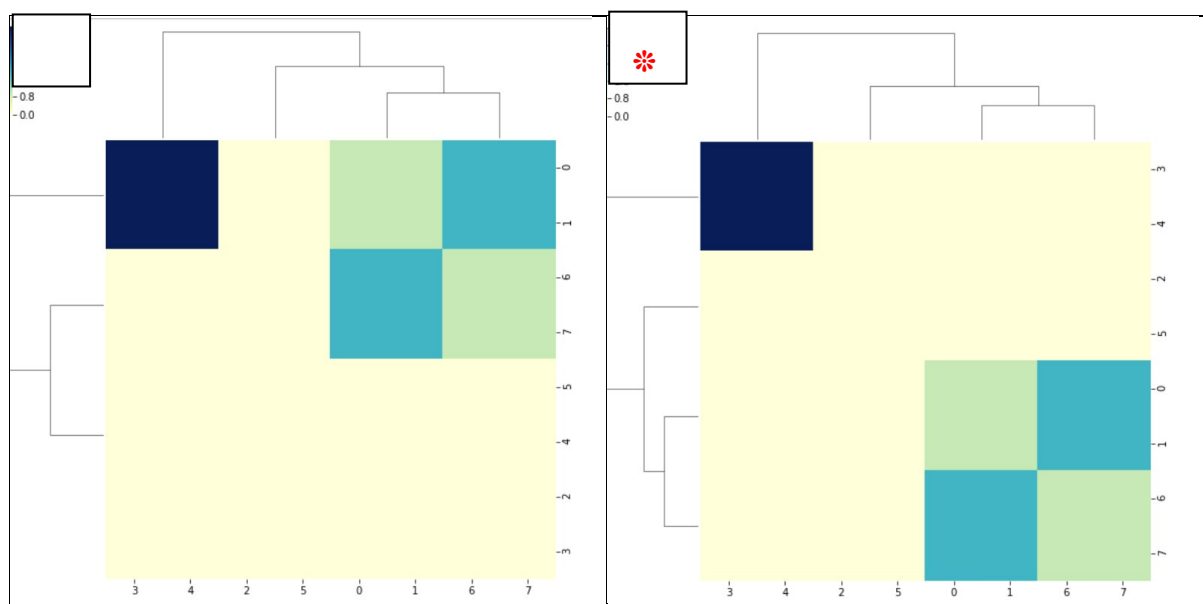
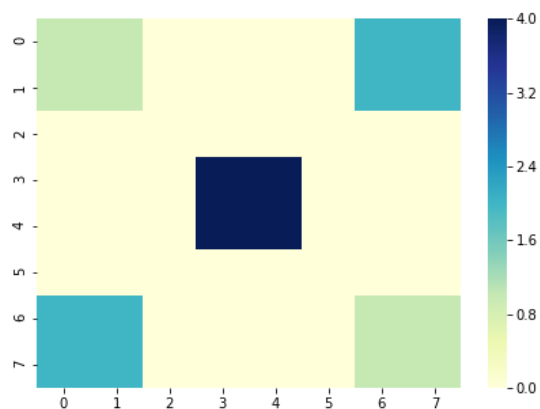
		A	W	N	N	T	C	S
	O							
W								
N								
S								
T								
S								
S								

	توالی 1
	توالی 2

		A	W	N	N	T	C	S
	0	-0.25	-0.5	-0.75	-1	-1.25	-1.5	-1.75
W	-0.25	-1.25	16.75	15.75	15.25	14.75	14.25	13.75
N	-0.5	-0.25	15.75	18.75	17.75	17.25	16.75	16.25
S	-0.75	0.5	15.25	17.75	19.75	18.75	17.25	18.75
T	-1	0.25	14.75	17.25	18.75	22.75	21.75	21.25
S	-1.25	0	14.25	16.75	18.25	21.75	22.75	23.75
S	-1.5	-0.25	13.75	16.25	17.75	21.25	21.75	24.75

Final Alignment = A W N N T C S  
- W N S T S S

2. در نقشه حرارتی (HeatMap) زیر حالت بدون اعمال dendrogram یک آرایه دو بعدی را مشاهده می کنید. حال پس از اعمال dendrogram کدام یک از نمودار ها ایجاد خواهند شد ؟  
(10 نمره مثبت، 3 نمره منفی)





3. جدول زیر تعداد خوانش ژن های 1 تا 4 را در گونه های A تا G در یک بررسی RNAseq نشان می دهد.

Read Count	G1	G2	G3	G4
A	0	14	25	19
B	4	4	13	24
C	16	5	2	21
D	3	7	14	17
E	1	2	1	27
F	3	1	9	20
G	0	3	2	30

از این دیتا برای Hierarchical Clustering این گونه ها استفاده کنید. 50 نمره

Clustering Method: Single-linkage

Distance Measurement Method: Euclidean

جدول اولین Distance Matrix برای شما رسم شده است. در صورت نیاز به بروزرسانی، هر مرحله را در پاسخ خود یادداشت کنید. (در صورت نیاز جدول های مراحل بعدی را رسم و تکمیل کنید).

Dendrogram نهایی را به همراه فاصله ها، رسم کنید.

تکمیل Distance Matrix : هر مورد 1 نمره (مجموعاً 21 نمره)

اعداد با خطای  $\pm 0.001$  پذیرفته می شوند.

Distance	A	B	C	D	E	F	G
A							
B	16.882						
C	29.496	16.583					
D	13.528	7.746	18.248				
E	28.018	12.884	16.462	17.263			
F	20.857	6.481	15.330	8.367	10.863		
G	27.767	13.191	18.466	18.385	3.464	12.728	

در این متود (Single-linkage)، نیازی به Distance Matrix دیگری نیست، تنها کافی است که زوج گونه ها را بر اساس فاصله های گونه های زوج، به ترتیب از کمترین تا بیشترین فاصله مرتب کنیم و بر اساس آن کلاسترینگ را انجام دهیم.

$E, G - B, F - B, D - D, F - E, F - G, F - B, E - B, G - A, D - C, F - E, C - B, C - A, B - E, D - C, D - D, G - C, G - A, F - A, G - A, E - A, C$

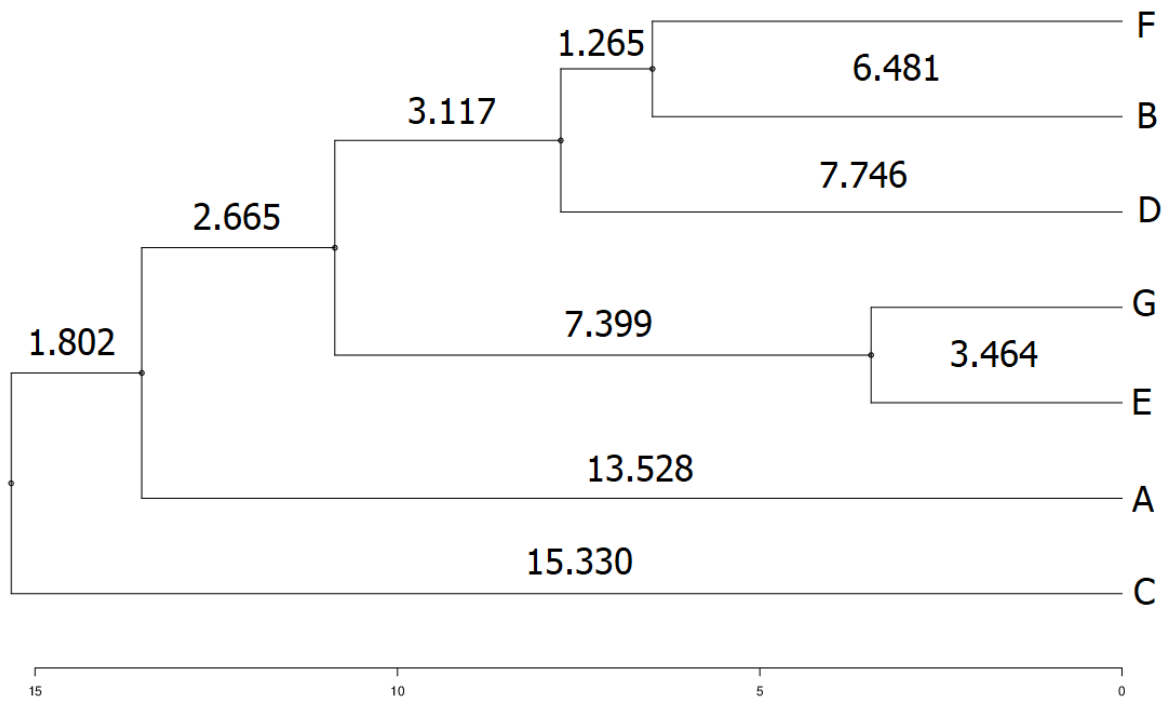
1. کلاستر کردن E و G فاصله 3.464
2. کلاستر کردن B و F فاصله 6.481
3. کلاستر کردن D با کلاستر (B,F:6.481) فاصله 7.746
4. تکراری ( بی اثر)
5. کلاستر کردن کلاستر (E,G:3.464) با کلاستر ((B,F:6.481),D:7.746) فاصله 10.863
6. تکراری ( بی اثر)
7. تکراری ( بی اثر)
8. تکراری ( بی اثر)
9. کلاستر کردن A با کلاستر ((B,F:6.481),D:7.746),(E,G:3.464):10.863) فاصله 13.528
10. کلاستر کردن C با کلاستر (((B,F:6.481),D:7.746),(E,G:3.464):10.863),A:13.528) فاصله 15.330
11. تکراری ( بی اثر)
12. تکراری ( بی اثر)
13. تکراری ( بی اثر)
14. تکراری ( بی اثر)
15. تکراری ( بی اثر)
16. تکراری ( بی اثر)
17. تکراری ( بی اثر)
18. تکراری ( بی اثر)
19. تکراری ( بی اثر)
20. تکراری ( بی اثر)
21. تکراری ( بی اثر)

نتیجه نهایی کلاسترینگ : (((((B,F:6.481),D:7.746),(E,G:3.464):10.863),A:13.528),C:15.330)

در صورت رسم و تکمیل Distance Matrix های اضافی، نمره ای به آن ها تعلق نمی گیرد.

## Dendrogram نهایی :

توپولوژی 19 نمره فاصله ها 10 نمره (مجموعاً 29 نمره)



## نحوه تصحیح Dendrogram :

این بخش به صورت By Data تصحیح خواهد شد. Dendrogram ای بر اساس مقادیر پاسخ شما در Distance Matrix توسط برنامه کامپیوتری طراحی شده و با پارامترهای مشخص شده در صورت سوال، ساخته شده و با پاسخ شما به این بخش مقایسه خواهد شد.

توپولوژی پاسخ کاملاً صحیح : 100 % نمره توپولوژی

یک اشتباه در توپولوژی پاسخ : 50 % نمره توپولوژی

دو اشتباه در توپولوژی پاسخ : 30 % نمره توپولوژی

بیش از دو اشتباه در توپولوژی پاسخ : 0 % نمره توپولوژی

هر مورد از فاصله ها ؟ نمره (مجموعاً 10 نمره)

بارم هر مورد از فاصله ها بر اساس تعداد آن ها در درخت ساخته شده برای هر فرد تنظیم خواهد شد (گرد تا 2 رقم اعشار).

اعداد با خطای  $\pm 0.003$  پذیرفته می شوند.

4. با توجه به ارتباط PDF (تابع چگالی احتمال) و CDF (تابع تجمعی احتمال) صحت گزاره های زیر را بررسی کنید. (هر مورد 5 نمره و نمره منفی برابر)

<b>T</b>	الف) شکل CDF تابع چگالی احتمال مرتبط با توزیع نرمال سیگمویید خواهد بود.
<b>F</b>	ب) با افزایش انحراف معیار توزیع، حداکثر مقدار شیب نمودار CDF افزایش خواهد یافت.
<b>F</b>	ج) مشتق دوم تابع PDF در مقدار امید ریاضی توزیع پواسون مثبت خواهد بود.
<b>T</b>	د) در بازه اعداد حقیقی، مشتق تابع PDF حداقل یکبار از محور X ها عبور خواهد کرد.
<b>F</b>	ه) مقدار مطلق تابع CDF می تواند بیشتر از عدد یک شود.