Practicals – 6

-BS19B032

-R. Vasantha Kumar

1)   Using the AL2CO server, I found the positional conservation scores for the aligned sequences. I have attached the output files in the submission along with the set1 and set2 sequences.

2)   For set1:

Highest score:

Position	Residue	Score
3	L	0.00
4	S	0.00
7	D	0.00
8	K	0.00
17	K	0.00
26	G	0.00
30	L	0.00
32	R	0.00
38	P	0.00
40	T	0.00

Lowest score:

Position	Residue	Score
117	K	-1.846
113	T	-1.720
70	L	-1.673
73	S	-1.594
84	E	-1.414
133	G	-1.414
36	V	-1.367

35	V	-1.295
9	A	-1.169
6	K	-1.160

For set2:

Highest score:

Position	Residue	Score
197	L	0.00
198	G	0.00
199	H	0.00
200	S	0.00
201	E	0.00
202	R	0.00
203	R	0.00
210	E	0.00
218	K	0.00
230	V	0.00

Lowest score:

Position	Residue	Score
133	T	-1.778
127	E	-1.738
164	Q	-1.631
247	A	-1.567
308	D	-1.537
149	L	-1.525
121	L	-1.525
123	V	-1.484
136	H	-1.484
190	Y	-1.470

3)   The program to find the conservation score is attached in submission.

4) Using MSA in clustal omega, MAFFT and MUSCLE, I have found the conservation scores using my code. The result files are attached in the submission.

For set1:

All are similar

For set2:

From 1 to 8 all are similar.

In positions 9 to 24 the MAFFT values are different from clustal omega and MUSCLE.

From 25 to 37 all are similar.

In positions 38 to 45 clustal omega values are different from other two.

At positions 46 to 52 each three of them differs.

From 53 to 115, there is a high mix-matches between three of them.

From 116, there is an interesting trend that, the values of MAFFT and MUSCLE are equal, but these values are equal to the succeeding values of clustal omega.

5) ***For set1:***

Position 9:

$$A=2 \quad freq(A)=2/11=0.18$$

$$S=2 \quad freq(S)=2/11=0.18$$

$$T=6 \quad freq(T)=6/11=0.54$$

$$G=1 \quad freq(G)=1/11=0.09$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$

$$Score = -1.16$$

Position 11:

$$I=3 \quad freq(I)=3/11=0.27$$
$$V=8 \quad freq(V)=8/11=0.72$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$
$$Score = -0.59$$

Position 20:

$$A=2 \quad freq(A)=2/11=0.18$$
$$S=1 \quad freq(S)=1/11=0.09$$
$$K=1 \quad freq(K)=1/11=0.09$$
$$G=7 \quad freq(G)=7/11=0.63$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$
$$Score = -1.03$$

Position 22:

$$A=9 \quad freq(A)=9/11=0.81$$
$$G=2 \quad freq(G)=2/11=0.18$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$
$$Score = -0.47$$

Position 30:

$$L=11 \quad freq(L)=11/11=0$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$
$$Score = 0.00$$

For set2:

Position 9:

$$L=1 \quad freq(L)=1/14=0.07$$
$$'-'= \quad freq(-)=13/14=0.92$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$

$$Score = -0.26$$

Position 11:

$$R=1 \quad freq(R)=1/14=0.07$$

$$'-'= \quad freq(-)=13/14=0.92$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$

$$Score = -0.26$$

Position 20:

$$V=1 \quad freq(V)=1/14=0.07$$

$$'-'= \quad freq(-)=13/14=0.92$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$

$$Score = -0.26$$

Position 22:

$$L=1 \quad freq(L)=1/14=0.07$$

$$'-'= \quad freq(-)=13/14=0.92$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$

$$Score = -0.26$$

Position 30:

$$R=1 \quad freq(R)=1/14=0.07$$

$$'-'= \quad freq(-)=13/14=0.92$$

Now, using formula,

$$Score = sigma(freq(a).ln(freq(a)))$$

$$Score = -0.26$$

6) Using the consurf server, I obtained the conservation score for 1BTM – A chain. I have also attached the result file(Q6) in submission.

The conservation scores of the result is normalized. It also contains colour scale for conservation score.