

Q1) On the next page is a multiple alignment of 35 human sequences. The sequences have been aligned around a donor splice. That site is indicated as the boundary between the 'Dark blue' and 'Dark red' colours.

Calculate the frequencies for positions shown in table below.

position	6	7	8	9	0	1	2	3	4	5
Counts A										
Counts T										
Counts C										
Counts G										
Frequency A										
Frequency T										
Frequency C										
Frequency G										

Q2) Calculate the Entropy (eq 1) and Information Content (eq 2) using the formula below

Eq.1 $H = -\sum_a p_a \log_2(p_a)$

Eq.2 $I = \sum_a p_a \log_2(p_a) + 2.0$

position	6	7	8	9	0	1	2	3	4	5
Entropy										
Information content										

Q3) Where does the constant 2.0 come from in Eq.2 ?

Q4) Draw an approximate Logo Plot by hand on the White board

Q5) [Optional - you can do this after we have done the computer exercise]

Submit the multiple alignment to the LogoLab server: <https://wernersson.dk/logolab/>

Make both the Logo plot and a frequency plot

Explain what you see on the two plots.

-----Exon><intron---

01234567890123456789

TATCACAATGGTAGGTAAC
TCAACCAGGAGTAAGTCTTG
GTTGCACCCTGTAAGTCTCA
TATCACAATGGTAGGTAAC
TCAACCAGGAGTAAGTCTTG
CTTGCGAGAGGTGTGACATG
GCTCTACTCGTAAGGTGAC
GCCTGGAGAGGTAATGACCC
CAAAACCATTGTGAGTAATC
GCCAGAGCAGGTAAAATATC
GAACAGTCAGGTCTGTTGCT
GAAGGCCCAGGTGAGCATAA
TCCTCTACAGGTGGGTACAT
GGCGTCCCGCGTAAGTATGG
CCTCGTGCAGGTAAGATTAA
TGCATGACAGGTGAGTGTTA
GAAATGTACAGTAAGTCTCT
GGTTCTCTGGTAAGTAGAG
AAATGTACAGGTGAGTACTG
ACCTCGCTTGGTACGTGGGA
AATCAGACAGGTATAGAAAC
AGGACAGAAGGTAATTTTCT
AACTATTTGGGTAGGTAGCA
AAACTTGAAGGTATGTTGTT
CTGGGATAAGGTAAAAGTAT
TTGCACCCAGGTAGTGGAT
ACTTCAATCGGTATGTTTTC
ACAGAGAAAAGTAAATTCCT
AATGGGAAAGGTAACAACAA
CATGCTACAGGTAGGTGAAT
GGCTAGGATGTTGAGGGCGC
CGACGCGGGCGTGAGAGGCG
CATTGAGAATGTGAGTTATT
AACAGAGCAGGTACTTGTAT
TGAACCAAAGGTGAAGACAT