

Making Sequence logos

Q1) Below 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.

```
-----Exon|intron-----
01234567890123456789
tatcacaATGGTAGGTAAGTACT
TCAACCAGGAGTAAGTCTTG
GTTGCACCCTGTAAGTCTCA
tatcacaATGGTAGGTAAGTACT
TCAACCAGGAGTAAGTCTTG
CTTGCAGAGGTGTGACATG
GCTCTACTCGGTAAGGTGAC
GCCTGGAGAGGTAATGACCC
CAAACCATTGTGAGTAATC
GCCAGAGCAGGTAATAATC
GAACAGTCAGGTCTGTTGCT
GAAGGCCAGGTGAGCATAA
TCCTCTACAGGTGGGTACAT
GGCGTCCCGCTAAGTATGG
CCTCGTGCAGGTAAGATTAA
TGCATGACAGGTGAGTGTTA
GAAATGTACAGTAAGTCTCT
GGTTCTCTGGGTAAGTAGAG
AAATGTACAGGTGAGTACTG
ACCTCGCTTGGTACGTGGGA
AATCAGACAGGTATAGAAAC
AGGACAGAAGGTAATTTTCT
AACTATTTGGGTAGGTAGCA
AAACTTGAAGGTATGTTGTT
CTGGGATAAGGTAAGTAT
TTGCACCCAGGTTAGTGGAT
ACTTCAATCGGTATGTTTTT
ACAGAGAAAAGTAAATTCCT
AATGGGAAAGGTAACAACAA
CATGCTACAGGTAGGTGAAT
ggctaggATGGTGGGGCGC
CGACGCGGGCGTGAGAGGCG
CATTGAGAATGTGAGTTATT
AACAGAGCAGGTAAGTAT
TGAACCAAAGGTAAGACAT
```

Calculate the frequencies for positions 6-5. You have each been assigned one column on the upper right corner of the handout.

Position	6	7	8	9	0	1	2	3	4	5
Counts A										
Counts T										
Counts C										
Counts G										
P(A)										
P(T)										
% C										
% G										

Q2) Calculate the Entropy and Information Content using the formula below

$$\text{Eq.1} \quad H(p) = -\sum_a p_a \log_2(p_a) = -\frac{1}{\log(2)} \sum_a p_a \log(p_a)$$

where \log_2 is the logarithm with base 2, and \log is the logarithm with base 10 (or any base for that sake)

$$\text{Eq.2} \quad I = 2.0 - H(p)$$

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

If you have internet-access

Q5) Submit the multiple alignment to the WebLogo server <http://weblogo.berkeley.edu/>

Make both the Logo plot and a frequency plot
Explain what you see on the two plots.