Assignment – II
(k-mer Analysis, Dotplots)

Deadline: 30th March

1. Simulate observations having the binomial distribution with p = 0.25 and n = 1000. What is the probability of observing at least 240 A's in such a sequence? [Hint: Obtain 10,000 simulations and compute the number of A's in each run]. Compare your result with the normal approximation to binomial distribution.

2. Suppose X has a binomial distribution with p = 0.3 and n = 10. Compute P(X=0), P(X=2), E(X) and VarX.

3. Briefly discuss the applications of k-mer analysis.

4. Show a dotplot of the following two sequences and give the conserved region:
(Make a n x m table and put '.' or 'x' for match)

GGCTGCAACTAGCTC
GGGTAAGCTTGC

5. Obtain the self-dotplot of the following sequence to identify repeat region:
TGGCACACTCACACCACACAGACAGTTA

6. Find self-complementary regions in the following RNA sequence:
AUGUGGCAUGCCAGG