### List of Figures

1.1 Central Dogma .................................................. 2
1.2 Prokaryotic Genome Structure ................................. 4
1.3 Multiple Sequence Alignment. (Black Arrow points to the highly conserved position) ............................................. 6
1.4 Sequence Logo. Source: http://www.lecb.ncifcrf.gov/toms .......... 10
1.5 Entropy Function For a Bernoulli Variable ..................... 13

3.1 Redundancy plot for HincII sites. a) 61 sites b) 17 sies c) 6 sites ... 36
3.2 Redundancy plots a) λ cl/Cro b) lexA c) T7 promoter ............. 37
3.3 Redundancy plot for Ribosome binding sites for different α values ... 38
3.4 Redundancy plot with error bars for correction for ribosome binding sites for α = 0.99 .............................................. 42
3.5 Redundancy plot with error bars for correction for ribosome binding sites for α = 0.5 ............................................... 43

4.1 Information content of E.coli HincII site .......................... 52
4.2 Distribution of entropy deviation values .......................... 53

6.1 Plots of (a) \( d_k^n \), (b) \( \tilde{d}_k^n \) for E.Coli promoter Data (471x101) for \( n = 6 \) .... 72
6.2 Plot of \( \tilde{d}_k^n \) randomly generated data of size 471x101 for \( n = 6 \) .... 73
6.3 Plots of \( \tilde{d}_k^n \) for E.Coli promoter Data (471x101) in the range \( n = 7 \) to 12 .... 74
6.4 Plots of \( \tilde{d}_k^n \) for Ribosome binding sites Data (1055x40), in the range \( n = 7 \) to 12 .... 74
6.5 Plots of \( \tilde{d}_k^n \) for Human promoter Data (4737x151), in the range \( n = 7 \) to 20 .... 75

x
6.6 Plot of the least value i.e., $\min_k \{d_k^p\}$ against word size for the E.Coli promoter Data ......................................................... 75

7.1 Inter-nucleotide distance distribution Plot of E.coli:Base "A" ........ 85
7.2 Truncated exponential fit for the distance for "A" and "T" ............ 89
7.3 Truncated exponential fit for the distance for "G" and "C" ............ 90
7.4 Fit of Lognormal distribution for "A" and "T" .......................... 91
7.5 Fit of Lognormal distribution for "G" and "C" .......................... 92
7.6 Distance Plot for the base "A" and "T" ................................. 93
7.7 Distance Plot for the base "G" and "C" ................................. 94
7.8 Inter-nucleotide distribution plot for E.coli and Plasmodium falciparum: Base "A" ....................................................... 96
7.9 Fit of probability distributions to distance data: Plasmodium falciparum - Base "A" ....................................................... 96
7.10 Fit of probability distributions to distance data: Plasmodium falciparum - Base "G" ....................................................... 97