

Contents

Page No.

1. Introduction	1
2. DNA	3
2.1 Transcription	6
2.2 Translation	7
2.3 Protein Structure	8
2.4 Gene Working	10
2.5 Genetic Code	11
3. Materials and Methods	
3.1 Markov Process	13
3.1.1 Markov Chain	14
3.1.2 Markov Model	16
3.2 Hidden Markov Model	17
3.3 Building an HMM	19
4. Nucleotide Sequence Samples	20
4.1 Heterogeneity	22
4.1.1 Genomic Signature	23
4.1.2 Mechanisms of Genomic Signature	23
4.2 Codon Signature	24
4.2.1 Codon Biases in Bacterial Genomes	24
4.2.2 Comparison of Codon Usage Between Different Gene Classes	25
5. Calculation of Relative Abundance	
5.1 Values of di-nucleotides at 2nd and 3rd Codon positions in top 100 high expression genes as well as whole genome	25
5.2 Calculation of relative abundance values of di-nucleotides in intergenic region	26
6. Results	27
7. Conclusion	34
8. References	35

List of figures

Page No

1.1 DNA sequence	1
2.1 DNA Double helix	4
2.2 Protein Synthesis	6
2.3 Introns and Exons in a gene	11
2.4 Genetic Code	12
3.1 Markov Process	15
3.2 Model for building an HMM	19
6.1 Relative Abundance Values at positions 2,3 in Top 100 HE genes	27
6.2 Relative Abundance Values at positions 2,3 in whole genome	28
6.3 Relative Abundance Values in Intergenic region	28
6.4 Relative Abundance Values at positions 3,1 in Top 100 HE genes	29
6.5 Relative Abundance Values at positions 3,1 in whole genome	29
6.6 Probability Values for Random Sequence	30
6.7 Preferred/Avoided Regions of different di-nucleotides	31