

List of Figures

1.1	Coding and non-coding regions in a single DNA-strand	2
2.1	Hidden Markov Model of order 0	9
3.1	Dinucleotides in Simulated sequences	19
3.2	Dinucleotides in Intergenic regions	20
3.3	Dinucleotides in Coding sequences	22
3.4	Dinucleotides in Coding sequences without codon bias	23

List of Tables

1.1	Genetic Code Table	6
2.1	List of Bacterial Genomes studied	8

Contents

1	Introduction	1
1.1	DNA Sequence	1
1.2	Genome signature	2
1.3	Genetic Code and Codon Usage	4
1.4	Objective	5
1.5	Motivation behind the objective	5
2	Materials and Methods	7
2.1	Nucleotide sequences of the inter-genic and the coding sequences of bacterial genomes	7
2.2	Simulated sequences of bacterial genomes	8
2.3	Coding sequences without codon bias	9
2.4	Calculation of RDFs at the second codon position of coding sequences and coding sequences without bias	9
2.5	Calculation of RDFs in intergenic regions and simulated sequences	10
2.6	Principal Coordinate Analysis	11
3	Experimental Results and Analysis	18

3.1	Dinucleotides in simulated DNA sequences are not under any constraint	18
3.2	Intergenic regions are constrained w.r.to dinucleotide abundance in bacterial genomes	20
3.3	Dinucleotides in the coding sequences are more constrained than in the intergenic regions	21
3.4	Dinucleotides in the coding sequences without any codon bias are less constrained than the coding sequences	23
4	Conclusion and Future Work	24
4.1	Conclusion	24
4.2	Future Work	24
	References	25