

Contents

List of figures	ix
List of tables	x
List of abbreviations	xi

1. Introduction.....	1
1.1 Central Dogma of Molecular Biology.....	1
1.2 DNA.....	2
1.3 RNA.....	2
1.3.1 Transcription and Translation.....	3
1.4 Genetic Code Table.....	4
1.5 Factor affecting Codon Usage in Different Organism.....	5
1.5.1 Mutation.....	5
1.5.2 Selection Mechanism.....	5
2. Measures of Codon usage bias.....	7
2.1 Frequency of Optimal codon.....	7
2.2 RSCU.....	7
2.3 CAI.....	8
2.4 Effective Number of Codon.....	9
2.5 MILC.....	9
3. Gene Expression.....	10
3.1 Measuring Actual Gene Expression.....	10
3.1.1 Microarray Experiment (Transcription Level Expression).....	10
3.1.2 Protein Level Expression.....	11
3.1.3 Method proposed in Research Literature for Predicting Gene Expression	11
3.1.3.1 CAI.....	11
4. Comparison of CAI with Gene Expression Different Organism.....	13
4.1 Method Used (ROC).....	13
4.2 Unevenness measure of Gene($Um(g)$).....	15
4.3 Result and Analysis.....	15
4.3.1 Area Under ROC curve.....	15
5. Conclusion and Future Work.....	21
6. Bibliography.....	22

List of Figures

1.1 DNA Replication Graph.....	2
1.2 Transcription and Translation Figure.....	3
1.3 TPR and FPR Figure.....	14
1.4 ROC Graph.....	16

List of Tables

1.1 Genetic Code table.....	4
1.2 AUC Table.....	18
1.3 Um(g) Table.....	19

List of Abbreviations

Standard abbreviations from the literature

1. Nucleotides

A	Adenine	R	A and G (puRine)
C	Cytosine	Y	C and T (pYrimidine)
G	Guanine	S	C and G (Strong interaction)
T	Thymine	W	A and T (Weak interaction)
U	Uracil	N	A, G, C and T (aNy)

2. Amino Acids

Ala	Alanine	Met	Methionine
Cys	Cysteine	Asn	Asparagine
Asp	Aspartic acid	Pro	Proline
Glu	Glutamic acid	Gln	Glutamine
Phe	Phenylalanine	Arg	Arginine
Gly	Glycine	Ser	Serine
His	Histidine	Thr	Threonine
Ile	Isoleucine	Val	Valine
Lys	Lysine	Trp	Tryptophan
Leu	Leucine	Tyr	Tyrosine

Non-standard abbreviations used in the thesis

CUB	Codon usage bias
FB(s)	Family box(es)
FDS(s)	Four-fold degenerate site(s)
NS(s)	Nucleotide substitution(s)
RDF(s)	Relative di-nucleotide frequency(ies)
SMD	Selection-Mutation-Drift
UdG	U difference Gly
UdR	U difference Arg