DECLARATION

I, **Himangshu Deka,** bearing Registration No. PhD/2421/13, Dt. 12/09/2013, hereby declare that the subject matter of the thesis entitled "**Analysis of synonymous codon usage pattern in Influenza A Virus Subtypes**" is the record of work done by me and that the contents of this thesis did not form the basis for award of any degree to me or to anybody else to the best of my knowledge. The thesis has not been submitted to any other University/ Institute.

Place: Assam University, Silchar

Date: Himangshu Deka

ACKNOWLEDGEMENTS

First of all, I thank the Almighty for the wisdom and perseverance that he has been bestowing upon me during the research period, and indeed, throughout my life.

I would like to express my deep and sincere gratitude to my supervisor, Prof. Supriyo Chakraborty, for his constant support throughout my PhD work. His generous support, advice and encouragement have made this work possible and his guidance helped me throughout this work.

I am also grateful to Dean, Hargobind Khurana School of Life Sciences, Head Department of Biotechnology and all the faculty members, technical and official staffs of Department of Biotechnology for their kind cooperation.

I would like to express my deepest gratitude to University Grants Commission (UGC), Govt. of India for awarding me the UGC-BSR Fellowship for pursuing my PhD work.

I owe my gratitude to all my lab mates and friends from the Department of Biotechnology. Particularly I would specially like to acknowledge my sincere gratitude to Arup Kr Malakar, Prosenjit Paul and Anupam Nath for their persistent help, kind cooperation and constant support throughout my research work.

My deepest gratitude goes to my family, especially my parents for their consistent love and support throughout my life. Their support and faith in me have sustained and inspired me throughout my academic endeavour. I owe my deep sense of gratitude to my extended family at Silchar, especially Shankar Debnath for their generous support, love and motivation during my PhD work.

Lastly, I would like to thank everyone who was important in the successful completion of this thesis. I apologize to all others who I cannot express my gratitude individually here due to paucity of space.

Date: (*Himangshu Deka*)
Place: Assam University, Silchar

CONTENTS

List of Figures **I-II**

List of Tables **III**

CHAPTER 1 INTRODUCTION **1-17**

1.1 The genetic code and codon usage bias 1-3

1.2 Factors affecting codon usage bias 3-6

1.2.1 Mutational bias and codon usage 3-4

1.2.2 Selection and codon usage 4

1.2.3 Gene expression 5

1.2.4 Mutation-selection-drift theory in codon usage bias 5-6

1.3 Applications of codon usage study 6-7

1.4 Influenza A virus 7-15

1.4.1 Influenza A virus subtypes infecting the humans 11-13

1.4.1.1 H1N1 11

1.4.1.2 H1N2 12

1.4.1.3 H2N2 12

1.4.1.4 H3N2 12-13

1.4.1.5 H5N1 13

1.4.2 Genetic structure of influenza A virus 13-15

1.5 Statement of the problem 15-16

1.6 Rationale of the study 16-17

1.7 Objectives 17

CHAPTER 2 REVIEW OF LITERATURE **18-29**

CHAPTER 3 MATERIALS AND METHODS **30-35**

3.1 Sequence data retrieval 30

3.2 Compositional analysis 30

3.3 Indices of codon usage study 30-35

3.3.1 Relative synonymous codon usage (RSCU) 30-31
3.3.2 Effective number of codons (Nc) 31
3.3.3 Dinucleotide odds ratio 31-32
3.3.4 Codon adaptation index (CAI) 32
3.3.5 MELP (MILC based expression measure) 33
3.3.6 Frequency of optimal codons (Fop) 33-34
3.3.7 Neutrality analysis 34
3.3.8 PR2 analysis 34
3.3.9 Codon pair context analysis 35
CHAPTER 4 RESULTS 36-64
4.1 Analysis of codon usage pattern in the genomes of the IAV subtypes 36-40
4.1.1 Effective number of codons (Nc) in the IAV genes 36
4.1.2 Relative synonymous codon usage (RSCU) 37-39
4.1.3 Frequency of optimal codons (Fop) 40-41
4.2 Compositional features in the coding sequences of IAV genes 41-48
4.2.1 Nucleotide composition 41-43
4.2.2 Amino acid usage 43-44
4.2.3 Dinucleotide analysis and CpG usage 45-46
4.2.4 Neutrality Analysis 47-48
4.3 Comparison of the codon usage patterns across the IAV subtypes 49-59
4.3.1 PR2 bias plot analysis 49-52
4.3.2 Correspondence analysis 52-54
4.3.3 Codon-pair context analysis 54-59
4.4 Prediction of the expression of the IAV genes 60-62
4.4.1 Codon adaptation index 60
4.4.2 MELP 61-62
4.5 Characterization the basic biochemical properties of proteins 62-64

CHAPTER 5 DISCUSSION **65-76**

5.1 Codon usage bias and molecular evolution 66-68

5.2 Codon usage in influenza A virus subtypes 68-70

5.3 Compositional features of the IAV genes among the subtypes 70-71

5.4 Role of compositional constraint in codon usage 71-72

5.5 Comparative account of the codon usage patterns across the IAV subtypes 72-73

5.6 Trends in codon usage variation across the subtypes 73

5.7 Codon context analysis among the IAV subtypes 73-74

5.8 Prediction of expression of the IAV genes 74

5.9 Codon usage and basic biochemical properties of IAV proteins 75

5.10 Future directions of the present study 75-76

SUMMARY **77-81**

Key findings 78-81

BIBLIOGRAPHY **81-102**

Appendix A List of Abbreviations 103-104

Appendix B List of Publications 105-106

Appendix C List of workshops and conferences atttended 107

Appendix D Accession numbers of the genes enrolled in the study 108-148

List of Figures

Fig. 1.4.1	Global burden of the 2009 H1N1 subtype of influenza A virus in the first year of infection	8
Fig. 1.4. 2	Schematic representation of genetic drift and genetic shift giving rise to novel IAV strains	9
Fig. 1.4. 3	Evolutionary timeline of the pandemic influenza A viruses	10
Fig. 4.1. 1	Heatmap of RSCU in the genes of the five IAV subtypes	37
Fig. 4.1.2	Rare codons (in blue color) in 5 IAV subtypes enrolled in the study	39
Fig. 4.1.3	Relationship between Fop and Nc values in IAV genes	41
Fig. 4.2.1	Nucleobase composition at synonymous positions in the IAV genes	43
Fig. 4.2.2	Overall dinucleotide usage in the IAV genes	45
Fig. 4.2.3	Neutrality plot (GC12 vs GC3) analysis in the five IAV subtypes	48
Fig. 4.3.1.1	PR2 bias analysis in H1N1 showing the overall as well as degeneracy level-specific PR2-fingerprints	49
Fig. 4.3.1.2	PR2 bias analysis in H1N2 showing the overall as well as degeneracy level-specific PR2-fingerprints	50
Fig. 4.3.1.3	PR2 bias analysis in H2N2 showing the overall as well as degeneracy level-specific PR2-fingerprints	51
Fig. 4.3.1.4	PR2 bias analysis in H3N2 showing the overall as well as degeneracy level-specific PR2-fingerprints	51
Fig. 4.3.1.5	PR2 bias analysis in H5N1 showing the overall as well as degeneracy level-specific PR2-fingerprints	52
Fig. 4.3.2.1	Row plot of correspondence analysis showing allocation of genes based on their corresponding codon choices	53
Fig. 4.3.2.2	Column plot of correspondence analysis showing allocation of the 59 codons	53
Fig. 4.3.2.3	Scree plot from correspondence analysis showing the contribution of the individual axes	54

Fig. 4.3.3.1	Average matrix plot of 5' codon-pair contexts (64x64) in the IAV genes	55
Fig. 4.3.3.2	Codon pair context preferences based on the dinucleotides at the junction of two adjacent codons	57
Fig. 4.3.3.3	Comparison of the codon-pair contexts between the IAV subtypes	58
Fig. 4.3.3.4	Histogram of residuals (3' contexts) of the preferred (green) as well as rejected (red) contexts among the subtypes of IAV	59
Fig. 4.4.1	Predictive expressivity of the genes (MELP and CAI) belonging to different IAV subtypes	60
Fig. 4.4.2	Plot of the two gene expression measures MELP and CAI as a function of Axis1 of correspondence analysis	61

List of Tables

Table 4.1.1	Average Nc values of all the genes across the subtypes	36
Table 4.1.2	Frequency of optimal codons (Fop) in the selected IAV genes	40
Table 4.2.1	Compositional features of the genes in the IAV subtypes	42
Table 4.2.2	Amino acid usage in the major genes of the IAV subtypes	44
Table 4.2.3	Dinucleotide usage in the IAV genes	46
Table 4.3.3.1	Ten most preferred contexts in the IAV subtypes	56
Table 4.3.3.2	Frequency of amino acids as tandem repeats in the IAV subtypes	56
Table 4.5.1	Basic biochemical properties of the IAV proteins	63
Table 4.5.2	Correlation between the basic biochemical properties with axis1 of correspondence analysis	63