“To succeed in your mission, you must have single-minded devotion to your goal”.

-Dr. A. P. J. Abdul Kalam: an inspiration to millions

Declaration

I hereby declare that **"A Study on Single Nucleotide Variations in Different Regions of *Escherichia coli* Genome Sequences"** has been submitted to Tezpur university in the Department of Molecular Biology and Biotechnology under the School of Sciences in partial fulfilment for the award of the degree Doctor of Philosophy in Molecular Biology and Biotechnology. This is an original work carried out by me. Further, I declare that no part of this work has been reproduced elsewhere for award of any other degree.

Date: 27·05·2024

Place: Tezpur

Pratyush Kumar Beura

(Pratyush Kumar Beura)

Reg. no.: TZ23000709

TEZPUR UNIVERSITY
(A central university established by an Act of Parliament)
DEPARTMENT OF MOLECULAR BIOLOGY AND BIOTECHNOLOGY
TEZPUR-784028, ASSAM, INDIA

Dr. Suvendra Kumar Ray
Professor

Ph. No.: +91-3712-275406(O)
+ 91-99544-72151(M)
Email: suven@tezu.ernet.in

CERTIFICATE OF THE SUPERVISOR

This is to certify that the thesis entitled **"A Study on Single Nucleotide Variations in Different Regions of *Escherichia coli* Genome Sequences"** submitted to the School of Sciences, Tezpur University in partial fulfilment for the award of the degree Doctor of Philosophy in Molecular Biology and Biotechnology is a record of original research work carried out by Mr. Pratyush Kumar Beura under my supervision and guidance.

All helps received by him from various sources have been duly acknowledged. No part of this thesis has been reproduced elsewhere for award of any other degree.

Suvendra Kr Ray

Date: 27/05/2024
Place: Napaam, Tezpur

(Suvendra Kumar Ray)
Supervisor

TEZPUR UNIVERSITY
(A central university established by an Act of Parliament)
DEPARTMENT OF MOLECULAR BIOLOGY AND BIOTECHNOLOGY
TEZPUR-784028, ASSAM, INDIA

Dr. Siddhartha S. Satapathy
Professor

Ph. No.: +91-94359-79648(O)
Email: ssankar@tezu.ernet.in

CERTIFICATE OF THE CO-SUPERVISOR

This is to certify that the thesis entitled **"A Study on Single Nucleotide Variations in Different Regions of *Escherichia coli* Genome Sequences"** submitted to the School of Sciences, Tezpur University in partial fulfilment for the award of the degree Doctor of Philosophy in Molecular Biology and Biotechnology is a record of original research work carried out by Mr. Pratyush Kumar Beura under my supervision and guidance.

All helps received by him from various sources have been duly acknowledged. No part of this thesis has been reproduced elsewhere for award of any other degree.

S. S. Satapathy

Date: 27-05-2024
Place: Tezpur

(Siddhartha S. Satapathy)
Co-Supervisor

TEZPUR UNIVERSITY
(A central university established by an Act of Parliament)
DEPARTMENT OF MOLECULAR BIOLOGY AND BIOTECHNOLOGY
TEZPUR-784028, ASSAM, INDIA

CERTIFICATE OF THE EXTERNAL EXAMINER AND ODEC

This is to certify that the thesis entitled **"A Study on Single Nucleotide Variations in Different Regions of *Escherichia coli* Genome Sequences"** submitted by Mr. Pratyush Kumar Beura to Tezpur University in the Department of Molecular Biology and Biotechnology under the School of Sciences in partial fulfilment for the award of the degree Doctor of Philosophy in Molecular Biology and Biotechnology has been examined by us on _____ and found to be satisfactory.

The Committee recommends for the award of the degree of Doctor of Philosophy.

Signature:

Supervisor	**Co-supervisor**	**External Examiner**
Date:	**Date:**	**Date:**

Acknowledgement

As I write the final words of my PhD thesis, foremost, I owe my deepest gratitude to each and everybody being there during the journey. I am pleased to have the opportunity to thank everyone whose efforts have contributed to every single word written in this thesis.

First and foremost, I would like to thank Tezpur University and Department of Molecular Biology and Biotechnology for providing the ambiance and opportunity for the smooth conductance of my research work during my PhD tenure and making this journey unforgettable.

Since my graduation, I was always fascinated by Molecular Biology, it has always been my favourite part of classical biology since those days. Since, I was selected for the PhD Course after my interview, I came across few research articles of My Supervisor Prof. Suvendra Kumar Ray. The articles were enriched with key words like Chargaff's Parity rules, distribution of genes in Leading and Lagging strands and more importantly about the genetic code table which were obviously fascinating and slowly I realised, the subject was not just about only Biology. Fortunately, the course work classes actually shaped my interest into the core areas of Molecular Evolution. My supervisor Prof. Suvendra Kumar Ray used to take those classes in online mode, and I was so excited about the subject since then. His philosophy like "Genetics is short term evolution and Evolution is long terms genetics", "Beyond Molecular Biology the only exciting subject is Astronomy" gave stability to my wavering mind.

I sincerely owe my gratitude to my Supervisor Prof. Suvendra Kumar Ray, for always being there. I am obliged to my supervisor for introducing me into the world of scientific research. His constant guidance and motivation for the work have always been a support for me during my work professionally and personally. I got the wonderful opportunity to work under his supreme guidance and his lively discussions in the lab have always been helpful for the completion of my work.

I extend my sincere gratitude to my Co-Supervisor Dr. Siddhartha Sankar Satapathy for his constant guidance since the beginning of this journey. His support in computational approaches like development of the software and teaching other tools have actually helped me to work ergonomically. The weekly lab meeting on Saturdays conducted by him is the real point of my progress in the work. His consistency in the work has encouraged me.

I am thankful to my Doctoral Committee members for their support and kind gesture during my tenure. I would like to thank Respected Prof. Ramesh Ch. Deka (Chemical Sciences and VC, Cotton University), Dr. Rupak Mukhopadhyay (Current Head, MBBT), Dr. Nima D Namsa (MBBT) and Dr. Aditya Kumar (MBBT) for their kind support.

I am thankful to Prof. Edward Feil (University of Bath, UK) for providing us with the bacterial genome sequences which helped us to carry out this work.

This work would not have been possible without the support and kindness of a senior. Since day one, Dr. Piyali Sen, my senior, has been so helpful since sharing our lab papers to be online whenever I required any help during my work. Her efforts in making me understand the methodology and most importantly her kind efforts in providing me with the code helped me to do this work. Her sincere efforts are always commendable.

I would like to highlight the vital contributions in my work. Our MSc juniors Chayanika Chetia, Saurav Das, Varsha Mahabal, Rajeev Barala and Sauranil Guha's humble efforts during my work were helpful for me. I worked with all of them and the discussions with all of them were highly enlightened.

I would also like to thank our lab-members Mr. Subham, Subhra, Lukapriya, Shuhada and Monika for their kind support during my PhD tenure.

I extend my gratitude to all the teaching and non-teaching members of the Department of Molecular Biology and Biotechnology.

I am thankful to Tezpur University and DBT-funded Multi institutional collaborative National Network Project of ACTREC-TMC, Navi Mumbai for the financial support received during my tenure.

*I am thankful to the linguistic improvement service provided by open AI platforms in different sections of the thesis (The Results are **not** AI generated).*

My heart-felt gratitude to Dharitri and Diganta Bhaiya for being so supportive in my personal sphere of life. My personal bonding with both of them really worked as a stressbuster for me.

It would be an injustice not to acknowledge my family members. They have actually witnessed my struggle. Since the beginning of my education my mother has always been an integral part of my life. Her morale support and my father's faith in my capabilities are the unseen forces behind this journey. My sister's support during my hard times is unforgettable.

Lastly, my heart-felt regards to all of them for being so kind and supportive during this journey.

I feel fortunate to be a part of this journey.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

%	Percent
A	Adenine
T	Thymine
C	Cytosine
G	Guanine
U	Uracil
N	Any nucleotide
R	Purine (A & G)
Y	Pyrimidine (T & C)
K	Keto (G & T)
M	Amino (A & C)
ti	Transition
tv	Transversion
S	Synonymous
NS	Non-synonymous
Sti	Synonymous transition
Stv	Synonymous transversion
Nti	Non-synonymous transition
Ntv	Non-synonymous transversion
CUB	Codon usage bias
PTC	Pre-termination codon
SNVs	Single nucleotide variations
Phe	Phenylalanine
Leu	Leucine
Ile	Isoleucine
Met	Methionine
Val	Valine
Ser	Serine
Pro	Proline
Thr	Threonine
Ala	Alanine
Tyr	Tyrosine
His	Histidine
Gln	Glutamine
Asn	Asparagine
Lys	Lysine
Asp	Aspartic acid
Glu	Glutamic acid
Cys	Cystine

Trp	Tryptophan
Arg	Arginine
Gly	Glycine
FFD	*Four-fold degenerate site*
TFD	*Two-fold degenerate site*
SFD	*Six-fold degenerate site*
ZFD	*Zero-fold degenerate site*
FB	Family box
SB	Spilt box
Fig.	Figure
CAI	Codon Adaptive Index
DNA	Deoxyribonucleic Acid
obs	Observed
exp	Expected
IGRs	Intergenic regions
E. coli	*Escherichia coli*
S. enterica	*Salmonella enterica*
CSM	Codon substitution model
JC model	Jukes-Cantor model
ML method	Maximum Likelihood method
e	Estimated
o	Observed
RBS	Ribosome binding site
SD sequence	Shine-Dalgarno sequence
DNAP	DNA polymerase
RNAP	RNA polymerase
HKY	Hasegawa-Kishino-Yano
GTR	General Time Reversible