STUDY ON GLUCOSE-6-PHOSPHATE DEHYDROGENASE (G6PD) VARIANTS AND ITS ASSOCIATION WITH HAEMOGLOBINOPATHIES AMONG THE TRIBAL POPULATION OF MALARIA ENDEMIC INDO-BHUTAN BORDER DISTRICTS OF BTR, ASSAM, INDIA.



A Thesis submitted to Bodoland University for the Degree of Doctor of Philosophy in Biotechnology in the Faculty of Science and Technology, 2023

By

NOYMI BASUMATARY

Registration no.- FINAL/03BIO0022 of 2017-18

Department of Biotechnology

Bodoland University, Kokrajhar-783370

Assam, India

DECLARATION

I do hereby declare that the research work embodied in this thesis entitled "Study on Glucose-6-Phosphate Dehydrogenase (G6PD) variants and its association with haemoglobinopathies among the tribal population of malaria endemic Indo-Bhutan border districts of BTR, Assam, India" has been carried out by me under direct guidance and supervision of Prof. (Dr.) Jatin Sarmah, Department of Biotechnology, Bodoland University, Kokrajhar, Assam, India.

The work is original and has not been submitted in part or in full for any degree or diploma to any university.

Date: 22/01/24 Place: Bedeland University

Nopui Basuaratary

(NOYMI BASUMATARY)



BODOLAND UNIVERSITY, KOKRAJHAR

Debargaon, P.O. Rangalikhata Kokrajhar - 783370, BTAD, Assam.

From,

Date: 22,01.2024

Prof. (Dr.) Jatin Sarmah Department of Biotechnology Bodoland University Kokrajhar, Assam Email: jatinsarmahindia@gmail.com

CERTIFICATE

This is to certify that the thesis entitled "Study on Glucose-6-Phosphate Dehydrogenase (G6PD) variants and its association with haemoglobinopathies among the tribal population of malaria endemic Indo-Bhutan border districts of BTR, Assam, India", which has been submitted by Miss Noymi Basumatary for the award of the degree of Doctor of Philosophy under Bodoland University is a record of original research works carried out by her under my direct guidance and supervision. She has fulfilled all the requirements for submitting the thesis.

The results embodied in the thesis have not been submitted to any other university or institution for the award of any degree or diploma.

01

(JATIN SARMAH) Professor, Department of Biotechnology Bodoland University Kokrajhar, Assam, India

> DR. JATIN SARMAH Professor, Department of Biotechnology Bodoland University, Kokrajhar Assam, India.

ACKNOWLEDGEMENT

Towards the end of this journey, I take this opportunity to express my heartfelt gratitude to the people who stood by me and extended their support in various ways during the tenure of my study. Above all, I thank Almighty God for his showers of blessings throughout my Ph.D. journey.

I express my sincere gratitude to Prof. Laishram Ladoo Singh, Honourable Vice-Chancellor, Dr. Subung Basumatary, Registrar, and Dr. Manjil Basumatary, Academic Registrar, Bodoland University for providing the facilities and environment to pursue my work.

I owe my sincere gratitude to my Ph. D. Supervisor, Prof. (Dr.) Jatin Sarmah, Department of Biotechnology, Bodoland University. I am deeply indebted to him for his constant guidance, encouragement and support throughout the journey of my Ph.D. work.

I convey my sincere gratitude to Prof. Sandeep Das, Head, Department of Biotechnology Bodoland University, for providing the necessary departmental facilities and rendering all possible assistance during the course of my research. I also thank Sir for allowing me to access the facilities of DBT-Biotech Hub and Technology Incubation centre, Bodoland University.

Acknowledgement is due to all faculty members of the department namely, Dr. Sharmistha Brahma Kour, Dr. Silistina Narzari, Dr. Arvind Kumar Goyal, Mrs. Jonali Owary and Dr. Monoranjan Boro for their words of wisdom and encouragement throughout the study period.

I am indebted to the Lady Tata Memorial Trust, Mumbai, for providing me financial assistance for completion of my work. I also thank the BTR Government for providing financial support during the initial stage of my work.

My heartfelt gratitude to Prof. Paresh Kumar Sarma, Department of Medicine, Dhubri Medical College and Hospital, Dhubri and Dr. Dipankar Baruah, Associate Professor, Department of Pathology, Gauhati Medical College and Hospital, Guwahati for their constant guidance and support in analyzing the medical reports of the patients.

I express my gratitude to the Superintendents and Co-ordinators of Hindlabs laboratories of Udalguri Civil Hospital and Kokrajhar Civil Hospital for allowing me to perform the haematological tests. I also thank the lab technicians of Hindlabs laboratories of both the hospitals for helping me to carry out the tests.

I acknowledge Dr. Anil Agarwala, Pathologist, Mr. Ritupal Baruah and Mr. Gyanendra Barman, Lab technicians of Arya Wellness Centre, Guwahati for helping me in Haemoglobin typing of the samples.

I also express my sincere thanks to Dr. Rajiv Chandra Dev Goswami and Mr. Kamal Gauli, Biotech Park, Guwahati for allowing and helping me to perform part of my research work at their centre.

My acknowledgement is also due to Dr. Manab Deka, Associate Professor, Arya Vidyapeeth College, Guwahati for helping me in statistical analysis.

I wish to acknowledge the Medical Officers, Managers and medical staffs of the tea estates and the village heads for allowing and helping me in collection of blood samples. I also thank Mr. Ankur Sarma and Miss Bidang Basumatary for helping me in collection of blood samples.

I gratefully acknowledge my labmates, Miss Trisha Sonowal, Miss Rijumoni Daimari, Mr. Nerswn Basumatary, Miss Seema Khakhalary and all my fellow scholars for their constant kind support throughout my Ph.D. journey.

I express my heartfelt thanks to Mr. Chandan Brahma, Mr. Hirak Moni Borgoyary and Mr. Paban Ray for their help and support.

Finally, I take this opportunity to pay my sincerest thanks to my parents, Mr. Guneswar Basumatary and Mrs. Sabita Basumatary for their sacrifices and tireless efforts towards me which can never be replenished by any words of gratitude. I also extend my special thanks to my sister, Miss Niru Basumatary, my friends, Mr. Dominic Narzary, Miss Manita Daimari, Miss Nitisha Boro, Mrs. Abhijita Daimari for their constant love, support and encouragement throughout the course of my research.

Nophi Basunalary

(Noymi Basumatary)

LIST OF TABLES

Table no.	Title	Page no.
1	Classification of G6PD variants by WHO (1989).	7
2	List of drugs to be avoided by G6PD deficient patients (Frank, 2005).	8
3	List of drugs to be used with caution in G6PD deficient patients	9-10
	(Beutler, 2008; Younster et al., 2010; Bubp et al., 2015).	
4	Types of structural disorders of haemoglobin.	12-13
5	Types of thalassaemia syndromes.	13-15
6	G6PD variants reported from Southeast Asia.	24-29
7	G6PD variants and mutations identified among the Indians.	30-31
8	Types of mutations involved in G6PD deficiency.	34
9	Location of mutations reported in <i>g6pd</i> gene.	34-35
10	District wise villages and tea estates covered in the study.	46-47
11	Interpretation of the device's reading as per WHO recommended	49
	range.	
12	Details of primers used for exon-wise amplification of g6pd gene.	52-53
13	Standardization of PCR settings for each exon.	53
14	Details of restriction endonucleases used for detection of mutation.	54-55
15	Result of G6PD screening in four districts of BTR, Assam.	61
16	Results of Hb-typing of anaemic subjects.	63
17	Mean ± SD of Hb types in different categories of	64
	haemoglobinopathies.	
18	Mean \pm SD values of RBC parameters and platelets, and Pearson	65
	correlation coefficient with G6PD along with p-values.	
19	Mean \pm SD values of WBC parameters and Pearson correlation	65-66
	coefficient with G6PD along with p-values.	
20	Cross tabulation between G6PD deficiency and gender.	66
21	Binding affinities of the antioxidants with the 3D structures of the	93-94

G6PD variants.

22	Drug-likeness and ADME properties of the compounds.	95-97
23	Toxicity analysis of the ligands.	98-101
24	Amino acid residues of the ligands interacting with the variants.	102-107
25	Mean \pm SD values of RMSD, RMSF, Rg and SASA of G6PD	115
	variants.	
26	Mean \pm SD values of RMSD, RMSF, Rg and SASA of G6PD variant	125
	complexed with the ligand.	

LIST OF FIGURES

Figure no.	Title	Page no.
1	Diagrammatic representation of NADPH generation via Pentose phosphate pathway.	3
2	Role of G6PD in defense mechanism against oxidative attack (Luzzatto <i>et al.</i> , 2016).	3
3	Location of g6pd gene (Gomez-Manzo et al., 2016).	5
4	Crystal structure of G6PD enzyme in dimeric form (Gomez-Manzo et al., 2016).	5
5	Graphical representation of $g \delta p d$ gene and mutations from India.	7
6	Map of the study site.	48
7	G6PD status of the study subjects.	62
8	Images of genomic DNA run on 0.8% agarose gel electrophoresis. M \rightarrow 1kb DNA ladder and lanes 1-12 \rightarrow genomic DNA of the G6PD deficient samples.	68
9	Agarose gel electrophoresis images of PCR products of exon 3. M \rightarrow 100bp ladder, lanes 1-12 \rightarrow 352bp PCR products of exon 3.	68
10	Agarose gel electrophoresis images of PCR products of exon 4. M \rightarrow 100bp ladder, lanes 1-12 \rightarrow 314bp PCR products of exon 4.	68
11	Agarose gel electrophoresis images of PCR products of exons 4-5. M \rightarrow 100bp ladder, lanes 1-12 \rightarrow 701bp PCR products of exons 4-5.	70
12	Agarose gel electrophoresis images of PCR products of exons 6-7. M \rightarrow 100bp ladder, lanes 1-12 \rightarrow 545bp PCR products of exons 6-7.	70
13	Agarose gel electrophoresis images of PCR products of exon 9. M \rightarrow 100bp ladder, lanes 1-12 \rightarrow 276bp PCR products of exon 9.	70
14	Agarose gel electrophoresis images of PCR products of exon 10. M	72

 \rightarrow 100bp ladder, lanes 1-12 \rightarrow 342bp PCR products of exon 10.

- 15 Agarose gel electrophoresis images of PCR products of exon 11. M 72 \rightarrow 100bp ladder, lanes 1-12 \rightarrow 214bp PCR products of exon 11. 16 Agarose gel electrophoresis images of PCR products of exon 12. M 72 \rightarrow 100bp ladder, lanes 1-12 \rightarrow 227bp PCR products of exon 12. 17 Agarose gel electrophoresis of PCR products digested with HaeIII 74 (Orissa 131C>G). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 3, lanes 4, lanes 2, 3, 5, 6, $10-12 \rightarrow 131C>G$ mutation is present, lanes $7-9 \rightarrow 131C>G$ mutation is absent. 18 74 Agarose gel electrophoresis of PCR products digested with NlaIII (Namoru 208T>C). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 4, lanes $2-12 \rightarrow \text{mutation } 208T > C$ was absent in all samples. 19 Agarose gel electrophoresis of PCR products digested with FokI (A⁺ 74 376A>G). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exons 4-5, lanes 5, $6 \rightarrow 376$ A>G mutation is present, lanes 2-4, 7-12 \rightarrow 376 A>G mutation is absent. Agarose gel electrophoresis of PCR products digested with NlaIII 20 76 (A⁻²⁰² G>A). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exons 4-5, lanes $2-12 \rightarrow$ mutation 202G>A was absent in all samples. 21 76 Agarose gel electrophoresis of PCR products digested with HindIII (Mahidol 487G>A). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exons 6-7, lanes 2, $3 \rightarrow 487$ G>A mutation is present. Agarose gel electrophoresis of PCR products digested with MboII 22 76 (Mediterranean 563C>T). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exons 6-7, lanes 2-4 & 6 \rightarrow 563C>T mutation is present, lanes 5 & 7 \rightarrow 563C>T mutation is absent. 23 Agarose gel electrophoresis of PCR products digested with BstUI 78 (Acores 595 A>G). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exons 6-7, lanes $2-12 \rightarrow 595$ A>G mutation is absent.
- Agarose gel electrophoresis of PCR products digested with MnII 78

(Kalyan-Kerala/ Jamnagar / Rohini 949G>A). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 9, lanes 2, 3 \rightarrow 949 G>A mutation is present, lanes 4-10 \rightarrow 949G>A mutation is absent.

- 25 Agarose gel electrophoresis of PCR products digested with BstXI 78 (Chatham (1003G>A). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 9, lanes 2-12 \rightarrow 1003G>A mutation is absent.
- 26 Agarose gel electrophoresis of PCR products digested with HhaI 81 (Guadalajara 1159C>T). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 10, lanes 2-12 \rightarrow 1159C>T mutation was absent in all samples.
- 27 Agarose gel electrophoresis of PCR products digested with HhaI 81 (Union 1360C>T). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 11, lanes 2-12 \rightarrow 1360C>T mutation was absent.
- 28 Agarose gel electrophoresis of PCR products digested with AfIII 81 (Canton 1376G>T). M \rightarrow 100bp DNA ladder, lane 1 \rightarrow undigested PCR product of exon 12, lanes 2-12 \rightarrow 1376G>T mutation was absent.
- Agarose gel electrophoresis of PCR products digested with NdeI
 Kaiping 1388G>A). M → 100bp DNA ladder, lane 1 → undigested
 PCR product of exon 12, lanes 2-12 → 1388G>A mutation was absent.
- 30Frequency of G6PD variants detected in the study.83

83

- 31 Gender-wise distribution of G6PD variants.
- 32 Modeled structure and validation of G6PD Orissa variant. (a) G6PD 86 Orissa modeled using I-TASSER. (b) Validation of G6PD Orissa by ProTSAV.
- of G6PD 33 Modeled structure and validation Kalyan-86 Kerala/Jamnagar/Rohini variant. (a) G6PD Kalyan-Kerala/Jamnagar/Rohini modeled using I-TASSER. (b) Validation of G6PD Kalyan-Kerala/Jamnagar/Rohini by ProTSAV.

34	Modeled structure and validation of G6PD Mahidol variant. (a) G6PD Mahidol modeled using I-TASSER. (b) Validation of G6PD Mahidol by ProTSAV.	87
35	Modeled structure and validation of G6PD A^+ variant. (a) G6PD A^+ modeled using I-TASSER. (b) Validation of G6PD A^+ by ProTSAV.	87
36	2D structures of antioxidant compounds downloaded from Pubchem.	89-91
37	3D and 2D visualization of docking of Orissa and Myricetin. (a) 3D interaction between Orissa and Myricetin. (b) 2D interaction between Orissa and Myricetin.	109
38	3D and 2D visualization of docking of Kalyan-Kerala and Apigenin.(a) 3D interaction between Kalyan-Kerala and Apigenin.(b) 2D interaction between Kalyan-Kerala and Apigenin.	109
39	3D and 2D visualization of docking of Mahidol and Catechin. (a) 3D interaction between Mahidol and Catechin. (b) 2D interaction between Mahidol and Catechin.	110
40	3D and 2D visualization of docking of A^+ and Daidzen. (a) 3D interaction between A^+ and Daidzen. (b) 2D interaction between A^+ and Daidzen.	110
41	RMSD of the WT protein and the G6PD variants.	113
42	RMSF of the WT protein and the variants.	113
43	RG of the WT protein and the variants.	114
44	SASA of the WT protein and the variants.	114
45	RMSD of the WT protein, Orissa and Orissa complexed with Myricetin.	117
46	RMSF of the WT protein, Orissa and Orissa complexed with Myricetin.	117
47	RG of the WT protein, Orissa and Orissa complexed with Myricetin.	118
48	SASA of the WT protein, Orissa and Orissa complexed with	118

Myricetin.

49	RMSD of the WT protein, Kalyan-Kerala and Kalyan-Kerala complexed with Apigenin.	120
50	RMSF of the WT protein, Kalyan-Kerala and Kalyan-Kerala complexed with Apigenin.	120
51	Rg of the WT protein, Kalyan-Kerala and Kalyan-Kerala complexed with Apigenin.	121
52	SASA of the WT protein, Kalyan-Kerala and Kalyan-Kerala complexed with Apigenin.	121
53	RMSD of the WT protein, Mahidol and Mahidol complexed with Catechin.	123
54	RMSF of the WT protein, Mahidol and Mahidol complexed with Catechin.	123
55	RG of the WT protein, Mahidol and Mahidol complexed with Catechin.	124
56	SASA of the WT protein, Mahidol and Mahidol complexed with Catechin.	124
57	RMSD of the WT protein, A^+ and A^+ complexed with Daidzen.	126
58	RMSF of the WT protein, A^+ and A^+ complexed with Daidzen.	126
59	RG of the WT protein, A^+ and A^+ complexed with Daidzen.	127
60	SASA of the WT protein, A^+ and A^+ complexed with Daidzen.	127

ABBREVIATIONS

G6PD:	Glucose-6-Phosphate Dehydrogense
PPP:	Pentose Phosphate Pathway
NADPH:	Nicotinamide Adenine Dinucleotide Phosphate Hydrogen
ROS:	Reactive Oxygen Species
GSH:	Glutathione
RBC:	Red Blood Cell
WHO:	World Health Organization
P. vivax:	Plasmodium vivax
P. ovale:	Plasmodium ovale
Hb:	Haemoglobin
HbS:	Haemoglobin S
HbE :	Haemoglobin E
HbC:	Haemoglobin C
HbSS:	Haemoglobin S homozygous
HbCC:	Haemoglobin C homozygous
Hb AE:	Haemoglobin E heterozygous
Glu:	Glutamine
Lys:	Lysine
Arg:	Arginine
Trp:	Tryptophan
His:	Histidine
Met:	Methionine
Thr:	Threonine

Ala:	Alanine
Gly:	Glycine
Ile:	Isoleucine
Cys:	Cystein
Phe:	Phenylalanine
Val:	Valine
Tyr:	Tyrosine
Ser:	Serine
Pro:	Proline
Asp:	Aspartic acid
Asn:	Asparagine
NR:	Not reported
HbD:	Haemoglobin D
NHM:	National Helth Mission
HPFH:	Hereditary Persistence of Foetal Haemoglobin
HbQ:	Haemoglobin Q
HbK:	Haemoglobin K
NAC:	N-acetyl-cysteine
LC:	L-cysteine
TCA:	Tricarboxylic acid
ME1:	Malic enzyme 1
IDH1:	Isocitrate dehydrogenase 1
CADD:	Computer-aided drug design
MD:	Molecular dynamics
BTR:	Bodoland Territorial Region

CBC:	Complete Blood Count
WBC:	White Blood Cell
MCV:	Mean Corpuscular Volume
MCH:	Mean Corpuscular Haemoglobin
MCHC:	Mean Corpuscular Haemoglobin Concentration
PCV:	Packed Cell Volume
RDW:	Red Cell Distribution Width
SPSS:	Statistical Package for Social Sciences
Buffer AL:	Lysis buffer
Buffer AW1:	Wash buffer 1
Buffer AW2:	Wash buffer 2
Buffer AE:	Elution buffer
DF:	Dilution factor
bp:	Base pair
Taq:	Thermus aquaticus
PCR:	Polymerase Chain Reaction
HaeIII:	Haemophilus aegyptus III
NlaIII:	Neisseria lactamica III
FokI:	Flavobacterium okeanokoites I
HindIII:	Haemophilus influenza III
MboII:	Moraxella bovis II
BstUI:	Bacillus stearothermophilus I
BstXI:	Bacillus stearothermophilus X I
HhaI:	Haemophilus haemolyticus I
AfIII:	Anabaena flosaquae II

NdeI:	Neisseria denitrificans I
RCSB PDB:	Research Collaboratory for Structural Bioinformatics Protein Data Bank
ADT:	AutoDock Tool
MS:	Molecular Surface
ADMET:	Absorption, Distribution, Metabolism, Excretion and Toxicity
GROMACS:	GROningen Machine for Chemical Simulations
GROMOS:	GROningen Molecular Simulation
SPCE:	Simulation Program with Integrated Circuit Emphasis
NaCl:	Sodium Chloride
NTP:	Number of particles, pressure and temperature
RMSD:	Root Mean Square Deviation
RMSF:	Root Mean Square Fluctuation
RG:	Radius of Gyration
SASA:	Solvent Accessibility Surface Area
HbA ₀ :	Adult haemoglobin
HbA ₂ :	Smaller component of adult haemoglobin
SD:	Standard deviation
TLC:	Total leucocyte count
LYM:	Lymphocytes
A ₂₆₀ :	Absorbance at 260nm
A ₂₈₀ :	Absorbance at 280nm
3D:	Three dimensional
2D:	Two dimensional
MW:	Molecular weight
TPSA:	Topological polar surface area

HbA:	Hydrogen bond acceptor
HbD:	Hydrogen bond donor
LD50:	Lethal dose
CNSHA:	Chronic non-spherocytic hemolytic anaemia
Ki:	Inhibition constant
Km:	Michaelis constant

STANDARD UNITS

km ² :	Square kilometre
mg:	Milligram
kg:	Kilogram
μ1:	Microlitre
U/g:	Units per gram
g/dL:	Gram per decilitre
ml:	Millilitre
Million/Cumm:	Miilion per cubic millimetre
pg:	Picogram
Cells/Cumm :	Cells per cubic millimeter
fL:	Femtolitre
Å:	Angstrom
mM:	Millimolar
nm:	Nanomolar
μg/ml:	Microgram per millilitre
kcal/mol:	Kilocalorie per mole
Da:	Dalton