

**ANALYSIS OF THE ANOPHELES GAMBIAE INSECTICIDE
RESISTANCE PROTEIN-PROTEIN INTERACTION NETWORK**

**SHEKARI, ABEDNEGO
(20PBF02178)**

FEBRUARY, 2023

**ANALYSIS OF THE ANOPHELES GAMBIAE INSECTICIDE
RESISTANCE PROTEIN-PROTEIN INTERACTION NETWORK**

BY

**SHEKARI, ABEDNEGO
(20PBF02178)**

B.Sc Biochemistry, Kaduna State University, Kaduna

**A DISSERTATION SUBMITTED TO THE SCHOOL OF
POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE AWARDS OF THE MASTER OF
SCIENCE (M.Sc) DEGREE IN BIOINFORMATICS IN THE
DEPARTEMENT OF COMPUTER AND INFORMATION SCIENCES,
COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT
UNIVERSITY, OTA**

FEBRUARY, 2023

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfillment of the requirement for the award of the degree of Master of Science in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.

Miss Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies)

Signature and Date

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies)

Signature and Date

DECLARATION

I, **SHEKARI, ABEDNEGO (20PBF02178)**, declare that this dissertation is a representation of my work, and is written and implemented by me under the supervision of Dr. Marion O. Adebisi of the Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

SHEKARI, ABEDNEGO

Signature and Date

CERTIFICATION

We certify that this dissertation titled “**ANALYSIS OF THE ANOPHELES GAMBIAE INSECTICIDE RESISTANCE PROTEIN-PROTEIN INTERACTION NETWORK**” is original research carried out by **SHEKARI, ABEDNEGO (20PBF02178)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Dr. Marion O. Adebisi. We have examined and found this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Bioinformatics.

Dr. Marion O. Adebisi
(Supervisor)

Signature and Date

Prof. Olufunke O. Oladipupo
(Head of Department)

Signature and Date

Dr. Victor Odumuyiwa
(External Examiner)

Signature and Date

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies)

Signature and Date

DEDICATION

I dedicate this project to God Almighty, the source of all wisdom, inspiration, and understanding, for his constant grace and faithfulness to me towards completing this research work. Indeed, it was a journey of faith, and God has been forever faithful toward helping me achieve this remarkable feat at the prestigious Covenant University. Thank you, my priority. Ascribe all glory, honor, and adoration to you, Lord.

ACKNOWLEDGEMENTS

I want to acknowledge God for the wisdom, grace, and enablement he has given me from day one of this journey of faith. Indeed, the journey would not have been a success without the help of God Almighty. Even when I thought I had no one to depend on for support, he brought people to my aid beyond my expectations.

I will like to especially appreciate my supervisor Dr. Marion Adebisi for her motherly love, total support, encouragement, patience, and guidance in making this research work a reality. Indeed, without her immense support and encouragement, this research would not have been possible. I would like to thank the Head of the Department of Computer and Information Sciences, Prof. Olufunke O. Oladipupo for her advice. My sincere appreciation goes to all other members of staff of the department.

My immense gratitude goes to my parents, Mr. & Mrs. Shekari Bako, my siblings, Mrs. Rebecca Samuel, Mrs. Regina Zamani, Mrs. Comfort Mathew, Mrs. Alheri Augustine, Mrs. Gloria Hananiah, Mrs. Tabitha Sunday, Mr. and Mrs. Joshua Shekari, Mr. Nehemiah Shekari and Uncle Elisha Bako for their encouragement and support, I pray God to bless you immensely. Mum, I know you are proud of me, and I want you to know that you are forever in my heart. I love you so dearly, continue to rest well. Also, my appreciation goes to my fiancé', Grace Amisi Iliya for her total support, encouragement, and patience.

My appreciation to the Covenant Applied Informatics and Communication Africa Centre of Excellence (CApIC-ACE) for the privilege given to me. I would love to also appreciate my colleagues who made my stay fun, Miracle Enwere, Itunu Ajiboye, Precious Adebola, Emmanue Alagbe, Hannah George, and Erika Biguerel. I love you all.

Finally, I would like to appreciate my friend Abraham for his financial support and encouragement from the very beginning of this journey to the end. Indeed, without your support, this journey would not have been a reality. I will also like to appreciate Zeal Bala Sunday, Caleb Wayo Silas, Moses Ogaba, and Mr. and Mrs. Jerry Emmanuel Dauda for their support and encouragement.

TABLE OF CONTENT

CONTENTS	PAGES
COVER PAGE	i
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
DEDICATION	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENT	viii
LIST OF FIGURES	xii
LIST OF TABLES	xiii
LIST OF ABBREVIATIONS	xiv
ABSTRACT	xv
CHAPTER ONE: INTRODUCTION	1
1.1 Background to the Study	1
1.2 Statement of the Problem	3
1.3 Aim and Objectives of the Study	4
1.4 Justification for the Study	4
1.5 Scope of the Study	5
CHAPTER TWO: LITERATURE REVIEW	6
2.1 The Overview of Anopheles Gambiae	6
2.2 Insecticides	7
2.3 Insecticide Resistance	8
2.4 Mechanism of Insecticide Resistance	9
2.4.1 Knockdown Mutation	10
2.4.2 Metabolic Mutation	10
2.4.3 Other Mechanism	11
2.5 Vector Control Intervention Programs	12
2.5.1 Long-lasting Insecticide-treated Nets	12

2.5.2	Indoor Residual Spray	12
2.6	Omics Technology Driven Information	13
2.6.1	Genomics	13
2.6.2	Transcriptomics	13
2.6.3	Proteomics	14
2.6.4	Metabolomics	14
2.6.5	Interactomics	15
2.7	System Biology	15
2.7.1	The Prospects of System Biology	16
2.7.2	Experimental System Biology	16
2.7.3	Computational System Biology	18
2.8	Network Biology	19
2.9	Categories of Biological Network	19
2.9.1	Data Driven Network	19
2.9.2	Knowledge Derived Network	20
2.10	Framework of Biological Networks	20
2.11	Molecular Interaction Networks	21
2.11.1	Protein-Protein Interaction Network	21
2.11.2	Regulatory Network	22
2.11.3	Metabolic Network	23
2.11.4	Signalling Network	24
2.12	Topology-based Approaches In Network Biology	24
2.13	Centrality Measures	25
2.13.1	Degree centrality	26
2.13.2	Closeness centrality	27
2.13.3	Eigenvector centrality	27
2.13.4	Betweenness centrality	27
2.13.5	Eccentric centrality	28
2.14	Network Clustering	28
2.14.1	Graph-based and Topology-based Techniques	28
2.15	Network Annotation	29
2.16	Interaction Databases	30

2.17	Review of Related Studies	32
CHAPTER THREE: METHODOLOGY		36
3.1	Collection of Insecticide Related Resistance Gene	36
3.2	Construction of PPI Network and Integration	37
3.3	Centrality Analysis	37
3.3.1	Degree Centrality	37
3.3.2	Closeness Centrality	38
3.3.3	Eigenvector Centrality	38
3.3.4	Betweenness Centrality	38
3.4	Clustering Analysis	39
3.5	Gene Ontology and KEGG Pathway Enrichment Analysis	39
CHAPTER FOUR: RESULTS AND DISCUSSIONS		41
4.1	Results	41
4.1.1	Protein-Protein Interaction Network	41
4.1.2	Centrality Measures Analysis	42
4.1.3	Clustering Analysis	46
4.1.4	Functional Enrichment Analysis	47
4.2	Discussions	50
4.2.1	Anopheles gambiae insecticide resistance PPIN	50
4.2.2	Topological properties analysis	50
4.2.3	Network Analysis of Densely Connected Modules	53
4.2.4	Functional Enrichment Analysis	55
CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS		58
5.1	Conclusion	58
5.2	Contribution to Knowledge	58
5.3	Recommendation	59
5.4	Limitation of the Study	59
5.5	Areas for Further Research	59
REFERENCES		60

APPENDIX A	72
APPENDIX B	81
APPENDIX C	81

LIST OF FIGURES

FIGURES	TITLE OF FIGURES	PAGES
2.1	The life cycle of <i>An. gambiae</i> .	7
2.2	Information flow within a cell.	17
2.3	An example of a top-down network based on high throughput data.	20
2.4	An example of a bottom-up network based on genome-scale metabolic reconstructions.	20
2.5	An illustration of the general structure of biological network investigations using a flowchart	21
2.6	The interaction network biological centrality metrics.	26
3.1	A workflow for the analysis of resistance mechanisms in <i>Anopheles gambie</i>	36
4.1	<i>An gambiae</i> Insecticide resistance protein-protein interaction network.	42
4.2	Protein-protein interaction clusters formed after performing MCODE, OH-PIN, and IPCA clustering analysis.	47
4.3	First neighbors of the top hub genes	49
4.4	Bubble plot showing the result of GO enrichment analysis of insecticide resistance-related proteins in <i>An. gambiae</i> .	56
4.5	Barplot showing the result of GO enrichment analysis of insecticide resistance-related proteins in <i>An. gambiae</i> .	57

LIST OF TABLES

TABLES	TITLE OF TABLES	PAGES
4.1	List of seed proteins used to retrieve interacting proteins at stringency >0.5.	41
4.2	The proteins and DC value of the top 10 proteins in <i>An. gambiae</i> insecticide resistance	43
4.3	The proteins and CCx value of the top 10 proteins in <i>An. gambiae</i> insecticide resistance	44
4.4	The proteins and EC value of the top 10 proteins in <i>An. gambiae</i> insecticide resistance	45
4.5	The proteins and BC value of the top 10 proteins in <i>An. gambiae</i> insecticide resistance	46
4.6	The top 10 centrality ranking in <i>An. gambiae</i> insecticide resistance network	53

LIST OF ABBREVIATIONS

PPI	Protein-Protein Interaction
PPIN	Protein-Protein Interaction Network
CC	Cellular Component
BP	Biological Process
MF	Molecular Function
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
DAVID	Database for Annotation, Visualization, and Integrated Discovery
BC	Betweenness Centrality
CCx	Closeness Centrality
DC	Degree Centrality
EC	Eigenvector Centrality
LLINs	Long-Lasting Insecticide-Treated Nets
IRS	Indoor Residual Spraying
ITNs	Insecticide-Treated Nets

ABSTRACT

One of the deadliest diseases affecting millions of people annually, particularly in Africa, and has claimed millions of lives is malaria. It is caused by the malaria pathogen *Plasmodium falciparum*, during a blood meal by a female *Anopheles* vector. Although several control programs were established to enable the effective control and management of the disease by targeting the malaria vector using insecticides, there has been an increase in resistance to the currently existing insecticides. One of the challenges in effectively controlling malaria is because the molecular mechanisms by which malaria vector evades insecticide effect remain unclear. This could have enabled the development of target-specific insecticides. This study aims to elucidate the complex underlying mechanisms *Anopheles gambiae* evade currently existing insecticides. The *An. gambiae* PPIN was constructed using Cytoscape, and topology and clustering analysis was conducted to predict key proteins and pathways that are significant in the molecular mechanisms of *An. gambiae*. The network's GO and KEGG enrichment analysis was conducted using DAVID. RStudio was used to visualize the network. The network consists of 224 proteins (nodes) and 2115 interactions (edges). The network topology analysis reveals three essential proteins (GSTE2, AGAP002945, and AGAP004581) that are hub genes, and when either of them is removed, the network will crash. The GO and KEGG analysis showed three important biological processes (glutathione metabolism, protein folding and refolding, and chaperone-mediated protein folding) and three pathways (cytochrome P450s, glutathione, and protein processing pathways) were significantly enriched in the vector's insecticide resistance molecular mechanisms. The analysis showed that the proteins associated with insecticide resistance in *An. gambiae* were confined with the cytoplasm, cytosol, and plasma membrane. From the Study, five essential candidate target proteins (GSTE1, GSTE2, GSTD3, AGAP002945, and AGAP004581) and three molecular pathways (cytochrome P450s, glutathione, and protein processing pathways) were found to be enriched in *An. gambiae*. It is anticipated that this will guide future research into developing novel target-specific insecticides to combat malaria.

Keywords: *Insecticide resistance, Protein-protein interaction network, Centrality analysis, Clustering analysis, Anopheles gambiae.*