

**MUTATION PROFILING OF *RBCL* AND *MATK* GENES AND
PHYLOGENETIC RELATIONSHIPS OF SELECTED MEMBERS OF
THE FABACEAE FAMILY**

**OBADIRE, TOBI TEJUMADE
(22PCO02402)**

**B. Tech, Agricultural Technology, Federal University of Technology Akure,
Ondo State, Nigeria**

AUGUST, 2024

**MUTATION PROFILING OF *RBCL* AND *MATK* GENES AND
PHYLOGENETIC RELATIONSHIPS OF SELECTED MEMBERS OF
THE FABACEAE FAMILY**

BY

**OBADIRE, TOBI TEJUMADE
(22PCO02402)**

**B. Tech, Agricultural Technology, Federal University of Technology Akure,
Ondo State, Nigeria**

**A DISSERTATION SUBMITTED TO THE SCHOOL OF
POSTGRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE (M.Sc)
DEGREE IN BIOLOGY IN THE DEPARTMENT OF BIOLOGICAL
SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT
UNIVERSITY, OTA, OGUN STATE, NIGERIA**

AUGUST, 2024

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Master of Science in Biology in the Department of Biological 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, **OBADIRE, TOBI TEJUMADE (22PCO02402)**, declare that I carried out this research under the supervision of Prof. Olawole O. Obembe of the Department of Biological Sciences, College of Science and Technology, Covenant University, Ota, Nigeria. I attest that the dissertation has not been presented wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

OBADIRE, TOBI TEJUMADE

Signature and Date

CERTIFICATION

We certify that this dissertation titled “**MUTATION PROFILING OF RBCL AND MATK GENES AND PHYLOGENETIC ANALYSIS OF SELECTED MEMBERS OF THE FABACEAE FAMILY**” is an original research work carried out by **OBADIRE, TOBI TEJUMADE (22PCO02024)** in the Department of Biological Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Prof. Olawole O. Obembe. We have examined and found this work acceptable as part of the requirements for the award of Master of Science in Biology (BIOTECHNOLOGY AND BIOINFORMATICS).

Prof. Olawole O. Obembe
(Supervisor)

Signature and Date

Dr. Opeyemi I. Ayanda
(Head of Department)

Signature and Date

Prof. Adegoke E. Adegbite
(External Examiner)

Signature and Date

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

Signature and Date

DEDICATION

I dedicate this work to the Almighty God, my help in ages past, and to my parents.

ACKNOWLEDGMENTS

All thanks and praises be to God Almighty, the giver of life with whom I made it this far; his name alone is highly exalted.

My appreciation goes to the Chancellor, Bishop David Oyedepo; the current Vice-Chancellor, Professor Abiodun H. Adebayo; the Acting Registrar, Mrs Tobi-David and the entire management and staff of Covenant University for providing me with the facilities for the successful completion of my program.

To the current Dean of the School of Postgraduate Studies, Professor Akan Williams, and the Postgraduate School staff, I appreciate your diverse training and programs organized towards capacity building for successful postgraduate research. Thank you to the Head of the Department of Biological Sciences, Dr Opeyemi I. Ayanda, and the department staff for their constant help.

My most profound gratitude and heartfelt appreciation goes to my supervisor, Prof. Olawole O. Obembe, for his guidance and patience throughout this research. I am forever grateful to you, Sir.

I want to acknowledge Dr. Adetutu Bello and Dr. Rotimi Aworunse for their guidance, constant help, and support during this period.

I appreciate my parents, Mr and Mrs. Obadire, for their unrelenting support, prayers and constant encouragement. I also appreciate my uncles, the Oladoyins, my aunty, Mrs Elufisan, and my grandma, Mrs Dorcas Oladoyin, for their financial support in attaining this degree. I want to thank all my friends and loved ones whose love, help, input and support during this time were unwavering. I am grateful to you all; thank you.

Finally, I thank all members of the Department of Biological Sciences for their support and tutelage, and also, special thanks go to my colleagues and senior colleagues; God bless you all.

TABLE OF CONTENTS

| CONTENTS | PAGES |
|--|----------|
| ACCEPTANCE | iii |
| DECLARATION | iv |
| CERTIFICATION | v |
| DEDICATION | vi |
| ACKNOWLEDGMENTS | vii |
| TABLE OF CONTENTS | viii |
| LIST OF APPENDIXES | xi |
| LIST OF FIGURES | xiii |
| LIST OF PLATES | xiv |
| ABSTRACT | xv |
| | |
| CHAPTER ONE | 1 |
| INTRODUCTION | 1 |
| 1.1 Background of the Study | 1 |
| 1.2 Statement of the Research Problem | 3 |
| 1.3 Research Questions | 4 |
| 1.4 Aim and Objectives | 4 |
| 1.5 Justification of the Study | 5 |
| 1.6 Scope of the Study | 5 |
| 1.7 Limitations of the Study | 5 |
| | |
| CHAPTER TWO | 6 |
| LITERATURE REVIEW | 6 |
| .1 Fabaceae Family | 6 |
| 2.2 Subfamilies of the Fabaceae | 7 |
| 2.3 Grain Legumes | 8 |
| 2.3.1 Minor Grain Legume | 9 |
| 2.3 Domestication | 10 |
| 2.4 Studied Genera and Species | 13 |
| 2.4.1 Genus <i>Vigna</i> | 14 |
| 2.4.2 Genus <i>Phaseolus</i> | 18 |
| 2.4.3 Genus <i>Canavalia</i> | 20 |
| 2.4.5 Hyacinth Bean | 23 |
| 2.4.6 Genus <i>Macrotyloma</i> | 24 |
| 2.4.7 Velvet Bean | 26 |
| 2.4.8 Genus <i>Sphenostylis</i> | 27 |
| 2.9.4-Winged Bean | 29 |
| 2.4.10 Genus <i>Cajanus</i> | 29 |
| 2.5 Morphological Properties of Fabaceae | 31 |
| 2.6 Molecular Studies of Fabaceae | 31 |
| 2.7 Phylogenetic Studies | 33 |
| 2.7.1 Morphological Marker | 33 |
| 2.7.2 Phylogenetic Markers | 34 |

| | |
|--|-----------|
| 2.7.3 Molecular Markers | 34 |
| 2.8 Barcoding | 35 |
| 2.8.1 Ribulose-Bisphosphate Carboxylase Gene (rbcL) | 36 |
| 2.8.2 Maturase Kinase | 37 |
| 2.9 Mutation | 37 |
| 2.9.1 Single Nucleotide Polymorphism (SNP) | 38 |
| 2.9.2 Single Amino Acid Substitution | 38 |
| 2.9.3 Related Work and Research Gaps | 38 |
| CHAPTER THREE | 41 |
| 3.0 MATERIALS AND METHOD | 41 |
| 3.1 Materials: | 41 |
| 3.1.1 Source of Samples | 41 |
| 3.1.1 Plant Studied | 41 |
| 3.1.2 Equipment | 41 |
| 3.1.3 Hardware | 41 |
| 3.1.4 Reagents | 42 |
| 3.2 Methods: | 42 |
| 3.2.1 Data Mining | 42 |
| 3.2.2 Planting | 42 |
| 3.2.3 Genomic DNA Extraction | 43 |
| 3.2.4 Primer Designing | 44 |
| 3.2.5 Quantification | 44 |
| 3.2.6 Gene Amplification (Polymerase Chain Reaction) | 44 |
| 3.2.7 Gel Electrophoresis | 44 |
| 3.2.8 Sequencing | 44 |
| 3.3 Models/ Softwares Employed in the Study | 45 |
| 3.3.1 Sequence Analysis | 46 |
| 3.3.2 Mutation Profiling | 47 |
| 3.3.3 Phylogenetic Analysis | 49 |
| 3.4 Methods of Statistical Analysis | 50 |
| CHAPTER FOUR | 51 |
| 4.0 RESULTS | 51 |
| 4.1 Data Exploration and Pre-Processing | 51 |
| 4.2 Blast Search Result for <i>Macrotyloma geocarpum</i> rbcL gene | 51 |
| 4.3 Blast Search Result for <i>Macrotyloma geocarpum</i> matK gene | 51 |
| 4.4 Gel electrophoresis result | 52 |
| 4.5 Conservation and Mutation Investigation | 54 |
| 4.6 Investigation of the Effect of SAAS on Protein Function | 67 |
| 4.7 Phylogenetic Studies | 69 |
| CHAPTER FIVE | 75 |
| 5.0 DISCUSSION | 75 |
| CHAPTER SIX | 78 |

| | |
|--------------------------------------|------------|
| CONCLUSION AND RECOMMENDATION | 78 |
| 6.1 Summary of the Study | 78 |
| 6.2 Conclusion | 78 |
| 6.3 Recommendation | 79 |
| 6.4 Contributions to Knowledge | 79 |
| REFERENCES | 80 |
| APPENDIX | 101 |

LIST OF APPENDIXES

| APPENDIXES | LIST OFAPPENDIXES | PAGES |
|------------|---|-------|
| 1: | Consensus sequences of matK gene amino acid sequences retrieved | 101 |
| 2: | Consensus sequences of rbcL gene amino acid sequences retrieved | 103 |
| 3: | Genomic nucleotide and translated amino acid sequence of the matK gene of <i>M. geocarpum</i> | 104 |
| 4: | Genomic nucleotide and translated amino acid sequence of the rbcL gene of <i>M. geocarpum</i> | 105 |

LIST OF TABLES

| TABLES | LIST OF TABLES | PAGES |
|---------------|--|--------------|
| 2.1: | A summary of the taxonomic information of the studied plants | 13 |
| 2.2: | Sequenced genome of some members of the Fabaceae family | 32 |
| 2.3: | Gaps identified in the literature and expected to be filled | 40 |
| 3.1: | Primer Sequences | 45 |
| 3.2: | Condition for Polymerase Chain Reaction | 45 |
| 4.1: | List of plant species and number of retrieved sequences used in this study | 53 |
| 4.2: | List of species with Refseq and complete chloroplast genome | 54 |
| 4.3: | Mutation profiling of matK gene | 64 |
| 4.4: | Mutation profiling of rbcL gene | 65 |
| 4.5: | Prediction of the SAASs in the matK gene using three prediction software | 68 |
| 4.6: | Prediction of the SAASs in the rbcL gene using three prediction software | 69 |

LIST OF FIGURES

| FIGURES | LIST OF FIGURES | PAGES |
|---------|---|-------|
| 4.1: | PCR amplified rbcL and matK genes fragment of <i>M. geocarpum</i> | 52 |
| 4.2: | SAAS of matK gene in <i>Lablab purpureus</i> | 56 |
| 4.3: | SAAS of matK gene in <i>Cajanus cajan</i> | 56 |
| 4.4: | SAAS in matK gene in <i>Macrotyloma uniflorum</i> | 57 |
| 4.5: | SAAS in matK gene in <i>Phaseolus lunatus</i> | 58 |
| 4.6: | SAAS in matK gene in <i>Phaseolus vulgaris</i> | 58 |
| 4.7: | SAAS in matK gene in <i>Vigna heterophylla</i> | 59 |
| 4.8: | SAAS in rbcL gene in <i>Mucuna pruriens</i> | 59 |
| 4.9: | SAAS in rbcL gene in <i>Phaseolus lunatus</i> | 60 |
| 4.10: | SAAS in rbcL gene in <i>Phaseolus vulgaris</i> | 60 |
| 4.11: | SAAS in rbcL gene in <i>Phosphocarpus tetragonolobus</i> | 61 |
| 4.12: | SAAS in rbcL gene in <i>Vigna heterophylla</i> | 61 |
| 4.13: | SAAS in rbcL gene in <i>Vigna radiata</i> | 62 |
| 4.14: | SAAS in rbcL gene in <i>Vigna unguiculata</i> | 62 |
| 4.15: | SAAS in rbcL gene in <i>Macrotyloma uniflorum</i> | 63 |
| 4.16: | Gene substitution rate and distribution | 66 |
| 4.17: | Neighbor-joining tree of selected Fabaceae species based on matK+rbcL sequences | 71 |
| 4.18: | Neighbor-joining tree of selected Fabaceae species based on chloroplast genome | 71 |
| 4.19: | Neighbor-Joining tree of selected Fabaceae species based on matK dataset | 72 |
| 4.20: | Maximum Likelihood tree of selected Fabaceae species based on matK dataset | 73 |
| 4.21: | Maximum Likelihood tree of selected Fabaceae species based on matK dataset | 73 |
| 4.22: | Neighbor-Joining tree of selected Fabaceae species based on rbcL dataset | 73 |
| 4.23: | Maximum Parsimony tree of selected Fabaceae species based on rbcL dataset | 74 |
| 4.24: | Maximum Likelihood tree of selected Fabaceae species based on rbcL dataset | 74 |

LIST OF PLATES

| PLATES | LIST OF PLATES | PAGES |
|---|----------------|-------|
| 2.1: Photographs of Cowpea plant | | 16 |
| 2.2: Photograph of Mung bean | | 17 |
| 2.3: Photograph of Bambara groundnut | | 18 |
| 2.4: Photograph of Common/Kidney bean | | 19 |
| 2.5: Photograph of Lima bean | | 20 |
| 2.6: Photograph of Jack bean | | 21 |
| 2.7: Photograph of Sword beans | | 22 |
| 2.8: Photograph of Groundnut | | 23 |
| 2.9: Photograph of Hyacinth Beans | | 24 |
| 2.10: Photograph of Kersting groundnut | | 25 |
| 2.11: Photograph of Horse gram | | 26 |
| 2.12: Photograph of Velvet bean | | 27 |
| 2.13: Photograph of African Yam Bean | | 28 |
| 2.14: Photograph of Winged bean | | 29 |
| 2.15: Photograph of Pigeon pea | | 30 |
| 3.1: Photograph of <i>Macrotyloma geocarpum</i> planted | | 42 |

ABSTRACT

Gene mutation profiling involves detecting the type and effect of mutation on a gene. The *rbcL* and *matK* genes are chloroplast genes involved in photosynthesis; mutations on these genes can be deleterious or beneficial to the plant. Integrating underutilized legumes into genetic studies alongside major legumes using barcoding methodologies is essential for their genetic improvement. This research aimed to provide detailed mutation information on the *rbcL* and *matK* genes of the selected legumes and gain insight into their classification based on DNA barcodes. The sequences of sixteen legumes were used in the study. Five amino acid sequence entries from these species, reference sequences and complete chloroplast genomes were retrieved from the National Center for Biotechnology Information database. The gene sequences were aligned with their reference sequences using MULTALIN to identify amino acid substitutions in the genes. The type and effect of these mutations on protein function and stability were predicted using the SIFT, PROVEAN, and I-MUTANT v.2.0 online servers. The *rbcL* and *MatK* genes of *Macrotyloma geocarpum* were isolated, amplified, and sequenced. Consensus sequences of the sequence entries from 15 species were generated and aligned with those sequenced using the Muscle algorithm. Phylogenetic analysis was conducted using the neighbor-joining algorithm in MEGA 11 software. Results revealed 32 and 22 amino acid substitutions in the *matK* and *rbcL* genes, respectively. Five substitutions in the *matK* gene across three species were predicted to be deleterious by all three predictive software, and one substitution in the *rbcL* gene was predicted to be beneficial. The double barcodes (*matK+rbcL*) and single barcode (*rbcL*) classified the selected legumes randomly, however, *matK* gene classified the legumes down to the genus and species levels. It also revealed that *Cajanus cajan*, an underutilized legume, is the closest to *Vigna unguiculata*, implying the possibility of genetically improving *C. cajan* using the *V. unguiculata* genome. The *matK* gene also gave similar result to that of chloroplast genome. Consequently, *matK* gene is recommended as a core barcode sequence in the Fabaceae family.

Keywords: *Mutation, rbcL, matK, Underutilized legumes, Phylogenetic Analysis*