

***IN SILICO* MUTATION PROFILING OF *RBCL* GENE AND
CLASSIFICATION OF SELECTED EDIBLE MEMBERS OF THE
CUCURBITACEAE FAMILY**

**OLORUNSOLA, HONEY AANU
(20PCO02179)**

JULY, 2022

***IN SILICO* MUTATION PROFILING OF *RBCL* GENE AND
CLASSIFICATION OF SELECTED EDIBLE MEMBERS OF THE
CUCURBITACEAE FAMILY**

BY

**OLORUNSOLA, HONEY AANU
(20PCO02179)**

B.Sc. Plant Science and Biotechnology, Ekiti State University, Ado-Ekiti

**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 APPLIED
BIOLOGY AND BIOTECHNOLOGY IN THE DEPARTMENT OF
BIOLOGICAL SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY,
COVENANT UNIVERSITY**

JULY, 2022

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 Applied Biology and Biotechnology in the Department of Biological Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.

Mr. Taiwo B. Erewunmi
(Secretary, School of Postgraduate Studies)

Signature and Date

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

Signature and Date

DECLARATION

I, **OLORUNSOLA, HONEY AANU (20PCO02179)** declare that this research was carried out by me 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 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.

OLORUNSOLA, HONEY AANU

Signature and Date

CERTIFICATION

We certify that this dissertation titled “**IN SILICO MUTATION PROFILING OF *RBCL* GENE AND CLASSIFICATION OF SELECTED EDIBLE MEMBERS OF THE CUCURBITACEAE FAMILY**” is an original research work carried out by **HONEY, AANU OLORUNSOLA (20PCO02179)** 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 Applied Biology and Biotechnology.

Prof. Olawole O. Obembe
(Supervisor)

Signature and Date

Prof. Solomon U. Oranusi
(Head of Department)

Signature and Date

Prof. Abiodun E. Ayodele
(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, Mr Emmanuel Igbani, and the entire management and staff of Covenant University for providing me 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 you for the diverse trainings and programmes organised towards capacity building for successful postgraduate research. And, a very big thank you goes to the Head of the Department of Biological Sciences, Professor Solomon U. Oranusi and the staff for the constant help.

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

I want to acknowledge Dr. Opeyemi Soremekun, Dr. Isewon Itunu, Dr. Oduselu Gbolahan, Dr. Jacob Popoola and Mr. Rotimi Aworunse for their guidance, constant help and support during this period.

I appreciate my parents, Mr. Gbenga Olorunsola and Mrs. Modupe Olorunsola for their unrelenting support, prayers and encouragement at all times.

I want to appreciate all my friends and loved ones whose love, help, input and support during this time was unwavering. I am grateful to you all, thank you.

Finally, I say a big thank you to 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

CONTENT	PAGES
COVER PAGE	i
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
DEDICATION	vi
ACKNOWLEDGMENTS	vii
TABLE OF CONTENTS	viii
LIST OF APPENDIX	xii
LIST OF FIGURES	xiii
LIST OF TABLES	xiv
ABSTRACT	xv
CHAPTER ONE: INTRODUCTION	1
1.1 Background of the study	1
1.2 Statement of research problem	9
1.3 Research questions	9
1.4 Aim and objectives	10
1.5 Research justification	10
1.6 Scope of study	10
1.7 Limitation of study	11
CHAPTER TWO: LITERATURE REVIEW	12
2.1 Cucurbitaceae	12
2.1.1 Domestication	12
2.2 Subfamilies of the Cucurbitaceae	13
2.3 Studied genera	13
2.3.1 Genus <i>Cucumis</i>	15
2.3.1.1 Melon	15
2.3.1.2 Cucumber	16
2.3.2 Genus <i>Cucurbita</i>	16
2.3.2.1 Pepo	17
2.3.2.2 Winter squash	18
2.3.2.3 Butternut squash	19
2.3.3 Genus <i>Citrullus</i>	19

2.3.3.1	Watermelon	20
2.3.3.2	Colocynth	21
2.3.4	Genus <i>Lagenaria</i>	22
2.3.4.1	Bottle gourd	22
2.3.5	Genus <i>Telfairia</i>	23
2.3.5.1	Fluted pumpkin	23
2.3.6	Genus <i>Coccinia</i>	24
2.3.6.1	Ivy gourd	25
2.3.7	Genus <i>Momordica</i>	25
2.3.7.1	Bitter gourd	26
2.4	Morphological properties of the Cucurbitaceae family	27
2.5	Ethnobotanical importance of the Cucurbitaceae Family	27
2.6	Other economic importance of the Cucurbitaceae family	28
2.7	Molecular studies of the Cucurbitaceae family	28
2.8	Phylogenetic studies	29
2.8.1	Phylogenetic markers	29
2.8.1.1.	Morphometric markers	30
2.8.1.2.	Molecular markers	30
2.8.1.2.1.	Advantages of molecular markers	30
2.9.	Barcoding	31
2.9.1.	Why <i>rbcL</i> ?	32
2.10.	Mutation	33
2.10.1.	Single nucleotide polymorphism	34
2.10.2.	Single amino acid substitution	35
2.10.3.	Mutation profiling in other plants	36
2.11.	Related works and research gaps	37
CHAPTER THREE: MATERIALS AND METHODS		39
3.1.	Materials	39
3.1.1.	Equipment	39
3.1.2	Hardware	39

3.1.3	Reagents	39
3.2.	Data mining	39
3.3.	Methods	39
3.3.1.	Sources of Sample and Identification	39
3.3.2	Sample Collection and Preparation	40
3.3.3	DNA Extraction and Quantification	40
3.3.4.	Gene Amplification	40
3.3.5	Gel electrophoresis	44
3.3.6	Sanger sequencing	44
3.3.7	Data analysis and data mining	44
3.4	Models/ software employed in the study	46
3.4.1	Molecular Evolutionary Genetics Analysis (MEGA) software	46
3.4.2	Plant Protein Variation Effect Detector (PPVED) software	46
3.4.3.	I-mutant server	47
3.4.4.	BLASTp	47
3.4.5.	I-tasser server	47
3.4.6.	Molprobit server	47
3.4.7.	Swiss model server	48
3.5	Methods of statistical analysis	48
CHAPTER FOUR: RESULTS		49
4.1	Sample analysis	49
4.2	Conversion of sequence	49
4.3	Blast search result for <i>Telfairia occidentalis rbcL</i> gene	49
4.4	Data exploration and pre-processing	49
4.5	BLASTp and comparism analysis	49
4.6	Conservation and mutation investigation	49
4.7.	SAAS type investigation	56
4.8.	Structure prediction	59
4.9.	Structure validation	63

4.10. Phylogenetic studies	65
CHAPTER FIVE: DISCUSSION	68
5.1 Data mining and processing	68
5.2 <i>RbcL</i> conservation studies	68
5.3 Implications of mutation on the <i>rbcL</i> gene	69
5.4 Evolutionary relationship and position of <i>T. occidentalis</i>	70
CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS	71
6.1 Summary of findings	71
6.2 Conclusion	71
6.3 Contributions to Knowledge	72
6.4 Recommendations	72
REFERENCES	73
APPENDIX	94

LIST OF APPENDIX

Appendix	Title of Appendix	Pages
1	Details on the amino acid sequences retrieved from the NCBI	94
2	Genomic and translated amino acid sequence of the <i>rbcL</i> gene of <i>T. occidentalis</i>	103
3	Sequence information of the database match of <i>T. occidentalis</i>	104
4	Prediction result of the SAAS A343X for all 19 amino acid substitution using two different softwares	105
5	Prediction result of the SAAS D10X for all 19 amino acid substitution using two different softwares	106
6	Prediction result of the SAAS I44X for all 19 amino acid substitution using two different softwares	107
7	Prediction result of the SAAS Q375X for all 19 amino acid substitution using two different softwares	108
8	Prediction result of the specific SAAS using two different softwares	109
9	Models validation of predicted structure for the <i>rbcL</i> of <i>T. occidentalis</i> using SAVES 6.0	110
10	Models validation of predicted structure for the database match using SAVES 6.0	111
11	Structure validation result of <i>T. occidentalis</i> using the Molprobit on Swiss model server	112
12	Structure validation result of <i>T. occidentalis</i> database match using the Molprobit on Swiss model server	113

LIST OF FIGURES

Figure	Title of Figures	Pages
2.1:	Leaf, flower, fruit and stem of <i>Cucumis melo</i>	15
2.2:	Leaf, flower, fruit and stem of <i>Cucumis sativus</i>	16
2.3:	Leaf, flower, fruit and stem of <i>Cucurbita pepo</i>	18
2.4:	Leaf, flower, fruit and stem of <i>Cucurbita maxima</i>	18
2.5:	Leaf, flower, fruit and stem of <i>Cucurbita moschata</i>	19
2.6:	Leaf, flower, fruit and stem of <i>Citrullus lanatus</i>	20
2.7:	Leaf, flower, fruit and stem of <i>Citrullus colocynthis</i>	21
2.8:	Leaf, flower, fruit and stem of <i>Lagenaria siceraria</i>	22
2.9:	Leaf, flower, fruit and stem of <i>Telfairia occidentalis</i>	24
2.10:	Leaf, flower, fruit and stem of <i>Coccinia grandis</i>	25
2.11:	Leaf, flower, fruit and stem of <i>Momordica charantia</i>	27
4.1:	MEGA11 interface showing the SAAS (a) D10X and (b) M228I in <i>Citrullus lanatus</i>	51
4.2:	MEGA11 interface showing the SAAS (a) T36- and (b) A343X in <i>Lagenaria siceraria</i>	52
4.3:	MEGA11 interface showing the SAAS (a) I44X and (b) Q375X in <i>Cucumis melo</i>	53
4.4:	MEGA11 interface showing the SAAS (a) F108V in <i>Cucumis sativus</i>	54
4.5:	SAVES 6.0 validation for the sequenced plant	60
4.6:	SAVES 6.0 validation for the database match	61
4.7:	I-TASSER predicted structure for (a) the database match and (b) the sequenced plant	62
4.8:	Molprobitry results comparism of the two predicted structures	64
4.9:	Maximum likelihood tree showing evolutionary relationships in Cucurbitaceae species and the phylogenetic position of <i>T. occidentalis</i>	66
4.10:	Maximum likelihood tree showing evolutionary relationships in Cucurbitaceae species, its close relatives species and the phylogenetic position of <i>T. occidentalis</i>	67

LIST OF TABLES

Tables	Title of Tables	Pages
1.1:	Sequenced genomes of members of the Cucurbitaceae	8
2.1:	A summary of the taxonomic information of the studied plants	14
2.2:	Gaps identified in literature and filled	38
3.2:	Primer sequences of <i>rbcL</i> gene	41
3.3:	Thermal cycling PCR conditions	42
3.4:	PCR reaction mix for <i>rbcL</i> gene	43
3.5:	Population information of the studied plant species	45
4.1:	Mutation profiling of <i>rbcL</i> of the ten selected Cucurbits	55
4.2:	Prediction of the SAASs in the <i>rbcL</i> gene using the two prediction software	57
4.3:	Conserved and mutated entries of the studied plant species	58

ABSTRACT

The Cucurbitaceae, with about 965 genera in 95 species is ranked next to the Solaneaceae; the largest family of fruits and vegetables. Its members are among the world's most significant plant foods. As at 2022, the whole genome sequence of 18 species of this family have been published. Ribulose biphosphate carboxylase large (*rbcL*) is central to the global carbon cycle and is one of the single barcode genes of individual plants that have phylogenetic utility. However, there has not been a documented detailed investigation into the *rbcL* gene mutation of this family. This research aimed at providing information on Single amino acid substitution (SAAS) of the *rbcL* gene of selected cucurbits and gaining insight into their classification. The studied populations are entries of *Citrullus lanatus*, *Citrullus colocynthis*, *Cucurbita pepo*, *Cucurbita maxima*, *Cucurbita moschata*, *Cucumis sativus*, *Cucumis melo*, *Telfairia occidentalis*, *Momordica charantia*, *Lagenaria siceraria*, and *Coccinia grandis*. Five amino acid sequence entries of these members (including their reference sequence) were retrieved from the RefSeq database of the National Centre for Biotechnology Information. The sequences of each species were compared with their reference sequence using MUSCLE algorithm of the MEGA11 to check for the presence of SAASs in the gene. The type and effect of mutation on protein stability were predicted using the PPVED v1.0 and I-MUTANT v.3.0 online server, respectively. The *rbcL* genomic DNA isolation, amplification, purification, and sequencing of *Telfairia occidentalis* were carried out. Multiple sequence alignment of the translated amino acid sequence alongside the references of the other plant species was carried out using MUSCLE algorithm of MEGA11. A phylogenetic tree was constructed using Maximum-likelihood method. The gene sequences of six species were conserved while four species had the presence of SAASs. F108A was recorded in *Cucumis sativus*, D10X and M228I were recorded in *Citrullus lanatus*, T36- and A343X were recorded in *Lagenaria siceraria*, and I44X and Q375X were recorded in *Cucumis melo*. A343X and F108A were predicted to be functional mutations with a decrease in the stability of the protein binding site. T36- is well established to be a deleterious mutation. Only *Citrullus lanatus* had non- functional mutations. In total, 45 entries of 10 species were screened and only five entries of four species had mutations (10.9%). The phylogenetic tree revealed that *T. occidentalis* is positioned within the Cucurbitaceae family, although, it had the lowest bootstrap confidence compared to other members. To authenticate this positioning, *rbcL* sequences of ten species belonging to four closely related families within the Order Cucurbitales were added to those of the Cucurbitaceae, and a new tree was constructed. *Telfairia occidentalis* was centrally positioned between the Cucurbitaceae and the close relatives. The newly constructed tree also revealed that *T. occidentalis* was not only the closest to the root of the tree, but it was also closer to the relatives than to Cucurbitaceae. Conclusively, this study validated the conservation of *rbcL* gene sequence amongst the populations screened. It also demonstrated that *T. occidentalis* classification within the Cucurbitaceae cannot be ascertained but should be further investigated using other molecular markers.

Keywords: *RbcL*, Phylogeny, Cucurbitaceae, Mutation, *Telfairia occidentalis*.