

**SEED LONGEVITY EVALUATION AND GENOME-WIDE
ASSOCIATION STUDIES ON SOME NUTRITIONAL TRAITS OF
AFRICAN YAM BEAN**

**OLUWOLE, OLUBUSAYO OLAYEMI
(18PCO01853)**

MAY, 2023

**SEED LONGEVITY EVALUATION AND GENOME-WIDE
ASSOCIATION STUDIES ON SOME NUTRITIONAL TRAITS OF
AFRICAN YAM BEAN**

BY

**OLUWOLE, OLUBUSAYO OLAYEMI
(18PCO01853)**

**B.Sc Microbiology, Bowen University, Iwo
M.Sc Industrial and Commercial Biotechnology, Newcastle University,
Newcastle upon Tyne, United Kingdom**

**A Ph.D THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE
STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR
THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D)
IN BIOLOGY IN THE DEPARTMENT OF BIOLOGICAL SCIENCES,
COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT
UNIVERSITY, OTA, OGUN STATE, NIGERIA**

MAY, 2023

ACCEPTANCE

This is to attest that this thesis has been accepted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy in Biology, in the Department of Biological Sciences, College of Science and Technology, Covenant University, Ota, Ogun State

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, OLUWOLE, OLUBUSAYO OLAYEMI (18PCO01853), declare that I carried out this research under the supervision of Prof. Olawole O. Obembe and Prof. Michael T. Abberton of the Department of Biological Sciences, College of Science and Technology, Covenant University, Ota, Nigeria and International Institute of Tropical Agriculture, Ibadan, Nigeria, respectively. I attest that the thesis 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 thesis are duly acknowledged.

OLUWOLE, OLUBUSAYO OLAYEMI

Signature and Date

CERTIFICATION

We certify that this thesis titled “**SEED LONGEVITY EVALUATION AND GENOME-WIDE ASSOCIATION STUDIES ON SOME NUTRITIONAL TRAITS OF AFRICAN YAM BEAN**” is an original research work carried out by **OLUWOLE, OLUBUSAYO OLAYEMI (18PCO01853)**, 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 and Prof. Michael T. Abberton. We have examined and found this work acceptable as part of the requirements for the award of Doctor of Philosophy (Ph.D) degree in Biology.

Prof. Olawole O. Obembe
(Supervisor)

Signature and Date

Prof. Michael T. Abberton
(Co-Supervisor)

Signature and Date

Prof. Solomon U. Oranusi
(Head of Department)

Signature and Date

Prof. Akinola R. Popoola
(External Examiner)

Signature and Date

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

Signature and Date

DEDICATION

This research work is dedicated to the Almighty God and my parents, Engr. Olufemi Oluwole and Mrs Florence Oluwole.

ACKNOWLEDGMENTS

I want to appreciate the Most High God, the custodian of all wisdom, knowledge, and understanding, for His grace and favour throughout the duration of carrying out my Ph.D programme. I acknowledge the University Chancellor and Chairman Board of Regents, Dr. David O. Oyedepo who had the vision of creating a suitable academic environment. I would also like to acknowledge the Vice-Chancellor, Prof. Abiodun H. Adebayo; the Acting Registrar, Mrs. Regina A. Tobi-David; the Dean, School of Postgraduate Studies, Prof. Akan B. Williams; the Sub-Dean School of Postgraduate Studies, Dr. Emmanuel O. Amoo; the Dean, College of Science and Technology, Prof. Timothy A. Anake and the Dean, Student Affairs, Mrs. Olushola E. Coker.

I would like to acknowledge the Head of Department, Prof. Solomon U. Oranusi for his contribution during my programme. My sincere appreciation goes to my Supervisor Prof. Olawole O. Obembe and co-Supervisor Prof. Michael T. Abberton for their support and mentorship given to me all through the stages of my research work. I would like to appreciate the Head of Unit, Dr. Samuel A. Ejoh; the Postgraduate coordinator, Dr. Olayemi O. Akinola; the previous Postgraduate coordinators, Prof. Conrad A. Omonhinmin and Dr. Isaac O. Ayanda; Dr Eze F. Ahuekwe; Dr. Jacob O. Popoola and the entire staff of the Department of Biological sciences for their support.

I would like to express my profound gratitude to Crop Trust in conjunction with GRC, IITA for sponsoring my Ph.D programme. I also want to appreciate Dr. Olaniyi A. Oyatomi and Dr. Rajneesh Paliwal for their contributions and guidance during my research work. I appreciate staff and colleague of Genetic Resources Centre, International Institute of Tropical agriculture. My sincere appreciation goes to my colleagues: Mr. Oluseyi E. Olomitutu, Mrs. Olorunnisola F. Salvador, Mr. Rotimi Aworunse, Dr. Temitope Jones and Dr. Ademola I. Aina to name a few, whose contributions were invaluable all through my postgraduate studies. God bless and reward you all abundantly.

I would like to appreciate Prof. Samuel U. Eyesan who encouraged me to pursue my education/career up to Ph.D level since my undergraduate days. I appreciate my darling husband and bestfriend, Mr. Adesile I. Obasa, for always being there for me. I also appreciate my ever

supportive parents (Engr. Olufemi O. Oluwole and Mrs Florence M. Oluwole) and my brother (Oluyemi O. Oluwole) for their continuous support.

TABLE OF CONTENTS

CONTENTS	PAGES
COVER PAGE	i
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
ACKNOWLEDGMENTS	vii
TABLE OF CONTENTS	ix
LIST OF TABLES	xiii
LIST OF FIGURES	xv
LIST OF PLATES	xvi
LIST OF APPENDICES	xvii
LIST OF ABBREVIATIONS	xviii
ABSTRACT	xxi
CHAPTER ONE: INTRODUCTION	1
1.1 Background to the Study	1
1.2 Statement of Research Problem	5
1.3 Research Questions	6
1.4 Aim and Objectives:	6
1.5 Justification for the Study	6
1.6 Scope of Study	7
CHAPTER TWO: LITERATURE REVIEW	8
2.1 Legumes	8
2.1.1 Legume-rhizobium symbiosis	9
2.2 Legumes in Africa	9
2.3 Origin and Distribution of African yam bean (AYB)	10
2.4 African yam bean Taxonomic Classification and Morphology	12
2.5 Nutritional Value	12
2.5.1 Seed Protein Determination	14
2.5.2 Seed Oil Determination	15
2.5.3 Seed Starch Determination	18
2.6 Limitations in African yam bean	18
2.7 Reasons why legume cultivation is more sustainable than cultivation of other staple crops	19

2.8 Germplasm Conservation	21
2.8.1 <i>In Situ</i> Germplasm Conservation	22
2.8.2 <i>Ex Situ</i> Germplasm Conservation	22
2.8.3 The Position of Seed Banks in Crop Biodiversity Conservation	23
2.8.4 Seed storage and conservation	24
2.8.5 Seed longevity	25
2.9 The Plant Genome	30
2.10 The Utilization of Molecular Markers/Biotechnology in African yam bean	30
2.10.1 Random Amplified Polymorphic DNA (RAPD)	31
2.10.2 Amplified Fragment Length Polymorphism (AFLP)	32
2.10.3 Simple Sequence Repeats (SSR) and Inter-Simple Sequence Repeat (ISSR) Markers	33
2.10.4 Single Nucleotide Polymorphism (SNP)	33
2.10.5 DArT Seq	34
2.11 Association Mapping	34
2.11.1 Quantitative Trait Loci Mapping	35
2.11.2 Genome-wide Association Studies (GWAS)	35
2.11.2.1 Common Bean as a Reference Genome	38
2.11.3 Phenome-wide association study (PheWAS)	38
2.11.4 GPWAS	39
2.11.5 MAS	39
2.12 The potential limitations of applying biotechnological techniques	39
2.13 Challenges and prospects of improving nutritional traits	40
2.14 Strategies to promote AYB utilization and consumption	40
2.15 Research Gaps	42
CHAPTER THREE: MATERIALS AND METHODS	43
3.1 Materials, consumables, reagents and equipment	43
3.1.1 Primary plant materials	43
Passport data of AYB accessions	43
3.1.2 Equipment	51
3.1.3 Reagents	51
3.2 Methods	51
3.2.1 Molecular and biochemical analysis	51

3.2.1.1 Biochemical analysis	52
3.2.1.1.1 Seed oil analysis	52
3.2.1.1.2 Protein Analysis	53
3.2.1.1.3 Starch Analysis	53
3.2.1.2 Genotyping	53
3.2 Seed Longevity Test	55
3.3 Statistical Analysis	58
3.4 Candidate gene selection and expression analysis	59
CHAPTER FOUR: RESULTS	60
4.1 Nutritional Composition	60
4.1.1 Phenotypic variation and correlation coefficient of phenotypic data	60
4.2 Genome-wide Association Studies (GWAS)	67
4.2.1 Candidate gene analysis	75
4.2.1.1. Identification of candidate genes closely located to significant SNP markers for oil synthesis	75
4.2.1.2. Identification of candidate genes closely located to significant SNP markers for protein synthesis	79
4.2.1.3. Identification of candidate genes closely located to significant SNP markers for starch synthesis	81
4.2.1.4. Significant SNP markers not found on the legume database	83
4.2.1.5. Identification of candidate genes closely located to significant SNP markers for other traits observed in the genomic region	85
4.3 Seed Longevity	88
4.3.1 Seed drying	88
4.3.2 Survival Curve	95
CHAPTER FIVE DISCUSSION	102
5.1 Nutritional Analysis	102
5.2 Genome-wide Association Studies (GWAS)	103
5.2.1 Candidate genes associated with oil QTL region	104
5.2.2 Candidate Genes Associated with Protein QTL region	104
5.2.3 Candidate Genes Associated with Starch QTL region	105
5.3 Seed Longevity	106

CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS	109
6.1 Summary	109
6.2 Conclusion	110
6.3 Contributions to Knowledge	110
6.4 Recommendations	111
6.5 Limitations to this Study	111
REFERENCES	112
APPENDICES	132

LIST OF TABLES

TABLES	TITLE OF TABLES	PAGES
1.1:	Research gaps	45
3.1:	African yam bean Passport Data showing the country of origin, acquisition date and collector's name	47
4.1:	Descriptive statistics for seed protein, oil and starch content of African yam bean in two planting seasons	63
4.2:	Least square means for seed oil, protein and starch content of African yam bean in two planting seasons	65
4.3:	Association analysis of oil content using the mixed linear model (MLM) across 2018/2019 and 2019/2020 planting seasons	72
4.4:	Association analysis of protein content using the mixed linear model (MLM) across 2018/2019 and 2019/2020 planting seasons	74
4.5:	Association analysis of starch content using the mixed linear model (MLM) across 2018/2019 and 2019/2020 planting seasons	76
4.6:	Candidate genes associated with significant SNP markers for oil synthesis	79
4.7:	Candidate genes associated with significant SNP markers for protein synthesis	83
4.8:	Candidate genes associated with significant SNP markers for starch synthesis	85
4.9:	Unidentified SNP markers associated with some nutritional traits in AYB	86
4.10:	Genes associated with some other significant SNP markers in AYB	89
4.11:	Mean moisture content at harvest and after drying at two temperatures (17 °C and 45 °C) 2019/2020 planting season	93
4.12:	Mean moisture content after further drying at 17 °C for 7 days	94
4.13:	Effect of low (17 °C) and high temperature (45 °C) drying on AYB seed longevity (<i>p50</i>)	95

LIST OF FIGURES

FIGURES	TITLE OF FIGURES	PAGES
2.1:	Solid squares representing the geographical distribution of AYB in Africa	11
2.2:	Flowchart of conservation of crop genetic resources as the basis for food security	28
2.3:	Flow chart of GWAS from sample collection to QTL identification and gene characterization	37
3.1:	Flowchart of the steps involved in seed longevity experiment	56
4.1:	Manhattan and Q-Q plot for oil content	70
4.2:	Manhattan and Q-Q plots for protein content	72
4.3:	Manhattan and Q-Q plots for starch content	74
4.4:	AYB seeds with enhanced seed longevity when dried at 17 °C/15% RH for 7 days	97
4.5:	AYB seeds with enhanced seed longevity when dried at 45 °C/60% RH for 2 days	98
4.6:	AYB seeds with enhanced seed longevity when dried at 45 °C/60% RH for 4 days	99
4.7:	AYB seeds with enhanced seed longevity when dried at 45 °C/60% RH for 6 days	100
4.8:	AYB seeds with enhanced seed longevity when dried at 45 °C/60% RH for 8 days	101

LIST OF PLATES

PLATES	TITLE OF PLATES	PAGES
2.1:	African yam bean seeds with variation and tuber	14
2.2:	Fully automated Soxtec TM 8000 machine for rapid fat extraction	18

LIST OF APPENDICES

APPENDICES	TITLE OF APPENDICES	PAGES
1:	Pearson Correlation Coefficients for 2018/2019 planting season at $p < 0.05$	135
2:	ANOVA Table (2018/2019 planting season)	136
3:	Pearson Correlation Coefficients for 2019/2020 planting season at $p < 0.05$	137
4:	ANOVA Table (2019/2020 planting season)	138
5:	Pearson Correlation Coefficients for both planting season at $p < 0.05$	139
6:	ANOVA Table (2018/2019 and 2019/2020 planting season)	140
7:	List of some of the marker-trait association using the mixed linear model for all the traits	141
8:	Sequences of identified 41 SNP markers for seed oil, protein and starch content	149
9:	Duncan's Multiple Range Test for AYB Oil Content	153
10:	Duncan's Multiple Range Test for AYB Protein Content	167
11:	Duncan's Multiple Range Test for AYB Starch Content	181
12:	Kew Seed Information Database	195

LIST OF ABBREVIATIONS

AA: accelerated ageing

AFLP: Amplified Fragment Length Polymorphism

ANOVA: analysis of variance

AYB: African Yam Bean

Bp: Base pairs

BLUE: Best linear and unbiased estimators

BNF: Biological nitrogen fixation

CDT: controlled deterioration treatment

Chr: Chromosome

CMLM: Compressed mixed linear model

CTAB: Cetyl trimethylammonium bromide

DArT: Diversity array technology

DArTseq: Diversity array technology sequence (DArTseq)

DF: degree of freedom

DNA: Deoxyribonucleic Acid

EDTA: Ethylene Diaminetetraacetic Acid

FAO: Food and Agriculture Organization

FT-IR: Fourier-Transform Infrared Spectroscopy

gDNA: Genomic DNA

GHG: greenhouse gases

GI: glycemic index

GLM: General linear model

GRC: Genetic Resources Centre

GWAS: Genome-Wide Association Study

IITA: International Institute of Tropical Agriculture

ISSR: Inter-Simple Sequence Repeat

LEA: late embryogenesis abundant

MAF: Minor allele frequency

MAS: Marker assisted selection

Mbp: Mega base pair

MC: moisture content

MLM: mixed linear model

mRNA: Messenger ribonucleic acid

N/A: Not available

NGS: next generation sequencing

*p*50: the time (days) it takes for seed viability to reduce to 50%

PCA: Principal component analysis

PCR: Polymerase chain reaction

Q-Q: Quantile-quantile

QTLs: Quantitative Trait Loci

RAPD: Random Amplified Polymorphic DNA

Raf: Raffinose

Rep: replicate

RFO: Raffinose family oligosaccharides

RH: relative humidity

RPM: revolution per minute

SFE: Supercritical Fluid Extraction

SID: seed information database

SNPs: Single Nucleotide Polymorphisms

SSR: Simple Sequence Repeat

TASSEL: Trait Analysis by aSSociation, Evolution and Linkage

TE: Tris-Ethylene diaminetetraacetic acid

tRNA: Transfer ribonucleic acid

TSs: Tropical *Sphenostylis sternocarpa*

ABSTRACT

African yam bean (AYB) (*Sphenostylis stenocarpa* Hochst. Ex A. Rich) is a neglected legume that has potential to contribute towards alleviating food insecurity in sub-Saharan Africa. As an underutilised legume, AYB provides a great source of protein and starch, with low oil content. Seed nutritional content is a quantitative trait influenced by environmental effects, genes and an interaction of both. These seed nutritional contents (e.g. oil content) are known to influence seed longevity. Seed longevity is the period of time a seed remains viable during storage. Understanding the genetic basis of nutritional traits will contribute to adequate plant conservation as well as food and nutrition security. The aim of this research was to identify the candidate genes associated with AYB seed oil content, seed protein content and the seed starch content, and also assess the influence of these traits on seed longevity of African yam bean. A total of one hundred and thirty-four accessions of African yam bean were used for this research. The nutritional traits i.e., the seed oil, protein and starch contents were analysed using the Soxtec, Kjeltex and Phenol-sulphuric acid method, respectively. Twelve AYB accessions were selected from the one hundred and thirty-four accessions on the basis of high and low protein and oil content for the seed longevity evaluation. The seeds were planted and at harvest maturity, the seeds were harvested, divided into five seed lots and dried at two different temperatures (17 °C (control) and 45 °C). In this study, variation was observed in the seed lots subjected to initial high-temperature drying (45 °C) and their subsequent seed longevity. Ten out of twelve accessions, which were subjected to two-stage drying at 45 °C and 17 °C showed better seed longevity when compared to seeds dried at only 17 °C in the drying room. A significant positive association was observed between the seed longevity ($p50$) of seeds across all the treatments indicating that seed starch content has a role to play in AYB seed longevity. A genome-wide association study (GWAS) was also carried out on some nutritional traits (i.e., seed oil, seed protein and seed starch content) collected from two planting seasons (2018/2019 and 2019/2020) for one hundred and thirty-four accessions of African yam bean based on 2,485 SNP (single nucleotide polymorphism) markers. A total of five, six, and two SNP markers were found to be significantly associated with genes responsible for oil, protein and starch content, respectively. Also, one pleiotropic marker 100004767|F|0-39:A>T-39:A>T was observed to be associated with both AYB seed oil and protein content. This study identified the candidate genes that may be responsible for AYB seed oil, protein and starch content and this can expedite the molecular breeding process using marker-assisted selection. Also, the two-stage temperature drying revealed the importance of adequate seed drying for optimum ex situ conservation of seeds. The effect of high temperature drying (45 °C) on the seed longevity of AYB provided a more efficient protocol to be used in conserving genetic resources, which indirectly contribute to food security.

Keywords: *African yam bean, Genome-wide Association Studies, Seed longevity, Seed oil content, Seed protein content, Seed starch content, Sphenostylis stenocarpa.*