Journal of Advances in Biology & Biotechnology

15(3): 1-16, 2017; Article no.JABB.28482
ISSN: 2394-1081

Phylogeny and Bioinformatics Study on Leaf Tissue Genes for Selected Elite Cultivars of Plantain (*Musa paradisiaca L.*) in the Rain Forest Ecology of Nigeria

Godwin Michael Ubi[1*], Chioma M. Okolo Adilieje[2], Nneka Constance Ogbonna[3] and Success Kalu Eni[4]

[1]Department of Genetics and Biotechnology, University of Calabar, Calabar, Nigeria.
[2]Department of Biological Sciences, ATCOI College of Education, Enugu, Enugu State, Nigeria.
[3]Department of Biological Sciences, Abia State University, Uturu, Abia State, Nigeria.
[4]Department of Genetics and Biotechnology, University of Calabar, Calabar, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. Author GMU designed the study, performed the statistical analysis. Authors GMU, CMOA and NCO wrote the protocol and author GMU wrote the first draft of the manuscript. Authors GMU and SKE managed the analyses of the study. Author GMU managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JABB/2017/28482
Editor(s):
(1) Tapan Kumar Mohanta, School of Biotechnology, Yeungnam University, Gyeongsan 712-749, Republic of Korea.
Reviewers:
(1) T. Pullaiah, Sri Krishnadevaraya University, India.
(2) M. S. R. Krishna, KL University, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/21571

Original Research Article

Received 21st July 2016
Accepted 31st August 2016
Published 26th October 2017

ABSTRACT

Aims: To show that phylogenetic and bioinformatics study could be very crucial to the survival and adaptation of elite plantain species that are threatened by rapid changes in environment, urbanization, genetic erosion and climate in the rain forest ecology of Nigeria.
Study Design: The study is a survey of elite cultivars of plantain in the rain forest ecology of Nigeria, involving laboratory work and use of genetic analysis software for evaluation.
Place and Duration of Study: The leaf samples were sourced and collected from the rain forest of Nigeria in 2015 and outsourced to South Africa for laboratory DNA extraction and leaf tissue gene sequencing.
Methodology: Apical leaf samples less that 2 day old were collected from each of the 14 elite plantain cultivars, preserved in silica gel nylon and outsourced to South Africa for DNA extraction

**Corresponding author: E-mail: ubiology.gu@gmail.com;*

using the CTAB method and Leaf Tissue genes sequencing using the Sangers method. Sequenced DNA nucleotides were aligned using the Molecular Evolution and Genetic Analysis (MEGA 6.0) software for multiple alignment and ClustalW. Aligned sequenced were pasted in note pad, saved and later pasted in a new window for phylogeny tree construction using the maximum livelihood option. The GOR IV and PHYRE2 softwares were used to determine the secondary and 3D tertiary structure for each elite cultivar.

Results: Phylogenetic and bioinformatics study is very crucial to the survival of these elite plantain cultivars in the study geographical locations, which are being threatened by the rapid changes in environment, urbanization, climate change and emerging pests and disease conditions. The results revealed that the elite plantain cultivars have a rooted phylogenetic tree indicating a common ancestry for all. The phylogeny tree showed two internal ancestors and two descendant relatives for the cultivars. The phylogeny tree showed four clusters with four branches and leaves. There was an OUT group made up of cooking bananas with ABB genome which complemented the other cultivars with AAB genome to give the tree a root. Cultivars in the first cluster showed 100G consensus, those on the second cluster showed 69G consensus while those in the third cluster showed 53G consensus. Results of bioinformatics study revealed that the C-G content of 46.04% was higher in cooking bananas and least 35.26% in Ogoni Red cultivar. Number of amino acids was also higher 221 in cooking bananas and least in Ogoni red. However, number of nucleotide was higher in Enugu plantain and least in cooking bananas. The secondary and 3D tertiary structures of all the elite cultivars showed variations their fold structures as well as in their alpha and beta properties.

Conclusion: Bioinformatics study, alongside a complete knowledge on the phylogeny of the elite plantain cultivars will scale up and promote conservation efforts for plantain germplasm which do not possess seeds for short, medium and long term storage. The study will also provide baseline information on the phylogeny and bio information that will help in the establishment of biodiversity within the plantain germplasm for future research purpose, food security and income generation.

Keywords: Evolutionary changes; phylogenetic tree; operational taxonomic units; multiple sequence alignment; 3D tertiary structures; C-G content; Musa paradisiacal.

1. INTRODUCTION

Phylogenetics as a branch of study emerged in the early 20th century but didn't begin in earnest until the 1960s, with the advent of protein sequencing, PCR, electrophoresis, and other molecular biology techniques. As computers have become more powerful and more generally accessible, computer algorithms becoming more sophisticated, researchers are now able to tackle complicated problems that define evolution at the molecular level more effectively [1].

Within past decade, the field of phylogenetics has been further reenergized and redefined, as whole genome sequencing for complex organisms has become faster and less expensive. As genomic data becomes publically available, molecular phylogenetics continue to grow and find new and wider applications [2].

The similarity of biological functions and molecular mechanisms in living organisms strongly suggests that species descended from a common ancestor. Molecular phylogenetics uses the structure and functions of molecules and how they change over time to infer these evolutionary relationships and to recover the order of evolutionary events and represent them in evolutionary trees that graphically depict relationships among species or genes over time [3].

Evolution is a process by which the traits of a population change from one generation to another. Evolution is shaped by homology, which refers to any similarity due to common ancestry. Phylogenetic data sets can consist of hundreds of different species, each of which may have varying mutation rates and patterns that influence evolutionary changes. In *On the Origin of Species by Means of Natural Selection,* Darwin had proposed that, given overwhelming evidence from his extensive comparative analysis of living specimens and fossils, all living organisms descended from a common ancestor [4].

At a molecular level, evolution is driven by the same types of mechanisms Darwin observed at the species level [5]. One molecule undergoes diversification into many variations. One or more of those variants can be selected to be reproduced or amplified throughout a population over many generations. Such variations at the molecular level can be caused by mutations,

such as deletions, insertions, inversions, or substitutions at the nucleotide level, which in turn affect protein structure and biological function [6].

According to modern evolutionary theory, all organisms on earth have descended from a common ancestor, which means that any set of species, extant or extinct, is related. This relationship is called a phylogeny, and is represented by phylogenetic trees, which graphically represent the evolutionary history related to the species of interest [7].

Phylogenetics infers trees from observations about existing organisms using morphological, physiological and molecular characteristics. The "tree of life" represents a phylogeny of all organisms, living and extinct. Other, more specialized species and molecular phylogenies are used to support comparative studies and have been widely adopted in the test of biogeographic hypotheses, evaluate mode and timing of speciation, infer amino acid sequence of extinct proteins, track the evolution of diseases and provide evidence in criminal cases [8].

Trees can be made of a single gene from many taxa (a species tree) or multi-gene families (gene trees). A tree is considered to be "rooted" if there is a particular node or out group (an external point of reference) from which all OTUs in the tree arises [9]. The root is the oldest point in the tree and the common ancestor of all taxa in the analysis. In the absence of a known out group, the root can be placed in the middle of the tree or a rootless tree may be generated [10].

Branches of a tree can be grouped together in different ways. A monophyletic group consists of an internal LCA node and all OTUs arising from it. All members within the group are derived from a common ancestor and have inherited a set of unique common traits. A paraphyletic group excludes some of its descendants for example as in mammals. A polyphyletic group can be a collection of distantly related OTUs that are associated by a similar characteristic or phenotype, but are not directly descended from a common ancestor [11].

Interestingly, evolutionary divergence of different species may result in many variations of a protein, all with similar structures and functions, but with very different amino acid sequences. Phylogenetic studies can trace the origin of such proteins to an ancestral protein family or gene and hence have necessitated the need for this studies of leaf tissue genes in plantains (*Musa paradisiaca*).

Plantain (*Musa paradisiaca* L.) is a giant perennial herb of the genus *Musa*. It is one of the most versatile food crops in the tropics, where it is seen as a very important component of food security and provides a substantial amount of income to local farming communities through internal trade [12]. The crop is of extraordinary significance with a bright economic horizon in West and Central Africa and Latin America, which are the predominant plantain growing regions of the world. It is one of the few most important suppliers of dietary energy in parts of the humid forests and mid-altitude agro-ecologies of tropical regions of the world, where it is cultivated and utilized as a major starchy staple and consumed by nearly half a billion people in different parts of the world [13].

Plantain *(Musa paradisiaca* L.) is undoubtedly one of the most important staple food crop and perhaps, one of the oldest cultivated fruit crop in the humid tropics of Africa, Central Asia, South America and the West Indies [14].

Nigeria is one of the major plantain producing countries (FAO, 2006); fuelled in part by the tremendous rate of consumption in the country as a result of the rapidly increasing urbanization rate and the great demand for easy and convenient foods by the non-farming urban proletariat [15].

Nigeria is the largest producer of plantain in West Africa with an estimated annual production rate of about 12.4 million metric tons. Most part of the country is characterized by a tropical rain forest, even rainfall distribution and soil conditions that favours the cultivation and production of the crop [16].

The object of this study, was therefore design to study the phylogenetic characteristics and bioinformatics of 14 prominent elite plantain cultivars in the study area which should serve as a basis for molecular breeding, conservation effort, marker assisted breeding and selection and to ensure the sustenance of biodiversity of these elite cultivars. The study will also provide a platform for the development of gene bank for plantain and improvement of available cultivars of plantain for food security in Nigeria.

2. MATERIALS AND METHODS

2.1 Plant Materials Used for the Study

The procedure for the collection of plant samples used for the study are as stated below.

2.2 Plant Materials

Fourteen Elite plantain cultivars comprising of 14 cultivars collected from diverse locations in Nigeria was used for the study.

The plant materials were the apical leaf samples of less than 2 day old, collected from all the 14 elite cultivars from the different areas dried with silica gel using method for sampling as described by Ubi et al., 2016. The leaf samples were outsourced to South Africa for DNA extraction and Sequencing as described below;

2.3 DNA Extraction

The Cetyl Trimethyl Ammonium Bromide (CTAB) method was adopted for the extraction of sample total genomic DNAs. Young and tender leaf tissue (10g) was weighed and ground in liquid Nitrogen using mortar and pestle along with 50 µl of B – mercaptoethanol (BME) and a pinch of polyvinylpyrrolidone (PVP). The sample was further ground using excess of liquid Nitrogen and 4 mL of extraction modified CTAB buffer (2X) and the powder was transferred to a sterile 50-mL centrifuge tube containing 3 mL of pre-warmed extraction buffer. The homogenate was incubated for 30 minutes at 65℃ with intermittent mixing or votexing.

Equal volumes of chlorofoam and isoamyl alcohol at (24:1) was added and centrifuged at 10,000 rpm for 15 min at 4℃ and the supernatant collected. A 0.6 volume amount of chilled isopropanol was added to enable for the collection of the supernatant. This was followed by incubation at – 20℃ for 30 min. The precipitated DNA was pelleted by centrifugation at 10,000 rpm for 10 min at 4℃.

The DNA pellet was harvested and washed with 5 mM ammonium acetate in 70% ethanol and later with 100% ethanol. The DNA samples were air dried for 30 min at room temperature and dissolve in 100 µl of TE buffer. The DNA sample was further treated with RNase-A at 37℃ for 1 hour to remove any RNA contamination. The quantity and quality of genomic DNA were estimated using a Nano Drop ND 1000 spectrophotometer (NanoDrop Technologies Inc., USA) and the samples were diluted to a concentration of 50ng/µl. Isolated DNA was visualized for its quantity and quality by running them in 1% agarose gel electrophoresis.

2.4 DNA Amplification

PCR reaction was carried out in an Ependorf Master Thermocycler (Ependorf, USA). Each reaction tube was made of 100 ng of template DNAm 2.5 mM of MgCL2, 100 µM of dNTPs, 1X *Taq* buffer, 20 pM of 10-mer Primer, and 1 unit of *Taq* DNA polymerase made to a final volume of 20 µL.

DNA amplification was performed using the following thermal profile or sequence: 94℃ for 5 min (10 Cycle); 94℃ for 1 min 37℃ for 1 min, 72 ℃ for 2 min (10 cycles); final extension was performed at 72℃ for 10 min (10 cycle) and cooling of samples at 4℃.

Table 1. Latitudes, longitudes and altitudes of 14 elite plantain cultivars used for the study

S/N	Elite cultivars	Latitude (N)	Longitude (E)	Altitude (m.a.s.l)
1	Enugu Black Plantain	06°02.835’	008°41. 104’	21 0m
2	Ebi Egome Plantain	05°56.540’	008°50. 457’	131. 98m
3	Ogoni Red Plantain	06°54.583’	009°17. 799’	178m
4	Kigwa Brown Plantain	06°48.617’	009°15. 301’	18 3m
5	Ejorgom Plantain	06°30.723’	009°10. 687’	119m
6	Bakpri Plantain	04°97.778’	008°36. 013’	54m
7	Owomoh Plantain	05°55.882’	008°26. 391’	175m
8	Kainjen Plantain	05°58.200’	008°63. 520’	181m
9	Ikpobata Plantain	06°28.427’	009°08. 845’	97m
10	Mgbeghe Plantain	05°38.710’	008°46. 024’	119m
11	Kenkwa Plantain	06°04.445’	008°54. 776’	129.6m
12	Uhom Plantain	05°42.188’	008°03. 233’	56m
13	Ekumkwam Plantain	06°33.462’	008°52. 290’	110m
14	Ingwam Plantain	06°39.995’	008°51. 607’	92m

m.a.s.l = meters above sea level.

Fig. A = Ogoni Red French Plantain

Fig. B = Kigwa Brown False Horn Plantain

Fig. C = Enugu Black False Horn Plantain

Fig. D = Ebi Egome False Horn Plantain

Fig. E = Owomoh True Horn Plantain

Fig. F = kenkwa False Horn Plantain

Fig. G = Kainjen False Horn Plantain

Fig. H = Uhom False Horn Plantain

Fig. I = Ekumkwam French Plantain

Fig. J = Ikpobata French plantain

Fig. K = Mgbeghe False Horn Plantain

Fig. L = Ingwam French Plantain

Fig. M = Bakpri French Plantain

Fig. N = Ejorgom True Horn Plantain

2.5 Primers Design

A total of 30 arbitrary primers were designed for the study using the primer 3 in http//www.biotools.umassmed.edu/ procedures. The designed primers were sourced from 'wet lad' of Operon Technology and use for the study. The sourced primers were screened for polymorphism against plantain cultivars. Primers that produced distinct bands and patterns with

good quality of amplification and reproducibility were selected and used for evaluation study. The designed primers were of 18-22 bps, melting temperature of 54℃, annealing temperature 55 – 65℃ and higher G-C content of 40 – 60%.

2.6 Method for Leaf Tissue Genes Sequencing

The Sanger's method was adopted for gene sequencing in lieu of the post sangers or new generation sequencing because of the cost implication.

DNA sequencing is a method used for determining the order of nucleotide bases (A-C-G-T) in a molecule of DNA.

Automated DNA sequencing machines are now available but the specimens have to be prepared prior to loading them in the machines.

DNA was extracted from each elite plantain cultivar leaves, purified and concentration determined as earlier described above. 2 µl of Big Dye terminator (v 3.1) cycle sequencer kit was used. A 4 µl 5X sequencer buffer was prepared using 1 µl 3' primer (3.3 pmol) which was about 4X lower than the concentration of the normal PCR in 10 µl of distilled water and 3 µl of purified DNA sample to give a total of 20 µl which was vortex to mix properly. PCR program was run for 30 cycles. The sequencer samples were prepared by mixing -20µl of DNA sample extracted, 50 µl of 100% ethanol, 2 µl of 3M NaOAc, 2 µl of 125 mM EDTA and incubated at room temperature for 15 min. The entire mixture was centrifuged at 15 rpm for 20 min at 20℃ before discarding the supernatant.

A 70 µl of 75% ethanol was added, vortex and centrifuged for 5 min at 15000 rpm at 20℃ after which the supernatant was discarded, dried in a desiccator, re-suspended in a buffer for the sequencer.

In the DNA sequencer, capillary electrophoresis for size separation, detection and recording of dye fluorescence with data output was carried out as fluorescent peak trace chromatograms.

Protocol for sequencing of the GTP-binding nuclear protein Ran 3A genes was achieved by the following protocols,

a. The use of Single stranded DNA template, a DNA primer, DNA polymerase, radioactively labelled nucleotides and ddNTPs (dideoxynucleotides).
b. The Four separate reactions inside the sequencer. Each reaction was furnished with the polymerase enzyme, dNTPs, (dATP, dGTP, dCTP, dTTP) and one of the 4 ddNTPs.
c. The newly synthesized DNA was heat denatured, using Gel Electrophoresis to separate the DNA fragments to a resolution of just one nucleotide.
d. Each of the four reactions ran on separate lane in the sequencer where each DNA band was visualized by autoradiography, UV light and the DNA nucleotide sequence was read directly from the bottom of the sequencer machine.
e. A dark band in each lane in the sequencer indicated a DNA fragment that emanated from the chain termination after incorporation of a ddNTPs to each of the reaction.

2.7 Phylogenetic Tree Construction for 14 Elite Cultivars

To build phylogenetic trees, statistical methods are applied to determine the tree topology and calculate the branch lengths that best describe the phylogenetic relationships of the aligned sequences among the dataset. Many different methods for building trees exist and no single method performs well for all types of trees and datasets. The most common computational methods applied include distance-matrix methods, and discrete data methods, such as maximum parsimony and maximum likelihood. Reconstructed phylogenetic trees normally are weighted trees.

The Molecular Evolution Genetic Analysis (MEGA 6.0) software was used for aligning the nucleotide sequence using multiple sequence alignment (MSA) using the Clustalw option. The aligned nucleotide sequence was pasted and saved in word pad documents in the desktop before inserting into a new window of the MEGA software for phylogenetic tree construction. The MEGA 6.0) software was used to develop the phylogenetic tree for the 14 elite cultivars of plantain in the study. The bootstrap method and maximum livelihood method was adopted.

2.8 Bioinformatics Study of 14 Elite Cultivar of Plantain

The GENSCAN software was used for determining the number of nucleotides, number of amino acids, the percent Cysteine to Guanine (C-G %) for each elite cultivar of plantain.

The Phyre 2 software was used for the prediction of 3D tertiary structure for the 14 elite plantain cultivars. The extracted nucleotide sequence was first transformed into amino acids sequence using the GENSCAN software for each of the cultivar before pasting the translated amino acids sequence on FASTA format into the window of the phyre2 and emailed to http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index.

A protein data bank (PDB) file for each query was sent to our email and using the Raswin, the PDB file was saved. The ribbon structure in the PDB file was fine-tuned to display the 3D structure ribbon for each of the cultivars.

The GORIV software was used in determining the secondary structure for each cultivar and their attributes.

2.9 Data Scoring and Analysis

Each band in the RAPD banding pattern was considered as a separate locus. Only distinct, reproducible, well-resolved fragments were selected and scored for presence (1) and absence (0) of a band. The binary matrix of RAPD phenotypes was then assembled for analyses. A similarity matrix was constructed and subjected to cluster analysis following the un-weighted pair group method with arithmetical averages (UPGMA) of the computer program DARwin version 5.0. Measurement of genetic distance for pair-wise accessions was based on Nei's unbiased genetic distances (Nei 1978) using DARwin 5.0 software. The results for the RAPD analysis and other analysis so mentioned here were not included in this manuscript as there are kept as results for molecular characteristics which is currently under processing for future publication.

3. RESULTS AND DISCUSSION

3.1 Plantain Phylogenetic Tree

The evolutionary history and the relationship among the 14 elite plantain cultivars were constructed using the UPGMA which builds the maximum composite likelihood showing time of divergence. The phylogenetic tree rooted from a common ancestor having three major branches (3 clusters) and an OUT group and the same time of divergence of about 100 million years ago (MYA) as shown in Fig. 1. Each branch or cluster has similar nucleotide sequence, the secondary and tertiary protein sequence structure was model for the representative(s) of each elite cultivar as shown in each Figure below. The cultivars in cluster (blue) one are 100G consensus, cultivars in cluster (brown) two are 78G, 69G and 53G consensus while cultivars in cluster (green) three are 98G, 53G and 62G consensus. The OUT group is represented in cluster (red) four.

3.2 Evolutionary Developmental Relationship and Pathways for 14 Elite Cultivars of Plantain (*Musa paradisiaca* L.)

3.2.1 The root

The phylogenetic tree for the 14 elite cultivars of plantain is considered to be “rooted” because there is an out group (Cooking bananas group with ABB genome is an out group) linkage with other groups from which all OTUs in the tree arises.

The root is the oldest point in the tree and the common ancestor of all 14 elite cultivars used for the study. This shows that there is a common ancestor of all the operational taxonomic units or classes of cultivars characterized. It is represented here in red and represents the common ancestor linking the OUT group (Cooking bananas ABB genome) with the other operational taxonomic units with the AAB genomes. [1,2,4 and 17].

3.2.2 The hypothetical taxonomic units (HTUs) or nodes

These represent the ancestors internal nodes as shown in Fig. 1 and marked in purple and green marks. They correspond and represent the hypothetical ancestors showing 78G and 98G consensus in first ancestral evolutionary pathway and 62G and 55G consensus in second ancestral evolutionary pathway. In phylogeny, they are referred to as HTUs (hypothetical taxonomic units). The HTUs evolutionary pathways was only traced between and among the species in the second and third cluster groups only and not in the first and fourth clusters because slow genetic divergence [1,6,12,18 and 19].

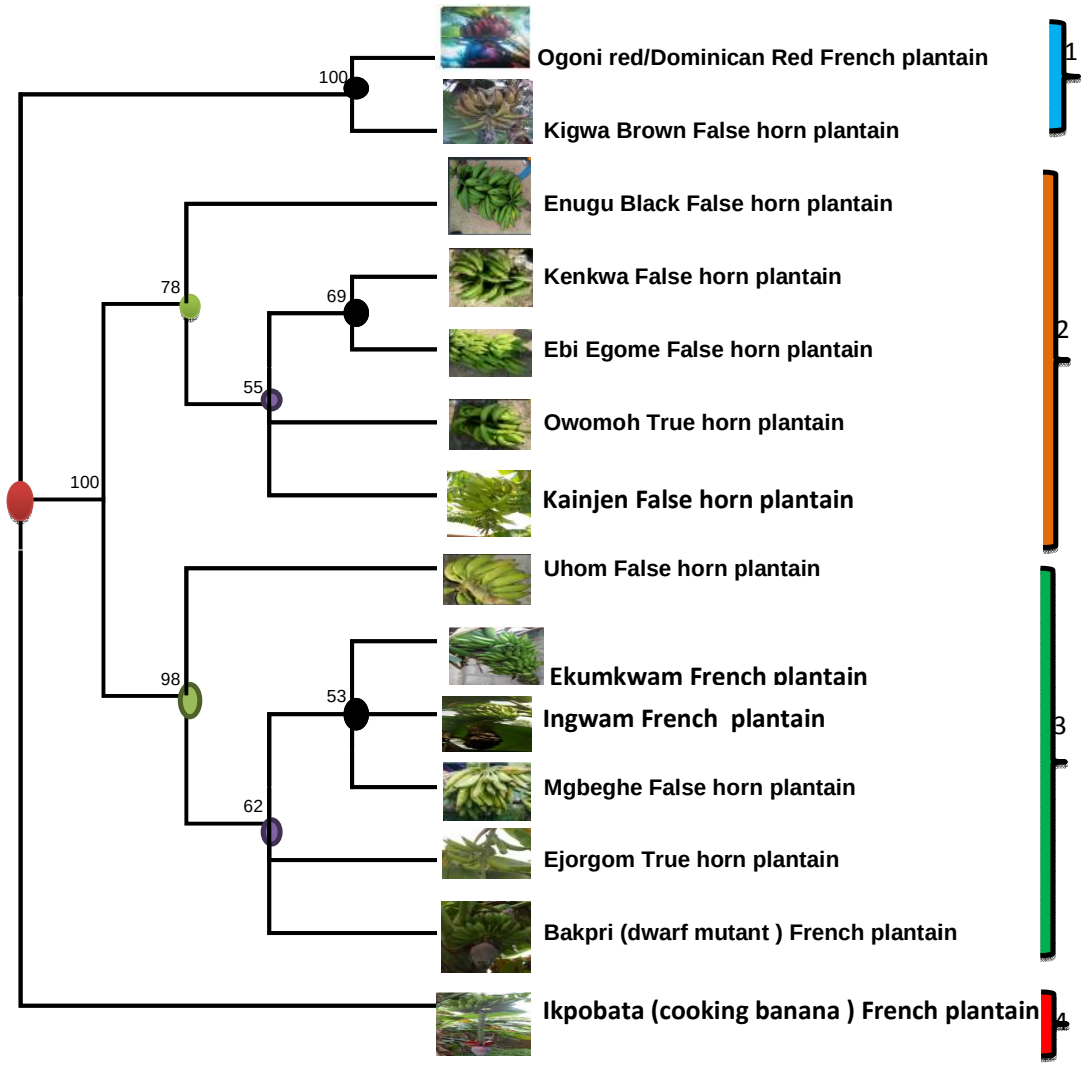


Fig. 1. Phylogenetic tree showing root, branches, HTUs, OTUs and OUT groups.

The phylogenetic tree above represents the evolutionary tree and pathway followed by the elite cultivars of plantain (Musa paradisiaca) in the study area. The phylogenetic tree composed of branches, also known as edges that connect and terminate at nodes.

The branches and nodes can be internal or external (terminal) as seen from the phylogenetic tree above. The branches length varies and represents the evolutionary pathway and time of evolution. Longer branches indicated longer period of evolutionary development and mutation while shorter branches indicated shorter period of evolutionary development between and among the plantain species.

The terminal nodes at the tips of the tree represent the operational taxonomic units (OTUs). OTUs correspond to the molecular sequences or taxa (species) from which the tree was inferred.

The internal nodes represent the last common ancestor (LCA) to all nodes that arise from that point.

The tree root emanated from a single gene from many taxa (Musa spp.) and through the evolutionary sequence and mutation changes developed into multi-gene families (gene trees).

Edge lengths represent mutation events which are supposed to have occurred on the evolutionary path. Differences in edge lengths in the above tree reflect the fact, that the rate at which mutations accumulate in the sequences vary among the cultivars of the taxa.

Species evolve in time. In the simplified tree model, we assume that a species evolves along an edge. Internal nodes reflect ancestral species and split into two new species. This is reflected by bifurcations in the binary tree [6,13,20].

3.2.3 The operational taxonomic units (OTUs) and the branches

The species or cultivars used here represents the leaves and are called taxa or OTUs (operational taxonomic units). Phylogeny was reconstructed for this set of taxa, using their respective genes or proteins sequence. The OTUs species in the first cluster have a 100G consensus with only two relatives, the species in the second cluster also have two relatives extending to develop into branches and have a 69G consensus while the third cluster group have three relatives extending to the branches with 53G consensus. OTUs can be used to build an un rooted phylogenetic tree that clearly depicts a path of evolutionary change [21-23].

3.2.4 The clades

These are a group of operational taxonomic units that included several sequences and their common ancestor nodes [24,25]. In the above phylogeny tree, two clades are represented in the tree comprising five cultivars of plantain as seen in cluster 2 and six cultivars of plantain as seen in cluster 3 above. In the first clade which represented as cluster 2, there are two branches (cultivars) with a first internal node or last common ancestor and a second descendant of its neighbouring close relative linked to the clade. The second clade which is represented as cluster 3, has three branches (cultivars) with first internal node or last common ancestor and a descendant of its neighbouring relatives linked to the clade.

3.2.5 The OUT group

Sometimes it is possible to obtain external information that a certain taxon is more distantly related to the other taxa than the other ones among themselves. Such a taxon is called an out group [26,27]. Adding a root node to the edge to the out group then allows interpreting bifurcations with respect to time [19,28]. An OUT group therefore in the study is for the purpose of finding the root of the tree [1,5,29]. In the phylogenetic tree above, the cooking bananas cultivar with ABB genome is distantly related the all other cultivars in the study with AAB genomes. This cultivar therefore was basically used to obtain the common root or ancestry for the 14 elite cultivars of plantain used for the study.

A tree is considered to be “rooted” if there is a particular node or out group (an external point of reference) from which all OTUs in the tree arises [14,30]. The root is the oldest point in the tree and the common ancestor of all taxa in the analysis. In the absence of a known out group, the root can be placed in the middle of the tree or a rootless tree may be generated [24,28,31].

Almost all phylogenetic tree reconstruction methods reconstruct an un - rooted binary tree which cannot be interpreted with respect to a time scale. In an un rooted tree, one does not know whether an internal node is the ancestor or the descendant of its neighbouring internal node. [7,32].

The monophyletic groups consists of an internal Last Common Ancestor (LCA) node and all Operational Taxonomic units (OTUs) arising from it. All members within the elite plantain cultivar groups here presented were derived from a common ancestor and have inherited a set of unique common traits [17,33].

The polyphyletic groups are a collection of distantly related OTUs that are associated by a similar characteristic or phenotype, but are not directly descended from a common ancestor. This does not apply in our findings as all elite cultivars have descended from a common ancestor. Evolution is shaped by homology, which refers to any similarity due to common ancestry. Similarly, phylogenetic trees are defined by homologous relationships. Paralogs are homologous sequences separated by a gene duplication event while Orthologs are homologous sequences separated by a speciation event (when one species diverges into two). Paralogs are created by gene duplication events [1,33].

Molecular phylogenetic trees are drawn so that branch length corresponds to amount of evolution (the percent difference in molecular sequences) between nodes. Once a gene has been duplicated, all subsequent species in the phylogeny will inherit both copies of the gene, creating orthologs [23,33] as seen in cluster 1, 2

and 3 above from our phylogeny tree construction.

Interestingly, evolutionary divergence of different species may result in many variations of a protein, all with similar structures and functions, but with very different amino acid sequences.

Phylogenetic studies can trace the origin of such proteins to an ancestral protein family or gene. Molecular sequence might evolve over time as a result of multiple mutations that results small, but evolutionarily important changes in a nucleotide sequence. At the protein level, these changes may not initially affect protein structure or function, but over time, they may eventually shape a new purpose for a protein within divergent species [8,13,33].

3.3 Results of Bioinformatics Study of 14 Elite Cultivars of Plantain

Results of bioinformatics study for the elite cultivars as presented in Table 2 and chart 1 below shows that all except the OUT group cultivar (cooking bananas) possess low C-G percent content, which is low compared the standard recommended rating of 40-60% G-C percent content. This may be responsible for their ease of mutation compared to the cooking bananas which did not evolve from any ancestral decent of node. The results further shows a high number of amino acids of 221 was found to be contained in cooking bananas (ABB genome) compared to all other cultivars with AAB genome with the least of 76 amino acids coming from four of the cultivars. Number of nucleotide sequence also varied greatly among the cultivars evaluated with Enugu black cultivar having the highest nucleotide sequence of 4145 and cooking bananas showing the least of 1185.

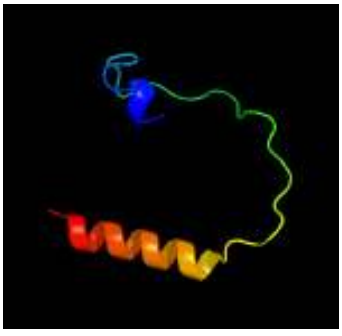
3.4 Results of Protein Sequence Modeling of 3D Tertiary Structures of Leaf Tissue Genes for 14 Elite Plantain Cultivars

The evolutionary history and the relationship among the 14 elite plantain cultivars were constructed using the UPGMA which builds the maximum composite likelihood showing time of divergence. The phylogenetic tree rooted from a common ancestor having three major branches (3 clusters) and the same time of divergence of about 100 million years ago (MYA) [2,26,34] as shown in Fig. 1. Each branch or cluster has similar nucleotide sequence, the secondary and tertiary protein sequence structure was model for the representative(s) of each elite cultivar as shown in each figure below. The cultivars in cluster (branch) one are 100G consensus, cultivars in cluster two are 69G and 53G consensus while cultivars in cluster three are 53G and 62G consensus. Secondary structures of some of the leaf tissue genes for the cultivars contained the alpha helix, beta sheet, extended strand and random coiling. Tertiary structures of protein for most of the cultivars showed gene mutation or mutant type of the cultivars.

3.5 3D Tertiary Structures for Elite Cultivars of Plantain

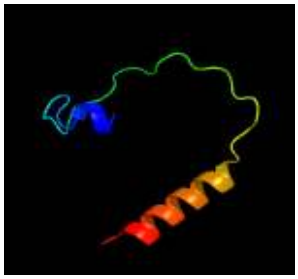
The 3D tertiary structure of a protein (amino acids) is a description of the complex and irregular folding of the peptide chain in three dimensions. It is essentially a picture of what the shape of the entire protein actually looks like. The 3D tertiary structure of a protein is a description of the way the whole chain (including the secondary structures) folds itself into its final 3-dimensional shape [28,34]. The results of 3D tertiary structure for the 14 elite cultivars of plantain studied are presented below

1



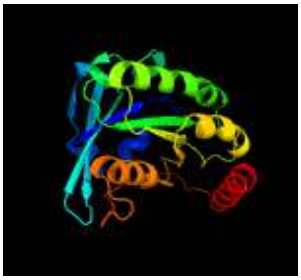
Ogoni Red/Dominican Red (AAB) French Plantain Cultivar 3D Structure

2



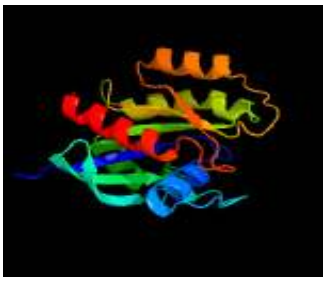
Kigwa brown (AAB) false horn plantain cultivar 3D structure

3



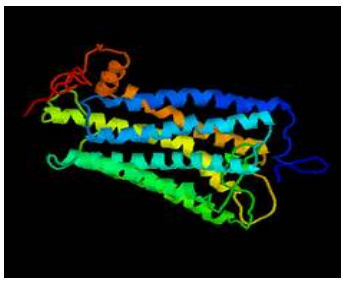
Enugu Black (AAB) false horn plantain cultivar 3D structure

4



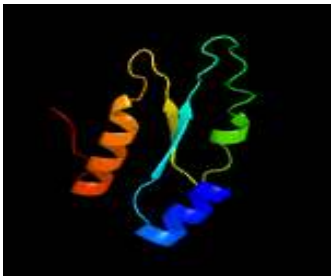
Owomoh (AAB) true horn plantain cultivar 3D structure

5



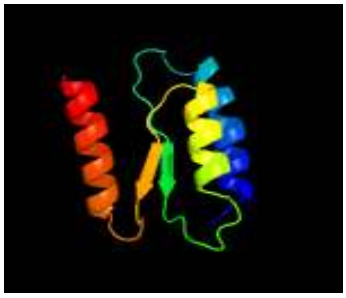
Ebi egome (AAB) false horn plantain cultivar 3D structure

6



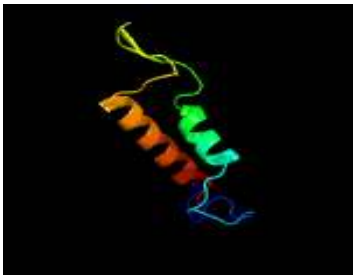
Kenkwa (AAB) false horn plantain cultivar 3D structure

7



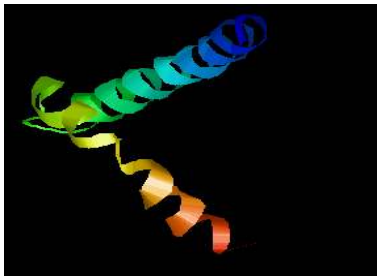
Mgbeghe (ABB) false horn plantain cultivar 3D structure

8



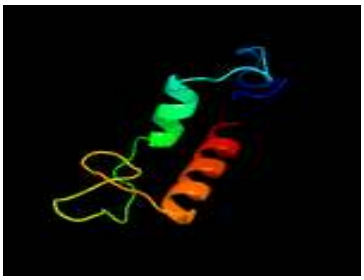
Ikpobata (Cooking banana) (ABB) French plantain cultivar 3D structure

9



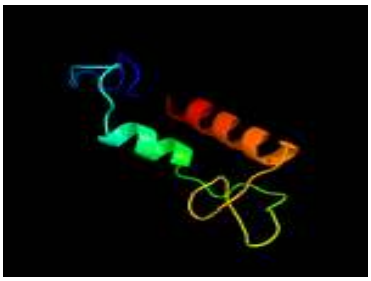
Bakpri (dwarf Mutant) (AAB) French plantain cultivar 3D structure

10



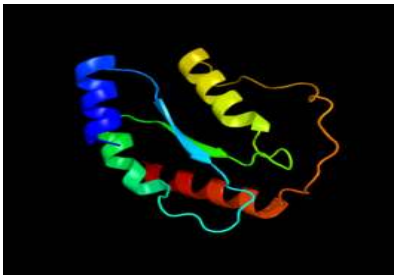
Ingwam (AAB) True horn plantain cultivar 3D structure

11



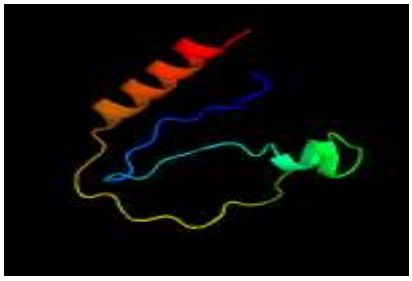
Ejorgom (AAB) French plantain cultivar 3D structure

12



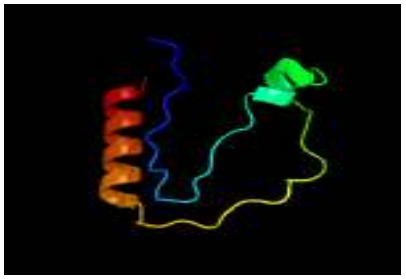
Kainjen (AAB) false horn plantain cultivar 3D structure

13



Uhom (AAB) True horn plantain cultivar 3D structure

14



Ekumkwam (AAB) False horn plantain cultivar 3D structure

Table 2. No. of Nucleotide sequence, Amino acid sequence and G-C content of leaf tissue genes for 14 elite cultivars of plantain

S/N	Name of Elite cultivar	No. of Nucleotide sequence	No. of Amino acid sequence	G-C content (%)	Genome
1	Enugu black plantain	4145	132	36.88	AAB
2	Uhom plantain	2832	76	35.39	AAB
3	Ebi Egome	6660	158	36.42	AAB
4	Ingwam	2869	76	35.26	AAB
5	Kenkwa	3862	105	37.00	AAB
6	Kainjen	3890	105	37.20	AAB
7	Mgbeghe	3864	105	37.27	AAB
8	Ogoni red	2802	76	35.63	AAB
9	Kigwa brown	2802	76	35.67	AAB
10	Owomoh	6617	158	36.69	ABB
11	Ejorgom	3478	83	37.25	AAB
12	Ekumkwam	3486	83	37.13	AAB
13	Bakpri (mutant dwarf)	1643	109	37.08	AAB
14	Ikpobata (cooking bananas)	1185	221	46.04	ABB

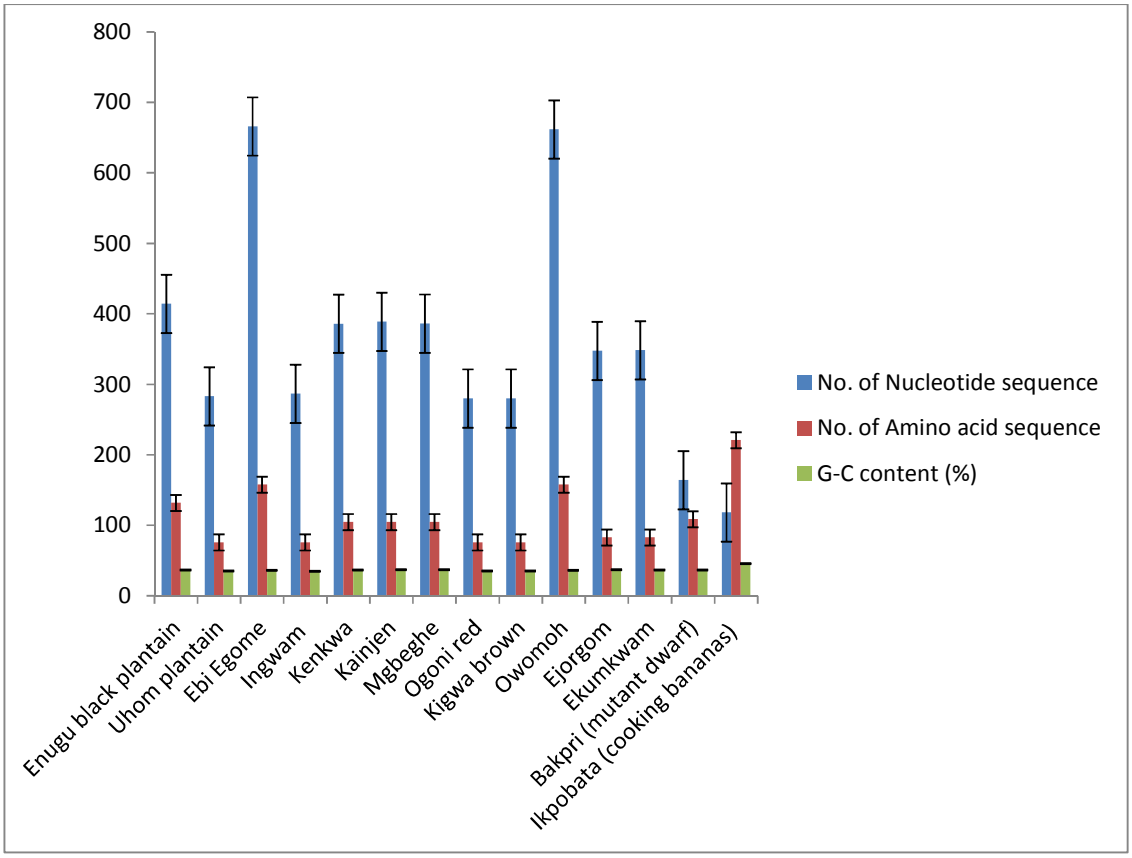


Chart 1. Showing no. of nucleotide sequence, No. of amino acids sequence and G-C contents of 14 elite plantain cultivars

4. CONCLUSION

Phylogenetic characterization is very crucial to the survival of these elite plantain cultivars in the study geographical locations, which are being threatened by the rapid changes in environment, urbanization, climate change and emerging pests and disease conditions. Bioinformatics study, alongside a complete knowledge on the phylogeny of the elite plantain cultivars will scale up and promote conservation efforts for plantain germplasm which do not possess seeds for short, medium and long term storage. The study will also provide baseline information on the phylogeny and bio information that will help in the establishment of biodiversity within the plantain germplasm for future research purpose, food security and income generation. Phylogeny and bioinformatics knowledge of the cultivars will help in the mutational breeding of the plant for associated benefits which will be made possible through a complete and in depth study and knowledge of the nucleotide and amino acids evolutionary pathway which can help in the recovery of lost desirable traits as well as development of new cultivars with desirable beneficial traits to man and his immediate environment.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Jones D, Peter R, James H, Jane S, (eds). "Phylogenetic". Merriam-Webster Dictionary. English Pronouncing Dictionary, Cambridge: Cambridge University Press. 2003;3-12-539683-2
2. Suárez-Díaz E, Anaya-Muñoz VH. History, objectivity, and the construction of molecular phylogenies. Stud. Hist. Phil. Biol. & Biomed. Sci. 2008;39(4):451–468.

3. Ahlquist JE, Charles G. Sibley. A commentary on 30 years of collaboration. The Auk. 1999;116(3):856–860.
4. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal.of Molecular. Biology. 1975;94(3):441–448.
5. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 1977;74(12):54637. Bibcode:1977PNAS...74.5463S.
6. Bast F. Sequence similarity search, multiple sequence alignment, model selection, distance matrix and phylogeny reconstruction. Nature Protocol Exchange. DOI: 10.10: 38/protex.2013.065
7. Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D, Penny D, ed. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLoS Biology. 2011;9(3):e1000602.
8. Avise JC. Molecular markers, natural history and evolution. Chapman & Hall, New York; 1994.
9. Doolittle WF. Phylogenetic classification and the universal tree. Science. 1999; 284:2124–2128.
10. Futuyama DJ. Evolutionary biology. 3rd edition. Sinauer, Sunderland, MA; 1998.
11. Hall BG. Phylogenetic trees made easy: Sinauer, Sunderland, MA; 2001.
12. Hartl DL, Clark AG. Principles of population genetics, 3rd edition. Sinauer, Sunderland, MA; 1997.
13. Hillis DM, Moritz C, Mable BK (eds). Molecular systematics, 2nd edition. Sinauer, Sunderland, MA; 1996.
14. Li WH, Graur D. Fundamentals of molecular evolution. Sinauer, Sunderland, MA; 1991.
15. Thornton JW, Desalle R. Gene family evolution and homology: Genomics meets phylogenetics. Annual. Review. Genomics Hum. Genetics. 2000;1:41–73.
16. Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R. The new animal phylogeny: Reliability and implications. PNAS. 2000;97(9):4453-4456.
17. Baldauf SL, Palmer JD. Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. Proceedings of the National Academy of Sciences. 1993;90:11558-11562.
18. Brown JR. 2002: Universal tree of life. In Encyclopedia of Life Sciences. Nature Publishing Group, Macmillan; 2002.
19. Hedges SB. The origin and evolution of model organisms. Nature Reviews. 2002; 3:838-849.
20. Penny D. Molecular evolution: Introduction. In Encyclopedia of Life Sciences. Nature Publishing Group, Macmillan; 2002.
21. Martin AP. Molecular clocks. In Encyclopedia of Life Sciences. Nature Publishing Group, Macmillan; 2002.
22. Tamas I, Klasson L, Canbäck B, Näslund A. Kristina, Eriksson, Ann-Sofie, Wernegreen, Jennifer J., Sandström, Jonas P., Moran, Nancy A., Andersson, Siv GE. 50 million years of genomic stasis in endosymbiotic bacteria. Science. 2002; 296:2376-2379.
23. Baldauf SL. Phylogeny for the faint of heart: A tutorial. Trends in Genetics. 2003; 19(6):345-351.
24. Brown D, Sjölander K. Functional classification using phylogenomic inference. PLos Computational Biology. 2006;2(6):0479-0483.
25. Cristianini N, Hahn M. Introduction to computational genomics: A case studies approach. Cambridge University Press: Cambridge; 2007.
26. Durbin R, Eddy S, Krogh A, Mitchison G. Biological sequence analysis. Cambridge University Press: Cambridge; 1998.
27. Ewens WJ, Grant R. Statistical methods in bioinformatics. Springer Science and Business Media: New York. 2005;12-15.
28. Ubi GM, Kinglsey EN, Onabe MB, Jemide JO, Egu CJ. Organoleptic and horticultural characterization of elite cultivars of plantain (*Musa paradisiaca* L.) for value addition and food security in Nigeria. Journal of Advances in Biology and Biotechnology. 2016;10(2):1-21.
29. Ubi GM, Jemide JO, Ebri MN, Ubi W, Essien IS. Assessment of phenological variability and nutritional value of the underutilize tropical jackfruit (*Artocarpus heterophyllus* Frost.) in Southern Nigeria. Journal of Advances in Biology and Biotechnology. 2016;10(1):1-18.
30. Finn RD, Tate J, Mistry PC, Coggill SJ, Sammut HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The Pfam protein families database. Nucleic Acids Research. 2008;36:D281-288.

31. Gabaldón T. Large-scale assignment of orthology: Back to phylogenetics? Genome Biology. 2008;9:235.1-235.6.
32. Gollery M. Handbook of hidden markov models in bioinformatics. CRC Press, Taylor & Francis Group: London. 2008;43.
33. Goodstadt L, Ponting CP. Phylogenetic reconstruction of orthology, paralogy and conserved synteny for dog and human. PLoS Computational Biology. 2006;2(9): 1134-1150.
34. Hall BG. Phylogenetic trees made easy: A how-to manual mathods 2nd ed. Sinauer Associates, Inc.: Sunderland, MA. 2004; 123.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/21571