# Simple Sequence Repeats-Based Fingerprinting Reveals the Genetic Architecture of Tea Plant (*Camellia sinensis* (L.) O. Kuntze) in Mambilla Plateau, Taraba State, Nigeria

LI Emeagi[1], TL Thomas[2], LE Uduak[1], TE Konyeme[3], OU Udensi[1]*

*[1]Department of Genetics and Biotechnology, University of Calabar, Calabar, Nigeria; [2]Department of Biological Sciences, Taraba State University, Jalingo, Nigeria; [3]Department of Biological Sciences, University of Delta, Agbor, Nigeria*

## ABSTRACT

The tremendous therapeutic, nutritional and economic importance of tea plant (*Camellia sinensis* (L.) O. Kuntze) cannot be overemphasized. However, the precise determination of its Deoxyribonucleic Acid (DNA) polymorphisms and genetic diversity, especially using DNA sequence dataset obtained through molecular characterization assay is an important strategy in planning and designing breeding programs for the Mambilla tea. In the present study, we reported here that the Simple Sequence Repeats (SSR) marker of *Histone* Gene (*H2A.Cs*) amplicons revealed distinct band profiles, mean Heterozygosity (He) were 0.277 ± 0.051, 0.214 ± 0.070, 0.274 ± 0.078, 0.385 ± 0.043 and 0.262 ± 0.078 respectively. Five tea populations showed narrow Genetic Distance (GD) and very close Genetic Identity (GI). Phylogenetic analysis showed that two distinct clusters were generated with 6 tea clones in cluster 1 and 38 tea clones in cluster 11, but there was no marked pattern of clustering as the different populations were mixed up together with bootstrap values showing common ancestral relationship. Analysis of Molecular Variance (AMOVA) result revealed 99.9% within population variation and 0.1% among population variation while PCoA generated six clusters with percentage of variation in the tea populations as 65.93%, 33.61% and 0.46% (1ˢᵗ, 2ⁿᵈ and 3ʳᵈ axes) respectively. Nucleotide frequencies were A (23.14%), T (32.53%), C (19.38%) and G (24.95%) while transition/transversion rate ratios are k1=14.562 for purines and k2=0.183 for pyrimidines. The overall transition/transversion bias Ratio (R)=3.414, R= (A$\cdot$G$\cdot$k1+T$\cdot$C$\cdot$k2)/((A+G)$\cdot$(T+C)). Our result also showed that only the gene sequence of tea clone obtained from Bangoba (BAN3) did not evolve with the same pattern of substitution with other gene sequences of other tea clones using Monte Carlo test. We also reported the degree of differentiation among the five tea populations, which indicated that Fixation Index (FST) estimation was 67.81% (0.6781) (Karaka population (germplasm))>61.33% (0.6133) (cultivated population)>Bangoba tea population (0.5984)>0.5810 (Kasalasa population)>(0.5032) (Kusuku population). Structure analysis revealed that the populations were genetically similar with percentage of membership of the sample in each of the 5 population clusters as Kusuku (19.4%), Bangoba (15.2%), Kasalasa (25.5%), Kakara (15.8%) and cultivated population was 24.2%, respectively. The present study suggests that the tea populations in the Mambilla Plateau is very homogeneous and needs introduction of new elite genotype/germplasm for breeding and improvement.

**Keywords:** *Camellia sinensis*; Molecualar fingerprinting; SSR; Genetic architecture; Mambilla plateau

## INTRODUCTION

Food and nutrition insecurity is one of the major challenges in the developing countries, especially in Sub-Saharan Africa (SSA). Another critical challenge in this region is the healthcare where high death rate as a result of some preventable diseases are prevalent. Among these killer diseases, cardiovascular disorder accounts for about 3.76% of total deaths in Nigeria as at 2018 and is projected to increase by 20% subsequently [1]. Interestingly, the medicinal properties of the tea plant (*Camellia sinensis* (L.) O. Kuntze) have scientifically been validated and known to treat a wide range of these health conditions, including cardiovascular diseases, obesity, immune enhancement, cancer, type 2 diabetes, neuralgic headaches, diarrhea, fatigue, sluggishness and asthma [2]. The tremendous therapeutic, nutritional and economic importance of tea plant should further strengthen the large-

scale cultivation, which is an incredibly important industry in several countries such as China, Japan, India, Vietnam and Indonesia. Despite these obvious significance of tea, negligence in breeding, improvement as well as lack of effective conservation strategies have robbed most African countries, especially Nigeria the economic and therapeutic benefits that comes with tea production and current clones may soon face genetic erosion if nothing strategic is done. Regrettably, of the 33 clones acquired by Cocoa Research Institute of Nigeria from Nigerian Beverage Production Company (NBPC) only 6 clones (BB35, 68, 143, 236, 318, 1212) are currently under cultivation in the Kakara Tea plantation, Taraba State [3]. However, vegetative means of propagating tea, which is currently being practiced in Kakara Tea Plantation and selection pressure for high yielding clones have posed serious problem leading to genetic erosion. Informatively, plant genetic resources are key to crop genetic improvement giving that the success of plant breeding and conservation is largely dependent on the amount and distribution of genetic variations present in plant collections. Among the various genetic markers such as the morphology, biochemical, cytogenetic as well as other DNA markers, microsatellites are considered one of the most informative. This is due to their co-dominance nature of inheritance, locus specificity, hyper-variability coupled with genome wide distribution, which has gained them considerable importance for evolutionary, plant genetics and breeding studies [4,5]. As been reported by and corroborated by our interviews, there is no information on the genetic analysis and improvement of tea plant using genomic/molecular tools in Nigerian tea germplasm. It is on this premise that this current research is hinged [6].

## MATERIALS AND METHODS

### Collection of plant materials and location of research

All tea clone samples collected covered the Kakara Highland Tea plantation of the Mambilla Plateau, Sardauna Local Government area, Taraba State located at coordinates range of 7020'N and 11043'E. Forty-four samples were used for the study (Table 1).

**Table 1:** Tea clones used for the study and their site of collection.

| S. No. | Tea clone/sample | Code | Source/location |
|---|---|---|---|
| 1 | Clone 1 | KUS1 | Kusuku community |
| 2 | Clone 2 | KUS2 | Kusuku community |
| 3 | Clone 3 | KUS3 | Kusuku community |
| 4 | Clone 4 | KUS4 | Kusuku community |
| 5 | Clone 5 | KUS5 | Kusuku community |
| 6 | Clone 6 | KUS6 | Kusuku community |
| 7 | Clone 7 | KUS7 | Kusuku community |
| 8 | Clone 8 | KUS8 | Kusuku community |
| 9 | Clone 9 | KUS9 | Kusuku community |
| 10 | Clone 10 | KUS10 | Kusuku community |
| 11 | Clone 11 | BAN1 | Bangoba community |
| 12 | Clone 12 | BAN2 | Bangoba community |
| 13 | Clone 13 | BAN3 | Bangoba community |
| 14 | Clone 14 | BAN4 | Bangoba community |
| 15 | Clone 15 | BAN5 | Bangoba community |
| 16 | Clone 16 | BAN6 | Bangoba community |
| 17 | Clone 17 | BAN7 | Bangoba community |
| 18 | Clone 18 | BAN8 | Bangoba community |
| 19 | Clone 17 | BAN8 | Bangoba community |
| 20 | Clone 20 | BAN10 | Bangoba community |
| 21 | Clone 21 | KAS1 | Kasalasa community |
| 22 | Clone 22 | KAS2 | Kasalasa community |
| 23 | Clone 23 | KAS3 | Kasalasa community |
| 24 | Clone 24 | KAS4 | Kasalasa community |
| 25 | Clone 25 | KAS5 | Kasalasa community |
| 26 | Clone 26 | KAS6 | Kasalasa community |
| 27 | Clone 27 | KAS7 | Kasalasa community |
| 28 | Clone 29 | KAS8 | Kasalasa community |
| 29 | Clone 31 | KAK1 | Germplasm (Kakara) |
| 30 | Clone 32 | KAK2 | Germplasm (Kakara) |
| 31 | Clone 33 | KAK3 | Germplasm (Kakara) |
| 32 | Clone 34 | KAK4 | Germplasm (Kakara) |
| 33 | Clone 35 | KAK5 | Germplasm (Kakara) |
| 34 | Clone 37 | KAK6 | Germplasm (Kakara) |
| 35 | Clone 38 | KAK7 | Germplasm (Kakara) |
| 36 | Clone 38B | KAK8 | Germplasm (Kakara) |
| 37 | Clone BB35 | KAK9 | Germplasm (Kakara) |
| 38 | Clone 39 | KAK10 | Germplasm (Kakara) |
| 39 | Clone 40 | KAK11 | Germplasm (Kakara) |
| 40 | Clone 68 | Clone 68 | Kakara commercial clone |
| 41 | Clone 143 | Clone 143 | Kakara commercial clone |
| 42 | Clone 236 | Clone 236 | Kakara commercial clone |
| 43 | Clone 318 | Clone 318 | Kakara commercial clone |
| 44 | Clone 1212 | Clone 1212 | Kakara commercial clone |

### Extraction of genomic DNA

DNA was isolated using DNeasy Plant Mini Kit (QIAGEN, www.qiagen.com). 100 mg of a ground sample of the leaves of each tea clone was weighed in a 1.5 ml micro-centrifuge tube in duplicate. 400 µl of Buffer AP1 and 4 µl of Ribonucleases (RNase) A was added to make a concentration of 100 g/ml and was vortexed vigorously until a homogenous mixture is obtained. This was incubated for 30 minutes at 65°C and mixed 2-3 times during incubation by inverting the tube. This step lyses the cells. One hundred and thirty micro litres of Buffer AP2 was added, mixed and incubated on ice for 5 minutes and later centrifuged for 15 minutes at 16100 xg. This step precipitates detergents, proteins and polysaccharides. Centrifuge for 15 minutes at 16,100 xg. The supernatant was added to the QIAshredder Mini Spin Column placed in a 2 ml collection tube and centrifuged for 2 minutes at 16,100 xg. The flow-through fraction was transferred to a new tube without disturbing the cell-debris pellet. 1.5 volumes of Buffer AP3/E were added to the cleared lysate and mixed immediately. 650 µl of the mixture was applied, including any precipitate that may be formed to the DNeasy Mini Spin Column sitting in a 2 ml collection tube. Centrifugation was done for 1 minute at

8000 xg and the flow-through was discarded. These procedures were repeated using the rest of the sample. The DNeasy Mini Spin Column was placed in a new 2 ml collection tube where 500 μl Buffer AW was added to the DNeasy Mini Spin Column and centrifuged for 1 minute at 8000 xg. The flow-through was discarded and the collection tube reused in the next step. 500 μl of Buffer AW was added to the DNeasy Mini Spin Column and centrifuged for 2 minutes at 16,100 xg to dry the membrane. The DNeasy Mini Spin Column was transferred to a 1.5 ml micro-centrifuge tube and 50 μl of Buffer AE pipetted directly on to the DNeasy membrane and incubated for 5 minutes at room temperature (15-25°C) and then centrifuge for 1 minute at 8,000 xg to elute. The previous step was repeated once to obtain a final volume of 100 μl. The samples were labelled and stored at -20°C for downstream analysis.

### Primer design for SSR markers

Computer simulation was performed to design primers using Integrated DNA Technologies (IDT) primer tool (www.eu.idtdna.com). The 'Oligonucleotide Properties Calculator' online tool was used to determine the appropriate melting Temperature (Tm), Guanine-Cytosine (GC) content and any possible formation of Hairpin's for all primers. The primers were designed according to the following standards; a Tm closest to 60°C, a GC content of at least 45%, a primer length of not less than 20 nucleotides and a minimum Polymerase Chain Reaction (PCR) product size of 700 base pair (bp). All primers had a preferable length of between 19 to 22 nucleotide bases. Once the primers are delivered in a pellet form, concentrated stock solutions of 100 μM (100 pmol/μl) was prepared by dissolving each lyophilized pellet in Tris Ethylenediamine Tetra-Acetic acid (EDTA) buffer (10 mMTris, pH 7.5; 1 mM EDTA, pH 8.0). 10 μM working solution was prepared for future use during PCR.

### Agarose gel electrophoresis

A 2% agarose gel was prepared by using electrophoresis grade agarose in 1x Tris, Acetic acid, EDTA (TAE) buffer (pH 8.0) and ethidium bromide (10 mg/ml) to a concentration of 15 mg/ml. The PCR product and 6x loading dye solution was mixed and loaded on the agarose gel. A GeneRuler of 1 kb DNA Ladder (New England's BioLabs® Inc.) was loaded in the first lane as a molecular size marker. The PCR products was separated at 100 Volts for 60 minutes and visualized on a gel documentation device known as "VacutecSyngene G-box" (Vacutech) under Ultraviolet (UV) light.

### SSR PCR-RFLP and data analyses

For PCR Restriction Fragment Length Polymorphism (RFLP) analysis, 10 μL of the amplified fragments were digested for 2 hrs with the restriction endonucleases HaeIII and HinfI, following the conditions recommended by the manufacturer (Thermo Scientific). Restriction fragments were analysed in 3% agarose gels stained with GelRed nucleic acid (Thermo Scientific) and visualised under UV light "Gel Doc EZ Gel Documentation System" (Bio-Rad). Band fragments were assigned as '0' or '1', where '0' meant absence and '1' represented the presence of a particular band [7].

GenAlEx 6.41 was used for computation [8]. The genetic distance produced in GenAlex was used to calculate the pair-wise Euclidean distance. Principal Coordinates Analysis (PCoA) was carried out using GenAlex 6.41, using the matrix of pair-wise distance. Genetic relationship among accessions was determined using Cluster analysis. Molecular Evolutionary Genetics Analysis (MEGA) was used to generate a phylogenetic tree using the maximum likelihood and visualized using Tree View Version 1.6.6. Further phylogenetic tree was drawn with *Histone* (*H2A*) gene sequences retrieved from National Center for Biotechnology Information (NCBI) database for other organisms to ascertain the evolutionary pattern using the *H2A* gene sequence of *C. sinensis* as a reference genome (Table 2). Analysis of Molecular Variance (AMOVA) was carried out to serve as an additional index of genetic diversity using GenAlex 6.41. Genetic structure for the five populations was analyzed using STRUCTUREv2.3.4 [9]. K-values from 2 to 5 populations were tested, at least 10 simulations were executed by each K-value, after which the burn-in length was set to 1,000,000 and the number of Markov Chain Monte Carlo (MCMC) repetitions after burn-in at 500,000. The result files from the runs were uploaded to online STRUCTURE-HARVESTER and then the deltaK (ΔK ad hoc) method described by was used to detect the most probable number of genetic clusters (K) [10].

**Table 2:** Organisms habouring *Histone* Gene (*H2A*) downloaded from the National Center for Biotechnology Information (NCBI) database.

| S. No. | Species name | Family | Phylum | Accession number | H2A variant |
|---|---|---|---|---|---|
| 1 | *Arabidopsis thaliana* | *Brassicaceae* | *Tracheophyta* | NM_115313.5 | Histone H2A 11 (HTA11), messenger Ribonucleic Acid (mRNA) |
| 2 | *Brassica napus* | *Brassicaceae* | *Tracheophyta* | XM_013872185.3 | Histone H2A variant 1 (LOC106431387), mRNA |
| 3 | *Camelina sativa* | *Brassicaceae* | *Tracheophyta* | XM_010428864.2 | Histone H2A variant 1-like (LOC104712061), transcript variant X3, mRNA |
| 4 | *Eutrema salsugineum* | *Brassicaceae* | *Tracheophyta* | XM_024158034.1 | Histone H2A variant 1 (LOC18021211), transcript variant X3, mRNA |
| 5 | *Capsella rubella* | *Brassicaceae* | *Tracheophyta* | XM_023782691.1 | Histone H2A variant 1 (LOC17885780), transcript variant X2, mRNA |
| 6 | *Pisum sativum* | *Fabaceae* | *Anthophyta* | M64838.1 | H2A mRNA |

| 7 | *Raphanus sativus* | *Brassicaceae* | *Tracheophyta* | XM_018636184.1 | Histone H2A variant 1-like (LOC108862140), transcript variant X2, mRNA |
|---|---|---|---|---|---|
| 8 | *Zea mays* | *Poaceae* | *Tracheophyta* | DQ244576.1 | clone 9352 mRNA sequence |
| 9 | *Tarenaya hassleriana* | *Cleomaceae* | *Embryophyta* | XM_010527048.2 | Histone H2A variant 1 (LOC104803165), transcript variant X2, mRNA |
| 10 | *Sorghum bicolor* | *Poaceae* | *Magnoliophyta* | XM_002465414.2 | Histone H2A (LOC8084912), mRNA |
| 11 | *Panicum virgatum* | *Poaceae* | *Angiospermophyta* | XM_039975633.1 | Histone H2A (LOC120692361), mRNA |
| 12 | *Setaria viridis* | *Poaceae* | *Tracheophyta* | XM_034715571.1 | Histone H2A (LOC117836195), mRNA |
| 13 | *Panicum virgatum* | *Poaceae* | *Angiospermophyta* | XM_039930729.1 | Histone H2A-like (LOC120652802), mRNA |
| 14 | *Panicum hallii* | *Poaceae* | *Tracheophyta* | XM_025946826.1 | Probable histone H2A.5 (LOC112881904), mRNA |
| 15 | *Oryza brachyantha* | *Poaceae* | *Magnoliophyta* | XM_006649799.3 | Probable histone H2A.5 (LOC102718440), mRNA |
| 16 | *Triticum aestivum* | *Poaceae* | *Tracheophyta* | XM_044516806.1 | Histone H2A-like (LOC123094949), mRNA |
| 17 | *Triticum dicoccoides* | *Poaceae* | *Tracheophyta* | XM_037586054.1 | Histone H2A (LOC119310245), mRNA |
| 18 | *Aegilops tauschii* | *Poaceae* | *Tracheophyta* | XM_020324854.2 | Subspecies *strangulata* histone H2A-like (LOC109766086), mRNA |
| 19 | *Hordeum vulgare* | *Poaceae* | *Magnoliophyta* | XM_045092015.1 | Subspecies *vulgare* histone H2A-like (LOC123397487), mRNA |
| 20 | *Lolium rigidum* | *Poaceae* | *Tracheophyta* | XM_047230454.1 | Histone H2A-like (LOC124697931), mRNA |
| 21 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028267242.1 | Histone H2A variant 1 (LOC114320128), transcript variant X5, mRNA |
| 22 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028267239.1 | Histone H2A variant 1 (LOC114320128), transcript variant X3, mRNA |
| 23 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028246852.1 | Probable histone H2A variant 1 (LOC114301900), mRNA |
| 24 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028200447.1 | *Camellia sinensis* sucrose synthase (LOC114260358), partial mRNA |
| 25 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028208381.1 | Histone deacetylase 19 (LOC114267341), transcript variant X2, mRNA |
| 26 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028208380.1 | Histone deacetylase 19 (LOC114267341), transcript variant X1, mRNA |
| 27 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028208383.1 | Histone deacetylase 19 (LOC114267341), transcript variant X4, mRNA |
| 28 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028208382.1 | Histone deacetylase 19 (LOC114267341), transcript variant X3, mRNA |
| 29 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028259483.1 | Histone H2A.1-like (LOC114313134), mRNA |
| 30 | *Cannabis sativa* | *Cannabaceae* | *Tracheophyta* | XM_030637134.1 | Probable histone H2A.5 (LOC115709095), mRNA |
| 31 | *Coffea Arabica* | *Rubiaceae* | *Magnoliophyta* | XM_027251918.1 | Histone H2A.1 (LOC113727647), mRNA |

| 32 | *Coffea eugenioides* | *Rubiaceae* | *Tracheophyta* | XM_027304509.1 | Histone H2A.1 (LOC113761496), mRNA |
|---|---|---|---|---|---|
| 33 | *Coffea Arabica* | *Rubiaceae* | *Magnoliophyta* | XM_027258862.1 | Histone H2A.1-like (LOC113732858), mRNA |
| 34 | *Camellia sinensis* | *Theaceae* | *Magnoliophyta* | XM_028254354.1 | Histone H2A (LOC114308704), mRNA |
| 35 | *Helianthus annuus* | *Asteraceae* | *Tracheophyta* | XM_022117558.2 | Probable histone H2A.5 (LOC110868406), mRNA |
| 36 | *Helianthus annuus* | *Asteraceae* | *Tracheophyta* | XM_022117557.2 | Histone H2A (LOC110868405), mRNA |
| 37 | *Ziziphus jujuba* var. spinose | *Rhamnaceae* | *Spermatophyta* | XM_048477318.1 | Histone deacetylase 19 (LOC107410367), transcript variant X8, mRNA |
| 38 | *Ziziphus jujuba* var. spinose | *Rhamnaceae* | *Spermatophyta* | XM_048477315.1 | Histone deacetylase 19 (LOC107410367), transcript variant X7, mRNA |
| 39 | *Fusarium thapsinum* | *Nectriaceae* | *Ascomycota* | MN737765.1 | Isolate A16_S15 internal transcribed spacer 1, partial sequence, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal Ribonucleic Acid (rRNA) gene, partial sequence |

## RESULTS

### Fragment analysis

The restriction enzyme digestion of HaeIII and HinfI of the SSR marker of *H2A.Cs* gene amplicons revealed distinct band profiles designated with letters for all 44 *Camellia sinensis* clones (Table 3). HaeIII enzyme produced five different band patterns; A (640 bp), B (230+200 bp), C (320 bp) and D (410+180 bp), while HinfI showed only two types of patterns; D (410+180 bp) and E (300+210+120 bp) (Table 3).

### Band patterns of the 5 tea clone populations sampled from the Mambilla plateau

The band pattern of the *H2A.Cs* gene sequences from the SSR markers revealed that tea clones from Kusuku, Bangoba, Kasalasa, Kakara and the cultivated produced 9, 6, 6, 9 and 6 bands respectively. Sequences of tea clones obtained from Kusuku and Kakara populations show common bands each. Mean Heterozygosity (He) were 0.277 ± 0.051, 0.214 ± 0.070, 0.274 ± 0.078, 0.385 ± 0.043 and 0.262 ± 0.078 for Kusuku, Bangoba, Kasalasa, Kakara and the cultivated populations with mean unbiased Heterozygosity (uHe) of 0.291 ± 0.053, 0.225 ± 0.074, 0.292 ± 0.083, 0. 403 ± 0.045 and 0.291 ± 0.086, respectively (Figures 1 and 2) (Table 4).

### Pairwise genetic distance and identity

Nei Genetic Distance (GD) is based on the population sampled. The result shows that the widest genetic distances was between tea clones obtained from Kakara germplasm and those from Kasalasa population (0.035) while the narrowest was between tea clones from Bangoba and those of Kasuku (0.004) (Table 5). However, for Nei Genetic Identity (GI), tea clones sampled from cultivated population shared close identity with that from Kasalasa population (0.998). Generally, the tea clones from the 5 populations showed very close identity. From unbiased genetic distance, the widest genetic distance was between tea clone populations from Kakara and those of Kasalasa (0.012). The unbiased genetic identity revealed that the closest was between tea clones obtained from the cultivated population and those of Kasalasa (1.031) while those from Kakara and Bangoba show low

identity of 0.988.

### Phylogenetic analysis

The phylogenetic analysis describing the hypothetical evolutionary relationship shows that two distinct clusters were generated with 6 tea clones in cluster 1 and 38 tea clones in cluster 11. Given the population studied- Kusuku, Bangoba, Kasalasa, Kakara and cultivated, there were no marked pattern of clustering as the different populations were mixed up together. However, the clones 1212, 318, 236, BB35 and 143 were grouped together in sub-cluster 11a. The test of reliability of phylogenetic tree at each of the node is such that the higher bootstrap values, the closer will be ancestral or evolutionary relatedness. For instance, clone 143 share 98% bootstrap value with KAK8 and KAK10 (Kakara population), KAS4 and KAS7 (Kasalasa population), BAN1 and BAN 10 (Bangoba population) and KUS8 (Kusuku population), respectively. CloneBB35 share 83% ancestry relationship with KUS2 (Kusuku population) and BAN2 (Bangoba population). In cluster 1, KAK6 (Kakara population) share evolutionary relatedness of 98% with KUS7 (Kusuku population). Tea clones obtained from Kasalasa population (KAS2) share 95% relatedness with KAS5.

Figure 2 is a phylogenetic analysis of *Histone* gene (*H2A.Cs*) sequences comprising of tea clones obtained from Mambilla plateau and other organisms from the NCBI database. The essence of this phylogenetic analysis was to trace the evolutionary origin of *H2A* gene. Sequences in the tea clones obtained from Mambilla plateau on the premise that the origin or the country of adoption is uncertain on one hand and the inheritance of this gene from other organisms on the other hand. The phylogenetic tree revealed 5 clusters, which was majorly based on species though there was some overlapping of organisms. For instance, tea clones obtained from Mambilla plateau were clustered in group IV but in cluster V, comprising of *H2A* gene sequences of other *Camellia sinensis* variants of the gene, including gene sequences of *P. sativum*, *C. sativum* and *Z. mays*, tea clones obtained from Kakara population (KAK2), Bangoba population (BAN7) and Kasalasa population (KAS6) were grouped together. This overlapping pattern into other clusters were observed in cluster II and cluster III (Figures 3 and 4) (Tables 6-8).

**Table 3:** Restriction fragment analysis of Polymerase Chain Reaction (PCR) amplified products of the different tea clones.

| Clones | Clone Population | PCR Amplified product base pair (bp) | Restriction fragments (bp) and individual profile (A-E) | |
| --- | --- | --- | --- | --- |
| | | | HaeIII processing bodies (pb) | HinfI (pb) |
| Clone 1 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 2 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 3 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 4 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 5 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 6 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 7 | Kusuku | 985 | 230+200 | 410+180 |
| Clone 8 | Kusuku | 985 | 230+200 | 300+210+120 |
| Clone 9 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 10 | Kusuku | 985 | 320 | 300+210+120 |
| Clone 11 | Bangoba | 985 | 320 | 300+210+120 |
| Clone 12 | Bangoba | 985 | 320 | 300+210+120 |
| Clone 13 | Bangoba | 985 | 320 | 300+210+120 |
| Clone 14 | Bangoba | 985 | 320 | 300+210+120 |
| Clone 15 | Bangoba | 985 | 640 | 410+180 |
| Clone 16 | Bangoba | 985 | 640 | 300+210+120 |
| Clone 17 | Bangoba | 985 | 640 | 300+210+120 |
| Clone 18 | Bangoba | 985 | 320 | 300+210+120 |
| Clone 19 | Bangoba | 985 | 640 | 300+210+120 |
| Clone 20 | Bangoba | 985 | 410+180 | 410+180 |
| Clone 21 | Kasalasa | 985 | 320 | 300+210+120 |
| Clone 22 | Kasalasa | 985 | 320 | 300+210+120 |
| Clone 23 | Kasalasa | 985 | 320 | 300+210+120 |
| Clone 24 | Kasalasa | 985 | 640 | 300+210+120 |
| Clone 25 | Kasalasa | 985 | 640 | 300+210+120 |
| Clone 26 | Kasalasa | 985 | 640 | 300+210+120 |
| Clone 27 | Kasalasa | 985 | 410+180 | 410+180 |
| Clone 29 | Kasalasa | 985 | 230+200 | 300+210+120 |
| Clone 31 | Kakara | 985 | 320 | 300+210+120 |
| Clone 32 | Kakara | 985 | 320 | 300+210+120 |
| Clone 33 | Kakara | 985 | 320 | 300+210+120 |
| Clone 34 | Kakara | 985 | 230+200 | 300+210+120 |
| Clone 35 | Kakara | 985 | 230+200 | 300+210+120 |
| Clone 37 | Kakara | 985 | 320 | 300+210+120 |
| Clone 38 | Kakara | 985 | 320 | 300+210+120 |
| Clone 38B | Kakara | 985 | 230+200 | 410+180 |
| Clone BB35 | Kakara | 985 | 230+200 | 410+180 |
| Clone 39 | Kakara | 985 | 320 | 300+210+120 |
| Clone 40 | Kakara | 985 | 640 | 300+210+120 |
| Clone 68 | Cultivated | 985 | 640 | 410+180 |

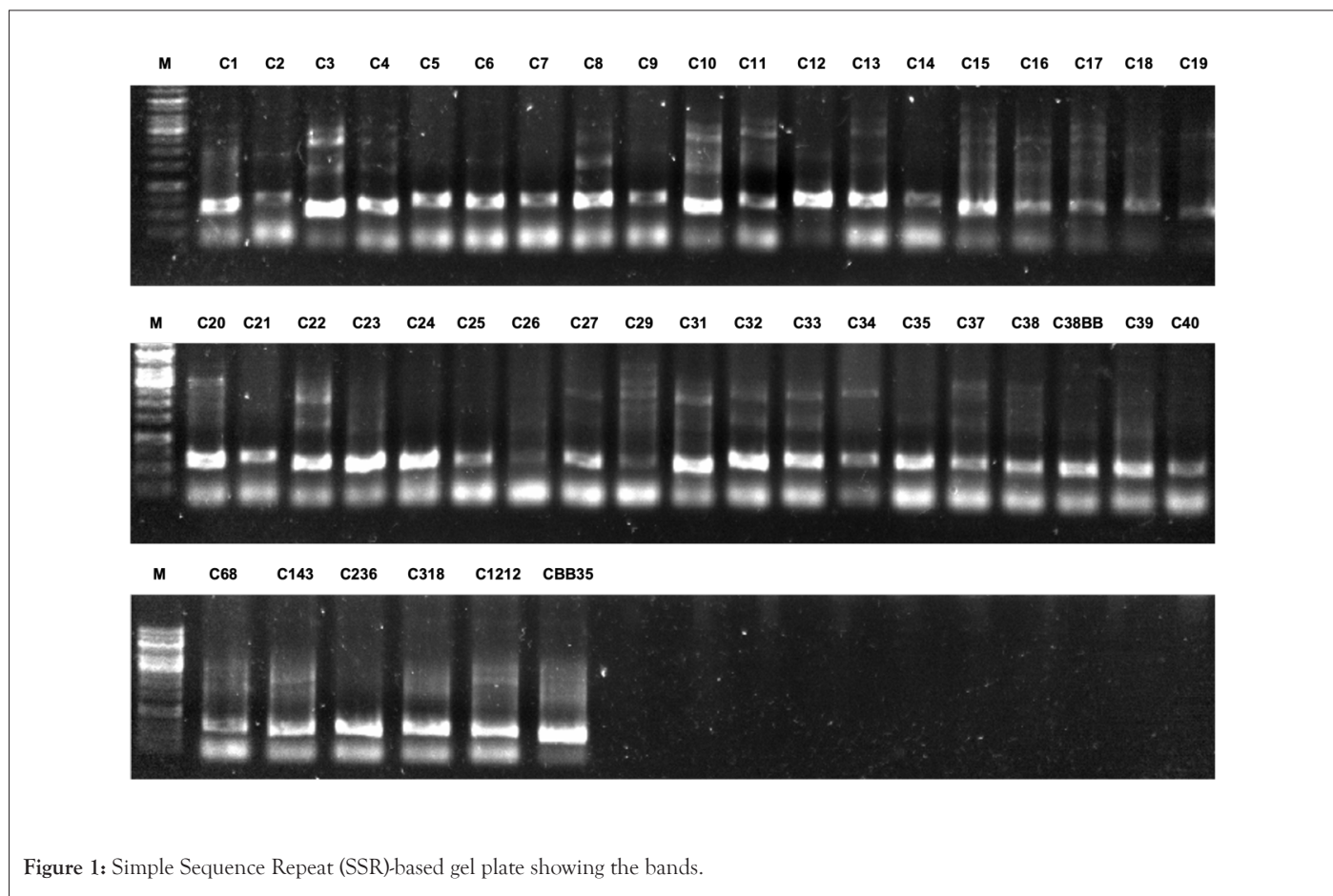| Clone 143 | Cultivated | 985 | 320 | 300+210+120 |
|-----------|-----------|-----|-----|-------------|
| Clone 236 | Cultivated | 985 | 640 | 300+210+120 |
| Clone 318 | Cultivated | 985 | 320 | 300+210+120 |
| Clone 1212 | Cultivated | 985 | 640 | 300+210+120 |



**Figure 1:** Simple Sequence Repeat (SSR)-based gel plate showing the bands.



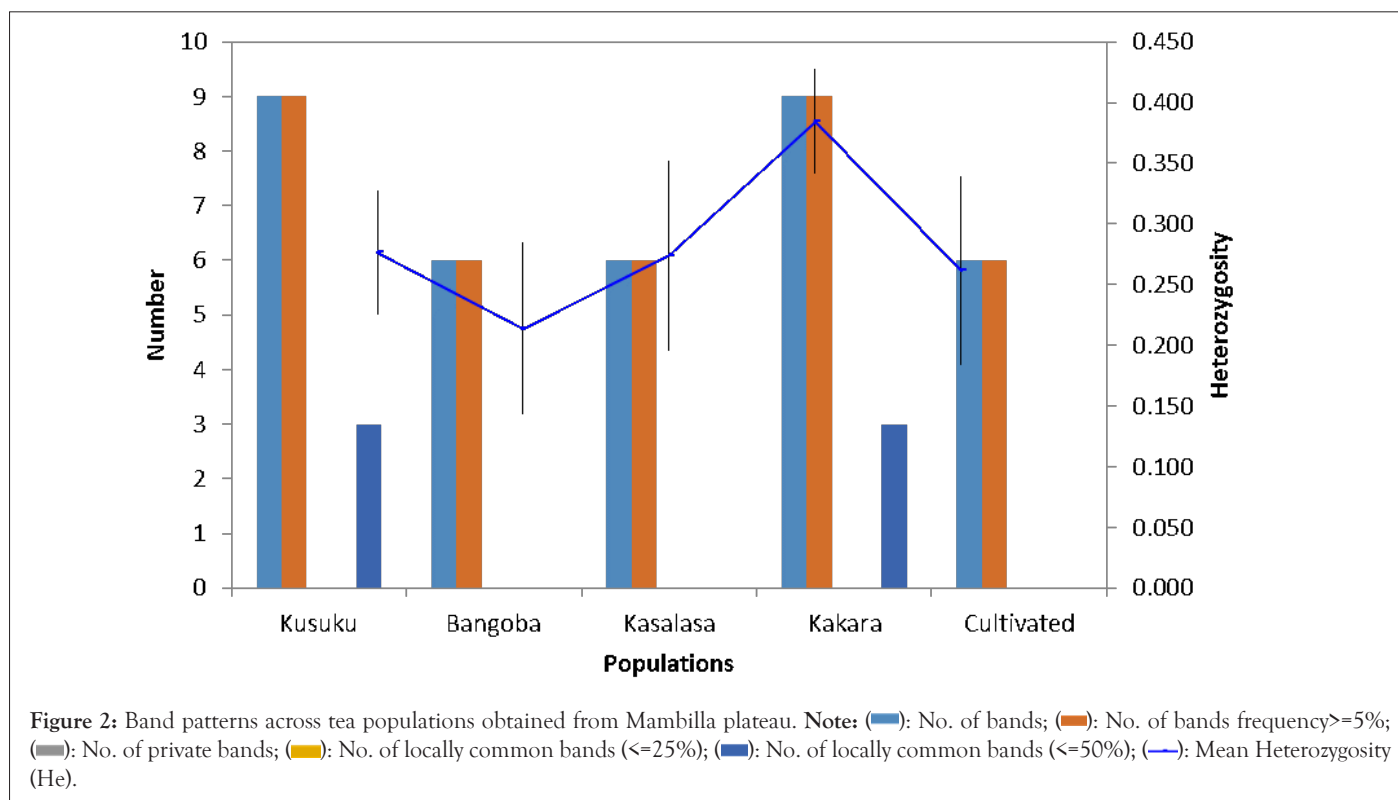**Figure 2:** Band patterns across tea populations obtained from Mambilla plateau. **Note:** (■): No. of bands; (■): No. of bands frequency>=5%; (■): No. of private bands; (■): No. of locally common bands (<=25%); (■): No. of locally common bands (<=50%); (——): Mean Heterozygosity (He).

**Table 4:** Total band patterns for binary (diploid) data by populations.

| Population | Kusuku | Bangoba | Kasalasa | Kakara | Cultivated |
|---|---|---|---|---|---|
| No. bands | 9 | 6 | 6 | 9 | 6 |
| No. bands frequency>=5% | 9 | 6 | 6 | 9 | 6 |
| No. private bands | 0 | 0 | 0 | 0 | 0 |
| No. of locally common bands (<=25%) | 0 | 0 | 0 | 0 | 0 |
| No. of locally common bands (<=50%) | 3 | 0 | 0 | 3 | 0 |
| Mean Heterozygosity (He) | 0.277 ± 0.051 | 0.214 ± 0.070 | 0.274 ± 0.078 | 0.385 ± 0.0443 | 0.262 ± 0.078 |
| Mean unbiased Heterozygosity (uHe) | 0.291 ± 0.053 | 0.225 ± 0.074 | 0.292 ± 0.083 | 0.403 ± 0.045 | 0.291 ± 0086 |

**Table 5:** Pairwise population matrix of Nei genetic distance.

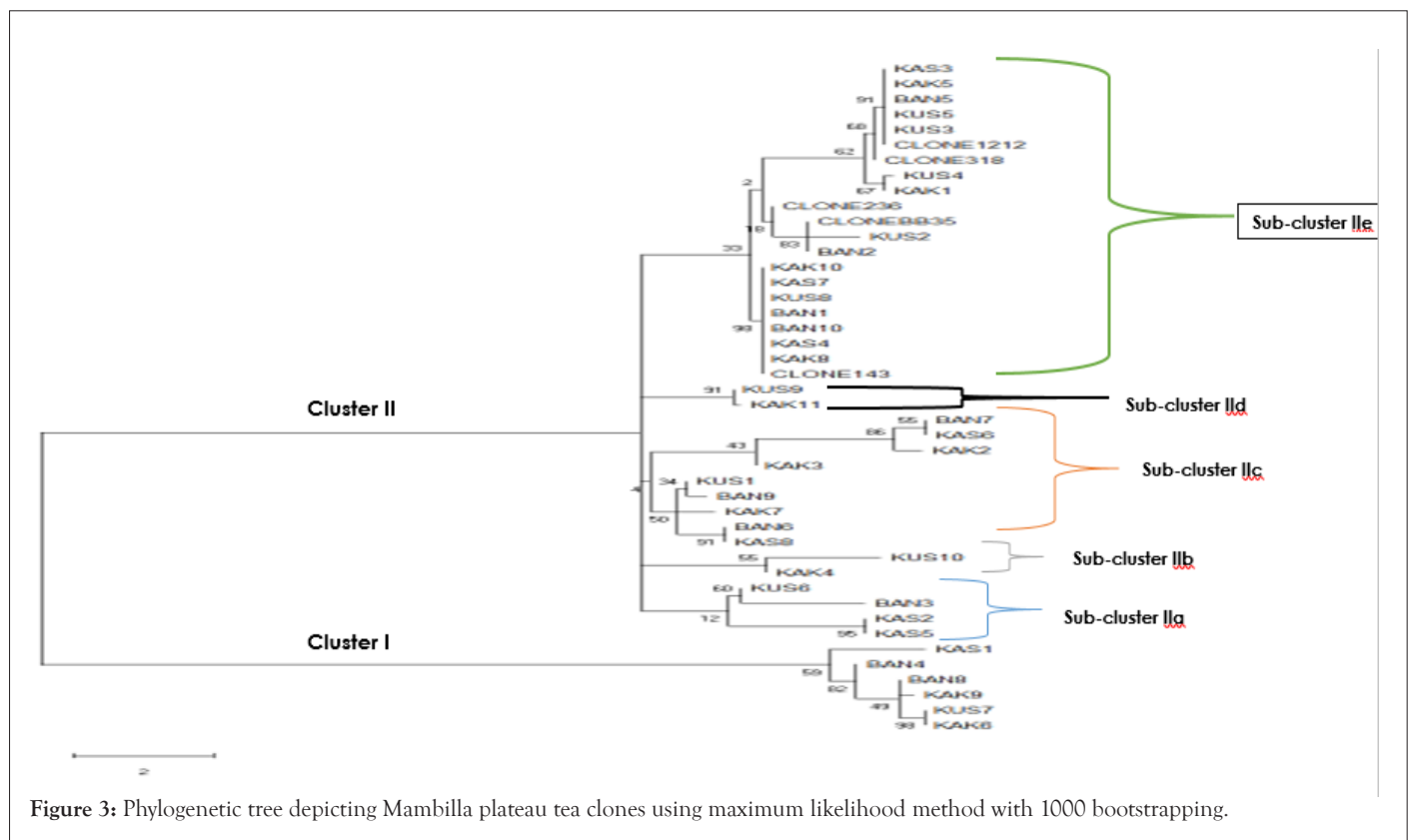| | Kusuku | Bangoba | Kasalasa | Kakara | Cultivated |
|---|---|---|---|---|---|
| Kusuku | 0 | - | - | - | - |
| Bangoba | 0.004 | 0 | - | - | - |
| Kasalasa | 0.028 | 0.021 | 0 | - | - |
| Kakara | 0.018 | 0.034 | 0.035 | 0 | - |
| Cultivated | 0.016 | 0.01 | 0.002 | 0.03 | 0 |

**Note:** Nei genetic distance: -1˚Ln (Nei identity).



**Figure 3:** Phylogenetic tree depicting Mambilla plateau tea clones using maximum likelihood method with 1000 bootstrapping.
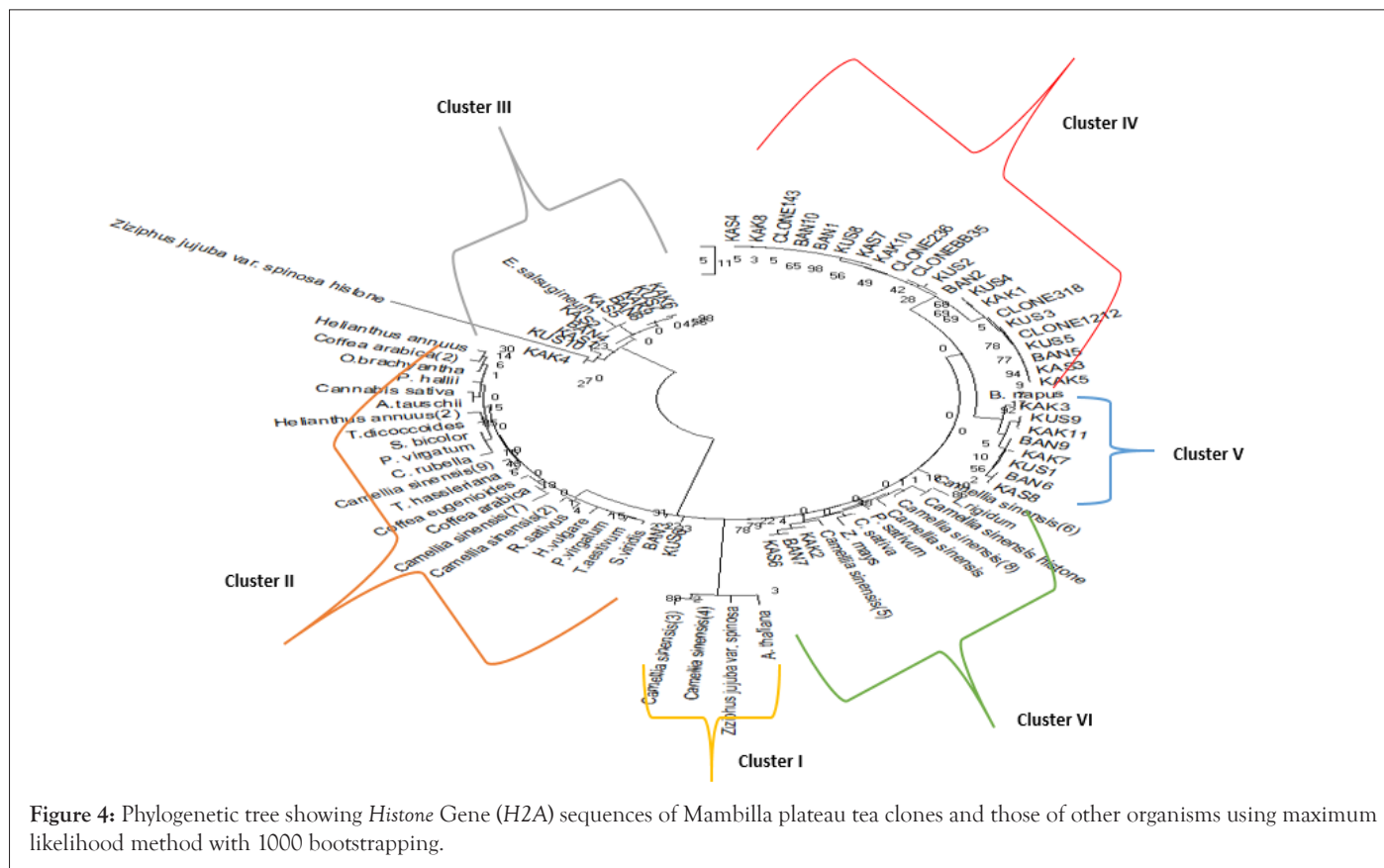
**Figure 4:** Phylogenetic tree showing *Histone* Gene (*H2A*) sequences of Mambilla plateau tea clones and those of other organisms using maximum likelihood method with 1000 bootstrapping.

**Table 6:** Pairwise population matrix of Nei genetic identity.

|  | Kusuku | Bangoba | Kasalasa | Kakara | Cultivated |
|---|---|---|---|---|---|
| Kusuku | 1 | - | - | - | - |
| Bangoba | 0.996 | 1 | - | - | - |
| Kasalasa | 0.972 | 0.979 | 1 | - | - |
| Kakara | 0.982 | 0.966 | 0.965 | 1 | - |
| Cultivated | 0.984 | 0.99 | 0.998 | 0.97 | 1 |

**Table 7:** Pairwise population matrix of Nei unbiased genetic distance.

|  | Kusuku | Bangoba | Kasalasa | Kakara | Cultivated |
|---|---|---|---|---|---|
| Kusuku | 0 | - | - | - | - |
| Bangoba | 0 | 0 | - | - | - |
| Kasalasa | 0.005 | 0.001 | 0 | - | - |
| Kakara | 0 | 0.012 | 0.008 | 0 | - |
| Cultivated | 0 | 0 | 0 | 0 | 0 |

**Note:** Nei unbiased genetic distance: -1˚Ln (Nei unbiased identity).

**Table 8:** Pairwise population matrix of Nei unbiased genetic identity.

|  | Kusuku | Bangoba | Kasalasa | Kakara | Cultivated |
|---|---|---|---|---|---|
| Kusuku | 1 | - | - | - | - |
| Bangoba | 1.013 | 1 | - | - | - |
| Kasalasa | 0.995 | 0.999 | 1 | - | - |
| Kakara | 1.008 | 0.988 | 0.992 | 1 | - |
| Cultivated | 1.014 | 1.017 | 1.031 | 1.005 | 1 |

**Note:** Nei genetic distance: -1˚Ln (Nei identity); Nei unbiased genetic distance: -1˚Ln (Nei unbiased identity).

## Analysis of Molecular Variation (AMOVA)

The AMOVA result shows that there was 99.9% within population variation and 0.1% among population variation (Table 9).

## Principal Coordinate Analysis (PCoA) of tea plant clones from 5 populations in Mambilla plateau

The analysis was done through covariant matrix with data standardization. Percentage of variation as explained by the first three axes showed that 1st, 2nd and 3rd axes had 65.93%, 33.61% and 0.46%, respectively with 100% cumulative percentage variance. Six clusters were generated from the tea plant clones under investigation. Clones 39 (Kakara population) BB35 (cultivated) were clustered together while clones 16 (Bangoba), 21 (Kasalasa), 29 (Kasalasa) and 143 (cultivated), were grouped together. However, clones 9 (Kusuku), 31, 35 and 37 (Kakara population) were grouped together while others were clustered together the location not withstanding (Figures 5-8) (Table 10).

## Maximum composite likelihood estimate of the pattern of nucleotide substitution

From our result, it was observed that the rate of transitional substitutions of Adenine (A) $\rightarrow$ Guanine (G) 39.93 was greater than G $\rightarrow$ A (37.04) while from Cytosine (C) $\rightarrow$ Thymine (T) (0.65) and T $\rightarrow$ C was (0.39). However, for transversional substitutions on the *H2A.Cs* gene sequences, from T $\rightarrow$ A or C $\rightarrow$ A was 2.54 and from T $\rightarrow$ G or C $\rightarrow$ G was 2.74, respectively. The highest transversional substitution was between A $\rightarrow$ T
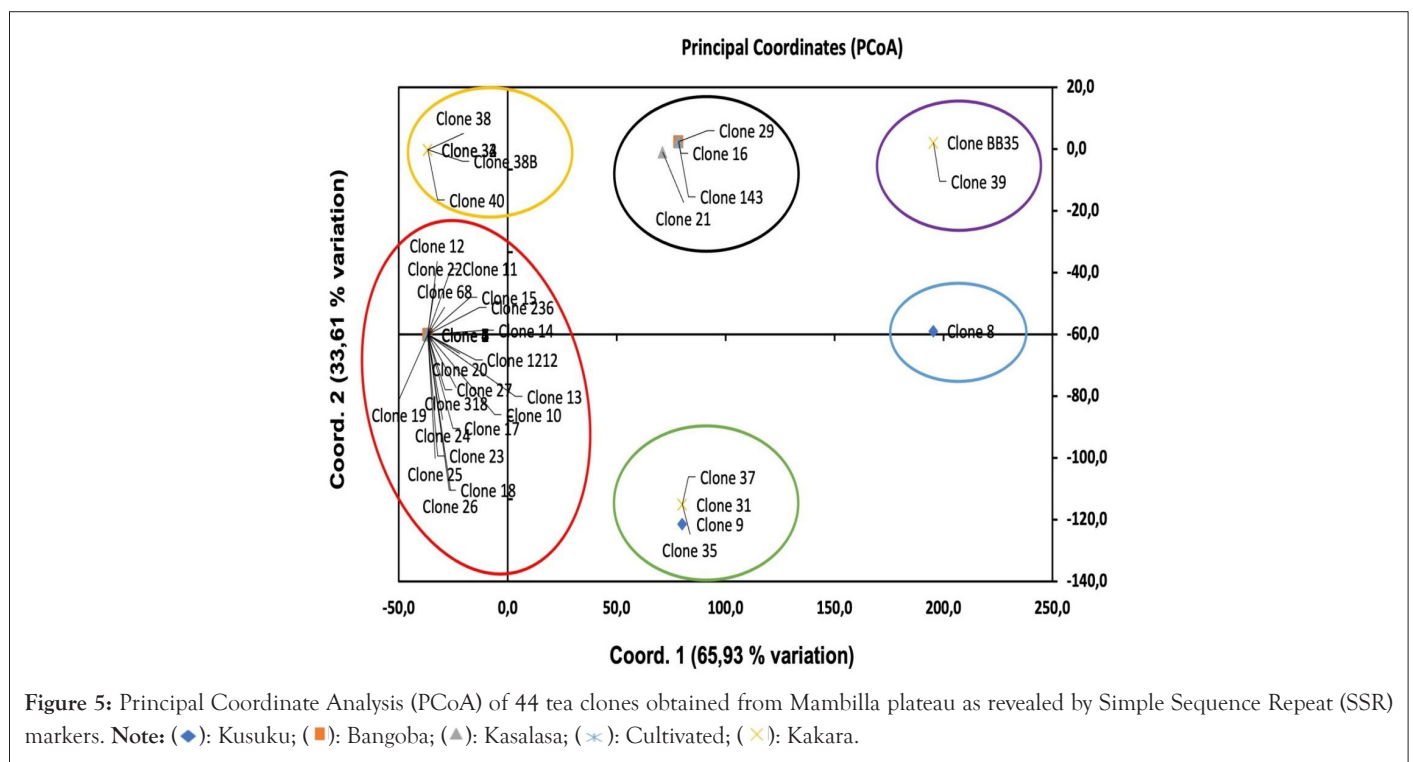
or G $\rightarrow$ T (3.58). Additionally, nucleotide frequencies were A (23.14%), T (32.53%), C (19.38%) and G (24.95%) while transition/transversion rate ratios are k1=14.562 for purines and k2=0.183 for pyrimidines. The overall Transition/Transversion bias (R)=3.414, R=(A*G*k1+T*C*k2)/((A+G)*(T+C)). On the other hand, when the *H2A* gene sequences from tea clones obtained from Mambilla plateau were merged with those retrieved from NCBI database, the results show that the nucleotide frequencies are 24.20% (A), 26.98% (T/U), 22.62% (C) and 26.20% (G). The transition/transversion rate ratios are k1=2.753 (purines) and k2=1.45 (pyrimidines). The overall transition/transversion bias is R=1.052, where R=(A*G*k1+T*C*k2)/((A+G)*(T+C)) (Tables 11 and 12).

## Test of the homogeneity of substitution patterns between sequences (Disparity index test)

We however, conducted disparity index test using a Monte Carlo test with 500 replicates to estimate the P-values. The probability of rejecting the null hypothesis that sequences have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences. When P<0.05, it is considered significant. Thus if significant, the alternative hypothesis would be accepted that gene sequences did not evolve with the same pattern of substitution. Our result showed that only the gene sequence of tea clone obtained from Bangoba (BAN3) did not evolve with the same pattern of substitution with other gene sequences of other tea clones.

**Table 9:** Analysis of Molecular Variance (AMOVA) in tea population obtained from Mambilla plateau, Taraba State.

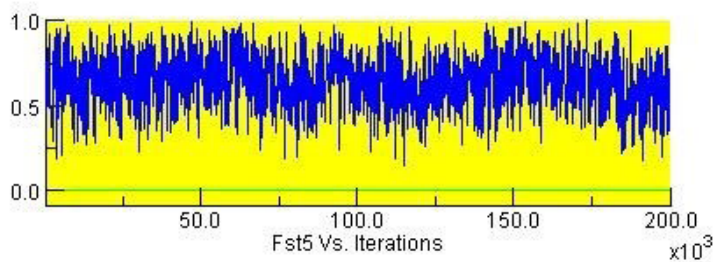| Source | Degrees of freedom (df) | Sum of Squares (SS) | Mean Squares (MS) | Estimated Variance | % |
|---|---|---|---|---|---|
| Among populations | 4 | 2.891 | 0.723 | 0.016 | 0.1 |
| Within populations | 40 | 30.109 | 0.984 | 0.984 | 99.9 |
| Total | 44 | 33 | - | - | 100 |



**Figure 5:** Principal Coordinate Analysis (PCoA) of 44 tea clones obtained from Mambilla plateau as revealed by Simple Sequence Repeat (SSR) markers. **Note:** (◆): Kusuku; (■): Bangoba; (▲): Kasalasa; (✳): Cultivated; (✕): Kakara.

**Figure 6:** Fixation Index (FST) analysis showing the degree of differentiation among tea populations obtained from Mambilla plateau, Taraba state.
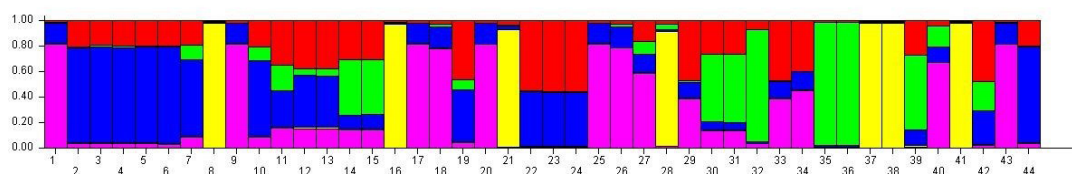


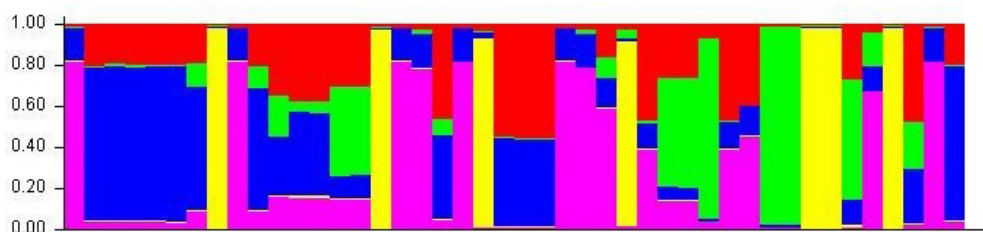**Figure 7:** Structure analysis bar plot of tea clones obtained from Mambilla plateau, Taraba state.



**Figure 8:** Simple Sequence Repeat (SSR) markers based structure analysis of k=5 bar plot of single lines.

**Table 10:** Tajima's neutrality test for the comparison of sequences of clones from Kakara and those from the genbank.

| M | S | $p_s$ | Θ | [] | D |
|---|---|---|---|---|---|
| 44 | 415 | 1 | 0.229885 | 0.644782 | 6.671491 |
| 82 | 414 | 1 | 0.200891 | 0.711333 | 8.768648 |

**Note:** M: Number of sequences; S: Number of segregating sites; $p_s=S/n$, n: Total number of sites; Θ: ps/a1; []: Nucleotide diversity; D: Tajima test statistic.

**Table 11:** Maximum composite likelihood estimate of the pattern of nucleotide substitution.

| | Adenine (A) | Thymine (T) | Cytosine (C) | Guanine (G) |
|---|---|---|---|---|
| A | - | 3.58 | 2.13 | **39.93** |
| T | 2.54 | - | **0.39** | 2.74 |
| C | 2.54 | **0.65** | - | 2.74 |
| G | **37.04** | 3.58 | 2.13 | - |

**Note:** Each entry shows the probability of substitution from one base (row) to another base (column). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in regular.

**Table 12:** Maximum composite likelihood estimate of the pattern of nucleotide substitution.

| | Adenine (A) | Thymine (T) | Cytosine (C) | Guanine (G) |
|---|---|---|---|---|
| A | - | 6.57 | 5.51 | **17.56** |
| T | 5.89 | - | **7.98** | 6.38 |
| C | 5.89 | **9.25** | - | 6.38 |
| G | **16.23** | 6.57 | 5.51 | - |

**Note:** Each entry shows the probability of substitution from one base (row) to another base (column). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in regular.

## Allele-frequencies of divergence among populations (Net nucleotide distance)

Our result also revealed allele frequencies as it pertains to divergence among the tea populations studied. It shows that the highest diverged population was Bangoba tea population and that of Karaka population (Germplasm) (0.5062), which was followed by the Karaka population and the cultivated population (0.3910). However, the least diverged was observed between Kusuku tea population and those from Kasalasa population (0.0196).

We also report the expected heterozygosity between individuals in same cluster/population. Kusuku population had an expected heterozygosity of 0.2407, Bangoba population had 0.1875 while Kasalasa population was 0.1567. Karaka population had 0.1529 and the cultivated population had 0.1644 (Table 13).

**Table 13:** Allele-frequencies of divergence among populations (Net nucleotide distance).

| | Kusuku | Bangoba | Kasalasa | Kakara | Cultivated |
|---|---|---|---|---|---|
| Kusuku | - | 0.1559 | 0.0196 | 0.3099 | 0.0716 |
| Bangoba | - | - | 0.1786 | 0.5062 | 0.231 |
| Kasalasa | - | - | - | 0.3084 | 0.0555 |
| Karaka | - | - | - | - | 0.391 |
| Cultivated | - | - | - | - | - |

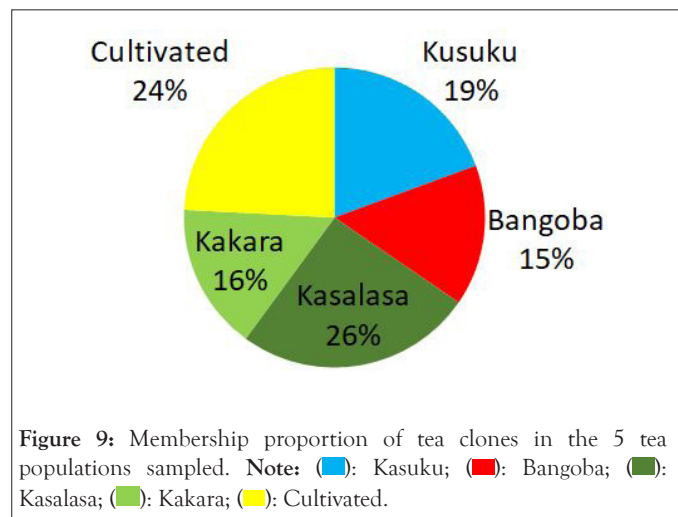## FST analysis (Degree of differentiation among tea populations)

The proportion of the total genetic variance contained in the sub populations FST were estimated for the five populations, which shows that the Karaka population (germplasm) had the highest mean value of FST of 67.81% (0.6781), followed by the cultivated population with 61.33% (0.6133). The remaining populations FST were such that Bangoba tea population>Kasalasa population>Kusuku population (0.5984; 0.5810; 0.5032).

## Structure analysis of tea populations from Mambilla plateau

We conducted a structure analysis using structure software and structure harvester to decipher the population structure of the tea clones obtained from Mambilla plateau. From the structure analysis, blue colour belonged to Kusuku tea population, Red colour cluster belonged to the Bangoba population while purple colour cluster belonged to the Kasalasa population. Green colour cluster belonged to the Kakara (geermplasm) population while yellow colour cluster belonged to the cultivated population.

Our observation from the population structure bar plot shows that there were no distinct group having only one solid colour as all the populations were shared. This implies that they were most probable genetically similar.

The percentage of membership of the sample in each of the 5 population clusters are Kusuku (19.4%), Bangoba (15.2%), Kasalasa (25.5%), Kakara (15.8%) and cultivated population was 24.2%, respectively (Figure 9).



**Figure 9:** Membership proportion of tea clones in the 5 tea populations sampled. **Note:** (■): Kasuku; (■): Bangoba; (■): Kasalasa; (■): Kakara; (■): Cultivated.

## DISCUSSION

The lack of genomic information on local genetic resources has impeded efficient conservation, utilization and improvement of crops, especially in tea plants in the Mambilla plateau and subsequent exploitation of its full agronomic and breeding potentials. Genetic diversity is fundamentally important in crop improvement as it provides plants with the capacity to meet the demands of changing environmental conditions. Thus establishing the genetic diversity and population structure of a species can guide the selection of appropriate conservation and sustainable utilization strategies [11]. Importantly, identification of genetic variation in crop species, especially in tea clones is essential for a robust and long-term success of breeding programmes and maximizes the use of germplasm resources. Regrettably, this lack of genomic information on Mambilla tea plants has impeded efficient conservation and improvement of the crop and harnessing of its full phenomics (agronomic and bioactive compounds).

Obviously, several methods have been adopted to evaluate genetic diversity in tea population ranging from the morphology, biochemistry, molecular markers as well as sensory traits [12-14]. However, the identification, classification and characterization of tea germplasm has been effectively and efficiently carried out using modern technologies involving the deployment of molecular markers, which has proven to be more robust and informative [11,12,15-17]. Though tea plant is an out-crossing species but is usually propagated through vegetative and released as clonal varieties, clonal identification is traditionally based on morphological descriptors-plant shape, stem width, leaf shape, young leaf type, fruit shape and unfortunately all these phenotypic traits show continuous variation pattern and high plasticity [16,18-21].

This might also make it difficult or contribute to the misidentification of tea clones based only on the morphology thus the need for high throughput technologies such as DNA markers (SSRs). It is known that genetic diversity is usually higher in cross-pollinated perennials than in selfed annuals [22]. With the breeding platform, high levels of genetic diversity for cultivated tea and its closely wild relatives is expected and have been reported in a number of previous studies [23,24]. Some reports point to high genetic diversity in tea population in Korea tea germplasm, while some had observed that there exists low genetic diversity in some tea plant germplasm [17,20,23,25].

Recall that gene diversity is otherwise known as expected heterozygosity, which is fundamentally crucial in the measurement of genetic variation in a population. Low value indicates homogeneous and admixture nature of the sub-populations. In our result, the expected Heterozygosity (He) among the populations ranged between 0.214-0.385, which show very low genetic diversity in contrast to the earlier reports of 0.792 across 21 SSR markers though higher than reports of (0.652) in 280 tea accessions; (0.640) in 450 tea accessions; (0.543) in 185 Chinese tea cultivars and (0.680) in 64 Sri Lankan tea cultivars [26].

According to Yao et al., it might be difficult to make comparison of the degree of genetic diversity given that the sampling schemes, number of SSR, sizes of SSR repeats as well as location of the SSR in the genome differ [23].

This notwithstanding, it gives a guide. Tea clones obtained from Kakara population having the highest level of heterozygosity (0.385) is expected given that they habour the *in situ* germplasm. The Nei Genetic Distance (GD) and Nei Genetic identity (GI) are important components of population genetic diversity estimates. The GD between Kakara tea population and Kasalasa population was 0.035, which was the widest; though narrow. GD is directly proportional to the GI. We report that the GI between the cultivated tea population and those obtained from Kasalasa population was 99.8% (0.998), suggesting high similarity. A robust interview with the tea breeders in the Kakara tea plantation, Mambilla plateau have it that tea farmers from the adjoining communities collect their seedlings from the Kakara nursery, which may have influenced the extent of homozygosity observed between the two populations. Our results also suggest that some workers in the tea plantation may have collected seedlings from the floor of the *in situ* germplasm, which had led to the level of heterozygosity observed in other tea populations.

A phylogenetic tree is an estimate of the relationship among taxa or sequences and their hypothetical common ancestry [27-29]. It is important to highlight the importance of the present phylogenetic analysis. Deciphering evolutionary patterns of gene sequences helps in understanding the functional and structural features governing their activity. First, the analysis was used to estimate relationships among five tea populations from the Mambilla plateau represented by the *H2A* gene sequences. Second, the analysis was used to understand the relationships among the *H2A* gene sequences of tea plants and those retrieved from the database. The phylogenetic tree showed only two distinct clusters though there was no traceable pattern of clustering in terms of location. However, the cultivated tea populations were seemingly clustered in cluster 2, which may suggest shared common ancestry origin.

The essence of bootstrapping is to measure the reliability of the nodes on the phylogenetic tree. Bootstrap percentage less than 70% is not reliable enough for testing evolutionary relatedness or common ancestry origin [29]. The higher the value, the closer the relationship or the higher the sequence similarity between or among the sequences or accessions compared. For instance, going by the submission of Hall, it means that KUS7 (Kusuku population) and KAK6 (Kakara population) sharing 98% bootstrap value is very similar in their gene sequence, suggesting common ancestry origin [29]. Additionally, clone 143 share bootstrap value of 96% with KAK10, KAS7, KUS8, BAN10, BAN1, KAS4 and KAK8 while clones 1212 and 318 share bootstrap value of 91% with KAS3, KAK5, BAN5, KUS5 and

KUS3. Interestingly, there are other tea clones that share more than 70% bootstrap values that was not location-dependent. Recall that the cultivated populations have phenotypic characteristics that distinguish them, which should be the underlying reason for their selection and cultivation. Other clones could be exploited for the purpose of breeding. When the *H2A* gene sequences of tea clones obtained from Mambilla plateau were clustered with those of other organisms, including cultivars of tea plants from other countries, the results show that though clustering was more organism-specific, however, some tea clones from the Mambilla plateau were grouped together with *H2A* gene sequences of others, which could be reasoned that this gene sequence may have been shared and inherited from a common ancestry origin. The Principal Coordinate Analysis (PCoA) result corroborated the observed clustering patter in the phylogenetic tree.

The allele frequencies of divergence among populations in terms of net nucleotide distance showed that Bangoba tea population was the highest diverged from Kakara population (0.5062), though this is not in tandem with other results presented in this research. One would have expected that the increase in the values of heterozygosity (gene diversity) should be in agreement with the degree of differentiation among tea populations. But this was not the case. There has also been a series of evolutionary bottlenecks as a result of the narrow genetic diversity, which can be widened by the introgression of genes from wild species landraces into the primary gene pool [30]. Most of the study mentioned previously explained that tea types show high levels of genetic diversity (He=0.71, range 0.644-0.760), which is consistent with previous studies using SSRs [16,23,24]. Park opined that the low genetic diversity observed in Korean tea may have stemmed from the fact that it was established from a limited gene stock from China [17]. This might also be the singular reason underlying the low genetic variation reported in our present work. Direct comparison could be misleading since these studies estimated the genetic diversity of tea cultivars based on sample grouping by country or region and not by tea type. The reason for any discrepancy in result of genetic diversity may be marker bias or the inclusion of a wider gene pool, although both are difficult to quantity given the lack of sampling details associated with their study [24]. It has been reported that understanding genetic diversity and the population structure of crops, especially tea plant germplasm is very critical for effective collection, conservation and utilization [31].

Lee reported 99% within population variation and 1% among variation using SSR markers on 410 tea accessions in Korea germplasm. However, our result on AMOVA showed 99.9% within population variation and 0.1% among population variation using 44 tea clones from Mambilla plateau, Nigeria. Our result and that of Lee et al., points to the fact that there was low genetic diversity in the tea collection [31]. Nucleotide diversity is a measure of genetic variation, which is synonymous with expected heterozygosity and directly measures the degree of polymorphisms within a given population. The neutral theory of molecular evolution holds that most evolutionary changes at the molecular level as well as most of the variation within and between species are attributable to random genetic drifts of mutant alleles, which are selectively neutral. Some mutations are deleterious and are removed by natural selection and thus do not contribute significantly to species' variation at a molecular level. The nucleotide diversity (π) from the present study was 0.644782. From the AMOVA result, especially for the among population variation, it implies that the mutation that may have occurred

did not affect significantly variability comparing the different populations.

It was observed that the estimate of the pattern of nucleotide substitution on the histone gene sequences was such that the highest transitional substitution was G → A (37.04) while the highest transversional substitution was A → T/G → T (3.58). Nucleotide frequencies were A (23.14%), T (32.53%), C (19.38%) and G (24.95%), implying that the highest was thymine. We also report transitional/transversional rate ratio of 14.562 for purines and 0.183 for pyrimidines with overall bias R of 3.414. Monte Carlo-based disparity test is the probability for rejecting the null hypothesis that gene sequences (*H2A*) have evolved with the same pattern of substitutions. Our result revealed that only tea clones from Bangoba population (BAN3) did not evolve with the same pattern of substitution with other *H2A* gene sequences of tea clones. The implication is that this population should be critically investigated to decipher their origin and probably trace them to the germplasm in Kakara.

When Tajima statistics is negative, it means that there is an excess of low frequency polymorphisms relative to expectation implying population size expansion. But when it is positive, it signifies low levels of both low and high frequency polymorphisms, indicating a decrease in population size/balancing selection. Our result on Tajima statistic (6.671491) suggests that there are haplotypes (more average heterozygosity) than that of the segregating sites. Recall that the number of segregating sites was 415, by extension means that there is more average heterozygosity. There are lack of rare alleles, which means there is a balancing selection and sudden population contraction. Balancing selection occurs when multiple alleles are maintained in a population, which can result in their preservation over long evolutionary period of time. The Tajima statistic obtained in our report showed that many multiple alleles are maintained in the tea plant population studied.

We report the highest degree of differentiation among the tea populations in Kakara population, 0.6781 (67.81%) while other populations ranged from 0.5032-0.6133. Interestingly, the Kakara tea population is the *in-situ* germplasm in the Mambilla plateau. This germplasm has been there for the past 3 decades and the cultivated tea clones in the Kakara tea plantation were generated from there. The FST analysis however, corroborated this fact as the FST of the cultivated population was 0.6133 (61.33%), which shows that they were derived from the germplasm (Kakara population). Structure based genetic clustering is a powerful tool to define genetically distinct groups [32-35]. Our result showed that structure analysis can be used to track genetic admixtures [36]. The implication of the structure analysis result means that the tea population in the Mambilla plateau originated from the same source thus the low genetic diversity observed. Our result on structure analysis showed that the 5 tea populations were to large extent genetically similar. Percentage of membership of the tea clones in each of the 5 populations revealed that Bangoba and Kakara tea populations had the least percentage admixture (15.2% and 15.8%), respectively. What it does suggest is that both the tea clones and those in the *in-situ* germplasm are an admixture from the original stock before introduction into Mambilla plateau while other populations are admixture within the germplasm. Our structure analysis showed an admixture of tea clones from different tea population. This may probably have indicated an artificial or spontaneous hybrid origin for these tea clones [37]. The other reason could be traceable to the fact that the adjoining tea growing communities where tea clones were sampled obtained their seedlings/cuttings from the Kakara Tea Plantation germplam/nursery of the cultivated clones, which are also derived from the Kakara germplasm [38-49].

## CONCLUSION

The genetic structure of species reflects the mutation, recombination, genetic drift and selection effects experienced by population during the course of evolution and thus population structure analyses are deemed important for gaining an accurate understanding of the genetic relationship among germplasms. In this study, three clustering methods were used (PCoA, Phylogenetic tree and Structure) to analyze the population structure of *C sinensis*. PCoA and phylogenetic tree were performed to classify the assessed materials based on their genetic similarity coefficients and genetic distance, which can clarify the intuitive relationship between populations. U ndoubtedly, to assist breeding programmes and conserve genotypes of tea cultivars going forward, germplasm collection should be assembled, which should include local cultivars and more recently germplasm exchanged between the main tea producing countries. For emphasis, there have been reports pertaining to the evaluation of genetic diversity and population structure of tea germplasm collection in China and India which is lacking in Nigeria. Clustering pattern, the geographical origin not withstanding could indicate the possible sharing of common genomic regions, which occurs across the accessions. Taking the results together, there is narrow genetic diversity in the tea population in Kakara Tea Plantation, Mambilla plateau, which needs urgent introgression of elite genotype.

## REFERENCES

1. Adegoke O, Awolola NA, Ajuluchukwu JN. Prevalence and pattern of cardiovascular-related causes of out-of-hospital deaths in Lagos, Nigeria. Afr Health Sci. 2018;18(4):942-949.

2. Zhen YS. Tea: bioactivity and therapeutic potential. CRC Press. 2002.

3. Olaniyi OO, Odeyemi OA, Adewale BD, Oloyede AA, Anagbogu CF, Adeigbe OO, et al. Tea (*Camellia sinensis*) breeding in Nigeria: past and present status. Int J Sci Res. 2014;4(9):1-4.

4. Gupta PK, Varshney RK. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. 2000;113(3):163-185.

5. Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M, et al. Comparative analysis of population genetic structure in *Athyrium distentifolium* (*Pteridophyta*) using AFLPs and SSRs from anonymous and transcribed gene regions. Mol Ecol. 2005;14(6):1681-1695.

6. Chen L, Zhou ZX, Yang YJ. Genetic improvement and breeding of tea plant (*Camellia sinensis*) in China: from individual selection to hybridization and molecular breeding. Euphytica. 2007;154:239-248.

7. Kim OG, Sa KJ, Lee JR, Lee JK. Genetic analysis of maize germplasm in the Korean Genebank and association with agronomic traits and simple sequence repeat markers. Genes Genom. 2017;39:843-853.

8. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012;28(19):2537-2539.

9. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-959.

10. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611-2620.

11. Udensi UO, Emeagi IL, Thomas TL, Ghosh S, Achilonu CC. Genetic Polymorphism and Lineage of Pigeon Pea [*Cajanus cajan* (L.) Millsp.] inferred from Chloroplast and Nuclear DNA gene regions. Arab J Sci Eng. 2021;46(6):5285-5297.

12. Wambulwa MC, Meegahakumbura MK, Chalo R, Kamunya S, Muchugi A, Xu JC, et al. Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. Tree Genet Genomes. 2016;12:1-10.

13. Li Y, Chen C, Li Y, Ding Z, Shen J, Wang Y, et al. The identification and evaluation of two different color variations of tea. J Sci Food Agric. 2016;96(15):4951-4961.

14. Li Y, Wang C, Huai Q, Guo F, Liu L, Feng R, et al. Effects of tea or tea extract on metabolic profiles in patients with type 2 diabetes mellitus: a meta-analysis of ten randomized controlled trials. Diabetes Metab Res Rev. 2016;32(1):2-10.

15. Chen L, Yamaguchi S. Rapd markers for discriminating tea germplasms at the inter-specific level in china. Plant Breed. 2005;124(4):404-409.

16. Fang W, Cheng H, Duan Y, Jiang X, Li X. Genetic diversity and relationship of clonal tea (*Camellia sinensis*) cultivars in China as revealed by SSR markers. Plant Syst Evol. 2012;298:469-483.

17. Young-Goo P, Kaundun SS, Zhyvoloup A. Use of the bulked genomic DNA-based RAPD methodology to assess the genetic diversity among abandoned Korean tea plantations. Genet Resour Crop Evol. 2002;49:159-165.

18. Lai JA, Yang WC, Hsiao JY. An assessment of genetic relationships in cultivated tea clones and native wild tea in Taiwan using RAPD and ISSR markers. Bot Bull Acad Sin. 2001;42:93-100.

19. Oh CJ, Lee S, You HC, Chae JG, Han SS. Genetic diversity of wild tea (*Camellia sinensis* L.) in Korea. Korean J Plant Res. 2008;21(1):41-46.

20. Chen L, Apostolides Z, Chen ZM, Jeong BC, Park YG. Tea plant (*Camellia sinensis*) breeding in Korea. Glob Tea Breed. 2012:263-288.

21. Gai Z, Wang Y, Jiang J, Xie H, Ding Z, Ding S, et al. The quality evaluation of tea (*Camellia sinensis*) varieties based on the metabolomics. Hort Sci. 2019;54(3):409-415.

22. Miller AJ, Gross BL. From forest to field: perennial fruit crop domestication. Am J Bot. 2011;98(9):1389-1414.

23. Yao MZ, Ma CL, Qiao TT, Jin JQ, Chen L. Diversity distribution and population structure of tea germplasms in China revealed by EST-SSR markers. Tree Genet Genomes. 2012;8:205-220.

24. Taniguchi F, Kimura K, Saba T, Ogino A, Yamaguchi S, Tanaka J. Worldwide core collections of tea (*Camellia sinensis*) based on SSR markers. Tree Genet Genomes. 2014;10(6):1555-1565.

25. Kaundun SS, Zhyvoloup A, Park YG. Evaluation of the genetic diversity among elite tea (*Camellia sinensis* var. *sinensis*) accessions using RAPD markers. Euphytica. 2000;115:7-16.

26. Karunarathna KH, Mewan KM, Weerasena OV, Perera SA, Edirisinghe EN, Jayasoma AA. Understanding the genetic relationships and breeding patterns of Sri Lankan tea cultivars with genomic and EST-SSR markers. Sci Hortic. 2018;240:72-80.

27. Nei M, Kumar S. Molecular evolution and phylogenetics. Oxford university press. 2000.

28. Penny D. Inferring Phylogenies.-Joseph Felsenstein, 2003, Sinauer Associates, Sunderland, Massachusetts. Sys Biol. 2004;53(4):669-670.

29. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013;30(5):1229-1235.

30. Udensi OU, Edu N, Ikpeme E, Onung O, Emeagi L, Nwanze B, et al. Genotyping of pigeon pea [*Cajanus cajan* (L.) Millsp.] accessions obtained from international institute of tropical agriculture (IITA) germplasm using random amplified polymorphic DNA. J Exp Agric Int. 2017;17(3):1-12.

31. Lee KJ, Lee JR, Sebastin R, Shin MJ, Kim SH, Cho GT, et al. Assessment of genetic diversity of tea germplasm for its management and sustainable use in Korea genebank. Forests. 2019;10(9):780.

32. Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, et al. Olive domestication and diversification in the Mediterranean Basin. New Phytol. 2015;206(1):436-447.

33. Wambulwa MC, Meegahakumbura MK, Kamunya S, Muchugi A, Möller M, Liu J, et al. Multiple origins and a narrow genepool characterise the African tea germplasm: concordant patterns revealed by nuclear and plastid DNA markers. Sci Rep. 2017;7(1):4053.

34. Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez G J, Doebley J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot. 2008;95(10):1240-1253.

35. Castillo L, Martínez AI, Gelis S, Ruiz-Herrera J, Valentín E, Sentandreu R. Genomic response programs of *Saccharomyces cerevisiae* following protoplasting and regeneration. Fungal Genet Bio. 2008;45(3):253-265.

36. Excoffier L, Estoup A, Cornuet JM. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics. 2005;169(3):1727-1738.

37. Sharma RK, Negi MS, Sharma S, Bhardwaj P, Kumar R, Bhattachrya E, et al. AFLP-based genetic diversity assessment of commercially important tea germplasm in India. Biochem Genet. 2010;48:549-564.

38. Hamrick JL, Godt MJ, Sherman-Broyles SL. Factors

influencing levels of genetic diversity in woody plant species. New For 6. 1992;42:95-124.

39. Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987;236(4803):787-792.

40. Ji PZ, Li H, Gao LZ, Zhang J, Cheng ZQ, Huang XQ. ISSR diversity and genetic differentiation of ancient tea (*Camellia sinensis* var. *assamica*) plantations from China: implications for precious tea germplasm conservation. Pak J Bot. 2011;43(1):281-291.

41. Chen J, Wang P, Xia Y, Xu M, Pei S. Genetic diversity and differentiation of *Camellia sinensis* L.(cultivated tea) and its wild relatives in Yunnan province of China, revealed by morphology, biochemistry and allozyme studies. Genet Resour Crop Evol. 2005;52:41-52.

42. Ben-Ying LI, You-Yong LI, Yi-Chun TA, Li-Yuan WA, Cheng H, Ping-Sheng WA. Assessment of genetic diversity and relationship of tea germplasm in Yunnan as revealed by ISSR markers. Acta Agr Sinica. 2010;36(3):391-400.

43. Fang WP, Meinhardt LW, Tan HW, Zhou L, Mischke S, Zhang D. Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. Hortic Res. 2014;1:14035.

44. Tan LQ, Peng M, Xu LY, Wang LY, Chen SX, Zou Y, et al. Fingerprinting 128 Chinese clonal tea cultivars using

SSR markers provides new insights into their pedigree relationships. Tree Genet Genomes. 2015;11:90.

45. Paul S, Wachira FN, Powell W, Waugh R. Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. Theor Appl Genet. 1997;94:255-263.

46. Balasaravanan T, Pius PK, Kumar RR, Muraleedharan N, Shasany AK. Genetic diversity among south Indian tea germplasm (*Camellia sinensis*, *C. assamica* and *C. assamica* spp. *lasiocalyx*) using AFLP markers. Plant Sci. 2003;165(2):365-372.

47. Karthigeyan S, Rajkumar S, Sharma RK, Gulati A, Sud RK, Ahuja PS. High level of genetic diversity among the selected accessions of tea (*Camellia sinensis*) from abandoned tea gardens in western Himalaya. Biochem Genet. 2008;46:810-819.

48. Roy SC, Chakraborty BN. Genetic diversity and relationships among tea (*Camellia sinensis*) cultivars as revealed by RAPD and ISSR based fingerprinting. Ind J Biotechnol. 2009;8(4):370-376.

49. Raina SN, Ahuja PS, Sharma RK, das SC, Bhardwaj P, Negi R, et al. Genetic structure and diversity of India hybrid tea. Genet Resour Crop Evol. 2012;59:1527-1541.