Original Article

Genetic relationships and Population Structure among Nigerian Ethnic Groups (Ibibio, Igbo, Hausa, Tiv and Yoruba) Based on Nine DNA Loci

Authors: Utom-Obong U. Akpan [*1,2], Oluwafemi D. Amusa[2], Olumide A. Adebesin[2], Peter Onaja[3], Kudighe Udoh[4], Nura Garba[5], Joy Okpuzor[2], and Khalid O. Adekoya[2]

1. Anatomy Department, Bowen University, Iwo, Osun State, Nigeria
2. Department of Cell Biology and Genetics, University of Lagos, Lagos, Nigeria
3. College of Medicine, Benue State University, Makurdi, Nigeria
4. Community Health, University of Uyo Teaching Hospital, Uyo, Nigeria
5. Department of Haematology, Aminu Kano Teaching Hospital, Kano, Nigeria

Corresponding Author: **Utom-Obong U. Akpan,** Anatomy Department, Bowen University, Iwo, Osun State, Nigeria
Phone numbers: +2348131841061
E-mail: utom-obong.akpan@bowen.edu.ng

Abstract:

Context: Examining the genetic relationships between populations presents an opportunity to observe patterns of genetic changes over time. By comparing populations, it is possible to see factors that may have sparked the evolution of a trait or genotype as well as the genetic diversity of the populations. Genetic diversity in the groups in Nigeria are still largely undefined. This can be achieved with the use of genetic markers.

Aims: To determine the genetic relationships between Igbo, Ibibio, Yoruba, Tiv, and Hausa ethnic groups using nine short tandem repeat DNA markers.

Settings and Design: Observational studies of 250 consenting participants

Methods and Material: Individuals of Igbo, Yoruba, Hausa, Ibibio, and Tiv were randomly sampled from their ethnic homes in Nigeria at participating institutions. DNA was extracted from blood samples using commercial DNA kits, and tested for nine autosomal short tandem repeat loci alleles using PCR and electrophoresis. The allele types and sizes were read and recorded for each locus for every individual

Statistical analysis used:

Allelic frequencies, population pairwise genetic distances (FST and RST, analysis of molecular variance (AMOVA), and principal component analysis (PCA) were calculated using GenAlEx v6.502.

Results:

Fixation index (FST) was between 0.001 and 0.500 in the five populations. Estimated variation was 99.98% within individuals and 0.10% among populations. Four heterogeneous clusters were observed with three axes accounting for 32.86% variation.

Conclusions: The genetic relationships closely mirrored linguistic groupings showing that language may have shaped early interaction and ultimately gene transfer between these groups.

Keywords: Relationship, Alleles, Ethnic, Population, Nigeria

Introduction:

Research into the genetic diversity of African populations has intensified in the last few decades. During this time, African populations have been observed to have the highest genetic diversity among other human populations. [1-4] Nigeria is located in Western Africa, bordering the Gulf of Guinea, between the Benin Republic, Niger Republic, Chad, and Cameroon, with a landmass of 923,768 sq. km and a population of approximately 220 million people. [5] Estimates of the number of distinct ethnic groupings vary from 250 to as many as 580, with the most prominent being Hausa, Edo, Fulani, Ibibio, Kanuri, Nupe, Tiv, Ijaw, Itsekiri, Urhobo, Aguleri, Umuleri, Jukun, Ogoni, Mambila, Banso, Kamba, Yoruba, and Igbo (Ibo).[5]

Few studies have investigated the population genetic structure of the Nigerian human population(s). Veeramah et al. [6] revealed the overall genetic homogeneity in the Cross River region in the face of language variation. Adeyemo et al. [7] reported the stratification of genetic structures in four West African population groups. Titilayo et al. [8] reported on the genetic variation among Igbo, Yoruba and Hausa samples. Recently, Joshi et al. [9] provided insights into the genetic diversity of some Nigerian populations. The diversity and structure of most of the 520

ethnolinguistic groups in Nigeria, including the Tiv, are still largely undefined. There is a need to study more groups to further elucidate the genetic diversity of Africans in general and Nigeria in particular. [6, 10]

The advent of genetic markers has made it possible to study population genetics at the molecular level, providing better diversity estimates. One such type is short tandem repeats (STRs). These are microsatellites with short sequences of DNA (1-10 base pairs) lying within genetic markers and short tandem repeats (STRs) end to end in a particular region of the genome (the loci). [11] Microsatellites are typically characterized by a high mutation rate and therefore a high level of polymorphism, resulting in different alleles in the population. [12,13] There are several thousand STR loci in the human genome. STRs on non-sex chromosomes are widely used as genetic markers in human identification, forensics, paternity investigations and other cases of kinship analysis. The loci are usually typed using polymerase chain reaction (PCR) and electrophoresis. When a sufficient number of loci are tested, a genetic profile is generated that statistically provides the discriminating power needed for human identification. [14]

Population structure appears as extensive allelic diversity and heterozygosity at the genomic level. Contemporary approaches use molecular markers, including SNPs and STR data, to reveal genetic diversity in populations. [15] Several assumptions and relationships exist between the properties of molecular markers, and these could be used to estimate genetic diversity in populations. These methods include principal component analysis (PCA), the fixation index (FST), analysis of molecular variance (AMOVA), and Nei's genetic distance, with varying degrees of precision. These measures of population structure and genetics are usually estimated using software programs. The list of programs can be obtained from different websites, including DuckDNA

(https://www.duckdna.org/softwares/) and the University of Washington (WU) popgen software page (https://courses.washington.edu/popgen/Software.htm).

Materials and Methods:

Ethical Approval

Approval for the research, consent forms, and sample collection methods was obtained from the Ethics Review Board of the Lagos University Teaching Hospital with reference ADM/DCST/HREC/1921 dated 08/09/2014.

Sample size determination

The sampling method was an initial purposive sampling of five selected ethnic groups from Nigeria, to represent the three largest groups (Igbo, Yoruba and Hausa) and two smaller groups (one Northern and one Southern - Tiv and Ibibio, respectively. This was followed by a convenience sampling of individuals from the chosen ethnic groups. The estimated effective population size that will capture all the alleles in the chosen ethnic groups was determined to be between 35 and 50 individuals based on previous reports. [16,17] 50 individuals were therefore sampled from each population

Participant Recruitment and Consent

Potential participants who self-identified as being from any of the Hausa, Yoruba, Igbo, Ibibio or Tiv ethnic/linguistic groups were recruited at the blood donor clinic of the participating hospitals. These hospitals are in regions with a high density of members of the ethnic groups included in the study. It includes the Lagos University Teaching Hospital (LUTH) and locations in Ilesha, Osun State and Ogbomoso, Oyo (for Yoruba), the University of Uyo Teaching Hospital (UUTH) (for Ibibio) and Aminu Kano Teaching Hospital (AKTH) (for Hausa), Anambra State University Teaching Hospital (ANSUTH), Amaku, and Awka communities in Anambra State (for Igbo) and Benue State University Teaching Hospital, Makurdi (for Tiv) (see Figure 1). Sociodemographic data were also collected from each individual,

including date of birth (or age), current residence, birthplace, self-declared cultural identity, first language, second language (when available) and ethnic affiliation. The ethnic origins of both paternal and maternal grandparents were also obtained. No names or addresses were obtained or recorded.

Figure 1: Principal ethnic groups of Nigeria, including the Ethnic groups of individuals that participated in this study. [18]

Collection of Samples

Blood samples were collected from participants who have consented to participate in the study. Whole blood (3-5 ml) was drawn from participating individuals. Blood was collected from each sample into an anticoagulant storage bottle labelled with a unique anonymous identifier, temporarily stored in mobile ice coolers, transported to the Department and stored in a -80 °C freezer before analysis.

Extraction and quantification of DNA

Genomic DNA was extracted from whole blood using Omega Biotek E.Z.N.A. Forensic DNA Mini Kit. Briefly, whole blood was lysed using a cocktail of lysis buffer. The supernatant was centrifuged, and proteins were removed using a protease (proteinase-K). The DNA was washed in spin columns and eluted using TE buffer. The recovered DNA was stored for use in quantification and PCR analysis. The quality of the DNA was assessed on a 1% agarose gel run at 70 mV for 45 minutes. The DNA was quantified using a NanoDrop spectrophotometer.

Locus Selection

The autosomal STR markers used in the study were selected from the lists of STRs with a high number of genotypes observed, high heterozygosity, high polymorphism information content value and good probability of identity values in African American, African and other black populations. [14,21,27] The ten (10) loci used and their primer sequences are presented in table 1. Based on these sequences, PCR primer oligonucleotides were purchased from Inqaba Biotec, South Africa. The nine

(9) loci were tested as four miniplexes of two 2-primer pairs and two 3-primer pairs based on product size. Primer-primer comparison to avoid excessive regions of complementarity between primers was performed using the software AutoDimer.[19] A visual schematic of the multiplex loci was prepared to provide a visual representation of the loci and highlight areas of possible size overlap. The miniplexes were then named M01, M02, M03 and M04 (Figure 2).

Figure 2: Schematic representation of loci in each multiplex based on PCR product size (bp). The loci are combined into groups, and each group is identified here by boxes of the same colour.

Table 1: Details of PCR loci used in the study

S/NO	LOCUS NAME	GENBANK ACCESSION	PRIMER SEQUENCE (5′ - 3′)	REFERENCES
1	D2S1338	AC010136	ACTGCAGTCCAATCTGGGT ATGAAATCAACAGAGGCTT GC	[20]
2	FGA alpha fibrinogen 3rd intron	M64982	GGCTGCAGGGCATAACATT A ATTCTATGACTTTGCGCTTC AGGA	[20]
3	D6S1017	AL035588	CCACCCGTCCATTTAGGC GTGAAAAAGTAGATATAAT GGTTGGTG	[21]

4	D8S1179	AF216671	ATTGCAACTTATATGTATTT TTGTATTTCATG ACCAAATTGTGTTCATGAGT ATAGTTTC	[20]
5	D9S2157	AL162417	CAAAGCGAGACTCTGTCTC AA GAAAATGCTATCCTCTTTGG TATAAAT	[21]
6	D10S1248	AL391869	TTAATGAATTGAACAAATG AGTGAG GCAACTCTGGTTGTATTGTC TTCAT	[21]
7	D18S51	AP001534	TTCTTGAGCCCAGAAGGTTA ATTCTACCAGCAACAACAC AAATAAAC	[20]
8	D21S11	AP001752	ATATGTGAGTCAATTCCCCA AG TGTATTAGTCAATGTTCTCC AGAGAC	[20]
9	D22S1045	AL033314	ATTTTCCCCGATGATAGTAG TCT GCGAATGTATGATTGGCAA TATTTTT	[21]

The PCR parameters for all the miniplexes is presented in table 2.

10	Amelogenin	M55418 M55419	ACCTCATCCTGGGCACCCTG G AGGCTTGAGGCCAACCATC AG	[22]

Table 2: PCR conditions for the four multiplexes

Loci		Initial Denaturation		Denaturation		Annealing		Extension		Final Extension		Cycles
Miniplex	Loci	Temp	Time	Temp	Time	Temp	Time	Temp	Time	Temp	Time	
M01	Amelogenin; D10S1248	95	15	95	30 s	62	1	72	1.40 min	72	10	30
M02	D6S1017; D18S51	95	15	95	30 s	59	1	72	1.40 min	72	10	35
M03	D22S1045; D21S11; D2S1338	95	15	95	30 s	57.4	1	72	1.40 min	72	10	35
M04	D9S2157; FGA; D8S1179	95	15	95	30 s	59.5	1	72	1.40 min	72	10	35

Polyacrylamide Gel Electrophoresis (PAGE):

The PCR products were mixed with a loading dye solution containing 10 mM NaOH, 95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol. The mixture was then heated (to denature) at 95 °C for 2-3 minutes and then subjected to 4-8% (depending on the size of the band) PAGE (37 mm long, 0.4 mm thick) containing 7 M urea and 0.5X Tris Borate EDTA buffer. Using a sequencing gel apparatus, the samples were then allowed to resolve at a constant temperature between 40-50 W for at least 75 minutes (and a maximum of 5-6 hours). The PAGE was then followed by silver staining using the method of Refaat et al. [23]. The gels were then viewed with a UV transilluminator, and images were captured with a camera for gel documentation. The gels were scored using GelAnalyzer (v. 23.1) [24].

Data Collection and Computation of Statistics

Allelic frequencies, population pairwise genetic distances (FST and RST, analysis of molecular variance (AMOVA) and principal component analysis PCA) were calculated using GenAlEx v6.502. The unweighted pair group method with arithmetic mean tree was built from the distance matrix (FST) using the option 'neighbour and draw tree' in the Phylip software package and visualized with Tree View software.

Results:

Fixation index (FST) and measures of population structure

The fixation index and other population genetic parameters were calculated for all the loci in all the populations and for the Nigerian summarized population (Table 3). The allele number (An), average number of different alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho),

expected heterozygosity (He), and fixation indices (F, FIS and FST) are all presented.

The observed number of alleles among the sampled populations ranged between 10.33 and 10.67. The observed heterozygosity (Ho) was greater than expected heterozygosity (He) for all the populations examined (Table 4). However, the highest observed heterozygosity was found in the Igbo (Ho = 0.88±0.01) population, while the lowest was observed in the Yoruba (0.85±0.02) population (Table 4).

Table 3: Summary of measures of population structure in the Nigerian population

Locus	**H_o**	**H_e**	***F_{IT}***	***F_{IS}***	***F_{ST}***
D21S11	0.86±0.00	0.83±0.00	-0.04±0.01	-0.04±0.01	0.001±0.01
D22S1045	0.83±0.01	0.84±0.00	0.01±0.01	0.01±0.01	0.002±0.00
D2S1338	0.88±0.01	0.87±0.01	-0.01±0.00	-0.01±0.01	0.003±0.01
D8S1179	0.83±0.02	0.79±0.01	-0.05±0.03	-0.05±0.00	0.003±0.01
FGA	0.88±0.00	0.87±0.00	-0.01±0.01	-0.01±0.00	0.001±0.02
D9S2157	0.90±0.00	0.85±0.00	-0.06±0.00	-0.06±0.00	0.001±0.00
D18S51	0.87±0.01	0.87±0.00	0.00±0.01	0.00±0.01	0.003±0.01
D6S1017	0.80±0.01	0.81±0.00	0.01±0.01	0.01±0.00	0.000±0.00
D10S1248	0.90±0.01	0.81±0.00	-0.11±0.01	-0.11±0.00	0.001±0.00
Mean ±SE	0.86±0.01	0.84±0.01	-0.03±0.01	0.03±0.01	0.002±0.00

SE: standard error, Ho: observed heterozygosity, He: expected heterozygosity, FST: Fixation index between gametes randomly drawn within subpopulations relative to gametes of the total population, FIS: Fixation index of individual relative to gametes of the subpopulation (inbreeding coefficient), FIT: Fixation index of individual relative to gametes of the total population

Table 4: Comparative measures of population structure in the 5 populations

Pop	N_a	N_e	H_o	H_e	F
Ibibio	10.56±0.69	6.35±0.42	0.87±0.01	0.84±0.01	-0.04±0.02
Igbo	10.67±0.67	6.52±0.42	0.88±0.01	0.84±0.01	-0.04±0.02
Yoruba	10.67±0.87	6.64±0.49	0.85±0.02	0.84±0.01	-0.01±0.02
Hausa	10.33±0.76	6.19±0.38	0.86±0.01	0.83±0.01	-0.03±0.02
Tiv	10.67±0.60	6.24±0.38	0.86±0.01	0.83±0.01	-0.04±0.01

Na: Observed number of alleles, Ne: Number of effective alleles, Ho: Observed heterozygosity, He: Expected heterozygosity F: Fixation

Pairwise Population Matrix

Pairwise population matrix of genetic similarity between populations estimates the genetic similarity between pairs of populations based on the alleles of the genetic markers. The results of the pairwise population matrix of genetic similarity between populations is usually between 0 and 1 and represents the probability of similarity between two populations, with higher values indicating greater similarity.

The results of the pairwise population matrix of genetic similarity between populations in this study is presented in Table 5. The pairwise population matrix value between Yoruba and Igbo was 0.993, which was the highest of the pairwise genetic similarity indices between the five populations. Yoruba-Ibibio was 0.990, Yoruba-Hausa was 0.992, and Yoruba-Tiv was 0.987. The Igbo-Ibibio similarity at these loci was the lowest at 0.895.

Table 5: Pairwise population matrix of genetic similarity between populations

Population	Ibibio	Igbo	Yoruba	Hausa	Tiv
Ibibio	1.000				
Igbo	0.985	1.000			

Yoruba	0.990	0.993	1.000		
Hausa	0.988	0.991	0.992	1.000	
Tiv	0.991	0.986	0.987	0.991	1.000

When a dendrogram was constructed for the five populations based on data from these 9 loci, Ibibio and Tiv clustered together. Igbo and Yoruba formed a more recent cluster and were joined by Hausa (Figure 3).

Figure 3: Phylogenetic tree of the relationship between the 5 populations based on the Nei Pairwise population matrix of genetic similarity between populations in Table 5

Analysis of Molecular Variance (AMOVA)

The results of the analysis of molecular variance at the 9 loci in the 5 populations is presented in Table 6. The estimated variation partitioned within individuals was 99.98%, with the remaining 0.10% attributed to variation among populations.

Table 6: Summary of AMOVA

Source	Df	SS	MS	Est. Var.	%
Among Pops	4	3.5460	0.8865	0.0004	0.10%
Among Indiv	245	927.3900	3.7853	0.0000	0%
Within Indiv	250	962.5000	3.8500	3.8500	99.98%
Total	499	1893.4360		3.8504	100%

df: degree of freedom, SS: sum of squares, MS: mean sum of squares, Est. Var.: Estimated variation

Principal component analysis

Principal component analysis (PCA) revealed four clusters (Figure 4). The clusters are labelled A-D. The coloured small shapes represent individuals, with each colour representing an ethnic group. The numbers, in addition to the shapes, represent the sample identities of the

individuals. The clusters are heterogeneous, with each cluster containing samples from all five (5) ethnic groups. The total variation explained by the PCA was 25.52%.

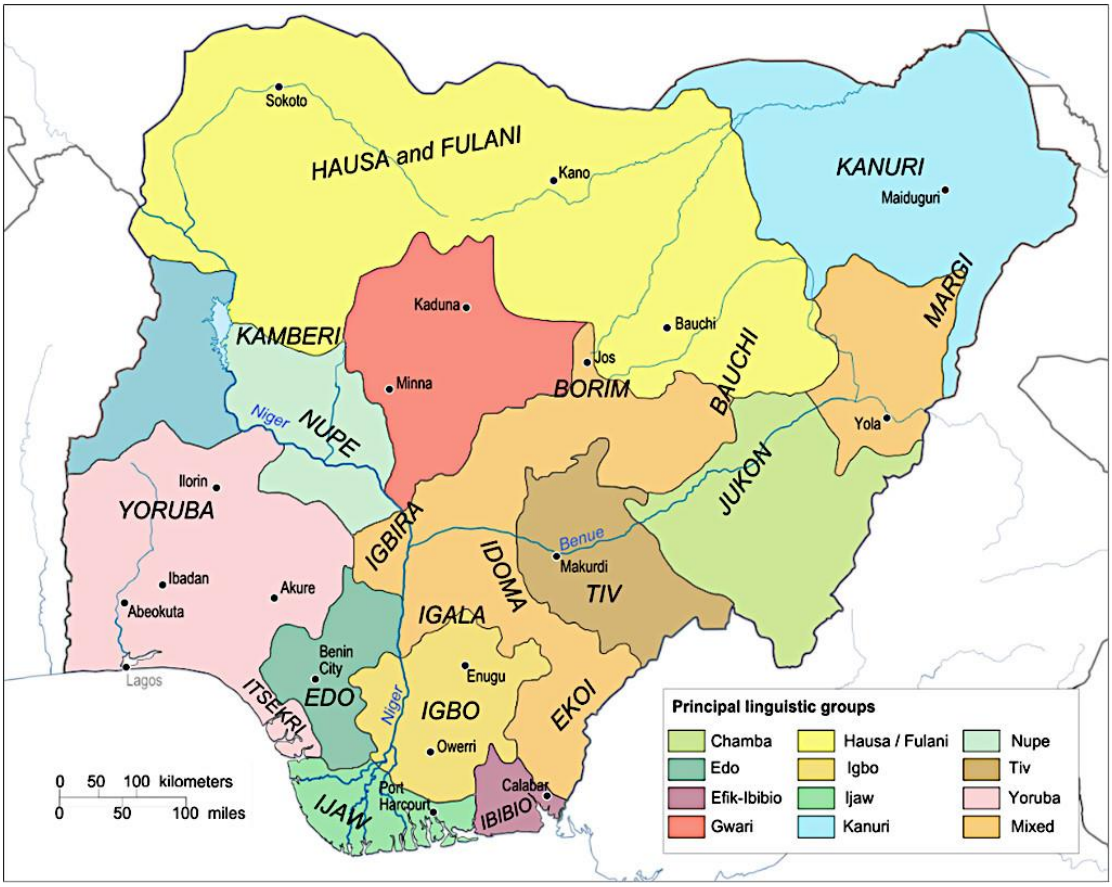

Figure 1: Principal ethnic groups of Nigeria, including the Ethnic groups of individuals that participated in this study

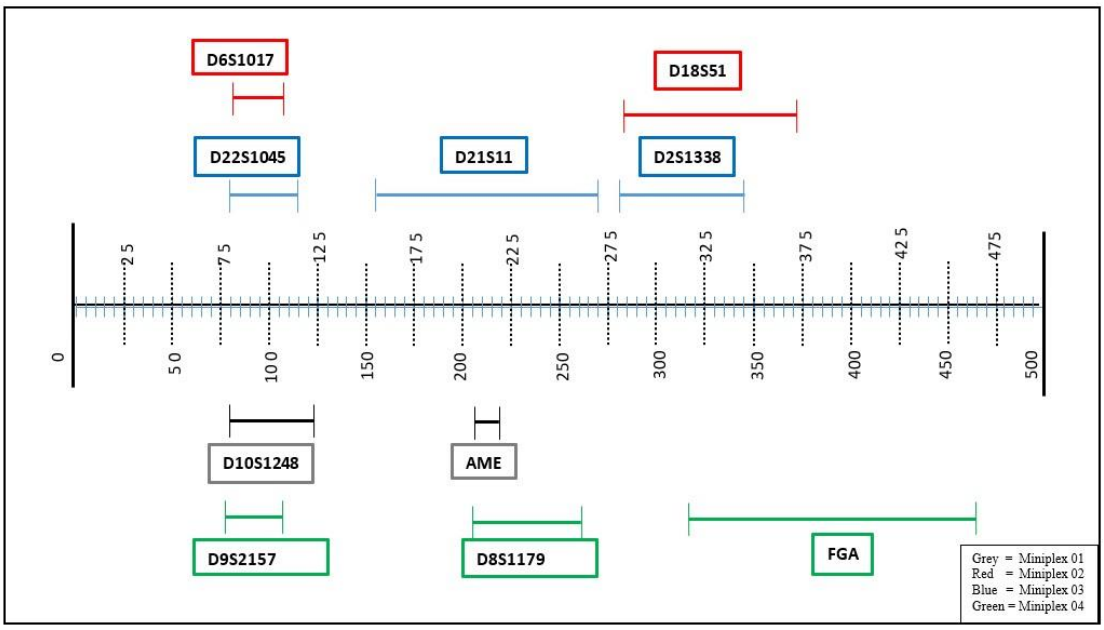

Figure 2: Schematic representation of loci in each multiplex based on PCR product size (bp). The loci are combined into groups, and each group is identified here by boxes of the same color.

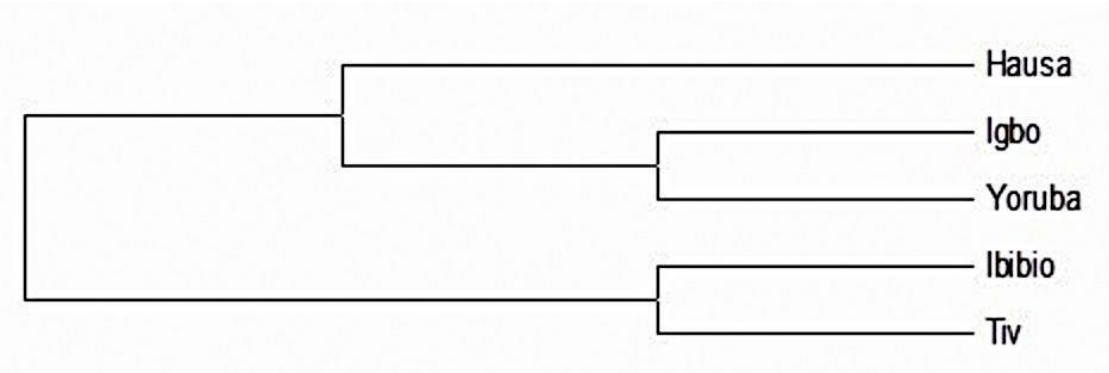

Figure 3: Phylogenetic tree of the relationship between the 5 populations based on the Nei Pairwise population matrix of genetic similarity between populations in Table 5

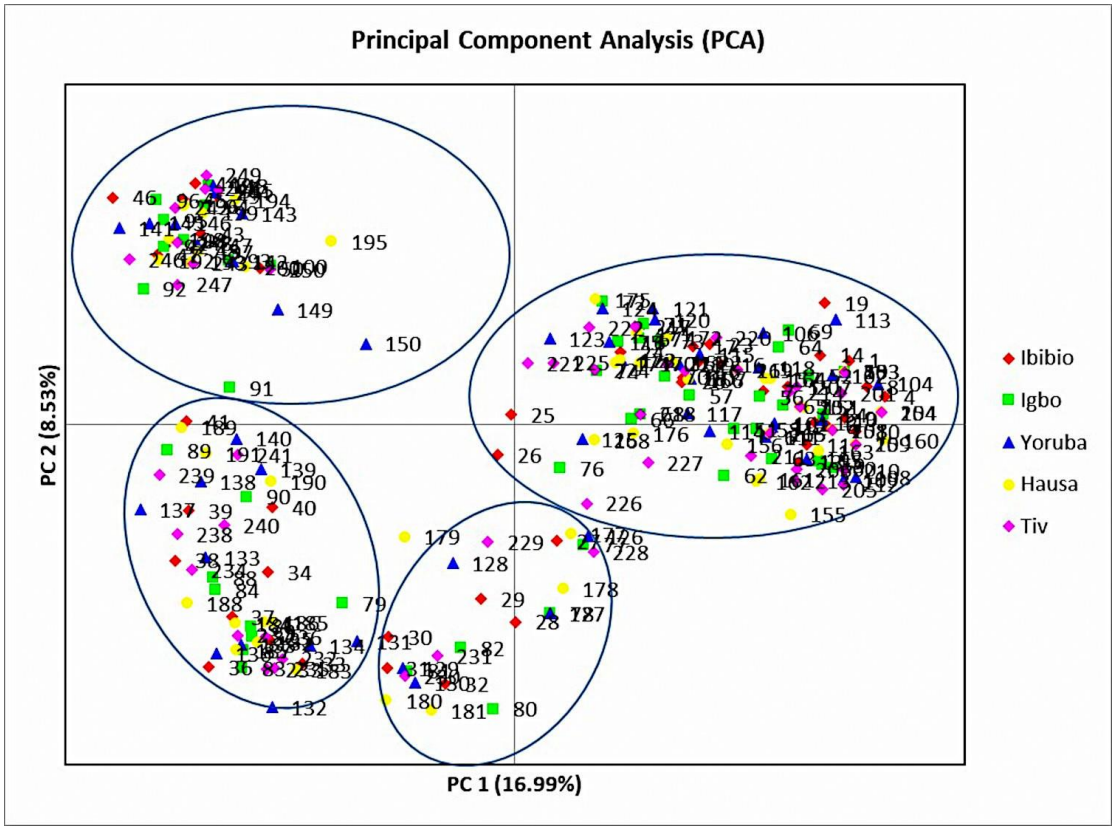

Figure 4: Principal component analysis for the variation in the five ethnic groups

Discussion:

The ethnic populations of Nigeria have occupied their present-day homes for several centuries, and as such, expected relationships between individuals from these ethnic groups are bound to exist. The dynamics of population substructure and the underlying genetics are the subject of several recent studies. In the last century, there has apparently been gene flow between ethnic populations, as evidenced by interethnic mating between individuals from these groups, although the extent and nature of genetic activity are unknown.

This study established that there is an excess of heterozygotes in all the examined loci for all the populations, as indicated by high values of observed heterozygotes compared to the expected heterozygotes. High observed heterozygosity is a frequent observation when human population samples are obtained at the ethnic home of

each population where genetic diversity is usually very high. Natural populations also usually violate the Hardy–Weinberg equilibrium (HWE) to some degree, which is the cause of allele frequency changes over time. Some of the loci in this study were not in HWE because the observed genotypes deviated from the expected genotypes. The main suspect for the deviation of some of the loci in any population is the population substructure. This excess heterozygosis points to the absence of consanguineous mating in these populations and that mate selection is mostly random. In fact, the inbreeding coefficient (FIS) observed in this study in the five Nigerian subpopulations was very low for all loci. The US National Research Council report recommends that a conservative FIS estimate of 0.01 is expected in populations examined for forensic data and suggests that a value of 0.03 indicates that cousin mating occurs in such populations [25].

Heterozygosity also shows that there is stabilizing or balanced selection where natural selection favours the heterozygote over homozygotes, leading to lower-than-average FST values for the selected loci. Five of the loci (D21S11, FGA, D9S2157, D6S1017, and D10S1248) had FSTs, indicating that there was stratification in the subpopulations. The extent of the variation in FST from one locus to another helps to establish the source of the variation being observed. In populations where natural selection is selectively neutral, the only expected force at play is drift, and the observed FST will be almost equal for all loci since drift depends only on demographic properties of the populations and not on any of the specific loci being studied [26]. This is because the populations are assumed to have evolved independently from each other, and each will produce constraints that affect the FST distribution. FST will still vary from locus to locus in each of the subpopulations, but the extent of variation will be similar in each population.

These stratifications seem to disappear when the Nigerian population is examined as a whole. The analysis of molecular variance in this study revealed an estimated 99.98% variation within individuals. The variation among populations was 0.10%. This shows that individuals could easily be identified using these markers, but it will be much more difficult to assign the individuals into populations because the variation among populations was low. The PCA analysis therefore presented four heterogenous clusters each containing samples from all five (5) ethnic groups and explaining 25.52% of the variation.

Although Igbo shares a common geographic boundary with Ibibio, the pairwise population matrix of genetic similarity for these loci shows that Igbo has greater similarity to Yoruba (0.993) than to every other population than to Ibibio. In fact, the Igbo-Ibibio similarity at these loci was the lowest (0.895) for all of the tested population pairs. These populations are actually the only populations that share a common geographic boundary, implying that this heightened dissimilarity for this population may be due to the alleles of these loci in individuals in this region diverging from each other. In fact, it has been previously established in other studies that the region of ethnic origin of any organism usually has the highest genetic diversity. The results of the pairwise population matrix of genetic similarity between populations also indicated that Ibibio and Tiv were more similar at these loci. The Hausa population had a similarity index that was similar for all populations: Yoruba (0.992), Igbo and Tiv each (0.991) and Ibibio (0.988).

These populations all share a common lingual ancestor, as they have all been classified into the Niger-Congo group of languages [5]. However, Igbo and Yoruba share linguistic similarity and are further classified together as YEAIs (non-Bantoid), whereas Ibibio and Tiv together belong to the Benue-Congo (Bantoid) subgroup of languages. The Hausa language is often

classified as a member of the Afro-Asiatic group of languages. A similar relationship was observed in this study when a dendrogram was constructed for the five populations based on data from these 9 loci. The dendrogram was able to further organize the variation observed in the AMOVA and PCA based on maximum linkage distance and average linkages similarity measures to show hierarchical relationships between the populations. Ibibio and Tiv clustered together while the Igbo and Yoruba formed a more recent cluster and were joined by the Hausa.

The Ibibio and Igbo diverge more from each other. The Tiv and Ibibio are more similar to each other than to other groups. These relationships mirrored linguistic groupings showing that language may have shaped early interaction and ultimately gene transfer between these groups.

References:

1. Yu N, Chen F-C, Ota S, Jorde LB, Pamilo P, Patthy L, et al. Larger genetic differences within Africans than between Africans and Eurasians. Genetics. 2002;161(1):269–74. Available from: http://dx.doi.org/10.1093/genetics/161.1.269

2. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet. 2002;3(8):611–21. Available from: http://dx.doi.org/10.1038/nrg865

3. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9(1):403–33. Available from: http://dx.doi.org/10.1146/annurev.genom.9.081307.164258

4. Choudhury A, Aron S, Sengupta D, Hazelhurst S, Ramsay M. African genetic diversity provides novel insights into evolutionary history and local adaptations. Hum Mol Genet.

2018;27(R2):R209–18. Available from: http://dx.doi.org/10.1093/hmg/ddy161

5. Eberhard DM, Gary F. Ethnologue: Languages of the World. Twenty-sixth edition. Fennig CD, editor. Dallas, Texas;

6. Veeramah KR, Connell BA, Ansari Pour N, Powell A, Plaster CA, Zeitlyn D, et al. Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. BMC Evol Biol. 2010;10:92. Available from: http://dx.doi.org/10.1186/1471-2148-10-92

7. Adeyemo AA, Chen G, Chen Y, Rotimi C. Genetic structure in four West African population groups. BMC Genet. 2005;6:38. Available from: http://dx.doi.org/10.1186/1471-2156-6-38

8. Titilayo OA, Samuel AO, David OO, Adewunmi TI. Genetic variations among three major ethnic groups in Nigeria using RAPD. Mol Biol Res Commun. 2018;7(2):51–8. Available from: http://dx.doi.org/10.22099/mbrc.2018.26098.1280

9. Joshi E, Biddanda A, Popoola J, Yakubu A, Osakwe O, Attipoe D, et al. Whole-genome sequencing across 449 samples spanning 47 ethnolinguistic groups provides insights into genetic diversity in Nigeria. Cell Genom. 2023;3(9):100378. Available from: http://dx.doi.org/10.1016/j.xgen.2023.100378

10. Tucci S, Akey JM. The long walk to African genomics. Genome Biol. 2019;20(1):130. Available from: http://dx.doi.org/10.1186/s13059-019-1740-1

11. Chatumal W, Markalanda D, Iiieperuwa RJ, Ranawaka G, Fernandopulle ND. Population study of combined DNA index system (CODIS) core loci D3S1358, D5S818, D8S1179 Short Tandem Repeats (STRs)

polymorphism in Sri Lanka. J Natl Sci Found Sri Lanka. 2010;38:65–8.

12. Zhu Y, Strassmann JE, Queller DC. Insertion, Substitution & the origin of microsatellite. Genetics Research. 2000;76:22–236.

13. Hardy OJ, Charbonnel N, Fréville H, Heuertz M. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics. 2003;163(4):1467–82. Available from: http://dx.doi.org/10.1093/genetics/163.4.1467

14. Butler JM, Hill CR. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Sci Rev. 2012;24(1):15–26.

15. Hannelius U, Salmela E, Lappalainen T, Guillot G, Lindgren CM, von Döbeln U, et al. Population substructure in Finland and Sweden revealed by the use of spatial coordinates and a small number of unlinked autosomal SNPs. BMC Genet. 2008;9(1):54. Available from: http://dx.doi.org/10.1186/1471-2156-9-54

16. Bashalkhanov S, Pandey M, Rajora OP. A simple method for estimating genetic diversity in large populations from finite sample sizes. BMC Genet. 2009;10(1):84. Available from: http://dx.doi.org/10.1186/1471-2156-10-84

17. Hale ML, Burg TM, Steeves TE. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS One. 2012;7(9):e45170. Available from: http://dx.doi.org/10.1371/journal.pone.0045170

18. Utexas.edu. Map of Nigeria. Available from: https://legacy.lib.utexas.edu/maps/africa/nigeria_linguistic_1979.jpg

19. Vallone PM, Butler JM. AutoDimer: a screening tool for primer-

dimer and hairpin structures. Biotechniques. 2004;37(2):226–31. Available from: http://dx.doi.org/10.2144/04372st03

20. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci. 2002;47(4):773–85. Available from: http://dx.doi.org/10.1520/jfs15445j

21. Hill CR, Kline MC, Coble MD, Butler JM. Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples: 26 Ministr loci. J Forensic Sci. 2008;53(1):73–80. Available from: http://dx.doi.org/10.1111/j.1556-4029.2008.00595.x

22. Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques. 1993;15(4):636–8, 640–1.

23. Refaat AM, El-Garf WT, Ramadan KS, Abdelfattah MM, El Awady MK. CTT Multiplex System is a quick & inexpensive method to exclude innocent suspects during criminal inquiries. Nature & Science. 2008;6(2):6–18.

24. Lazar I, Lazar I GelAnalyzer 23.1.1. Available from: http://www.gelanalyzer.com

25. Balding DJ. Evaluation of mixed-source, low-template DNA profiles in forensic science. Proc Natl Acad Sci U S A. 2013;110(30):12241–6. Available from: http://dx.doi.org/10.1073/pnas.1219739110

26. Khan MMH, Rafii MY, Ramlee SI, Jusoh M, Al Mamun M, Halidu J. DNA fingerprinting, fixation-index (Fst), & admixture mapping of selected Bambara groundnut (Vigna subterranea [L.] Verdc.) accessions using ISSR markers system. Sci Reports. 2021;(1). Available from:

http://dx.doi.org/10.1038/s41598-021-93867-5

27. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. (2003). Allele frequencies for 15 autosomal STR loci on U.S Caucasian, African American, and Hispanic populations. J Forensic Sci. 2003; 48(4):908-911.

List of Abbreviations:

Abbreviation	Definition
An:	Allele number
bp:	Base pair
AMOVA:	Analysis of Molecular Variance
DNA:	Deoxyribonucleic acid
F:	Fixation Index
FIS:	Fixation index of Individual relative to gametes of the Subpopulation
FIT:	Fixation index of Individual relative to gametes of the Total population
FST:	Fixation index between gametes randomly drawn within Subpopulations relative to gametes of the Total population
He:	Expected heterozygosity
Ho:	Observed heterozygosity
HWE:	Hardy-Weinberg Equilibrium
Na:	Number of alleles
Ne:	Number of effective alleles
PAGE:	Polyacrylamide gel electrophoresis
PCA/PCoA:	Principal component Analysis
PCR:	Polymerase Chain Reaction
STR:	Short Tandem Repeats

Tm: Melting temperature