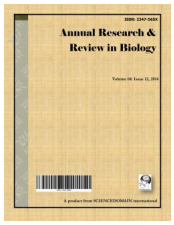

Annual Research & Review in Biology
12(4): 1-8, 2017; Article no.ARRB.32783
ISSN: 2347-565X, NLM ID: 101632869

SCIENCEDOMAIN *international*
www.sciencedomain.org

Ethnic Origin of Crime Scene Evidential Materials Determination in Three Main Ethno-linguistic Population Groups in Nigeria

B. U. Agbo[1], O. A. T. Ebuehi[1*] and A. A. Osuntoki[1]

[1]Department of Biochemistry, College of Medicine, University of Lagos, Lagos, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. Author BUA designed the study, performed the statistical analysis and interpretation, wrote the protocol and wrote the first draft of the manuscript. Authors OATE and AAO were involved in the planning, constant monitoring of the research and literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/ARRB/2017/32783
Editor(s):
(1) Iskra Ventseslavova Sainova, Institute of Experimental Morphology, Pathology and Anthropology with Museum to Bulgarian Academy of Sciences (IEMPAM - BAS) in Sofia, Bulgaria.
(2) J. David Puett, Department of Biochemistry and Molecular Biology University of Georgia, Athens, USA.
(3) George Perry, Dean and Professor of Biology, University of Texas at San Antonio, USA.
Reviewers:
(1) Anatole A. Klyosov, The Academy of DNA Genealogy, Moscow, Russia.
(2) Alex S. Prayson, National Council on Rehabilitation Education, USA.
Complete Peer review History: http://www.sciencedomain.org/review-history/19004

Original Research Article

Received 15th March 2017
Accepted 22nd April 2017
Published 10th May 2017

ABSTRACT

DNA analysis using autosomal short tandem repeat (microsatellite) polymorphism is a useful tool for forensic purposes, such as individual identification, stain analysis and paternity testing. Analyses of such materials are carried out by the comparison of profiles from questioned samples or crime scene with those from suspects or victims or from database. In some instances, the profiles generated will neither match that from suspects nor the database. The objective of the current study is to identify population specific markers that will show distinct genetic variability among the three main ethno-linguistic population groups in Nigeria. The profiles generated can be used to infer ethnic origin of test samples from the populations in an ethnically blinded test.

Allele frequencies for each ethnic group from 315 unrelated individuals representing the three populations; Ibo, Hausa and Yoruba were generated using 15 Microsatellite loci (STRs) from Applied Biosystems. Multi-locus genotype frequencies were utilized for testing conformity with Hardy-Weinberg equilibrium. Chi-square goodness of fit showed seven loci to be in Hardy-Weinberg

**Corresponding author: E-mail: ebuehi@yahoo.com, oebuehi@unilag.edu.ng, oatebuehi@cmul.edu.ng;*

equilibrium. However after Bonferoni correction all loci were found to be in conformity with Hardy-Weinberg equilibrium.
The allele frequencies generated for each population were tested in the determination of ethnic origin of twenty test samples randomly collected in an ethnically blinded test. The ethnic origins of the twenty samples were correctly determined with 99.5% success, by using the principle that the ratio of profile frequencies for the same profile in different ethnic groups is a likelihood ratio.

Keywords: Short tandem repeat; genetic variation; ethno-linguistic group; forensic parameters; heterozygosity; homozygosity.

1. INTRODUCTION

Nigeria is located at the Gulf of Guinea in the Western part of Africa with an estimated population of 183 million according to the July 2016 estimate of the American Central Intelligence Agency (CIA) world fact book [1]. It comprises of over 527 ethno-linguistic groups with 520 of them in existence and 7 extinct [2]. Each ethnic group is associated with a distinct language which belongs to the three main African Linguistic families; Niger-Kordofanian, Afro-Asiatic and Nilo-Saharan [3].

The three main ethnic groups in Nigeria, which are Hausa, Yoruba and Ibo constitute over 68% of its total population. The Hausa ethnic group is found in the North and it belongs to the Afro-Asiatic linguistic family in the subphylum 'Chadic' together with the Angas, Bachamas, Tangale, Bokkos and so on. They are reported to have migrated from North Africa in Upper Egypt. The Yoruba and Ibo ethnic groups are found in the West and East respectively. They belong to the Niger-Kordofanian linguistic family in the subphylum 'Kwa' together with Edo, Itsekiri, Tiv, Igala, Idoma and so on. Studies have it that they migrated from East Africa [4]. Linguistic diversity in Nigeria is attributed to this complex history of migration and settlements over a long period of time, which altered the linguistic landscape of the country resulting in rich cultural heritage (cultural, linguistic, ethnic, etc.) histories [5]. Populations separated from one another for a long time, through genetic drift and mutation will become statistically differentiated by their differing allelic frequencies [6]. These alleles may exist with large allele frequency differential referred to as population specific alleles (PSAs) [6,7].

Many studies have shown that most of the human variations, about 80% to 90% are observed within populations, and about only 10% to 20% are due to population differences [8,9].

Genetic markers found in one population and not in the other were for the first time used by Neel to estimate mutation rates, who referred to them as private alleles. Chakraborty called them "unique Alleles", and concluded that those with the largest allele frequency differences among populations are most useful for forensic and admixture mapping studies. Shriver referred to them as population specific alleles (PSA) and stated that they are markers, which exhibit large frequency differential of more than 50% [10].

The population specific alleles of forensic relevance found in populations can be used to establish ethnic origin of the crime scene evidential materials. Klinstar supported this conclusion and he further observed that the frequency differences between populations could be used to situate an individual with broad, geographically based groups by genetic clustering method. According to the same authors a forensic sample is associated with a population in which its profiles are most common [11]. In his work, Brenner established a possibility for determination of the ethnic origin of a sample by using the principle that the ratio of the profile frequencies for the same profile in different ethnic groups is a likelihood ratio [12].

2. MATERIALS AND METHODS

2.1 Populations

Twenty different communities in six states, within the three ethnic groups in Nigeria, were selected based on historical antecedents and ethnic affiliations. Six communities were selected from the Hausa ethnic group, while seven communities were selected from each of the Ibo and Yoruba ethnic groups. Genealogies of each participant were traced to the fourth generation of target population by self-declaration.

2.2 Short Tandem Repeat (STR) Markers

Fifteen microsatellite Loci (STR) were utilized in the analysis. They include the 13 core CODIS loci widely used in forensic DNA analysis and two other additional forensically relevant loci. All

the selected loci are tetra-nucleotide repeat motif and located on 13 different chromosomes. The loci CSFIPO and D5S818 are located on chromosome five and D2S328 and TPOX are located on chromosome two but they are unlinked, which makes them ideal for the study.

2.3 DNA Extraction

Mouth swabs were collected from three hundred and fifteen (315) consenting individuals. One hundred and five (105) from each ethnic group using DNAase free cotton swabs from Sirche Forensic Inc. DNA was extracted from each sample using Applied Biosystems PrepFiler® forensic DNA extraction kit according to the manufacturer's protocol.

2.4 Amplification and Genotyping

The extracted DNA was then amplified using Applied Biosystems GeneAmp 9700 Thermal Cycler. This was done in a single multiplex reaction for each individual using AmpF*L* STR®Identifiler® Direct Kit PCR reagents according to the manufacturer's protocol as validated by the laboratory in 26 cycles (Applied BiosystemAmpF*L* STR®Identifiler® Direct Kit PCR Manual, 2012) [13]. The amplified DNA fragments (amplicons) were separated by capillary electrophoresis using 310 Genetic Analyzer according to the validated method of the laboratory.

2.5 Statistical Analysis

Allele frequency for each locus was calculated from the number of genotype in the sample set by the method of gene count since the STR loci are autosomal. Conformity with Hardy Weinberg genotype frequencies were carried out by two tests; Chi-square and Exact test for multi-allelic loci [11]. Bonferoni correction was used to ascertain significant departures from Hardy-Weinberg equilibrium [14].

Ethnic origin inferences were carried out by pair-wise likelihood ratio analysis of the genotype frequencies of the twenty samples collected in Lagos using the different ethnic group allele frequencies generated by the study [12].

2.6 Allele Frequency Database Generation and Power of Discrimination (PD) Determination

From the DNA profiles obtained, genotype data were generated. This was used to calculate the allele frequency for the various ethno-linguistic groups across the 15 loci. A combined genotype and allele frequency for the three ethno-linguistic groups were also generated. The DNA profiles generated was used to calculated power of discrimination (PD) across the 15 loci using Power Stat from Promega Inc [15].

2.7 Determination of Ethnic Origin

Mouth swab were collected from randomly selected 20 unrelated consenting subjects without lineage screening for the three ethnic groups in Lagos in the South Western part of Nigeria. DNA profiles were extracted from the samples, amplified, separated and genotyped across the 15 loci. The allele frequencies obtained from the study were used to calculate observed genotype frequencies for the 20 samples for each individual across the 15 loci. Likelihood ratio of the genotype frequencies for each individual across the 15 loci was also calculated. Then, the percentage of the number of loci for an individual with the highest value for likelihood ratio was calculated for the three ethno-linguistic groups. An ethnic group is then assigned to an individual with more number of loci with highest value of the likelihood ratio.

3. RESULTS

Seven loci out of the 15 loci used in the current study were found to be in conformity with Hardy-Weinberg equilibrium using Chi-Square goodness of fit. (THO1, vWA, D2S1338, D3S1358, D16S539, D18S51, and D21S11). However, after Bonferoni corrections all the loci were found to be in conformity with Hardy-Weinberg equilibrium (Table 1). Exact test for multi-allelic loci showed all loci to be in conformity with Hardy-Weinberg proportion of genotype frequencies.

3.1 Allele Frequencies

The variation in allele frequency distribution across the 15 loci for the combined population was substantial with the population variance of 7.13257 ±0.2166 (Fig. 1). The Hausa ethnic group has the highest variation value of allele frequency distribution with population variance of 7.3742 ±1.045. The Ibo ethnic group has the least variation frequency with population variance of 6.8095 ±0.96, while the Yoruba ethnic group has population variance of 7.00581 ±0.99 (Fig. 2). These differences in variance between the ethnic groups are very small in value and not statistically significant.

3.2 Power of Discrimination (PD)

Power of Discrimination (PD), an important parameter for individual identification ranged from 0.829 in THO1 locus in the Ibo ethnic group to 0.975 in FGA locus in the Hausa ethnic group with population variance of 0.00126 ±0.0080 (Table 2).

Analysis of power of discrimination in each population across the 15 loci showed that the value in Hausa ethnic group ranged from 0.886 at D5S818 locus to 0.975 at FGA locus with population variance of 0.00097 ±0.00008. The Ibo ethnic group ranged from 0.829 at THO1 locus to 0.968 at both FGA and D2S1338 loci with population variance of 0.0014 ±0.00009. Power of discrimination (PD) on the other hand ranged from 0.846 at THO1 locus to 0.971 at FGA locus in the Yoruba ethnic group with population variance of 0.00141 ±0.00009.

Pair wise comparison of the population variance among the three ethno-linguistic groups showed that the Ibo and Yoruba ethnic groups have the lowest difference of 0.00001 (Fig. 3).

3.3 Ethnically Blinded Test

The sum of the Likelihood ratio of the genotype frequencies across the 15 loci in the ethnically blinded test gave correct ethnic origin for 19 of the 20 randomly collected samples from unrelated subjects from the three populations. The remaining one sample gave the same value for both Hausa and Ibo groups. However, ethnic lineage crosscheck on the questionnaire from the individual showed that the individual was an admixture with an Ibo Father and a Hausa mother (Fig. 4).

The percentage success of a locus in the determination of ethnic origin of a sample among the three-population groups showed CSFIPO, D3S1358 and D16S539 have equal success in both Ibo and Yoruba, D2S1338 and D13S317 had equal success in both Hausa and Ibo. D5S818 and D7S820 had the highest success in Ibo ethnic group. D13S317, D19S433, D21S11, FGA and TPOX had the highest success in Yoruba ethnic group. D8S1179, D18S51, THO1 and vWA had the highest success in the Hausa group (Fig. 5).

Table 1. Chi-square result and bonferoni corrections

Loci	X^2 calculated	Critical value 0.05%	df	No. of sample	Bonferoni. correction
D18S51	42.3416	43.2	62	64	0.0008
D16S539	23.0719	26.5	36	37	0.0016
D21S11	53.1247	67.5	51	52	0.0010
D3S1358	9.7158	13.1	23	24	0.0021
D2S1338	50.233	51.7	67	68	0.0007
THO1	9.2285	10.9	20	21	0.0024
vWA	22,7949	26.5	37	38	0.0013
CSFIPO	29.6213	18.5	32	33	0.0015
D19S433	43.6525	43.2	57	58	0.0009
D7S820	29.1995	13.8	24	25	0.002
D5S818	142.2317	18.5	34	35	0.0013
D8S1179	19.8807	17.7	29	30	0.0017
D13S317	18.006	11.6	21	22	0.0023
TPOX	37.612	17.7	29	30	0.0017
FGA	19.8807	17.7	29	30	0.0017

Table 2. Forensic parameter (Power of discrimination)

Locus	Ethnic groups		Locus		Locus		Locus		Locus	
D2S1338		PD	D21S11	PD	D18S51	PD	D13S317	PD	D16S39	PD
	HAUSA	0.973		0.943		0.964		0.887		0.934
	IBO	0.968		0.950		0.957		0.869		0.943
	YORUBA	0.966		0.053		0.962		0.849		0.932
CSFIPO			THO1		D19S433		D7S820		D8S1179	
	HAUSA	0.941		0.873		0.948		0.917		0.907
	IBO	0.917		0.829		0.955		0.922		0.886
	YORUBA	0.918		0.846		0.961		0.920		0.892
D3S1358			TPOX		D5S818		vWA		FGA	
	HAUSA	0.893		0.932		0.886		0.927		0.975
	IBO	0.904		0.918		0.924		0.928		0.968
	YORUBA	0.905		0.915		0.935		0.933		0.971

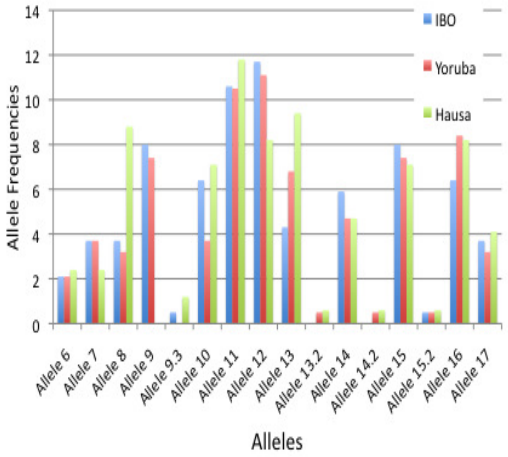

Fig. 1. Percentage allele frequencies among the 3 populations for alleles 6-17

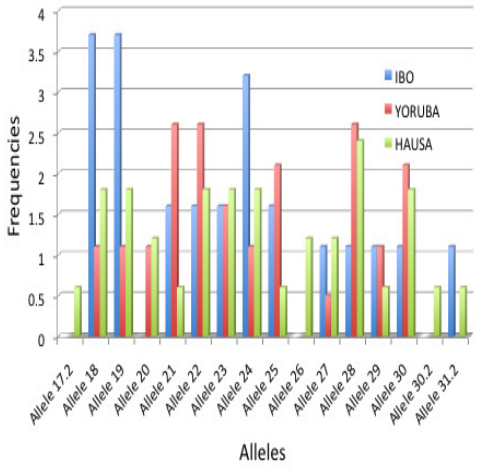

Fig. 1.(contd.) Percentage allele frequencies among the 3 populations from allele 17.2 to allele 32.2

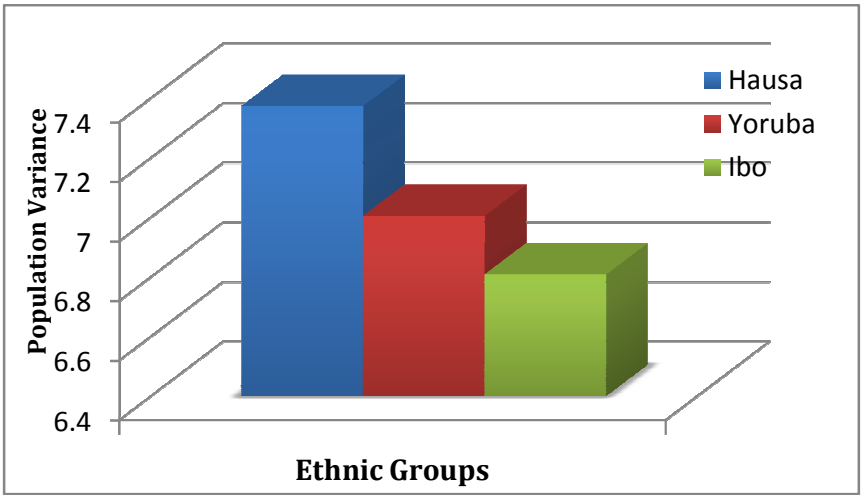

Fig. 2. Population variance for the 3 ethnic groups

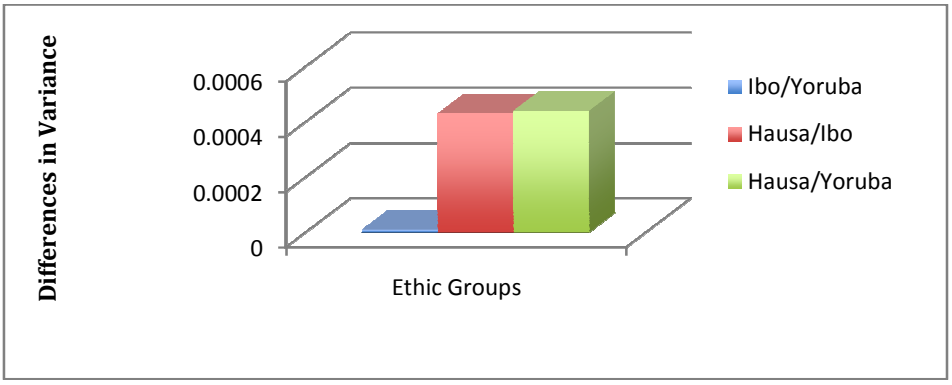

Fig. 3. Differences in pair wise population PD variance

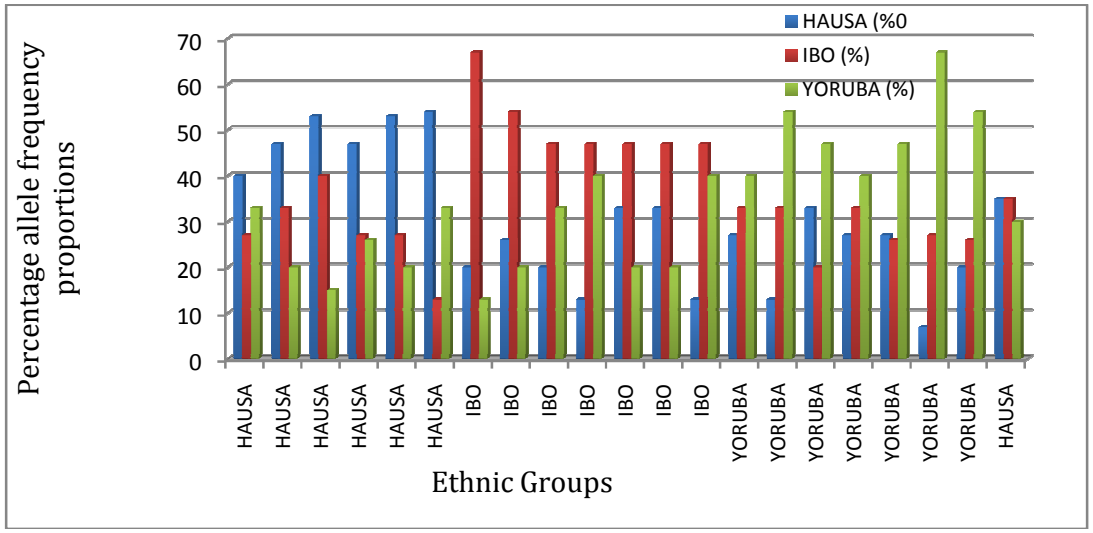

Fig. 4. Percentage allele frequency proportions of the twenty test samples

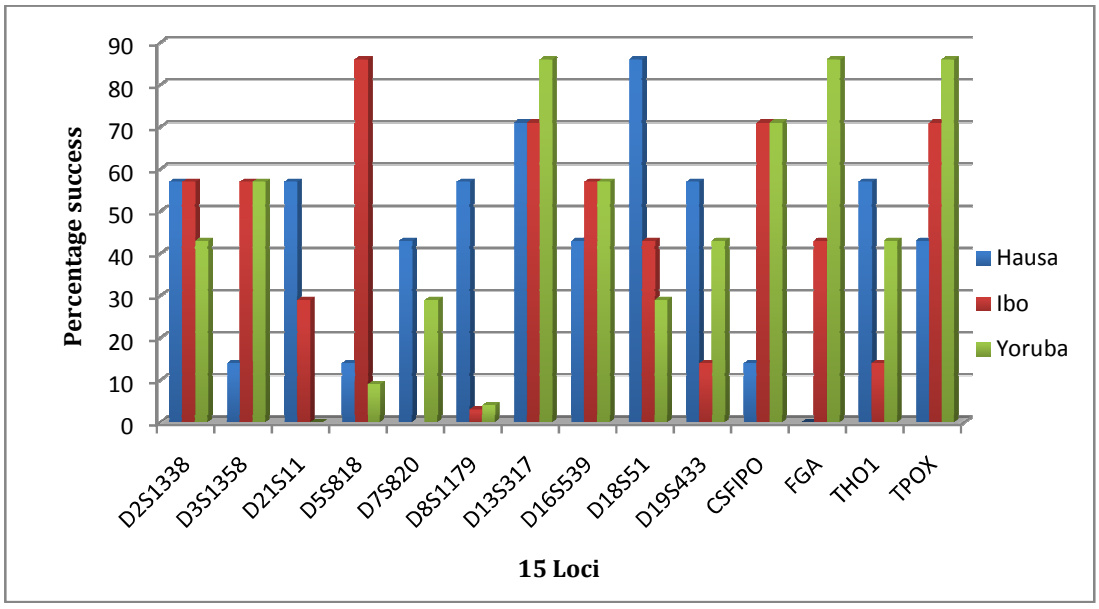

Fig. 5. Percentage Success of Ethnic determination locus-wise

4. DISCUSSION

Microsatellites are highly polymorphic and the most suitable genetic marker for studies in the field of population genetics [9]. In the current investigation, it has been used to infer ethnic origin of test samples. The observed deviations in the chi-square test for conformity of the genotype frequencies with Hardy Weinberg (HW) equilibrium are due to the sample size. This resulted in absence of some alleles in some loci, since it is practically impossible to capture all the alleles in the population in the sample drawn, as well as in allele combination of both high and low frequencies, which usually produce deviations for the equilibrium [6]. However, no individual locus deviated significantly from HW equilibrium after application of Bonferoni corrections for the 15 loci, which adjusted the P- value to 0.003. Exact test also showed all the loci to be in conformity with HW equilibrium. These results agree with earlier studies of such population structure [14,16].

Table 3. Calculated percentage allele proportions for each locus

Ethnic group	Donor	Hausa	IBO	YORUBA
Ha	Hausa Sample 1	40	27	33
Ha	Hausa Sample 2	47	33	20
Ha	Hausa Sample 3	53	32	15
Ha	Hausa Sample 4	53	27	20
Ha	Hausa Sample 5	53.4	13.3	33.3
Ha	Hausa Sample 6	47	27	26
Ibo	Ibo Sample 1	20	67	13
Ibo	Ibo sample 2	26.7	53.3	20
Ibo	Ibo Sample 3	20	47	33
Ibo	Ibo Sample 4	13	47	40
Ibo	Ibo Sample 5	33	47	20
Ibo	Ibo Sample 6	33	47	20
Ibo	Ibo Sample 7	13	47	40
Yoruba	Yoruba Sample 1	27	33	40
Yoruba	Yoruba Sample 2	13.3	33.3	53.3
Yoruba	Yoruba Sample 3	33.3	20	46.7
Yoruba	Yoruba Sample 4	27	33	40
Yoruba	Yoruba Sample 5	27	26	47
Yoruba	Yoruba Sample 6	7	27	67
Yoruba	Yoruba Sample 7	20	26.7	53.3
Hausa	Hausa Sample 7	35	35	29*

Analysis of inter-and intra- population diversity among the studied groups using allele frequency distribution and power of discrimination showed substantial variation across the 15 loci studied, an indication that each STR locus is substantially polymorphic. However, group wise comparison of variance of allele frequency distribution among the populations showed little difference, which is not statistically significant. Though, some relatively reduced genetic variability at some of the loci were observed in terms of power of discrimination, non-the less they exhibit substantial power of discrimination. The combined power of discrimination for the 15 loci was highly significant and that means that when the fifteen loci are used together, they can distinguish samples from different individuals in the population with a high level of probability.

The smallest value of difference between the population in Pair-wise comparison between the Ibo and Yourba observed with regard to power of discrimination is in agreement with the findings that the Ibo and Yoruba ethnic groups belong to the Niger-Kordofanian African linguistic family [13,14,17].

Ethnic origins of the twenty test samples were correctly determined using the allele frequencies generated by the present study. It was based on the principle that the ratio of profile frequencies for the same profile in different ethnic groups is a likelihood ratio [12]. It is equally supported by findings that when populations are isolated from one another for long enough time, then through genetic drift and mutation, they become statistically differentiated by their allele frequencies [18,19,20].

5. CONCLUSION

The determination of ethnic origin by use of the profiles generated was highly successful. The ethnic origin inferential test suggests that the population allele frequencies generated in the current study could be used for the correct prediction of ethnic origin of samples in the three populations. Therefore, this finding could become a tool in the hands of forensic scientists when faced with the need for investigative leads in the studied populations.

ACKNOWLEDGEMENT

The authors are grateful for the laboratory and technical assistance received from Prof. R. Chakraborty, Department of Molecular and Medical Genetics, Health Science Center, University of North Texas, Texas, USA.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Available:http://www.cia.gov/library/publications/the-world-factbook

2. Ethnologue, Ahan. An Ethnologue Country Report. Ethnologue: Languages of the World, 19th Edition, Dallas, USA. 2016;237. Available:http://www.ethnologue.com
3. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment ASM. Genetic structure and history of Africans and African Americans. Science. 2009;324:1035-1044.
4. Gordon, Raymong GJr. Ethnologue (15th Ed.) Dallas: SIL International; 2005.
5. Johnson JA. Ethnic differences in cardiovascular drug response: Potential contribution of pharmacogenetics. Circulation. 2008;118:851-861.
6. Venables SJ, Daniel R, Sarra SD, Soedarsona N, Sudayo H, Suryadi, H. Allele frequency data for 15 autosomal STR loci in eight Indonesian sub population. FSI: Genetics. 2016;20:45-52.
7. Shriver MD, Jin L, Ferrell RE, Deka R. Microsatellite data support an early population expansion in African genome. Res. 1997;7:586-591.
8. Butler JM, Hill CR. Biology and genetics of new autosomal STR loc useful for forensic DNA analysis. Forensic Sci DNA Analysis; 2012.
9. Francesco M, Boschi I, Trombelta F, Merigoni S, Anagnostou P, Baltagia, et al. Using forensic microsatellite to decipher the genetic structure of linguistic and geographic isolates: A survey in the Eastern Italian Alps. FSI. Genetics. 2012;6(6):827-833.
10. Klinstar M, Furedi S, Egyed B, Reiclenpfader, Klieider M. Estimating the ethnic origin (EEO) of individuals using short tandem repeat loci of forensic relevance. Int. Cong. Series. 2003;1239: 53-56.
11. Steele CD, Balding DJ. Statistical evaluation of forensic DNA profiles evidence. Annu. Rev. Stat. Its Appl. 2014;1:361-383.
12. Churchil JD, Shiedes SE, King JL, Budowle B. Evaluation of Illuminia® beta version forenseq™ DNA signature prep kit for use in genetic profiling. FSI. Genetic. 2016;20:20-29.
13. Applied Biosystems. PCR Direct Kit Manual. Foster City, CA; 2012.
14. Katarzyna B, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Wandebe C. Genome wide patterns of population structure and admixture in West Africans and African Americans Nat. Genet; 2009.
15. Available:http://promega.com//
16. Silva NM, Pereira L, Peloni ES, Curat M. Human neutral genetic variation andforensic STR data. PLOSONE. 2012;7:e49666.
17. Adeyemo AA, Guanjie C, Yuanxiu C, Rotimi C. Genetic structure in four West African population groups. BMC Genetics. 2005;6:38.
18. Brenner CH. Probable race of a stain donor: Proceedings from the sixth international symposium on human identification. Promega Madison. 1997;48-52.
19. Available:http://ghr.Alm.N.gov/handbook/genomicresearch/haplgroup
20. Available:http://www.dna.gov/research/alternatve markers/

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/19004