Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System

Benedico DP[2], Calderón S[3], Stojkovic O[4], Garcia-Bertrand R[1*], Rowold DJ[5], Herrera RJ[1]

[1] Department of Molecular Biology, Colorado College, Colorado Springs, CO, USA.
[2] Biology Department, Miami Dade College, Miami, FL, USA.
[3] College of Dentistry, New York University, New York, N.Y., USA.
[4] Institute of Forensic Medicine, School of Medicine, University of Belgrade, Belgrade, Serbia.
[5] Foundation for Applied Molecular Evolution, Gainesville, FL, USA.

Abstract

The aim of this investigation is to determine the capacity of the newly available Y-STR multiplex system, PowerPlex® Y23, to discriminate between populations of similar ancestry, specifically of African descent. Using network analysis, the partitioning of the 23-loci haplotypes was assessed in relation to Y-specific haplogroups. In the network projection, a number of Bantu haplogroups including E1b1a1a1a-M58, B2a1a-M109 and E2b-M98 as well as non-Bantu African haplogroups such as B2b2-M115 and A1b1b2b1-M118 segregate differentially based on Y-STR haplotypes. Further, we contrast population genetics parameters of the Bantu Southeast African and African American populations. Also, the genetic distance values illustrate the robust capacity of the PowerPlex® Y23 system to discriminate among populations. Noteworthy, we demonstrate that the two populations of African ancestry are as genetically different from each other as the African American population is from the Caucasian, Hispanic and Native American groups. For the first time, allelic and genotypic frequencies for the 23 Y-STR loci included in the PowerPlex® Y23 forensic system are provided for a continental Southeast African population, the Bantu from the Maputo Province.

Keywords: South East Africans; Forensic Analysis; Y-chromosome; Y-STRs; Population Genetics; Bantu.

*Corresponding Author:
Dr. Ralph Garcia-Bertrand,
Department of Molecular Biology, Colorado College, 14 East Cache La Poudre Street, Colorado Springs, CO 80903-3294, USA.
Tel: + 1 719 389 6402
Fax: + 1 719 389 6940
E-mail: rbertrand@ColoradoCollege.edu

Introduction

Southeast Africa is a region theorized to represent the southeast fringe of the Bantu expansion. Primarily, the term Bantu represent a Niger-Congo language group [1]. In more recent times, the name Bantu also has become synonymous with a culture. The Bantu people are thought to have originated in what is today northern Cameroon about 5,000 years ago (ya) [1]. Notably, this West African group of people is associated with one of the major human diasporas.

It is thought that the Bantu Expansion started 4,000 to 3,000 ya and by about as recent as 300 ya, it reached Southeast Africa [1]. Bantus disseminated their language as well as agriculture, domestication of animals and iron working over most of sub-Saharan Africa [2-6]. They exerted their influence not merely by acculturation and the spread of ideas but by massive movement of people as well. Although recently investigators have speculated on the possibility that the Bantus migrated south in a single wave along West Africa and then transverse eastward to populate Southeast Africa, the orthodoxy is still of the opinion that the Bantu demographic expansion took place in two parallel dispersals from their homeland, a southwestern course and a southeaster trajectory [4-6]. It has been theorized that limited resources and overpopulation were the primary driving forces for the Bantu dispersion [5]. It is thought that, for the most part, the Bantu expansion involved the dissemination of language, ideas and technology by peaceful means and not by the annihilation of the native people in the conquered lands. In this type of scenario, it is expected that the genetic imprint of Bantus would be evident in most of sub-Sahara Africa.

The available genetic data on sub-Saharan African populations including mtDNA and Y-specific markers [7-10] indicate various degrees of admixture between the invading Bantus and the autochthonous populations of sub-Saharan Africa, depending on the indigenous tribes involved, location and marker system employed. But, for the most part, Bantu DNA is overwhelmingly present in most sub-Equatorial African populations, including the Southeast African Bantu [11].

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol*. 3(11), 202-209

202

Although a number of forensic Y-STR studies [12-16] have been performed on Southeast African populations, it is still an area of research largely neglected. Also, in general, forensic and population genetics studies on sub-Saharan human groups suffer from the limited scope of the investigations, varying in the specific populations examined as well as the marker systems employed. This has made direct comparison among forensic studies difficult. Furthermore, this lack of correspondence makes it difficult to assess the equivalency of forensic databases.

It is known that African American populations derived primary from West African tribes although records of slaves from Southeast Africa have been reported [17]. Considering the historical connections between West African as well as East African tribes and African American populations resulting from the Trans-Atlantic Slave Trade, we embarked in the Y-STR DNA profiling of the Maputo Province Bantu population of Southeast Africa. To accomplish this, we employed the forensic PowerPlex® Y23 system and Y-SNP markers. In addition, here we discuss the results of a number of comparative population genetics analyses designed to ascertain the relationship of this Southeast African population to a number of key pertinent reference populations, including African Americans. We also examine the capacity of the 23-loci Y-STR data to partition Bantu and non-Bantu Y-specific haplogroups. The results presented in this article alleviate a vacuum of basic knowledge, allow for the direct comparison with forensically pertinent worldwide population databases, in particular of African descent, and extend the utility of this Y-STR system outside the USA to a continental African population. This investigation is the first to genotype a Southeast African population across an extended 23Y-STR multiplex using the new generation forensic PowerPlex® Y23 system.

Materials and Methods

Sample collection and DNA isolation

Buccal swabs were collected with informed consent from 78 unrelated Bantu males from the Maputo Province in Southeast Africa. The regional ancestry of each donor was assessed through genealogical information, which was recorded for at least two generations. Only unrelated individuals were sampled. Genomic DNA was isolated using the Gentra Buccal cell kit according to the manufacturer's specifications (Puregene, Gentra Systems, and Minneapolis, MN, USA). The DNA concentration of each sample was determined using a NanoDrop 2000 spectrophotometer (NanoDrop products, Wilmington, DE). The solutions were diluted to a final concentration of 1 ng/µl prior to STR analysis.

Samples were stored as stock solutions in 10 mM Tris–EDTA at -80°C.

Previously published reference populations employed for comparisons

Table 1 provides the geographical locations, abbreviations used to define each population throughout the article, number of individuals per population and the publications reporting on the populations in the literature. Allelic frequencies from a total of seven geographically targeted, forensically pertinent, reference populations (Table 1) were employed for comparison across the 23 Y-STR loci (DYS576, DYS389I/II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 and YGATAH4).

DNA amplification and Y-STR genotyping

The 78 Southeast African samples were amplified across the 23 Y-STR loci (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643, and Y-GATAH4) included in the PowerPlex® Y23 System (Promega Corporation, Madison, WI) following the manufacturer's recommended specifications. Y-STR profiling was accomplished as indicated in a previous publication from our research group [18]. Briefly, amplification reactions were carry out in an ABI PRISM1 GeneAmp1 9700 Silver block Thermal Cycler (Life Technologies) using the 9600 emulation mode for 30 cycles. All analyses employed the CC5 Internal Lane Standard 500 Y23 (ILS) and the allelic ladder mix provided with the PowerPlex1 Y23 System. PCR products were separated and detected on an ABI PRISM1 3130xl Genetic Analyzer (Life Technologies) following the manufacturer's recommendations, using the LIZ-120 internal size standard as a basis for comparison. Fragment sizes were determined using the software GeneMapper v3.1 (Applied Biosystems, Foster City, CA) and alleles were scored by comparison to an allelic ladder supplied by the manufacturer (Promega). It is important to note that for all analyses, alleles at DYS389b were determined by subtracting the repeat length recorded at the DYS389I locus from the number of repeats observed at DYS389II.

Accession number

The 23-loci haplotypes for all the individuals reported for the first time in this publication have been successfully submitted and are

Table 1. Populations Analyzed for the Present Study.

Region	Population	Abbreviation	Sample Size	References
USA	Caucasians	CAU	290	[12]
Southwest USA	Athabaskan	ATH	141	[12]
Alaska USA	Inupiat	INU	145	[12]
Alaska USA	Yupik	YUP	139	[12]
USA	USA Hispanics	HIS	270	[12]
Southeast Florida USA	Southeast Florida Hispanics	SFH	195	[13]
USA	African Americans	AFA	349	[12]
East Africa	Maputo Bantu	MAP	78	Present Study

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol.* 3(11), 202-209

203

now included in the YHRD database under the accession number YA004028.

Statistical and phylogenetic analyses

Allelic frequencies for the 23 Y-STR loci typed were estimated using the gene counting method [19]. Gene diversity (GD) values were calculated, on a per locus basis, with the Arlequin program v3.5 [20]. The same software package was utilized to compute haplotype diversities (HD) using both the PowerPlex® Y23-and the Y-Filer 17-loci datasets. The discriminatory capacity (DC) of each of the aforementioned datasets was calculated by dividing the number of different haplotypes (NDH) by the total number of individuals (N=78) in the population. The fraction of unique haplotypes (FUH), on the other hand, was determined by dividing the number of unique haplotypes (NUH) by the total number of samples.

The haplotype data gathered from the US-YSTR and YHRD databases was used for phylogenetic comparisons to the Southeast African population under investigation. In addition to the Bantu individuals reported in this study, a total of 1,529 samples from seven published, ethnically defined groups, including Native Alaskans [Athabaskans (ATH, N=141), Inupiat (INU, N=145) and Yupik (YUP, N=139)], African Americans (AFA, N=349), Caucasians (CAU, N=290), [USA-wide Hispanics (USH, N=270) [21] and Southeast Florida Hispanics (SFH, N=195)] [18], were utilized to generate population pair-wise Nei's genetic distances (Rst values) and corresponding P values with the aid of Arlequin v3.5 [20]. The statistical significance of each pair-wise comparison was assessed using a significance level of 0.05 and 1,000 permutations [22]. The Bonferroni correction ($\alpha = 0.05/21 = 0.00238$), which compensates for type I errors, was also applied. The Rst values were subsequently employed in the construction of a multidimensional scaling (MDS) plot using the statistical package for the social sciences (SPSS) program version 14.0 (SPSS, 2006). A correspondence analysis (CA) was performed with the NTSYSpc-2.02i software [23]. In addition, a Neighbor Joining (NJ) dendrogram based upon Fst distances was constructed using the POPTREE2 program [24]. Bootstrap analysis with 1,000 reiterations was used to assess the integrity of the phylogenetic relationships determined by the NJ tree. The tree viewing software TreeView v3.2 was subsequently utilized to visualize the dendrogram. Network analysis was performed utilizing the NETWORK 4.6.1.1 software program (www.fluxus-engineering.com) using all the Maputo and African American individuals. In the network calculations, the Y-STR loci were weighted inversely to the size variance. The simplest possible projections were obtained by subjecting the resulting MJ networks to post processing using maximum parsimony parameters. Y-specific haplogroup data for the Maputo individuals was obtained from the literature [11].

For all analyses, the duplicated DYS385 locus was excluded because it is not possible to discriminate between the DYS385a and DYS385b loci with the Powerplex 23 system. In addition, any samples carrying microvariants and/or null alleles were omitted.

Results

Allelic/haplotype frequencies and diversity indices

Allele frequencies and GD indices for the 23 Y-STR loci analyzed

in the Bantu population from Maputo are presented in Table 2. The allelic frequencies and GD values illustrate moderately high genetic heterogeneity with 11 out of 23 loci generating diversity values higher than 0.6 and 15 of the 23 loci generating values greater than 0.5. The most informative locus was DYS481, which exhibited a GD value of 0.8705, while DYS392 was the least polymorphic (GD = 0.0506). A total of 134 alleles were observed, with frequencies ranging from 0.0128 to 0.9744. The number of alleles at each locus differs, with DYS392 possessing the lowest number (2) and DYS481, the highest (10). The 10.2 microvariant allele at DYS391 was detected.

The complete 23-loci Y-STR haplotypes observed in the Maputo Province collection are listed in Supplementary Table 1. A total of 74 haplotypes were noted among the 78 individuals examined when the 17-loci Y-Filer dataset was considered. As anticipated, this number increased to 76 haplotypes with the inclusion of the six additional Y-STR markers (i.e., DYS481, DYS533, DYS549, DYS570, DYS576 and DYS643), with only two of the haplotypes being shared among the males at this resolution.

Population genetics parameters

Table 3 provides the population genetics parameters calculated for the Southeast African population using both the 17- and 23- loci datasets. These analyses were undertaken to ascertain the impact of the six additional markers in the discrimination capacity of the 23-loci Y-STR profiling system. Population genetics parameters of an African American populations using 23-loci were calculated in a previous study [14]. The number of haplotypes (NH), unique haplotypes (UH), fraction of unique haplotypes (FUH), discrimination capacity (DC) and haplotype diversity (HD) values increase when the Powerplex Y23 system is used instead of the Yfiler in the Southeast African Bantu population indicating greater sensitivity and a more robust level of genetic discrimination. The impact of the six additional loci in the Powerplex Y23 over the Yfiler is particularly noted in the DC index (97.44 versus 94.87 in the Maputo population). In comparison, the improvement observed for these parameters in the African American population is not as marked when the 23-loci system is employed (Table 3). The comparatively higher DC value in the African Americans compared to the Southeast Africans at the 23- and 17-loci resolution can be attributed to the multi-ethnic nature of African Americans.

In the Maputo Bantu population, only two out of 78 different 23-loci haplotypes are not unique within the population compare to double the number when the 17-loci is used. The greater number of unique haplotypes resulting from the higher resolution provided by the PowerPlex Y23 STR system can prove useful in discriminating between individuals belonging to similar paternal lineages within populations, and provide additional resolution to exclude innocent individuals in forensic cases.

Statistical analyses and genetic relationships among populations

Pairwise genetic distances (R$_{st}$) were calculated among the eight populations based on 21 Y-STR loci provided in the PowerPlex® Y23 system (Table 4). P-values indicate significant differences for all of the population pairs before ($\alpha = 0.05$) and after the Bonferroni correction. The greatest genetic distance is observed between the Maputo Southeast African population and the USA

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol.* 3(11), 202-209

204

Table 2. Allelic frequencies for the 23 Y-STR loci in the Maputo Southeast African population (n=78).

allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	GATA_H4	DYS481	DYS533	DYS549	DYS570	DYS576	DYS643	Genotype	DYS385a/b
8																					0.012821	11,17	0.01282
9						0.025641																12,16	0.03846
10				0.025641	0.025641			0.192308	0.012821							0.038462	0.051282	0.012821			0.025641	12,17	0.14103
10.2				0.012321																		12,18	0.05128
11		0.012821			0.782051	0.974359		0.782051	0.461538						0.346154		0.75641	0.74359		0.115385		13,14	0.01282
12		0.230769		0.166667			0.012821	0.012821	0.025641	0.333333		0.012821			0.435897		0.166667	0.192308		0.115385		13,15	0.01282
13	0.025641	0.589744		0.012821				0.602564		0.076923		0.141026			0.166667		0.025641	0.038462		0.397436		13,16	0.02564
14	0.076923	0.166667	0.012821				0.205128	0.923077		0.064103	0.012821	0.012821	0.012821							0.153846	0.282051	13,17	0.32051
15	0.641026		0.025641				0.141026	0.051282		0.025641		0.602564	0.115385						0.012321	0.25641	0.051282	13,18	0.17949
16	0.166667		0.064103						0.038462	0.01282		0.115385	0.397436						0.012821		0.397436	13,19	0.03846
17	0.089744	0.5										0.089744	0.269231	0.128205					0.230769	0.217949		14,15	0.01282
18		0.358974									0.012821	0.025641				0.012821			0.333333	0.153846		14,17	0.02564
19		0.038462										0.089744		0.025641					0.179487	0.064103		14,18	0.12821
20				0.012821							0.141026			0.076923					0.166667	0.038462			
21		0.692308									0.628205			0.602564					0.025641	0.012821			
22		0.025641									0.025641			0.089744		0.012821							
23		0.038462									0.089744			0.038462		0.102564							
24		0.153846												0.064103		0.076923							
25		0.076923														0.205128							
26																0.102564							
27																0.153846							
28																0.192308							
29																0.089744							
30																0.051282							
33																0.012821							
GD	0.5538	0.5784	0.6227	0.4952	0.3643	0.0506	0.5808	0.4496	0.3553	0.673	0.5758	0.6024	0.7326	0.6087	0.6693	0.8705	0.4019	0.4136	0.7822	0.8338	0.7419		0.8318

Table 3. Parameters of forensic interest.

Haplotypes	Maputo	African American
23-loci Y-STR Haplotype		
Sample Size	78	349
Number of haplotypes	76	347
Unique haplotypes	74	345
Fraction of unique haplotypes	94.87	98.85
Discrimination capacity	97.44	99.43
Haplotype diversity ± SD	0.9993 +/- 0.0022	0.9999 ± 0.0002
17-loci Y-STR Haplotype		
Sample Size	78	349
Number of haplotypes	74	344
Unique haplotypes	70	340
Fraction of unique haplotypes	89.74	97.42
Discrimination capacity	94.87	98.57
Haplotype diversity ± SD	0.9987 +/- 0.0023	0.9999 +/- 0.0002

Table 4. R*st* values based on 21 Y-STR loci (above diagonal) and associated p-values (below diagonal) between the Maputo Province Southeast Africa population and other pertinent populations (α = 0.05, 1,000 reiterations).

	ATH	INU	YUP	CAU	HisUS	SFH	AFA	MAP
ATH	*	0.15703	0.1518	0.04539	0.01444	0.03528	0.19543	0.3489
INU	0	*	0.02907	0.25281	0.14522	0.18028	0.28442	0.41809
YUP	0	0	*	0.26726	0.13949	0.18075	0.26908	0.40466
CAU	0	0	0	*	0.04124	0.03787	0.30374	0.48606
HisUS	0.00196	0	0	0	*	0.01545	0.19581	0.35726
SFH	0	0	0	0	0.00098	*	0.22069	0.37001
AFA	0	0	0	0	0	0	*	0.0753
MAP	0	0	0	0	0	0	0'	*

[a] Excluding the DYS385a/b loci.

Bonferroni<0.00238

Caucasian group (R*st* = 0.48606) while the smallest coefficient occurs between Athabascans and USA Hispanics (R*st* = 0.01444). Noteworthy is that the top six highest genetic distances are generated when the Southeast African Bantus are compared to the references populations.

The phylogenetic relationships between the Maputo Province population and seven reference populations were assessed using a MDS plot (Figure 1), a CA graph (Figure 2) and a NJ dendrogram (Figure 3). The MDS analysis (Kruskal's stress = 0.055) is based on R*st* distances derived from genotypic data, while the CA and NJ tree were generated from allelic frequencies. Yet, both the MDS and CA plots exhibit very similar partitioning of populations. The MDS and CA reveal the Native Alaskan populations,

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol.* 3(11), 202-209

205

Inupiat and Yupik, segregating close together while the Athabaskans, the two Hispanic populations and the Caucasians cluster together. In both MDS and CA, the Caucasian group plots more distant from the other three populations. Of particular interest is the very similar (in relation to the other populations) partitioning of the two populations of African descent in the two plots, with the Southeast Africans as extreme outliers in the upper right quadrant distantly separated from the African American group along the X-axis (first component, 49.37% of the total diversity). The NJ tree topology (Figure 3) is congruent with the distribution of the MDS and CA graphs with half of the bifurcations at bootstraps values of 100% and the other half above 50% values.

The network analysis based on Y-STR haplotypes of all Maputo Bantu and African American individuals exhibits a well-defined separation into three main clusters, from now on referred as top, middle and bottom (Figure 4). In addition, sub-clusters partition

a number of Bantus, non-Bantus Africans and Eurasian haplogroups (see sub-haplogroups indicated within brackets of Figure 4). The population from the Maputo Province segregates differentially into Bantu Africans (E1b1a1a1a-M58, B2a1a-M109 and E2b-M98), non-Bantu Africans (B2b2- M115 and A1b1b2b1-M118) and Eurasians (R1a1a-M198) Y haplogroups. Unfortunately, no haplogroup data is available for the African American population. The three main clusters in the network are separated by considerable distance indicating substantial differentiation among them. Briefly, the top conglomerate is made up mostly of Bantu B2a1a-M109 individuals. A sub-division of this top clade is made up of Bantu E2b-M98 samples. The ancient A lineage is represented in this cluster by a single A1b1b2b1-M118 individual. The majority of the samples in the middle aggregate belong to a number of well-resolved E1 haplogroups including E1b1a1a1f1a1c-P116 and E1b1a1a1g1c-M154. At the lower extreme of this middle cluster, the less resolved E1b1a1a1a-M58 haplogroup is found

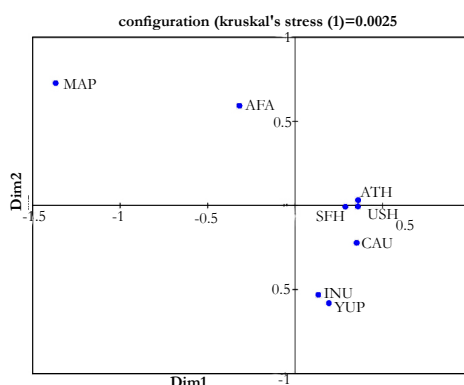**Figure 1. MDS of the Maputo Southeast African and reference populations.**
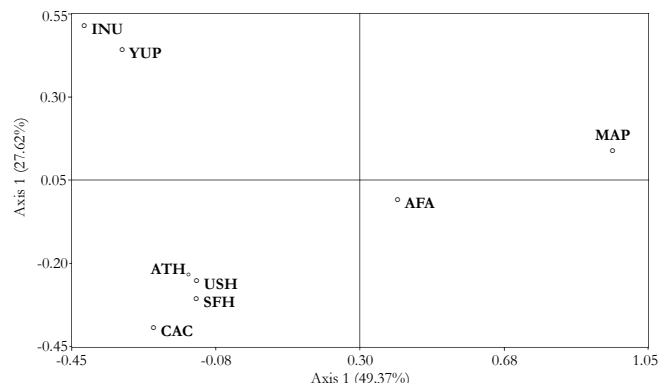

Figure 2. CA of Southeast Africans.

Figure 3. NJ of Southeast Africans.

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol.* 3(11), 202-209
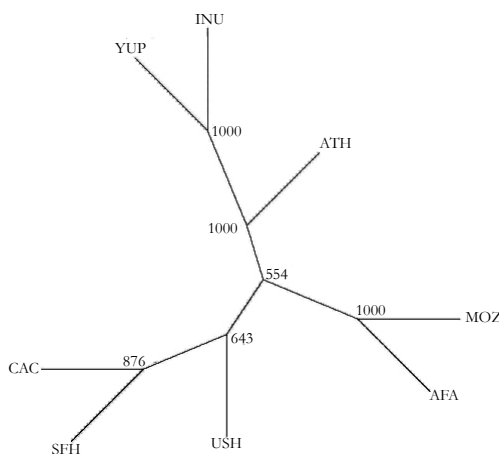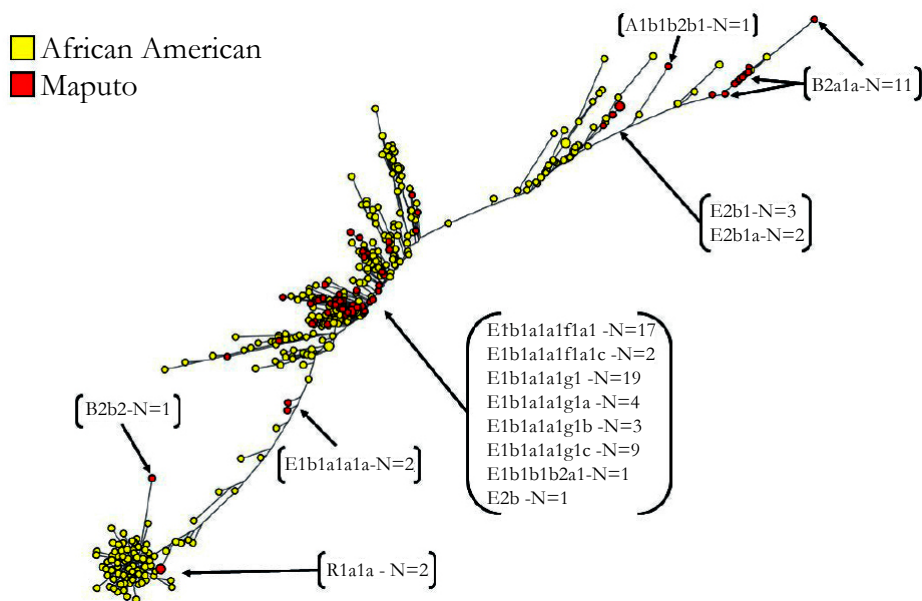
Figure 4. Network analysis of African American and Maputo individuals.

in two persons. The bottom conglomerate is composed of one individual containing the Pygmy B2b2-M115 haplogroup and two persons with the Eurasian haplogroup R1a1a-M198. This bottom cluster is genetically very homogeneous with samples clumping close together. Although the haplogroups of the African American individuals examined in this study are not available, it is likely that the African American individuals at the bottom duster are R1a1a-M198 of European decent. The presence of a substantial number of European haplogroups in the African American population is congruent with the admix nature of Americans of African ancestry. The two Maputo R1a1a-M198 individuals in the bottom cluster may be of recent European or West Asian ancestry. The low level of genetic differentiation among the samples in this bottom group suggests that the haplogroups within it have had limited time to differentiate from each other.

Discussion

Intra-population Diversity

The allelic frequencies in Table 2 reveal a moderately high degree of genetic diversity within the Maputo Southeast African population which can be attributed to the high variability characteristic of sub-Saharan African populations in general. Four of the six additional loci provided by PowerPlex® Y23 exhibit the highest number of alleles ((DYS481, 10 alleles; DYS576, 8; DYS570, 7; DYS643, 7). The high diversity of these four additional markers reflects, overall, the most polymorphic loci seen in the world wide and sub-Saharan African populations previously examined with the Y23 system [25]. The elevated diversity provided by these four highly polymormic loci is also reflected in their GD values. Except for DYS385a/b, the same four of the additional six loci (DYS481, 0.8705; DYS576, 0.8338; DYS570, 0.7822; DYS643, 0.7419) possess the highest GD values in the set of 23 markers (Table 2). Similarly, overall, the world wide and continental African populations (Benin, West Africa; Nigeria, West Africa; Zimbabwe, South Africa; South Africa, South Africa; and two from Kenya, Northeast Africa) recently genotyped with the Y23 system [25] possess high GD values for these same four markers. None of the sub-equatorial African populations examined in the study

[25] are from Southeast Africa.

The forensic parameters results for the Southeast African population indicate improvement when ascertained with the PowerPlex® Y23 as opposed to the 17 loci Yfiler® system (Table 3). The greater number of unique haplotypes resulting from the higher resolution provided by the PowerPlex® Y23 STR system as compared to the 17 loci kit can prove useful in discriminating between individuals belonging to related but different paternal lineages. Further, the additional six loci should also assist in population genetics studies as it will allow for higher resolution assessment of genetic differences, and may perhaps reveal novel genetic signatures of populations or demographic events that were hidden at the lower resolution provided by 17 or lower number of loci.

The HD observed for the Southeast Maputo population (0.9990) is comparable to the value of the African American group previously reported [18, 21] while the DC favors the African Americans (99.43 versus 96.15). This higher DC of the African American over the Southeast Africans may be the result of the highly admixed nature of the African Americans. When the HD and DC indices of the Southeast Africans are compared to the corresponding average values computed from the six continental African populations previously reported [25], the Maputo Bantus exhibit higher numbers (HD, 0.9990 versus 0.9987 and DC, 96.15 versus 84.72, respectively). These differences are likely due to the genetic diversity generated by the admixture experienced from numerous distinct ethnic groups (e.g., Khoisan, Zulu, Ndebele and Bantus) in Southeast Africa.

Genetic Distances

The pairwise genetic distances (R*st*) calculated among the eight populations based on the 21 Y-STR loci allowed for significant differences among all of the population pairs before and after the Bonferroni correction. The top highest genetic distances (R*st* = 0.48606 - 0.34890) are generated when the Southeast African population is compared to the six collections of non-African descent. The second highest genetic distances (R*st* = 0.30374 - 0.19543) are observed involving comparisons between African

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol*. 3(11), 202-209

207

Americans and the non-African reference populations. In both instances, the highest values represented comparisons between Caucasians versus the Southeast African and African American populations (Rst = 0.48606 and 0.30374, respectively). The genetic distance between Southeast Africans and African Americans was found to be greater than between Caucasians and Hispanic populations or Athabascans. These results underscore the genetic differences of populations of African ancestry from each other and from other ethnic groups.

Inter-Population Diversity

The MDS (Figure 1) and CA (Figure 2) graphs based on different types of raw data (genotypes and allelic frequencies, respectively) and algorithms, illustrate very similar partitioning of populations. This corroboration of results by different methods adds weight to the reliability of the phylogenetic data. In both projections, the two populations of African ancestry partition distant from the rest of the populations and from each other. In the CA analysis (Figure 2), the separation between the two African populations is comparable to the segregation between the African Americans from the rest of the reference groups. The considerable distance between the two groups of African descent suggests a lack of close genetic affinities between the two. Even though in both plots the African populations segregate distant from each other as well as from the rest of the reference collections, the distance between the African American and Southeast African populations is smaller in the MDS compared to the CA. The outlier positions of both African groups reflects their distant genetic affinities to the other reference collections while the smaller separation of the African Americans, compared to the Maputo Bantus, from Caucasians, Hispanics and Native Americans likely reflects the admixture of African Americans with Caucasians, Hispanic and Native Americans in the USA. The NJ dendrogram constructed, including bootstrap values, is congruent with the relationships displayed in the CA and MDS analyses. Our results suggest that as DNA fingerprinting technology becomes part of the arsenal of forensic laboratories in countries world wide, including continental Africa, these marked differences among populations of African ancestry will need to be considered possibly justifying independent databases.

Network Analysis

The network analysis based on the 23-loci STR haplotypes indicates a well-resolved topology exhibiting three distinct main clusters. The 23 STR loci in the PowerPlex® Y23 system provides for the resolution necessary to discriminate among the Bantu, non-Bantu (*e.g.*, Pygmy) and Eurasian Y chromosomes in the Maputo population of Southeast Africa. Also, the partitioning of high-resolution haplogroup markers such as E1b1a1a1f1a1c-P116 and E1b1a1a1g1c-M154 to a genetically homogeneous cluster (middle cluster, Figure 4) is indicative of the power of this Y23 STR system to differentiate among recent Y haplogroups. It is expected that the Maputo R1a1a-M198 individuals are underrepresented in the bottom cluster relative to African Americans. Although the haplogroup data for the African American population in the network is not available, it is likely that the African Americans in the bottom cluster possess R1a1a-M198 Y chromosomes of European descent. This notion is supported by the common practice in America of European masters fathering children with female slaves.

Conclusion

The findings of the present study include:

1. The six additional loci provided by the PowerPlex® Y23 system, compared to Yfiler® 17 loci, improve the forensic parameter values of the Bantu population from Southeast Africa.
2. Four of the six additional loci exhibit the highest number of alleles and GD values in the PowerPlex® Y23 system.
3. The higher discrimination capacity exhibited by African Americans over the Southeast African population may be the result of the highly admixed nature of the former.
4. The phylogenetic separation between the Southeast African and African American populations is comparable to the segregation between the African Americans and the rest of the reference populations.
5. The considerable genetic distance between the two populations of African ancestry suggests a lack of close genetic affinities between the two and argues for the use of independent databases.
6. Network analysis based on the 23 STR loci in the PowerPlex® Y23 system provides for the resolution necessary to discriminate among the Bantu, non-Bantu (*e.g.*, Pygmy) and Eurasian Y chromosomes in the Maputo population of Southeast Africa.

Acknowledgement

References

[1]. Greenberg JH (1972) Linguistic evidence regarding Bantu origins. J African History 13(2): 189-216.
[2]. Desmond J, Brandt S (1984) From Hunters to Farmers: the Causes and Consequences of Food Production in Africa. University of California Press, CA, USA.
[3]. Phillipson D (1993) African Archaeology. Cambridge University Press, Cambridge, MA, USA.
[4]. Newman J (1995) The Peopling of Africa: a Geographic Interpretation. Yale University Press, New Haven, CT, USA.
[5]. Diamond J (1997) Guns, Germs and Steel: the Fates of Human Societies. W. W. Norton & Company, New York, USA.
[6]. Berniell-Lee G, Bosch E, Bertranpetit J, Comas D (2006) Y-chromosome diversity in Bantu and Pygmy populations from Central Africa. International Congress Series 1288: 234-236.
[7]. Batini C, Coia V, Battaggia C, Rocha J, Pilkington MM, et al. (2007) Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. Mol Phylogenet Evol 43(2): 635-644.
[8]. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, et al. (2008) Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. PNAS 105(5): 1596-1601.
[9]. Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A (2005) The genetic legacy of western Bantu migrations. Hum Genet 117(4): 366-375.
[10]. Batai K, Babrowski KB, Arroyo JP, Kusimba C, Williams SR (2013) Mitochondrial DNA diversity in two ethnic groups in Southeastern Kenya: Perspectives from the northeastern periphery of the Bantu expansion. Am J Phys Anthropol 150(3): 482-491.
[11]. Rowold D, Garcia-Bertrand R, Calderon S, Rivera L, Benedico DP, et al. (2014) At the southeast fringe of the Bantu expansion: genetic diversity and phylogenetic relationships to other sub-Saharan tribes. Meta Gene 2: 670-685.
[12]. Pereira L, Gusmão L, Alves A, Amorin A, Prata MJ (2002) Bantu and European Y-lineages in Sub-Saharan Africa. Ann Hum Genet 66: 369-378.
[13]. Alves C, Gusmão L, Barbosa J, Amorim A (2003) Evaluating the informative power of Y-STRs: a comparative study using European and new African

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol.* 3(11), 202-209

208

haplotype data. Forensic Sci Int 134(2-3): 126-133.

[14]. Gusmão L, Sanchez-Diz P, Alves C, Quintáns B, García-Poveda E, et al. (2003) Results of the GEP-ISFG collaborative study on the Y chromosome STRs GATA A10, GATA C4, GATA H4, DYS437, DYS438, DYS439, DYS460 and DYS461: population data. Forensic Sci Int 135(2): 150-157.

[15]. Sanchez-Diz P, De La Fe T, Quintans B, Salas A, Lareu MV, et al. (2003) Y-chromosome STRs in populations of Bantu origin from Mozambique: male contribution to the Africa genetic pool and forensic implications. International Congress Series 1239: 419-424.

[16]. Carvalho M, Brito P, Lopes V, Andrade L, Anjos MJ, et al. (2010) Analysis of paternal lineages in Brazilian and African populations. Genet Mol Biol 33(3): 422–427.

[17]. Simms TM, Martinez E, Herrera KJ, Wright MR, Perez OA, et al. (2011) Paternal lineages signal distinct genetic contributions from British Loyalists and continental Africans among different Bahamian islands. Am J Phys Anthropol 146(4): 594-608.

[18]. Calderon S, Perez-Benedico D, Mesa L, Guyton D, Rowold DJ, et al. (2013) Phylogenetic and forensic studies of the Southeast Florida Hispanic population using the next-generation forensic PowerPlex® Y23 STR marker system. Leg Med (Tokyo) 15(6): 289-292.

[19]. Li CC (1976) Course in Population Genetics. Boxwood Press, Pacific Grove, CA, USA.

[20]. Excoffier L, Laval LG, Schneider S (2005) Arlequin ver.3.0: an integrated software package for population genetic data analysis. Evol Bioinform 1: 47-50.

[21]. Davis C, Ge J, Sprecher C, Chidambaram A, Thompson J, et al. (2013) Prototype PowerPlex® Y23 System: A concordance study. Forensic Sci Int Genet 7(1): 204-208.

[22]. Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, et al. (2003) Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. Genome Res 13(4): 624-634.

[23]. Rohlf F (2002) NTSYSpc Numerical Taxonomy and Multivariate Analysis System. Setauket, New York.

[24]. Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. Mol Biol Evol 27(4): 747-752.

[25]. Purps J, Siegert S, Willuweit S, Nagy M, Alves C, et al. (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet 12: 12-23.

Garcia-Bertrand R, et al., (2015) Genetic Diversity of Southeast African Bantus and African Americans using the PowerPlex Y23 System. *Int J Forensic Sci Pathol.* 3(11), 202-209

209