

7-15-2002

D1S80 DNA profiling in five African populations

Leslie R. Adrien

Florida International University

Follow this and additional works at: <http://digitalcommons.fiu.edu/etd>



Part of the [Medical Pathology Commons](#)

Recommended Citation

Adrien, Leslie R., "D1S80 DNA profiling in five African populations" (2002). *FIU Electronic Theses and Dissertations*. Paper 1138.
<http://digitalcommons.fiu.edu/etd/1138>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

DIS80 DNA PROFILING IN FIVE AFRICAN POPULATIONS

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

FORENSIC SCIENCE

by

Leslie R. Adrien

2002

To: Dean Arthur W. Herriott
College of Arts and Sciences

This thesis, written by Leslie R. Adrien, and entitled D1S80 DNA Profiling in Five African Populations, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Stephen Winkle

George Duncan

Rene J. Herrera, Major Professor

Date of Defense: July 15, 2002

The thesis of Leslie R. Adrien is approved.

Dean Arthur W. Herriott
College of Arts and Sciences

Dean Douglas Wartzok
University Graduate School

Florida International University, 2002

DEDICATION

I dedicate this thesis to my family and friends, whose support and love made it possible to keep my sanity while successfully completing my goals.

ACKNOWLEDGMENTS

My deepest indebtedness is to God whose endless grace and blessings enabled me to achieve my current feat.

I would like to thank all the people who have assisted me in accomplishing my objectives. First, to all my lab mates who were always willing to answer my questions and share their technical knowledge with me. I would also like to thank all my committee members for their support.

This research was partially supported by MBRS SCORE GM 08205, MBRS SCORE GM 61347, MARC U-STAR GM 08771.

ABSTRACT OF THE THESIS

D1S80 DNA PROFILING IN FIVE AFRICAN POPULATIONS

by

Leslie R. Adrien

Florida International University, 2002

Miami, Florida

Professor Rene J. Herrera, Major Professor

The highly polymorphic D1S80 locus has no known genetic function. This variable number of tandem repeat (VNTR) has been valuable in forensic identification. We have obtained allelic and genotypic frequencies for five African populations (Benin, Cameroon, Egypt, Kenya and Rwanda), which could be employed as databases to identify individuals.

The polymerase chain reaction, followed by vertical polyacrylamide gel electrophoresis and silver staining was our method of analysis. Allele frequencies were used to infer genetic associations using Phylip 3.5, Principal Component and G-test statistical programs. Tests for Hardy-Weinberg equilibrium were employed. F_{st} estimates and power of discrimination values were also determined for each of our populations.

Our analyses of 28 additional populations demonstrated that the D1S80 locus alone provided for the discrimination of major racial groups. Genetic homogeneity between the African groups was observed. We have generated a database useful for human differentiation and phylogenetic studies.

TABLE OF CONTENTS

CHAPTER	PAGE
INTRODUCTION	1
BACKGROUND INFORMATION	4
MATERIALS AND METHODS	
DNA Isolation	6
Amplification of DNA Template	6
Vertical Polyacrylamide Gel Electrophoresis	7
Population Sampling	8
Statistical Analysis	11
RESULTS	13
DISCUSSION	28
CONCLUSION	43
REFERENCES	44

LIST OF TABLES

TABLE	PAGE
1. Populations Included	10
2. Allelic Frequencies	14
3. Genotypic Frequencies	15
4. Observed/Expected Homozygosity and Heterozygosity Values	24
5. Test for Hardy-Weinberg Equilibrium and Power of Discrimination	24
6. G Scores with probability values between all 33 populations	25

LIST OF FIGURES

FIGURE	PAGE
1. Neighbor Joining Tree	18
2. UPGMA Tree	19
3. Maximum Likelihood Tree with Bootstrap Values	22
4. Principal Component Plot	23
5. Allelic Distribution Graph of the Benin Population	30
6. Allelic Distribution Graph of the Kenya Population	31
7. Allelic Distribution Graph of the Rwanda Population	32
8. Allelic Distribution Graph of the Tanta Population	33
9. Allelic Distribution Graph of the Cameroon2 Population	34
10. Allelic Distribution Graph of the Cameroon Population	35
11. Allelic Distribution Graph of the Congo Population	36
12. Allelic Distribution Graph of the Nigeria Population	37
13. Allelic Distribution Graph of the Zimbabwe Population	38
14. Allelic Distribution Graph of the African American Population	39
15. Allelic Distribution Graph of the Caucasian Population	40

INTRODUCTION

The D1S80 locus has been utilized in paternity testing, forensic applications and phylogenetic analysis, since its description by Nakamura [1]. It is the most characterized amplified fragment length polymorphism (AmpliFLP) [2]. As a single locus, this variable number of tandem repeat (VNTR) is highly polymorphic. Observed heterozygosity has been reported as high as 78% [3]. The D1S80 locus by itself enables the discrimination of races and ethnic groups [4]. The standard procedure for D1S80 analysis involves the polymerase chain reaction (PCR) followed by polyacrylamide gel electrophoresis (PAGE) and silver staining. D1S80 is amenable to the PCR since its size ranges from 369 to over 785 base pairs [5].

We decided to focus our study on African populations and their phylogenetic relationships to non-African populations. Sub-Saharan African populations are known for their high levels of diversity [6]. Our research is concentrated on five different populations. These five groups were collected in Benin, Cameroon, Egypt, Kenya and Rwanda. All populations, except the Egyptian population, are sub-Saharan. The majority of all sub-Saharan Africans are now Bantu speaking resulting from the Bantu expansion estimated to have occurred approximately three thousand years ago [6]. The Bantu diaspora was driven by the incipient agricultural revolution in Africa. These farmers/invaders migrated from the central-western coast of Africa eastward and southward. Bantus make up one of four main extant African groups. A brief look into the history of each population provided information on the genetic composition of each. The Sahara desert has been a migration barrier in recent human evolution. Berbers and

Arabs mainly influenced the genetic make-up of groups in northern Africa, while the heritage south of the Sahara are native Africans, with limited European admixture occurring relatively recently, in the late 15th century [7].

Forty-two separate ethnic groups populate Benin [8]. Due to its location on the west coast of Africa, it was highly involved in the Atlantic slave trade. The natives of this country endured both British and French authority over them for more than a century. The Bantu dispersal is thought to have originated in an area including the country of Benin [6].

Tutsi kings reigned in Rwanda from the late 18th century up to the 1960's [8]. The Hutus, a Bantu group, penetrated the region west of Lake Victoria during the Bantu dispersal. Ethnic strife exists between these two groups and is manifested through their battles for leadership. Belgium colonized the territory following World War I for a period of approximately 40 years. During this time, the requirement for each tribe to be distinguished with identity cards kept them isolated. Rwanda is known to be one of the most densely populated countries in Africa [9]. The Hutu group makes up 90% of this country [8].

The native people of Cameroon experienced authority from the Portuguese, British, Germans and French. Cameroon has the most ethnic groups of any other country in Africa, namely 200 [8]. The country is currently divided into a northern French zone and a southern British zone. Cameroon is, geographically, just south of the focus of the Bantu expansion.

Kenya, on the east coast of Africa, has influence of several Arab groups. The cultural and genetic impact of Omani traders is clearly visible in this country. The British

arrived in the 1800's and assigned different ethnic groups to different parts of the country to live, thereby further isolating these groups. There are more than 100 different ethnic groups established in Kenya [8]. A 1989 census indicated that Arabs, Asians and Europeans made up less than 1% of the total population [10].

The Egyptian samples were collected from the area of Tanta, which is located on the Nile River delta. This north African population is largely made up of Arabs and Berbers [6]. Berbers were the earliest known inhabitants in north Africa [7]. From 711 of the common era (CE) on, Arabs have dominated the Berbers within these areas [6].

In this study, we examined the D1S80 allelic frequencies of African populations from Benin, Cameroon, Egypt, Kenya and Rwanda and compared them to 28 worldwide reference populations. This information will be the basis of databases for forensic probability estimation and population genetic studies. Allelic and genotypic frequencies were determined by utilizing the gene count method [11]. Divergence from Hardy-Weinberg equilibrium (HWE) expectations were determined, with the Fisher exact test and the Chi-Square test, as well as F_{st} estimates and power of discrimination values in each of our populations. Observed and expected homozygosity and heterozygosity values were also generated. In addition, twenty-eight other populations were incorporated into our study to ascertain possible genetic similarities. Allele frequencies were used to infer genetic associations through phylogenetic, principal component and G-Test statistical analyses.

BACKGROUND INFORMATION

The D1S80 locus is located at the telomeric position of the p arm of chromosome 1. Nakamura and collaborators first described this locus in 1988 [3]. Its designation stems from the fact that this is a DNA locus on chromosome 1, a Single copy sequence and it was the 80th locus described on chromosome 1. D1S80 is a VNTR that is also referred to as an AmpliFLP. To date, no phenotypic trait has been associated with this locus [5]. AmpliFLPs are distinguished by differences in length between the alleles, not by sequence [3, 12], although considerable sequence differences have been reported. The overall length ranges of AmpliFLPs are from 200 base pairs (bp) to 1000 bp [3]. The D1S80 known allele size range, including the flanking regions, is from 369bp to 785bp [5]. D1S80 has a core repeat unit of 16 bps [3]. D1S80 alleles typically contain from 14 to 41 repeats of this 16 bp sequence [13]. D1S80 alleles are co-dominantly inherited [12]. Estimates of heterozygosity are generally greater than 70% [14, 15]. Although this is a single locus, it is highly polymorphic due to the large number of alleles it contains (more than 27). As many as twenty-nine different alleles have been found within one population [5, 13]. The number of possible genotypes can be calculated by the formula $G = n(n+1) / 2$ [3]. According to this formula, 435 genotypes are possible when 29 alleles are identified. A high number of alleles most likely result from high mutation rates and a lack of selection pressure. Alleles with greater than 41 repeats and those smaller than 14 have both been reported [16-19]. The alleles with the highest recurring frequency are those with 18 and 24 repeats [12, 20-22]. The high frequency of these alleles suggests

that these are ancestral alleles [23]. Allele 15 is not common, and therefore is not incorporated into D1S80 allelic ladders.

MATERIALS AND METHODS

DNA Isolation

All samples were collected as whole blood in EDTA Vacutainer tubes from unrelated individuals. These individuals were identified by biographical information traced back at least two generations. The appropriate ethical guidelines, as dictated by Florida International University, were followed during the collection of samples. Each collection was arranged through the leaders of the regions and supervised by the same. The blood samples were lysed and leukocyte nuclei were separated from the rest of the blood components. Several washes in lysis buffer were performed to rinse and isolate the leukocyte nuclei. The DNA was then purified using proteinase-K digestion and standard organic phenol-chloroform extraction. Ethanol precipitation followed to pellet the DNA. DNA was then dissolved in TE (Tris-Ethylenediamine Tetraacetate) buffer. All samples were stored in -80°C when not in use.

Amplification of DNA

PCR was performed in a Perkin-Elmer 480 thermal cycler. Amplification parameters were 1 minute at 94°C, 1 minute at 65°C and 1 minute at 72°C for twenty-nine cycles. An additional 10-minute extension at 72°C followed [13]. PCR products were stored at 4°C until they were loaded onto a polyacrylamide gel. PCR amplifications were carried out in 25µl reactions containing 16.3µl of water, 2.5µl of 10X buffer containing

15mM MgCl₂, 2.5µl (0.15mM) dNTPs, 1.25µl (10µM) of each primer, 0.2µl (5U) *Taq* DNA and 10-50 nanograms of DNA sample. The sequence of the forward primer was 5' GAAACTGGCCTCCAAACACTGCCCCGCCG 3' and that of the reverse primer was 5' GTCTTGTTGGAGATGCACGTGCCCTTGC 3' [24]. The Benin samples were DOP (degenerate oligonucleotide primer) pre-amplified according to Buchanan and collaborators [25]. This method of PCR is used when the DNA is of poor quality or low yield. From the DOP-PCR product, 6µl were used to specifically amplify for D1S80. Samples were overlaid with two drops of mineral oil to protect them from evaporation throughout the amplification procedure.

Vertical Polyacrylamide Gel Electrophoresis

A 39:1 acrylamide stock solution was used to make an 8% polyacrylamide gel. The longer glass plate used for casting the slab gel was treated with an adherent to retain the gel as the shorter plate is detached. Gels were 35 x 45 centimeters in size with a 0.8-millimeter thickness or 38 x 46 centimeters in size with a 1-millimeter thickness. An Applied Biosystems AmpliFLP D1S80 Allelic Ladder was electrophoresed every four samples. The gels were run at 1000-1200 volts for five to six hours, depending upon the migration of the xylene-cyanol tracking dye. Electrophoresis was stopped 45 minutes after the xylene-cyanol dye ran out of the bottom of the gel. All gels were silver stained to visualize the alleles.

Silver staining consisted of a 5-minute slow agitation in 1% nitric acid. Following a water rinse, the gel was placed in fresh 0.2% silver nitrate and shaken for 20

minutes. Another water rinse was done prior to the development step in fresh 0.28 M sodium carbonate, 0.05% formaldehyde solution. Development consisted of several rinses in the sodium carbonate solution until it no longer became “milky”. Once the desired band intensities were achieved, the gel was placed in 10% acetic acid stop solution for 5 minutes. Gels were then placed in water for another 5 minutes. After drying, gels were photographed under visible light. All gels were scored to ascertain genotypes independently by at least two people. Genotyping was performed directly from the gels by comparison to the allelic ladder. Any discrepancies were repeated.

Population Sampling

A total of five African populations were analyzed in our study. They were Benin, Cameroon, Egypt, Kenya and Rwanda. All the sub-Saharan populations belong to the Niger-Congo language family. A total of one hundred samples were analyzed from Benin. These individuals belong to the Fon ethnic group. They were collected in the southern part of the country from the Zagnanado population. The Kenya group consisted of 106 Bantu individuals from small villages 100 kilometers northeast of Nairobi. The Rwanda population was made up of 100 individuals of the Hutu tribe. Hutus are a Bantu group. The Egyptian group was collected in the region of Tanta located at the center of the Nile delta. Tanta is approximately 94 kilometers north of Egypt’s capital city, Cairo. This population is primarily Arab/Berber in origin. Arabia and north Africa both share the Afro-Asiatic language. This north African population shares their ancestry to a lesser extent with Greeks, Turks and other Mediterranean Basin people. Tantas are therefore

Caucasians. There were 100 samples from Tanta processed. The Cameroon group consisted of 16 Bantu individuals from villages 50 kilometers southwest of the city of Yaounde, in the southern region of the country. In our study, these 16 samples were analyzed together with the 34 individuals reported by Araujo da Silva in 1999 [26]. Another twenty-eight populations were analyzed along with these in order to ascertain genetic relationships. These are all extracted from published information as indicated in Table 1.

Table 1. Populations included

Population	3 Letter Code*	N	Location	Reference
Andalucia	Anc	120	Southern province of Spain	[27]
Andalusian	Ans	147	Southern province of Spain	[16]
ArabMoslem	Ara	94	Gaza Strip, Judaea, Samaria, Israel	[28]
Australia	Aus	250	Victoria	[29]
Bahama	Bah	88	Throughout Bahamas	[4]
Bari	Bar	24	NE Colombia, S. America	[4]
Basque	Bas	257	North Central Spain	[28]
Benin	Ben	100	S. Benin, West Africa	Present Study
BWHAlaska	BWH	109	Bethel-Wade Hampton, Alaska	[30]
Cameroon	Cam	34	Yaounde City	[26]
Cameroon2	Ca2	16	SW of Yaounde City, W. Africa	Present Study
CanaryIsl	Can	123	General population from Canary Islands	[16]
Chimila	Chm	46	NE Colombia, S. America	[4]
China (Han)	Chn	216	Xian & Shijiazhuang districts of China	[28]
Congo	Con	34	Lubumbashi City	[26]
Denmark	Den	210	General population from Denmark	[28]
DubaiArab	Dub	93	United Arab Emirates	[31]
Egypt	Tan	100	Tanta, N. Africa	Present Study
Galicia	Gal	149	NW province of Spain	[28]
Greece	Gre	107	Cyprus	[32]
Haiti	Hai	83	Caribbean	[28]
Kenya	Ken	106	NE of Nairobi, E. Africa	Present Study
Korea	Kor	116	Seoul	[28]
MapucheArg	Map	61	Argentina, S. America	[33]
Navajo	Nav	28	New Mexico, USA	[4]
Nigeria	Nig	67	W. Africa	[28]
N.SlopeAlas	NSA	92	North Slope Borough, Alaska	[30]
Philippines	Phi	103	Metro Manila	[34]
Rwanda	Rwa	100	E. Africa	Present Study
SaudiArabia	Sau	220	Riyadh	[35]
Taiwan (Han)	Tai	105	General population of Taiwan	[28]
Turkey	Tur	112	General population of Turkey	[36]
Zimbabwe	Zim	101	Mashonaland Province	[28]
Totals	33	3611		

*These codes are used in Table 6.

Statistical Analysis

The gene counting method [11] was used to generate genotypic and allelic frequencies. Allelic frequencies were analyzed with PHYLIP 3.5 [37], Numerical Taxonomy and Multivariate Analysis System (NTSys) principal component [38], and Carmody's G-test [39]. The depiction of homogeneous clustering within PHYLIP and NTSys suggests adequate population sampling. The SEQBOOT, CONSENSE, GENDIST, CONTML, UPGMA, NJ and ML programs are all within the PHYLIP software. Phylogenetic trees were used to visualize the general phylogenetic relationships between the populations and as a proof of the integrity of the data. Three distance-based analyses were used to generate dendrograms, the Unweighted Pair Group Method with Averages (UPGMA), Neighbor Joining (NJ) and Maximum Likelihood (ML). UPGMA assumes that the rates of mutation are equal across time for all lineages [40]. Nei genetic distances were calculated through the use of the GENDIST program. This tree was then created with the UPGMA program, 1000 bootstrap replications were generated with SEQBOOT and CONSENSE configured a consensus tree. The NJ method, on the other hand, produces an un-rooted network. When rooted, a constant rate of evolution is assumed. First, GENDIST was used to calculate Nei's genetic distances. The NJ program was then run, SEQBOOT generated 1000 bootstrap replications and a consensus tree was built with CONSENSE. The NJ method offers rapid analysis with large data sets. The final tree was done with the ML method, which does not assume constant evolutionary rates, thereby enhancing the validity of the relationships in the other trees. SEQBOOT produced a large number (1000) of bootstrapped data sets that

were then inputted into the CONTML program. The CONTML and CONSENSE programs configured the best-fit tree directly from the allele frequencies. The principal component (PC) test was performed to generate a two dimensional plot of PC 1 and 2. In PC analysis, allelic frequencies are inputted into the NTSys program to segregate and cluster groups within by plotting according to their variability. PCs 1 and 2 usually contain the most amount of variability, with subsequent PCs decreasing in value. The G-test calculates likelihood ratios according to the obtained frequencies. It contrasts the observed and expected frequencies [15]. One thousand simulations were done for each 2 x 2 contingency table. The G-test analyzes each 2 x 2 table and indicates which populations are homogeneous with each other. Probability values less than 0.050 indicate significantly different populations.

In addition to these phylogenetic and statistical analyses, several other parameters were computed to describe the five African populations. Expected homozygosity and heterozygosity values were compared to those reported in the literature. Conformance to Hardy-Weinberg equilibrium was tested by use of the Fisher exact and Chi-Square tests within the Genetic Data Analysis (GDA) software [41]. This software was also used to calculate theta-P values which are equivalent to F_{st} values [41]. This analysis estimates the probability that two alleles are identical by descent [42]. It is an attempt to offset the substructure effects on heterozygote genotypic frequencies [43]. Assumptions are no mutation rate and the sampling of an effective population size. If these assumptions are met, these values range between 0 and 1. This measure indirectly indicates gene flow [44].

RESULTS

Table 2 shows the allelic frequencies for the five African populations. The 24 allele was most frequent in the Benin, Kenya, Rwanda and Tanta populations with frequencies of 0.2150, 0.2028, 0.2200 and 0.4150, respectively. In the Cameroon population the 34 allele was most common with a frequency of 0.3125. African groups have been shown to have the highest frequencies for the 24, 28 and 34 alleles [45]. All five populations shared alleles 17, 20, 21, 23, 24, 25, 28, 31 and 34. Only within the Rwandan population was allele 16 detected and allele 36 was found in a single Cameroon sample.

There were 16 different alleles and 50 different genotypes found in the Benin population. Within the Kenyan and Rwandan groups, 17 various alleles and 49 various genotypes were distinguished. The Tanta samples exhibited 18 alleles and 40 genotypes. There were 11 alleles and 14 genotypes discerned in the Cameroon samples. All genotypic frequencies are contained in Table 3. Within the Benin population, the 22/24 genotype was most frequent representing 6% of the samples. In the Kenya population, 24/28 was the most recurring genotype representing 7.55% of the total. Both 24/34 and 17/24 were most abundant within the Rwandan samples, each with a frequency of 6%. The homozygous genotype, 24/24, was highly repetitive within the Tanta population, displaying a frequency of 18%. Cameroon also had two most recurrent genotypes. These were 21/34 and 25/34, both with a frequency of 12.5%. The number of alleles and genotypes observed were not exceptionally high compared to the number of genotypes that are possible. Utilizing the formula $G=n(n+1)/2$, the most genotypes attainable for

Table 2. Allelic Frequencies

Alleles	Benin (N=100)	Kenya (N=106)	Rwanda (N=100)	Tanta (N=100)	Cameroon2 (N=16)
14
15
16	0.0100
17	0.0550	0.0283	0.1050	0.0150	0.0625
18	0.0550	0.0613	0.0900	0.1400	...
19	0.0050	0.0050	...
20	0.0550	0.0236	0.0100	0.0150	0.0313
21	0.1100	0.0991	0.1000	0.0300	0.1563
22	0.1000	0.0896	0.1200	0.0450	...
23	0.0300	0.0283	0.0050	0.0350	0.0313
24	0.2150	0.2028	0.2200	0.4150	0.2188
25	0.0650	0.0377	0.0300	0.0600	0.0625
26	0.0050	0.0142	...	0.0300	...
27	0.0100	0.0236	0.0550	0.0200	...
28	0.1100	0.1981	0.0700	0.0650	0.0313
29	0.0200	0.0330	0.0100	0.0500	...
30	0.0100	0.0050	...
31	0.0800	0.0330	0.0400	0.0400	0.0313
32	0.0050	0.0283	0.0200	0.0050	...
33	...	0.0047	...	0.0050	0.0313
34	0.0750	0.0802	0.1000	0.0200	0.3125
35
36	0.0313
37
38
39
40	...	0.0142	0.0100
41

...=allele not detected; N=total number of individuals

Theta-P values: Benin=0.032, Kenya=0.027, Rwanda=0.029, Tanta=0.009, Cameroon2=0.021

Table 3. Genotypic Frequencies

Genotype	Benin		Kenya		Rwanda		Tanta		Cameroon2		Genotype	Benin		Kenya		Rwanda		Tanta		Cameroon2	
	N	Freq	N	Freq	N	Freq	N	Freq	N	Freq		N	Freq	N	Freq	N	Freq	N	Freq	N	Freq
16/16		1	0.010		20/24	2	0.0200	2	0.018	1	0.0100	2	0.0200		...
17/17	3	0.0300		...	3	0.030		20/25	2	0.0200	
17/18		1	0.010		20/29	1	0.0100		...	1	0.0100	
17/21		3	0.030		20/31	1	0.0100	1	0.009	
17/22	2	0.0200	1	0.0094	2	0.020		20/32	1	0.0100	
17/24	1	0.0100	2	0.0189	6	0.060	2	0.0200	1	0.0625	20/33		1	0.0625
17/25		1	0.010		20/34		...	2	0.018	
17/28		...	2	0.0189		21/21	1	0.0100		1	0.0100	1	0.0625
17/29		...	1	0.0094		...	1	0.0100		...	21/22	3	0.0300	3	0.028	5	0.0500	
17/31	1	0.0100		21/24	5	0.0500	5	0.047	5	0.0500		...	1	0.0625
17/34	1	0.0100		...	2	0.020		...	1	0.0625	21/25		...	2	0.018	
18/18	1	0.0100		...	2	0.020	5	0.0500		...	21/27		...	1	0.009	
18/21		...	2	0.0189	2	0.020		21/28	5	0.0500	6	0.056	2	0.0200	
18/22		...	1	0.0094	3	0.030		21/29		2	0.0200	
18/23	1	0.0100		21/31	1	0.0100		...	1	0.0100	
18/24	4	0.0400	3	0.0283	4	0.040	11	0.1100		...	21/32		...		1	0.0100		
18/25	1	0.0100	1	0.0094		...	2	0.0200		...	21/34	5	0.0500	2	0.018	2	0.0200	1	0.0100	2	0.1250
18/27		1	0.010	1	0.0100		...	22/22	3	0.0300		1	0.0100		
18/28	2	0.0200	4	0.0377		...	3	0.0300		...	22/23		...	1	0.009	
18/29		1	0.0100		...	22/24	6	0.0600	2	0.018	4	0.0400	4	0.0400		...
18/31	1	0.0100		...	1	0.010		22/25	1	0.0100		...	2	0.0200	
18/32		...	1	0.0094		22/27		...		3	0.0300	1	0.0100		...	
18/34		...	1	0.0094	2	0.020		22/28	1	0.0100	5	0.047	1	0.0100	
19/22		1	0.010		22/29		...	2	0.018	
19/24		1	0.0100		...	22/31		...	2	0.018		...	2	0.0200		...
20/20	1	0.0100		22/32		...		1	0.0100		
20/21	1	0.0100		22/34		...	2	0.018	1	0.0100	
20/22	1	0.0100		...	1	0.010		23/23	1	0.0100	

Table 3. Genotypic Frequencies (Con't)

Genotype	Benin		Kenya		Rwanda		Tanta		Cameroon2		Genotype	Benin		Kenya		Rwanda		Tanta		Cameroon2	
	N	Freq	N	Freq	N	Freq	N	Freq	N	Freq		N	Freq	N	Freq	N	Freq	N	Freq	N	Freq
23/31		...	1	0.0094	1	0.0100					26/31		...			1	0.0100				
23/32		...	1	0.0094							27/27		...			1	0.0100	1	0.0100		
23/34	1	0.0100	1	0.0094							27/33		...	1	0.009						
24/24	2	0.0200	3	0.0283	3	0.0300	18	0.1800	1	0.0625	27/34	1	0.0100	1	0.009	2	0.0200				
24/25	2	0.0200	3	0.0283	2	0.0200	4	0.0400			28/28	1	0.0100	4	0.037	2	0.0200	1	0.0100		
24/26	1	0.0100	2	0.0189		...	2	0.0200			28/29		...	3	0.028		...	1	0.0100		
24/27	1	0.0100	2	0.0189	3	0.0300					28/30	2	0.0200				...	1	0.0100		
24/28	5	0.0500	8	0.0755	1	0.0100	5	0.0500			28/31	1	0.0100			1	0.0100				
24/29	2	0.0200	1	0.0094	2	0.0200	3	0.0300			28/32		...			1	0.0100				
24/31	5	0.0500	1	0.0094	3	0.0300	2	0.0200			28/34		...	2	0.018	3	0.0300		...	1	0.0625
24/32		1	0.0100			28/40		...	1	0.009						
24/33		1	0.0100			29/34	1	0.0100								
24/34	3	0.0300	4	0.0377	6	0.0600	2	0.0200	1	0.0625	31/31	1	0.0100								
24/36		1	0.0625	31/32		...	2	0.018	1	0.0100			
24/40		...	2	0.0189	1	0.0100					31/34	1	0.0100				...	1	0.0100	1	0.0625
25/25		1	0.0100	1	0.0100			31/40		...			1	0.0100				
25/26		...	1	0.0094		...					32/34		...	2	0.018						
25/28	4	0.0400	1	0.0094	1	0.0100	1	0.0100			34/34	1	0.0100			1	0.0100		...	1	0.0625
25/31	3	0.0300		1	0.0100													
25/34		2	0.1250											
26/26		1	0.0100													
26/29		1	0.0100													

...=allele not detected; N=Number of genotypes

each population is calculated to be 136 for Benin, 153 for the Kenyan and Rwandan populations, 171 for Tanta and 66 for Cameroon.

The observed homozygosity for Benin, Cameroon, Kenya, Rwanda and Tanta were 0.1500, 0.1880, 0.0660, 0.1500 and 0.2800, respectively. The corresponding heterozygosity values were 0.8500, 0.8130, 0.9340, 0.8500 and 0.7200 (Table 4). These values did not vary greatly from the expected ones. These high levels of heterozygosity are consistent with genetic diversity within African populations at other loci [46].

Figure 1 shows an NJ tree using Nei distances [47]. Four main clusters of populations can be discerned. Native Americans are seen in a cluster at one end of the tree. Caucasians are grouped in the adjacent clade. The Mapuche Argentin population segregated with the Caucasians. Hutz and collaborators also placed this population with Caucasians [33]. The following clade exhibits all the African groups as well as the Saudi Arabia and Moslem Arab populations. The last clade clusters the Oriental populations.

In the UPGMA depiction (Figure 2), there are two main branches after the bifurcation between the Andalusian population sample and all the other groups. These two branches have Denmark and Australia versus the rest of the populations in our study. Five clusters can be found here. Dubai Arabs head off the first cluster that incorporates other southwest Asian populations along with two southeast European populations, Tanta and Mapuche Argentina. Once again, Tanta appears to be most closely related to those countries across the adjoining Mediterranean Sea. This would suggest that southern European and southwest Asian countries neighboring the Mediterranean Sea contributed greatly to the genetic constitution of this group. The next cluster holds

Figure 1. Neighbor Joining Tree

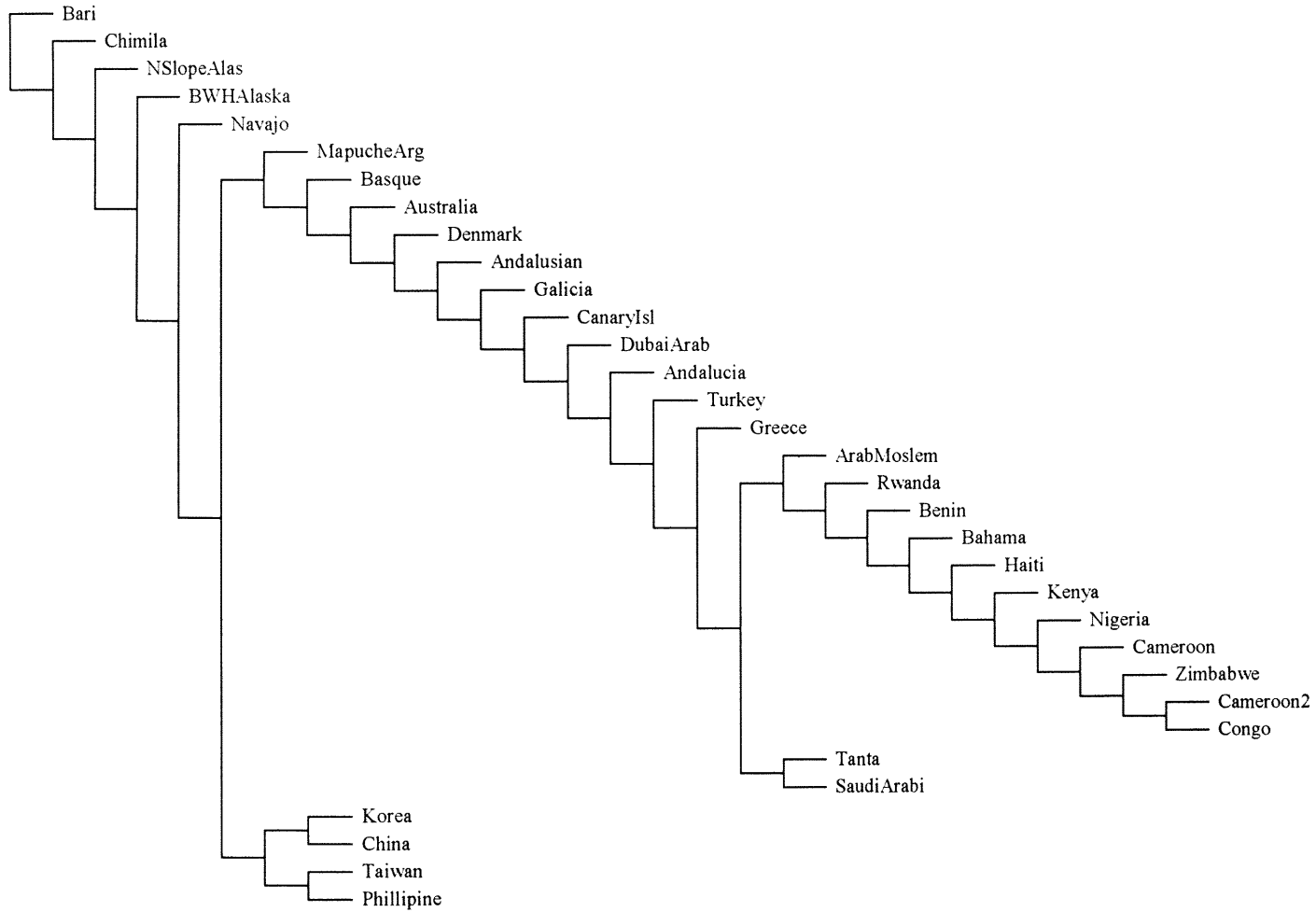
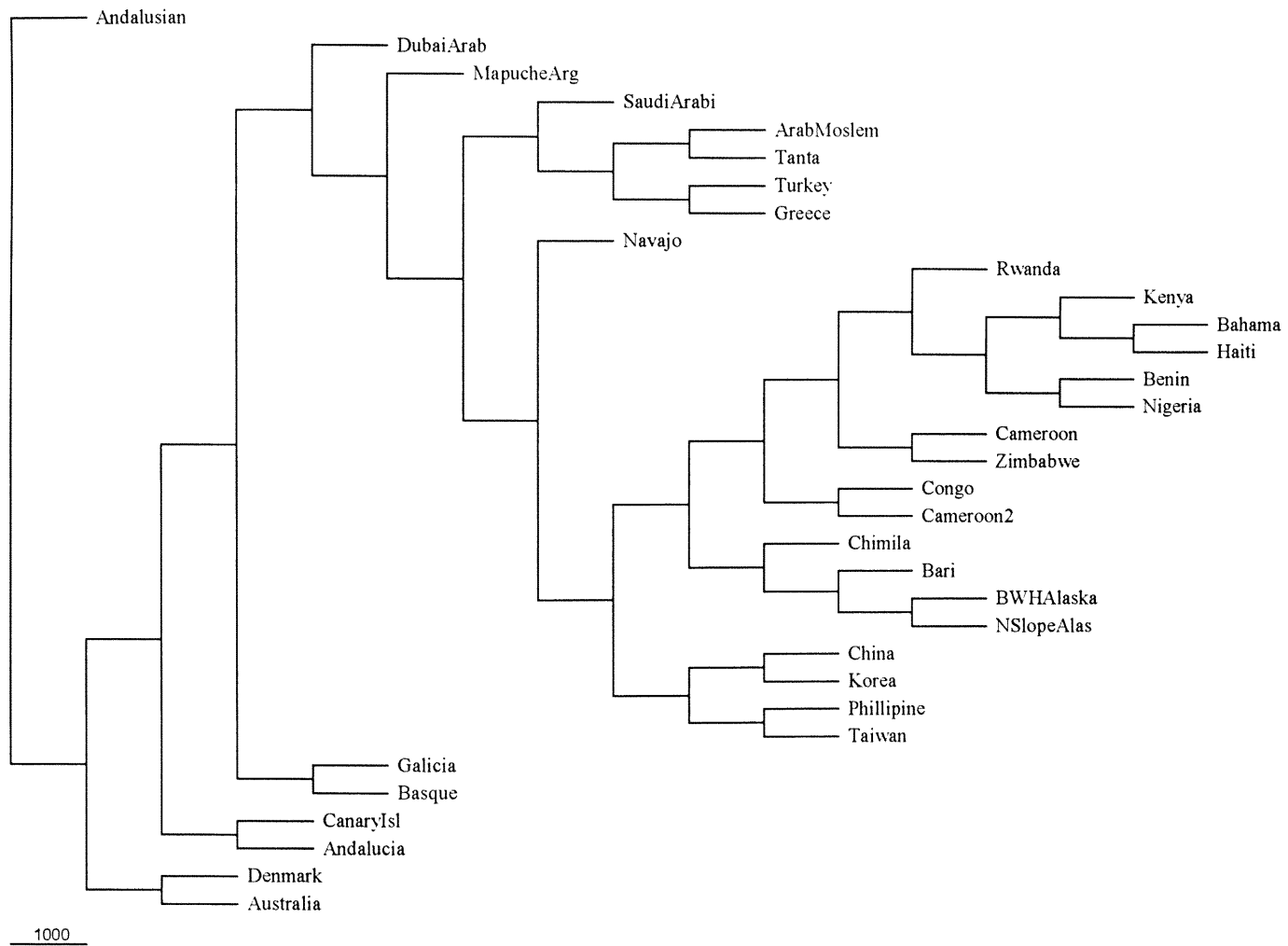


Figure 2. UPGMA Tree



the Navajo population branched with the Africans and other Mongoloids. Lastly, the rest of the Caucasians are depicted.

Figure 3 shows a radial representation of the Maximum Likelihood Tree. This tree's primary bifurcation is between the southwest Asian populations and all other populations. All of the African groups are clustered together along with the Greek population. The Caucasians also segregate together. The Mongoloids make up the next assemblage. The bootstrap values in this figure indicate the likelihood that the depicted relationships are statistically significant. Bootstrap values greater than 50 percent (indicated as 500 or greater in the figure) are considered credible indicators of phylogenetic relationships.

Figure 4 displays the variations in all 27 alleles amongst the 33 populations included in the study. Each allele frequency is standardized upon input into the NTSys program. Two by two contingency tables were generated and eigenvalues and eigenvectors were calculated and then plotted. PC 1 indicated 21% of the variability while 14% was held in PC 2. PC 1 shows delineation between African populations and all other populations. PC 2 separates west African populations from the rest of the African groups and the southeast Asians from the American Indian groups. There is also a separation between the southwest Asians and European groups.

The data represented in Table 5 indicates the results of tests for adherence to Hardy-Weinberg equilibrium expectations. A Fisher exact test, along with a Chi-Square test was performed. The Benin population did not conform to Hardy-Weinberg equilibrium expectations according to the Fisher exact test. Its probability (p) value was

0.019. The Rwanda population was also shown to be marginally out of Hardy-Weinberg equilibrium by the Chi-Square test, with a p-value of 0.036.

The G-test analysis of the 33 different populations (Table 6) provided data consistent with the other statistical tests previously mentioned. In the 2 x 2 comparisons, statistically insignificant values and low G-scores indicate homogeneity among populations, while statistically significant values with high G-scores are indicative of distinctiveness. G-score p-values are significant below 0.05 and indicate non-homogeneity among populations. Generally, groups within the major races were not found to be significantly different. Notable exceptions were the Tanta population when compared to the sub-Saharan African groups and the Mapuche and Navajo populations when compared to other Native Americans. No significant differences were seen between the Tanta population and the Caucasoid groups. Similarly, the Mapuche and Navajo populations were not significantly different when compared to the Caucasian groups.

Figure 3. Maximum Likelihood Tree with Bootstrap Values

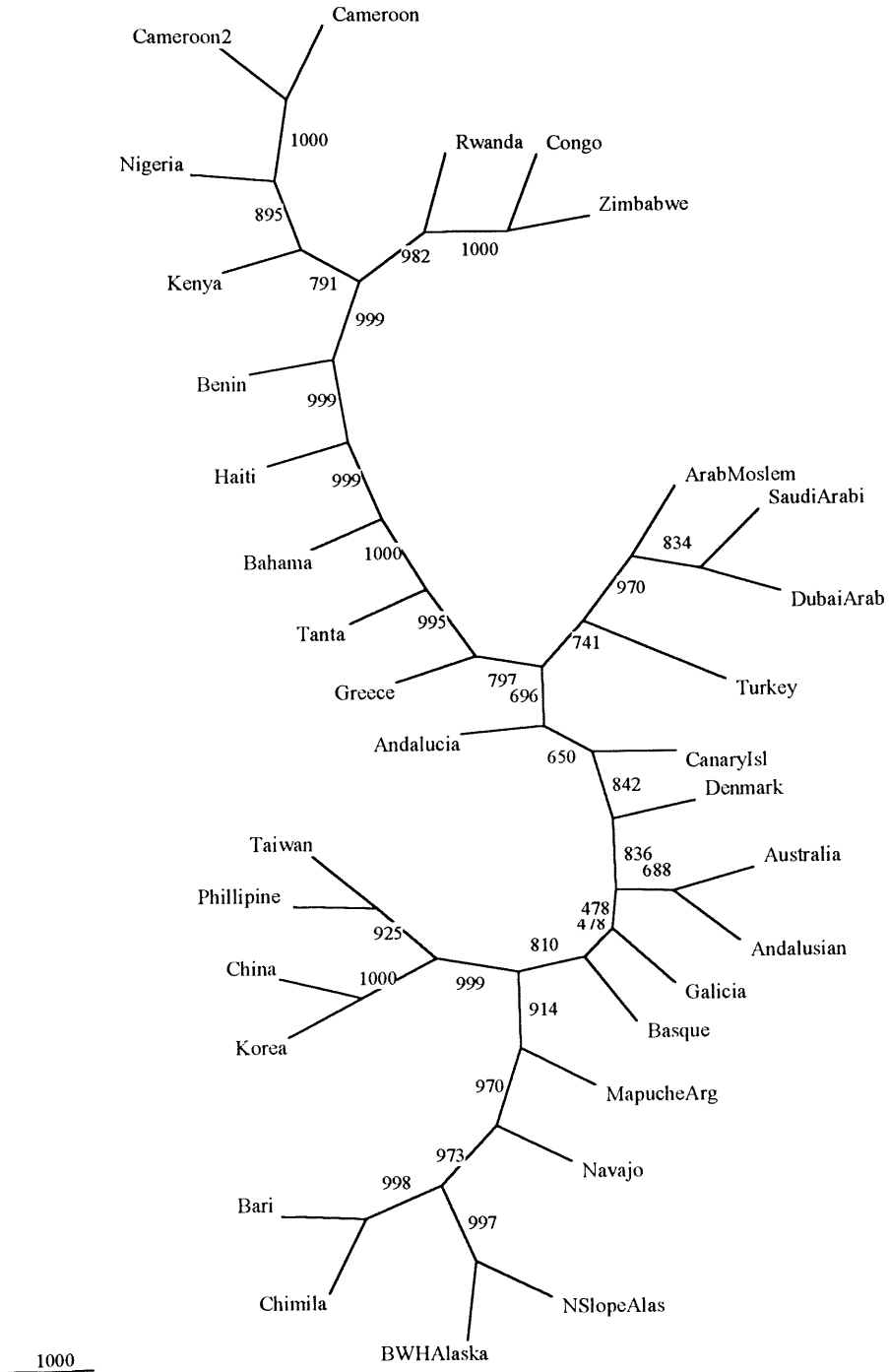


Figure 4. Principal Component Plot

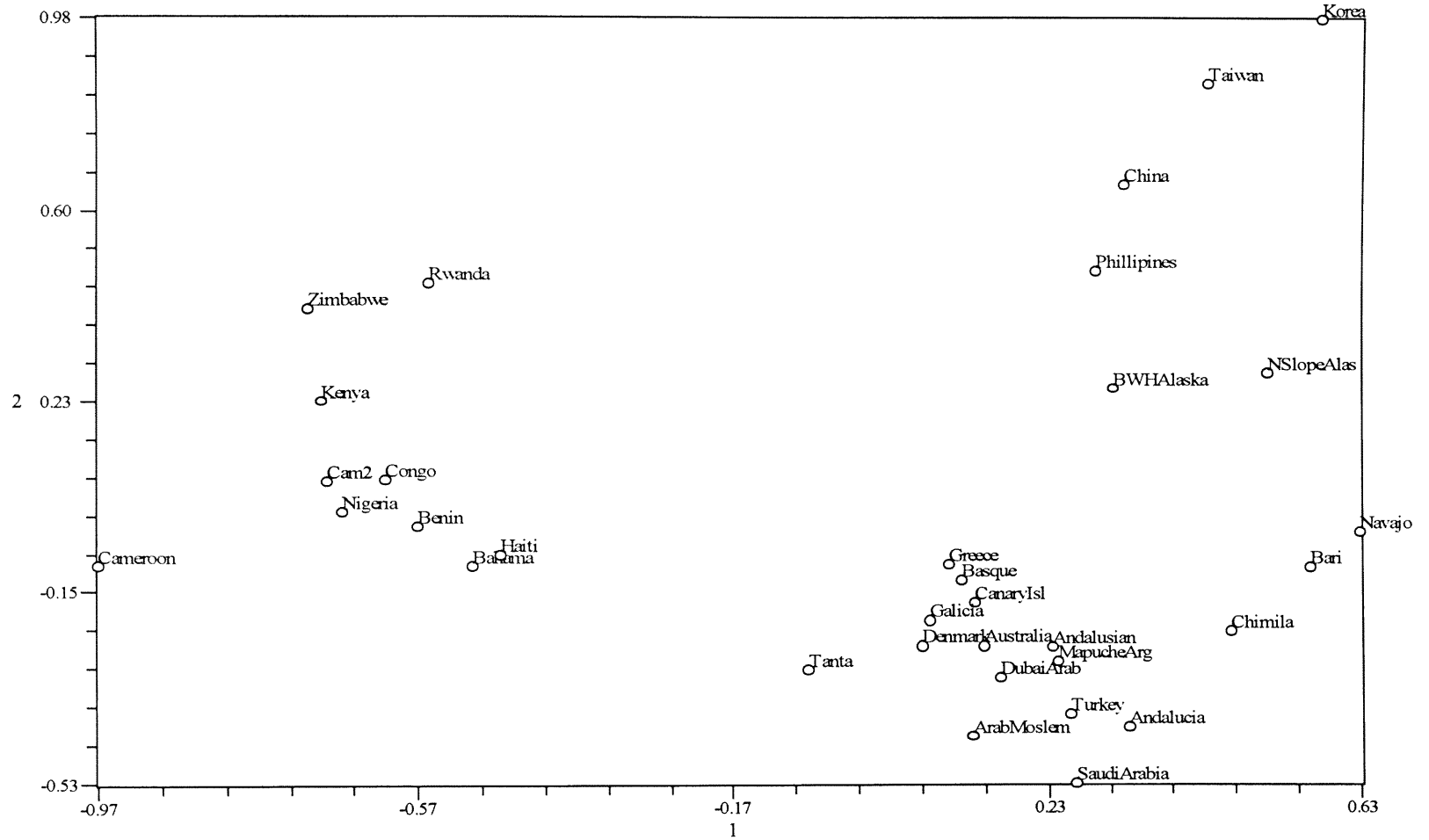


Table 4. Observed/Expected homozygosity and heterozygosity values

Population	N	Observed	Expected	Observed	Expected
		Homozygosity %	Homozygosity Unbiased (%)	Heterozygosity %	Heterozygosity Unbiased (%)
Benin	100	15	10.28	85	89.72
Kenya	106	6.6	11.18	93.4	88.82
Rwanda	100	15	10.87	85	89.13
Tanta	100	28	20.62	72	79.38
Cameroon	16	18.8	15.73	81.3	84.27

Table 5. Test for Hardy-Weinberg Equilibrium and Power of Discrimination

Population	N	Fisher Exact	Chi Square	Power of Discrimination
		(3200 shufflings) (p)	(3200 shufflings) (p)	%
Benin	100	0.019	0.109	96.98
Kenya	106	0.704	0.212	97.03
Rwanda	100	0.247	0.036	97.06
Tanta	100	0.049	0.377	93.48
Cameroon	16	0.879	0.492	92.19

Table 6. G Scores with probability values between all 33 populations

	Anc	Ans	Ara	Aus	Bah	Bar	Bas	Ben	BWH	Cam	Ca2	Can	Chm	Chn
Andalucia		39.607	37.366	38.200	113.841	110.568	44.118	143.180	152.635	130.007	100.841	26.710	242.916	123.671
Andalusian	0.904		47.932	25.949	93.360	117.457	49.165	113.126	144.274	111.033	77.863	28.462	212.149	131.290
ArabMoslem	0.895	0.732		58.012	75.265	129.990	78.713	102.207	179.643	100.351	74.700	33.556	230.491	146.251
Australia	0.937	1.000	0.322		130.877	118.601	38.410	172.272	162.712	150.267	108.843	30.595	275.305	156.095
Bahama	0.000	0.000	0.008	0.000		151.487	140.302	17.390	162.303	31.186	43.138	98.566	231.274	163.705
Bari	0.000	0.000	0.000	0.000	0.000		103.104	167.631	65.014	141.848	106.582	113.677	65.327	117.950
Basque	0.803	0.744	0.012	0.971	0.000	0.001		178.190	141.540	152.011	110.707	42.610	274.587	152.081
Benin	0.000	0.000	0.000	0.000	1.000	0.000	0.000		176.910	31.453	37.415	121.300	251.940	218.647
Bethel-Wade	0.000	0.000	0.000	0.000	0.000	0.009	0.001	0.000		162.327	111.761	160.778	171.059	144.319
Cameroon	0.000	0.000	0.000	0.000	0.967	0.000	0.000	0.958	0.000		21.447	119.943	180.598	163.019
Cameroon2	0.000	0.000	0.000	0.000	0.677	0.000	0.000	0.812	0.000	1.000		94.847	121.522	120.043
CanaryIsland	0.996	0.998	0.963	0.995	0.000	0.000	0.888	0.000	0.000	0.000	0.000		233.634	109.141
Chimila	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000		279.437
China	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	
Congo	0.000	0.000	0.000	0.000	0.176	0.000	0.000	0.425	0.000	0.934	0.993	0.000	0.000	0.000
Denmark	0.994	0.995	0.294	1.000	0.000	0.000	0.993	0.000	0.000	0.000	0.000	0.997	0.000	0.001
DubaiArab	0.674	0.723	0.975	0.575	0.000	0.000	0.157	0.000	0.000	0.000	0.000	0.935	0.000	0.000
Galicia	0.902	0.990	0.146	0.976	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.998	0.000	0.001
Greece	0.814	0.629	0.830	0.112	0.000	0.000	0.424	0.000	0.000	0.000	0.000	0.999	0.000	0.000
Haiti	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.992	0.773	0.000	0.000	0.000
Kenya	0.000	0.000	0.000	0.000	0.995	0.000	0.000	0.957	0.000	0.975	0.645	0.000	0.000	0.000
Korea	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.998
Mapuche	0.714	0.773	0.661	0.827	0.000	0.000	0.689	0.000	0.001	0.000	0.000	0.565	0.000	0.034
Navajo	0.366	0.697	0.112	0.753	0.001	0.004	0.588	0.000	0.166	0.000	0.004	0.289	0.000	0.079
Nigeria	0.000	0.000	0.009	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.960	0.000	0.000	0.000
NorthSlope	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.319	0.000	0.000	0.000	0.000	0.000
Philippines	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.243
Rwanda	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.566	0.000	0.349	0.678	0.000	0.000	0.000
SaudiArabia	0.054	0.001	0.329	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.005	0.000	0.001
Taiwan	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.002	0.000	1.000
Tanta	0.944	0.978	0.995	0.843	0.134	0.000	0.799	0.005	0.000	0.002	0.032	0.966	0.000	0.000
Turkey	0.999	0.694	0.999	0.751	0.000	0.000	0.247	0.000	0.000	0.000	0.000	0.983	0.000	0.000
Zimbabwe	0.000	0.000	0.000	0.000	0.227	0.000	0.000	0.546	0.000	0.922	0.891	0.000	0.000	0.000

* G Scores are read with the vertical population columns at the side named first then the horizontal population columns at the top.

Probability values are read with the horizontal population columns at the top first, followed by the vertical population columns at the side.

Table 6. G Scores with probability values between all 33 populations (Con't)

	Con	Den	Dub	Gal	Gre	Hai	Ken	Kor	Map	Nav	Nig	NSA	Phi	Rwa
Andalucia	146.468	30.276	45.073	40.870	40.924	125.557	153.487	126.254	41.104	49.716	131.656	169.899	108.519	168.469
Andalusian	111.572	31.112	45.934	30.928	47.290	106.217	135.584	138.957	43.099	43.987	109.624	157.981	120.831	142.548
ArabMoslem	97.702	56.502	31.172	62.444	39.725	89.025	99.273	139.397	41.504	56.448	78.158	192.991	120.772	117.153
Australia	171.063	20.131	50.561	34.893	65.816	155.570	191.049	168.103	41.717	42.383	156.745	172.030	147.608	223.569
Bahama	56.588	121.455	97.947	115.624	91.023	9.918	26.481	130.172	84.808	80.533	20.551	227.178	159.925	65.632
Bari	142.303	123.785	112.306	109.924	120.636	170.148	181.735	120.155	76.470	61.784	175.901	59.775	115.429	166.623
Basque	179.871	32.600	62.875	25.339	53.222	167.673	209.024	158.468	45.585	46.337	168.817	159.639	128.989	223.925
Benin	46.469	148.110	128.501	140.188	110.725	18.432	32.126	183.707	107.236	92.984	18.197	229.203	174.025	46.187
Bethel-Wade	187.132	156.341	150.532	165.211	151.856	173.671	221.464	147.171	75.158	51.168	193.508	46.936	160.884	207.418
Cameroon	33.435	136.752	113.046	128.385	105.311	27.216	30.811	137.238	90.682	112.572	22.213	206.225	157.942	50.547
Cameroon2	29.011	101.393	81.840	84.684	93.787	37.972	42.955	100.573	77.951	71.980	32.567	118.487	106.827	41.962
CanaryIsland	119.099	26.296	35.715	29.656	26.248	112.040	126.744	119.247	44.023	51.156	107.602	160.159	81.753	132.269
Chimila	179.004	280.336	228.222	256.023	232.589	249.148	276.121	254.587	147.515	93.234	239.257	186.523	223.738	264.791
China	186.581	152.396	131.487	148.493	117.489	180.447	221.648	32.800	72.374	63.737	194.085	159.601	60.732	230.538
Congo	0.000	161.252	117.265	130.518	108.194	55.783	52.436	152.371	125.675	102.051	35.139	191.858	142.699	38.328
Denmark	0.000	0.604	48.415	29.092	57.369	141.032	179.522	166.526	50.298	47.231	140.014	157.979	130.925	196.173
Dubai-Arab	0.000	0.000	40.832	45.101	110.938	126.084	130.129	45.437	55.877	102.245	145.521	128.256	122.977	157.906
Galicia	0.000	1.000	0.851	52.790	135.106	161.797	137.335	62.296	47.986	134.122	156.447	121.664	157.906	107.389
Greece	0.000	0.260	0.595	0.384	101.656	113.479	123.019	48.039	64.533	95.708	177.959	96.639	107.389	67.002
Haiti	0.127	0.000	0.000	0.000	0.000	21.703	142.693	95.959	93.759	20.347	238.417	171.147	67.002	49.228
Kenya	0.294	0.000	0.000	0.000	0.000	1.000	177.234	119.720	113.156	13.760	282.176	202.894	49.228	185.059
Korea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	96.236	71.044	162.908	171.779	98.165	185.059	136.517
Mapuche	0.000	0.468	0.611	0.121	0.365	0.000	0.000	0.000	33.734	109.913	103.612	81.819	136.517	105.015
Navajo	0.000	0.547	0.132	0.544	0.020	0.000	0.000	0.040	0.770	101.281	53.518	56.840	105.015	30.837
Nigeria	0.926	0.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	234.626	166.791	30.837	228.179
NorthSlope	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.135	0.000	138.991	228.179	173.564
Philippines	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.103	0.000	0.000	173.564	0.000
Rwanda	0.850	0.000	0.000	0.000	0.000	0.047	0.425	0.000	0.000	0.000	0.983	0.000	0.000	0.000
Saudi-Arabia	0.000	0.001	0.361	0.001	0.027	0.000	0.000	0.000	0.001	0.004	0.000	0.000	0.000	0.000
Taiwan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.693	0.072	0.304	0.000	0.000	0.995	0.000
Tanta	0.000	0.971	0.894	0.918	0.989	0.010	0.001	0.000	0.617	0.211	0.017	0.000	0.000	0.000
Turkey	0.000	0.687	0.766	0.433	0.973	0.000	0.000	0.000	0.788	0.145	0.000	0.000	0.000	0.000
Zimbabwe	0.969	0.000	0.000	0.000	0.000	0.415	0.646	0.000	0.000	0.000	0.942	0.000	0.000	0.183

* G Scores are read with the vertical population columns at the side named first then the horizontal population columns at the top.
Probability values are read with the horizontal population columns at the top first, followed by the vertical population columns at the side.

Table 6. G Scores with probability values between all 33 populations (Con't)

	Sau	Tai	Tan	Tur	Zim
Andalucia	65.720	103.030	36.641	26.075	222.281
Andalusian	104.940	103.976	33.689	48.917	193.150
ArabMoslem	51.768	107.591	29.973	27.009	159.421
Australia	128.394	120.909	42.397	46.275	285.404
Bahama	175.476	121.110	59.671	147.989	54.610
Bari	176.081	122.218	133.725	119.068	191.230
Basque	137.352	114.901	44.065	59.029	298.696
Benin	218.823	145.400	75.585	152.219	44.397
Bethel-Wade	296.155	126.507	148.347	117.485	245.616
Cameroon	159.837	133.847	80.703	139.143	35.580
Cameroon2	110.536	97.612	65.794	106.624	36.584
CanaryIsland	76.757	82.041	33.659	32.125	189.462
Chimila	344.003	230.828	232.619	241.360	273.669
China	308.712	23.733	129.085	137.124	272.791
Congo	173.634	136.409	98.742	137.854	31.104
Denmark	99.333	119.295	33.970	46.820	265.064
Dubai:Arab	47.847	120.199	37.462	45.368	207.168
Galicia	97.748	117.993	38.109	54.235	233.191
Greece	67.310	86.571	27.537	32.431	175.021
Haiti	183.290	133.163	71.018	144.069	48.656
Kenya	228.369	166.715	81.392	156.318	43.631
Korea	277.317	48.672	133.840	149.635	213.142
Mapuche	81.088	61.267	45.510	39.556	162.977
Navajo	74.869	50.194	56.984	56.483	119.820
Nigeria	164.973	143.123	70.595	131.079	34.258
NorthSlope	303.174	140.537	179.921	186.407	296.454
Philippines	209.880	27.685	111.666	115.354	208.422
Rwanda	236.932	159.017	100.531	163.188	55.842
Saudi:Arabia		213.788	46.167	49.613	342.655
Taiwan	0.000		96.118	107.290	179.009
Tanta	0.568	0.000		36.962	147.280
Turkey	0.497	0.000	0.937		225.073
Zimbabwe	0.000	0.000	0.000	0.000	

* G Scores are read with the vertical population columns at the side named first then the horizontal population columns at the top. Probability values are read with the horizontal population columns at the top first, followed by the vertical population columns at the side.

DISCUSSION

To aid in our comparison of allelic frequencies, we generated allelic frequency graphs from Table 2 for each of our five African populations (Figures 5-9). Graphs were also created for African American and Caucasian samples from the United States (Figures 14 & 15) as well as the four previously data-banked African populations (Figures 10-13).

The African populations examined, including the African Americans, exhibited alleles 18, 24, 28, 31 and 34, indicating genetic similarity. Exceptions to this include both Cameroon samples (Cameroon at allele 31 and Cameroon2 at allele 18), which may be due to sampling error. The two alleles most prevalent in US Caucasians were the same two that were found most frequently in the Tanta population.

The observed heterozygosity for the five African populations in our study was greater than 70%. These heterozygosity values are indicative of the genetic diversity of this locus. Of the five African populations, the Kenya population appears to be the most diverse with a heterozygosity value of 93.4%, compared to reported heterozygosity of 87.6% in the African American population [48]. The discriminatory power of the D1S80 locus was evident, with values above 90% for all five populations. Previous reports have indicated that the discrimination power for this locus is between 94 and 98% [49].

Of the five populations in our study, only the Benin population was significantly out of conformance to Hardy–Weinberg equilibrium expectations, with a Fisher exact test p-value of 0.019. Preferential amplification of specific alleles does not seem to be an explanation since a low number of samples tested were homozygous (15 of the 100

samples). Our Theta values ranged from 0.009 in the Tanta population sample to 0.032 in the Benin samples. High values of theta indicates greater genetic substructure within a population. The overall theta-P value, indicating the variance present in the combination of all the alleles detected in the 5 African populations, for the D1S80 locus was 0.023. This was comparable to published values of 0.019 [50]. The National Research Council II recommends the use of such a factor in all population studies. Recommendation 4.1 of this council's 1996 report suggests an inbreeding coefficient of 0.01 be used for United States populations, while an increased value of 0.03 would better suit more isolated populations [51].

The topology of the NJ tree illustrates the similarities between different populations by clustering them into clades. The Oriental and American Indian groups were separated from Caucasians and sub-Saharan Africans. Caucasians and sub-Saharan Africans are also separated from each other. This tree further discriminated between east Asians, with the two northern samples (Korea and China) grouped independently of the two southern ones (Taiwan and Philippine), and American Indians. The American Indian groups segregate together even though they may not be geographically close to each other, their origins ranging from North America to South America. The sample from Mapuche Argentina is genetically grouped with the Caucasians. The Tanta and Saudi Arabia populations segregated together indicating the historic influence that Saudi Arabian populations had on Tanta.

The UPGMA tree also clustered the populations based on their genetic similarities. Within this tree, the southeastern Europeans (Turkey and Greece) and the southwestern Asians (Saudi Arabia, Dubai Arab and Arab Moslem) were split from the

Figure 5. Allelic Distribution Graph of the Benin Population

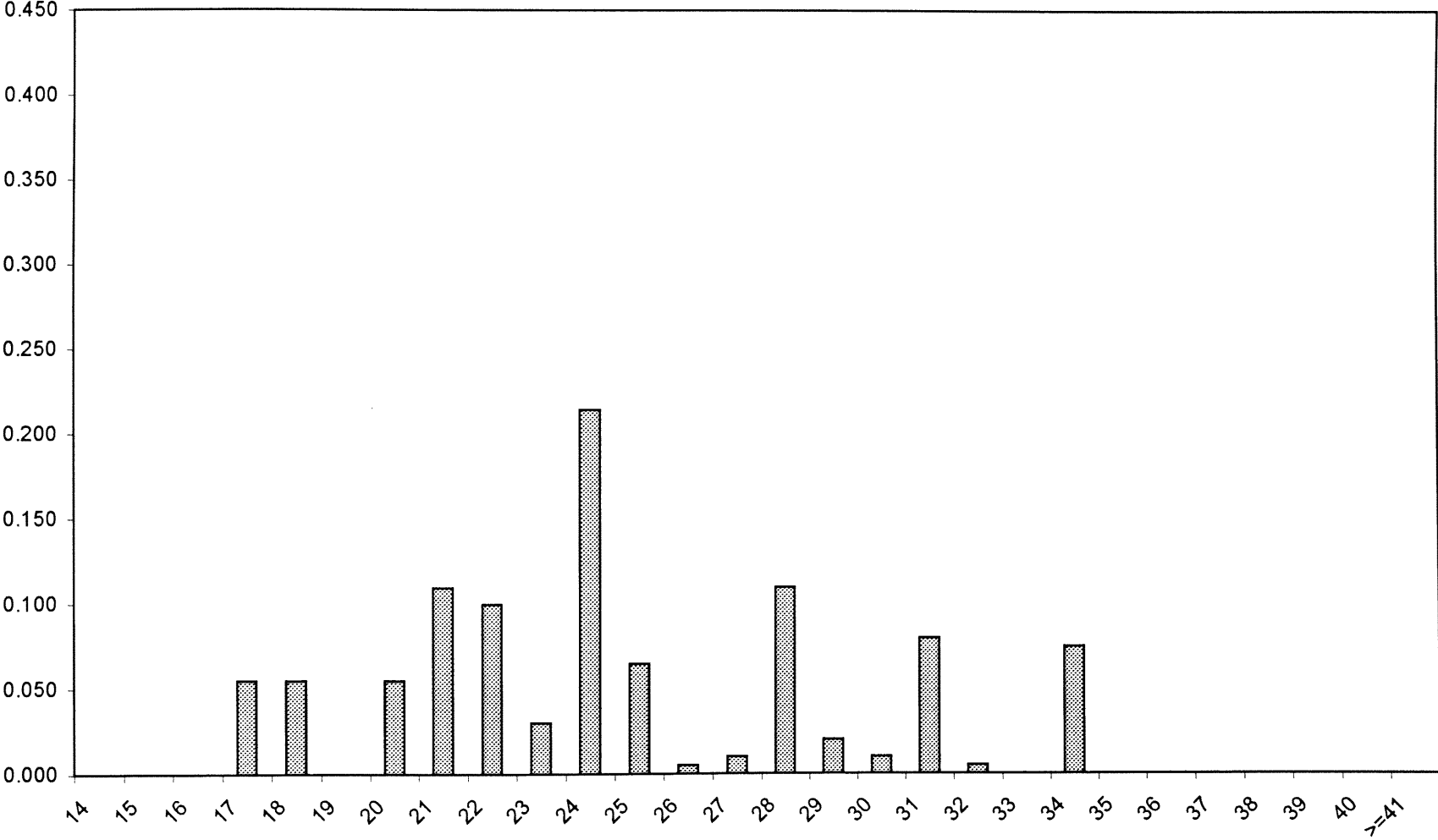


Figure 6. Allelic Distribution Graph of the Kenya Population

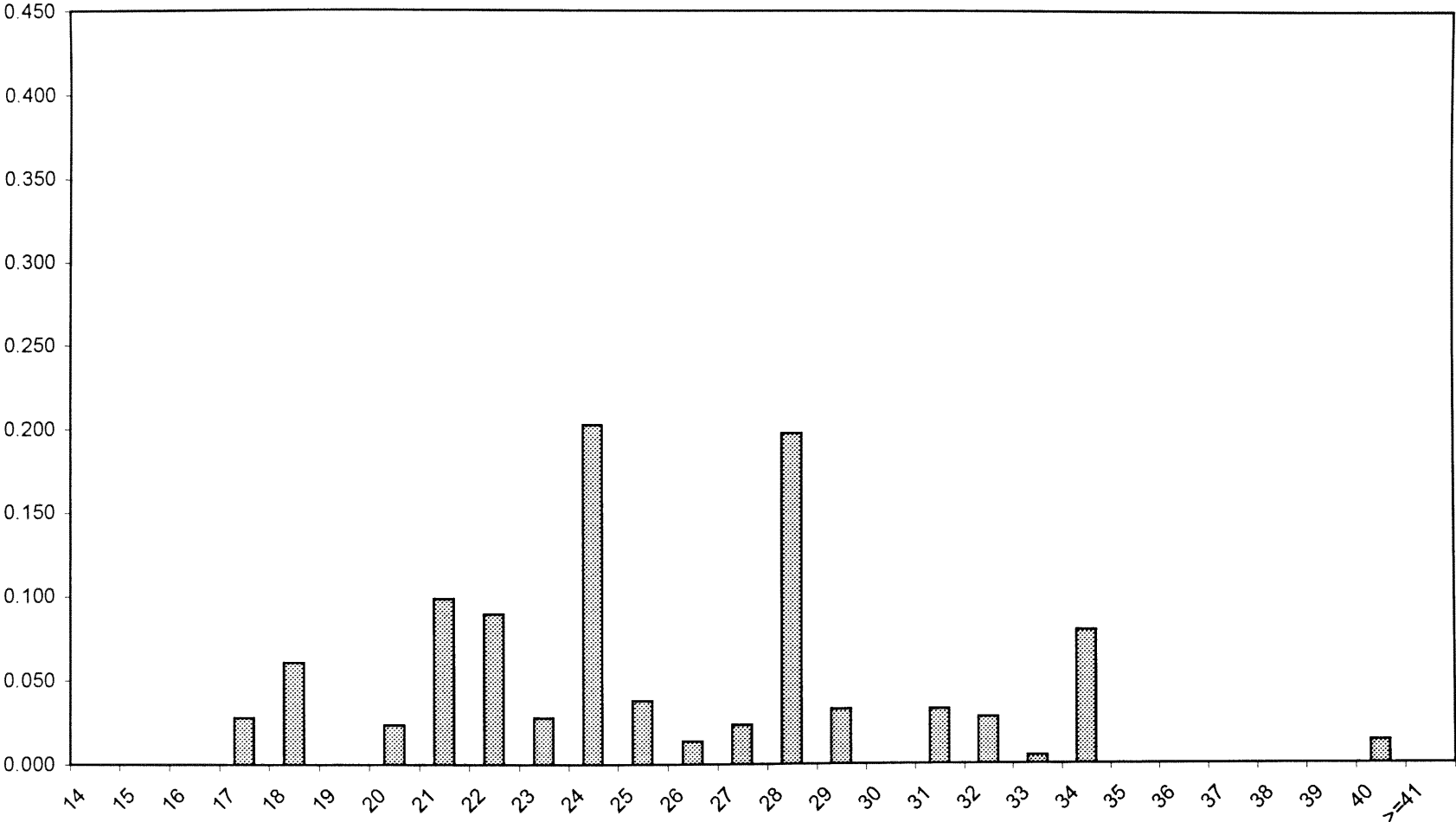


Figure 7. Allelic Distribution Graph of the Rwanda Population

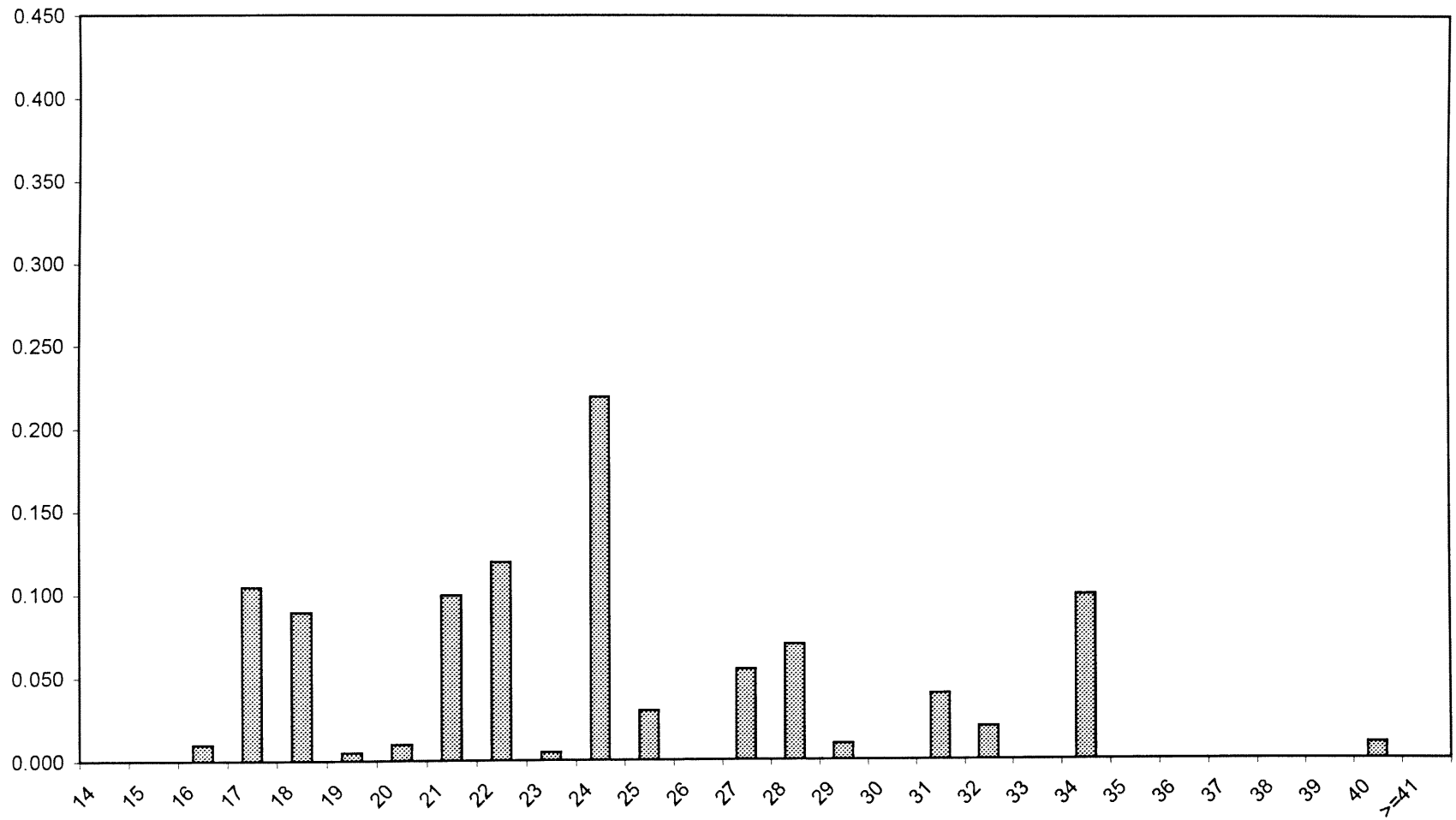


Figure 8. Allelic Distribution Graph of the Tanta Population

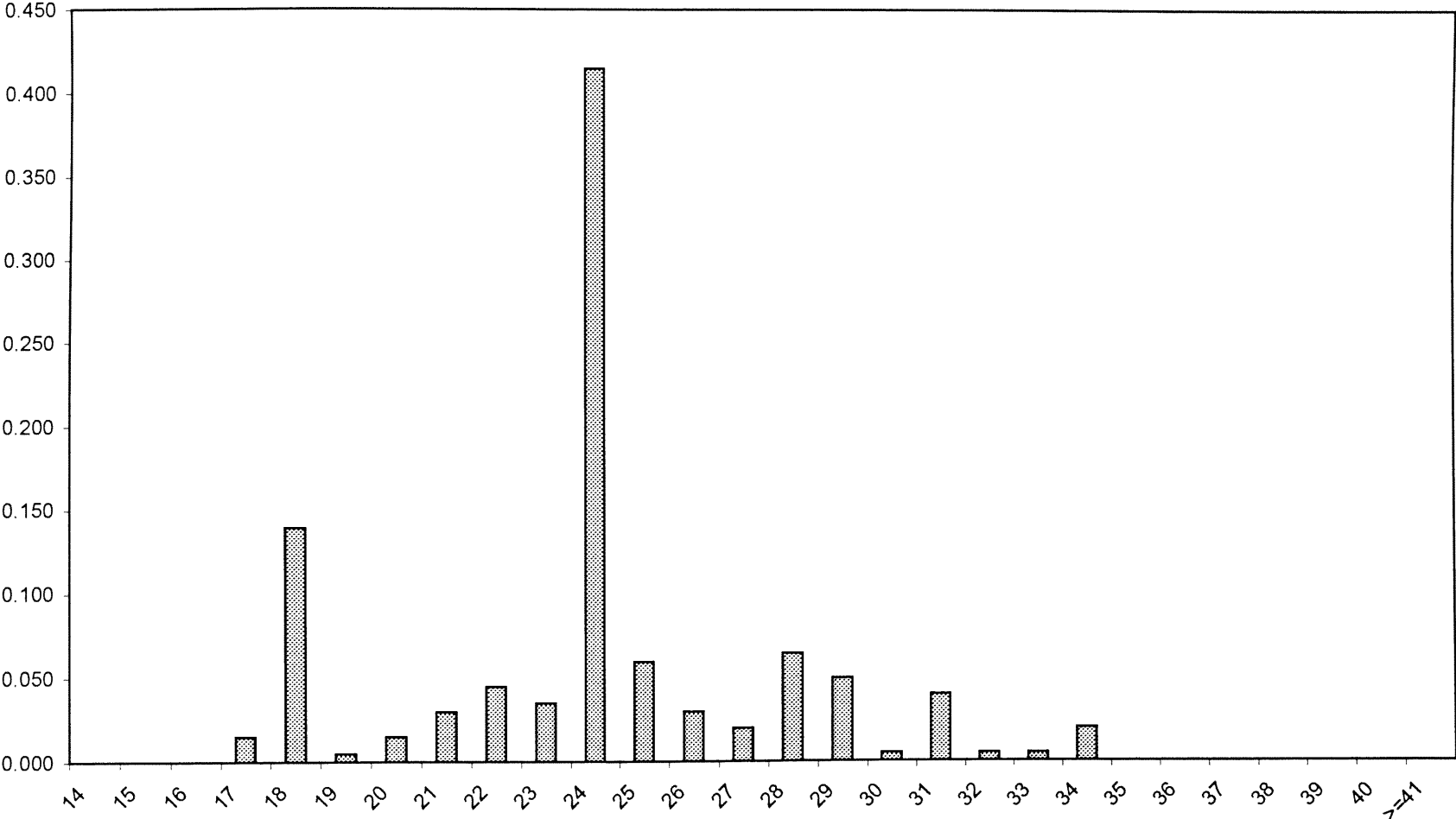


Figure 9. Allelic Distribution Graph of the Cameroon2 Population

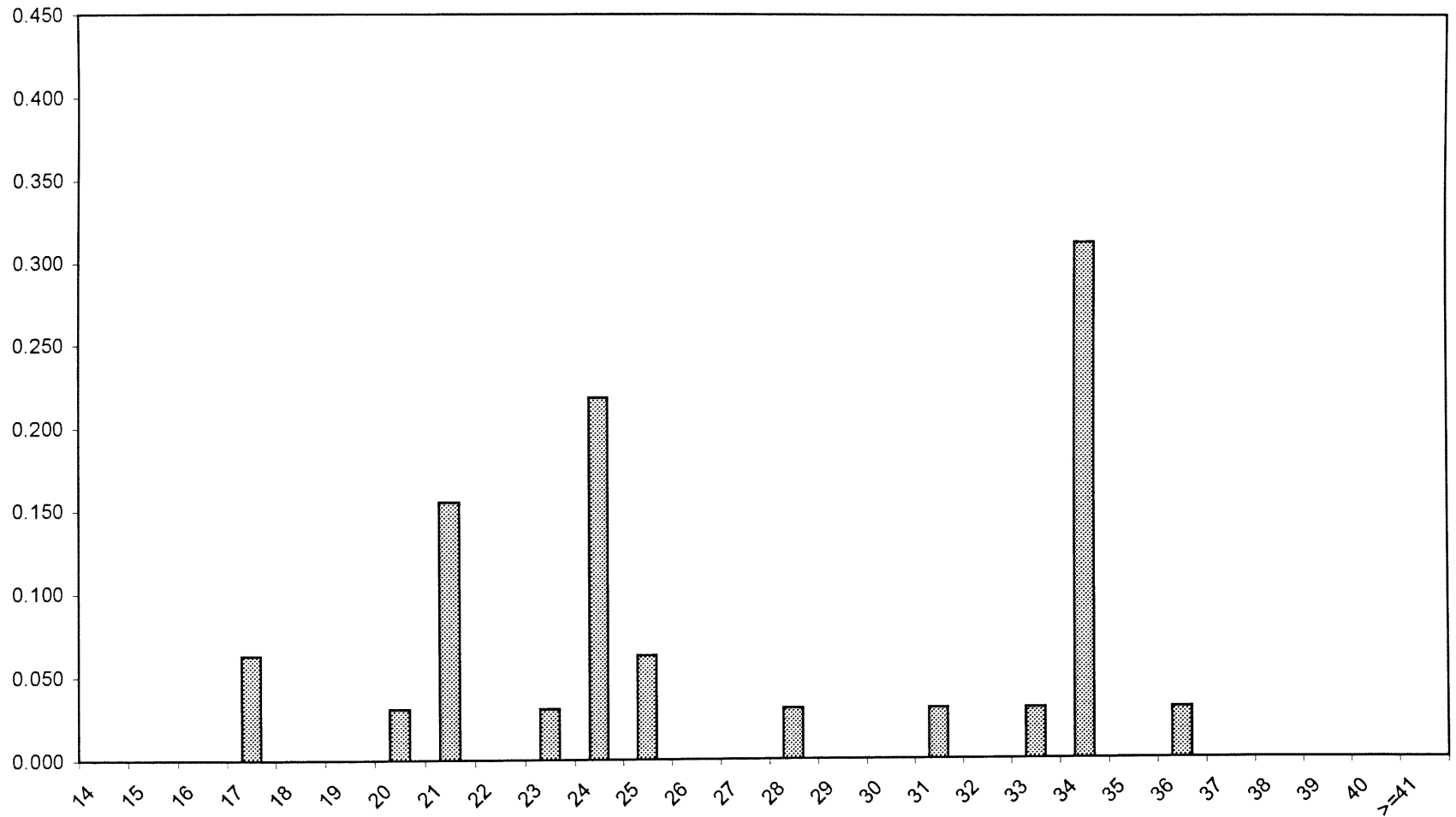


Figure 10. Allelic Distribution Graph of the Cameroon Population [26]

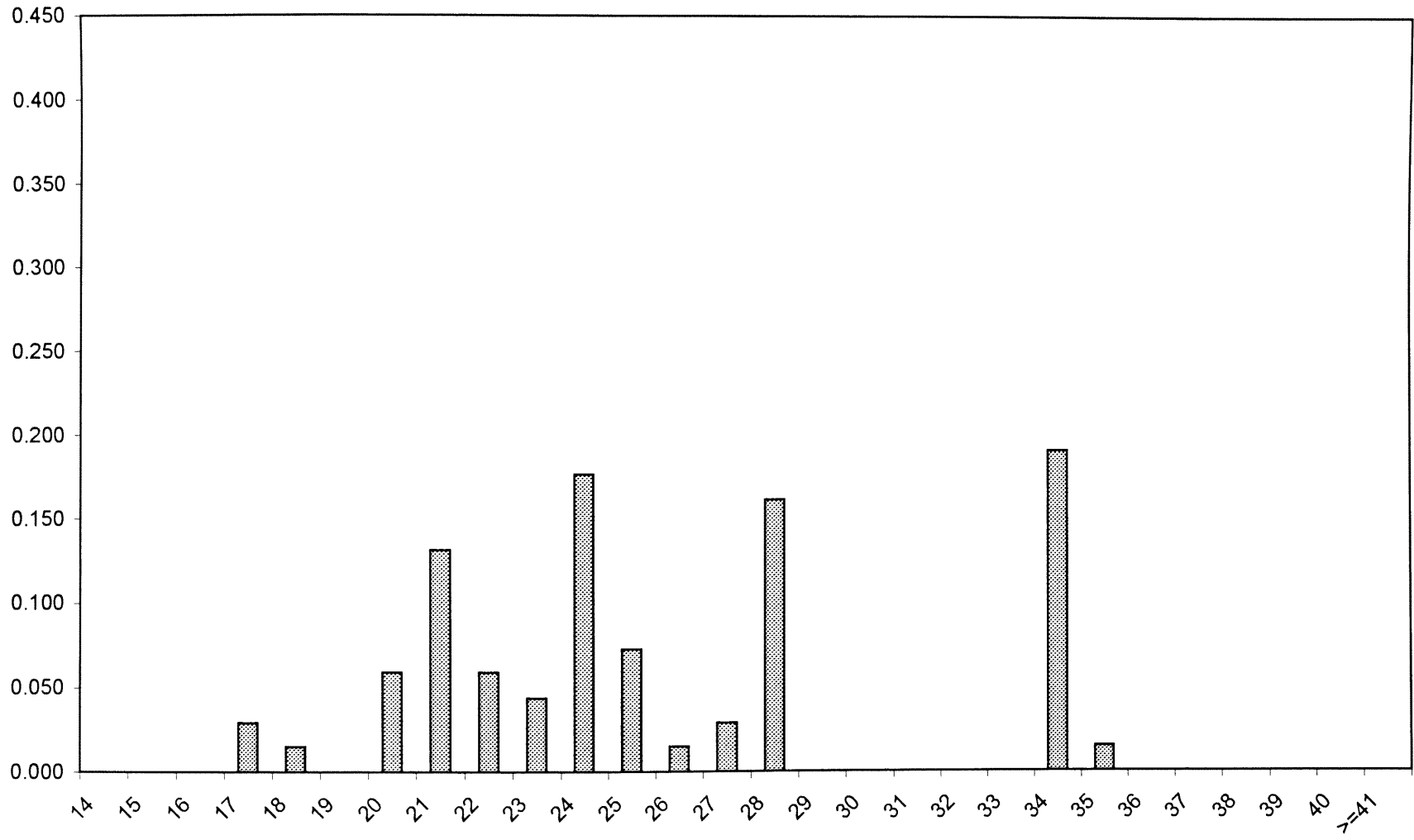


Figure 11. Allelic Distribution Graph of the Congo Population [26]

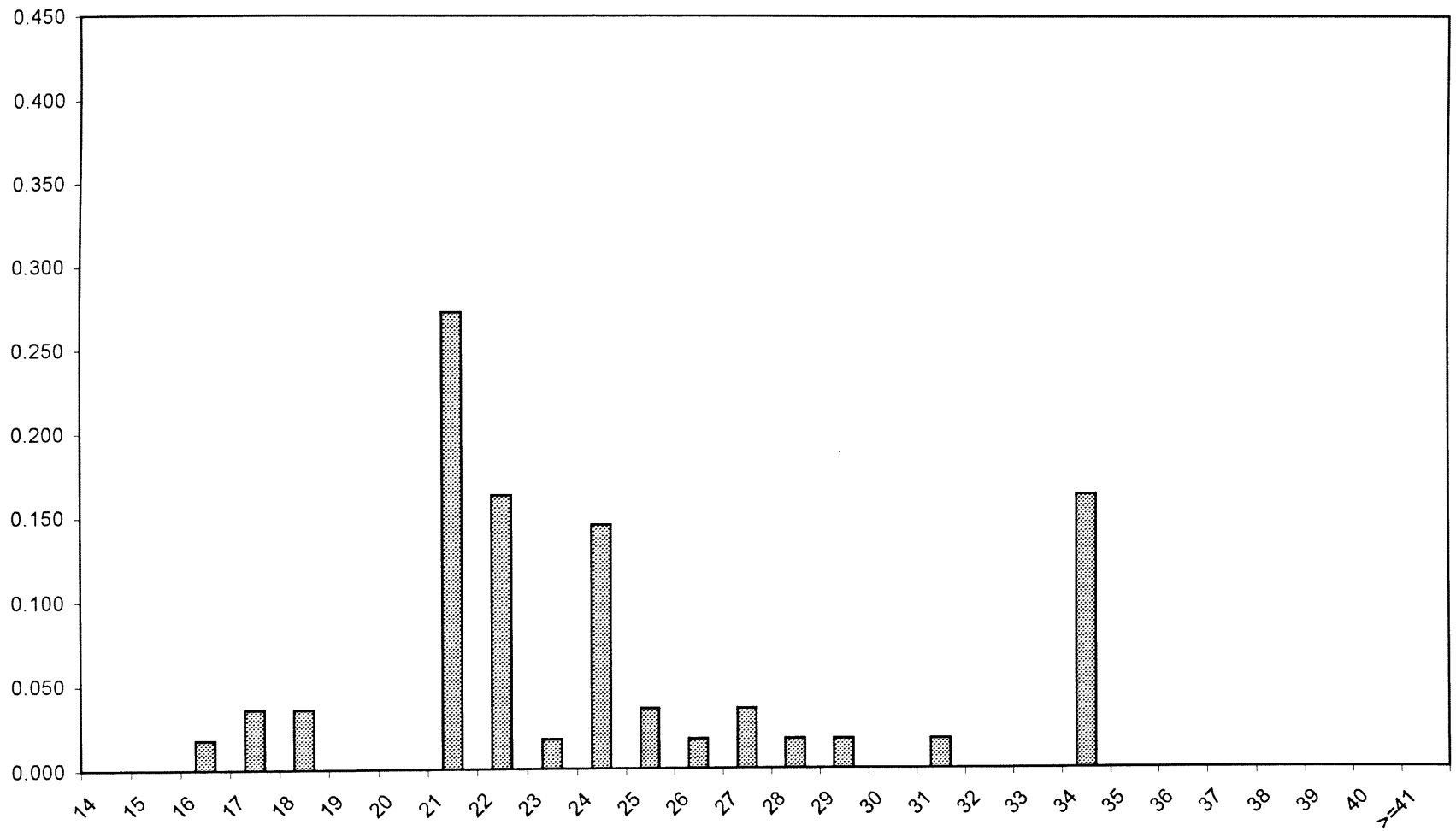


Figure 12. Allelic Distribution Graph of the Nigeria Population [28]

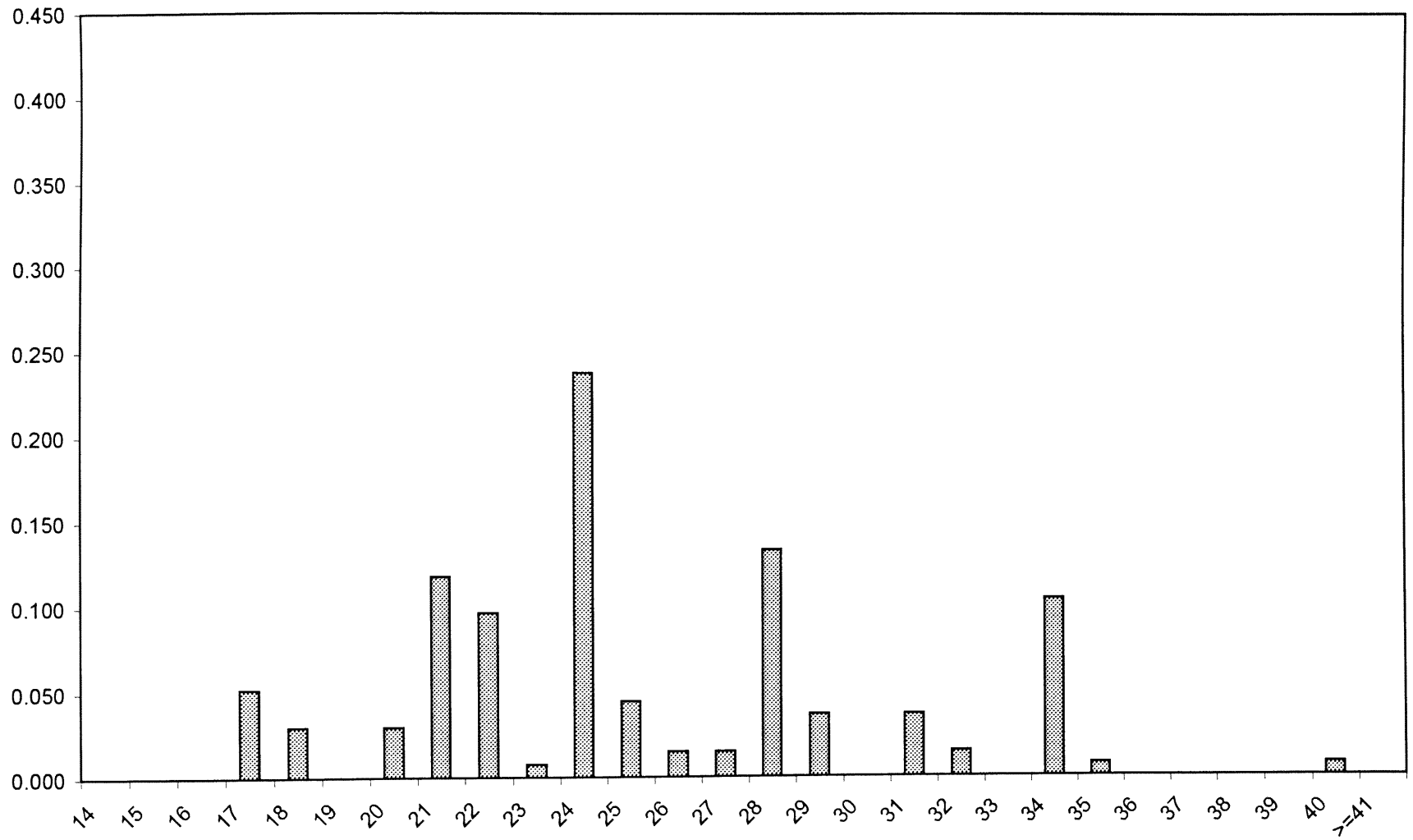


Figure 13. Allelic Distribution Graph of the Zimbabwe Population [28]

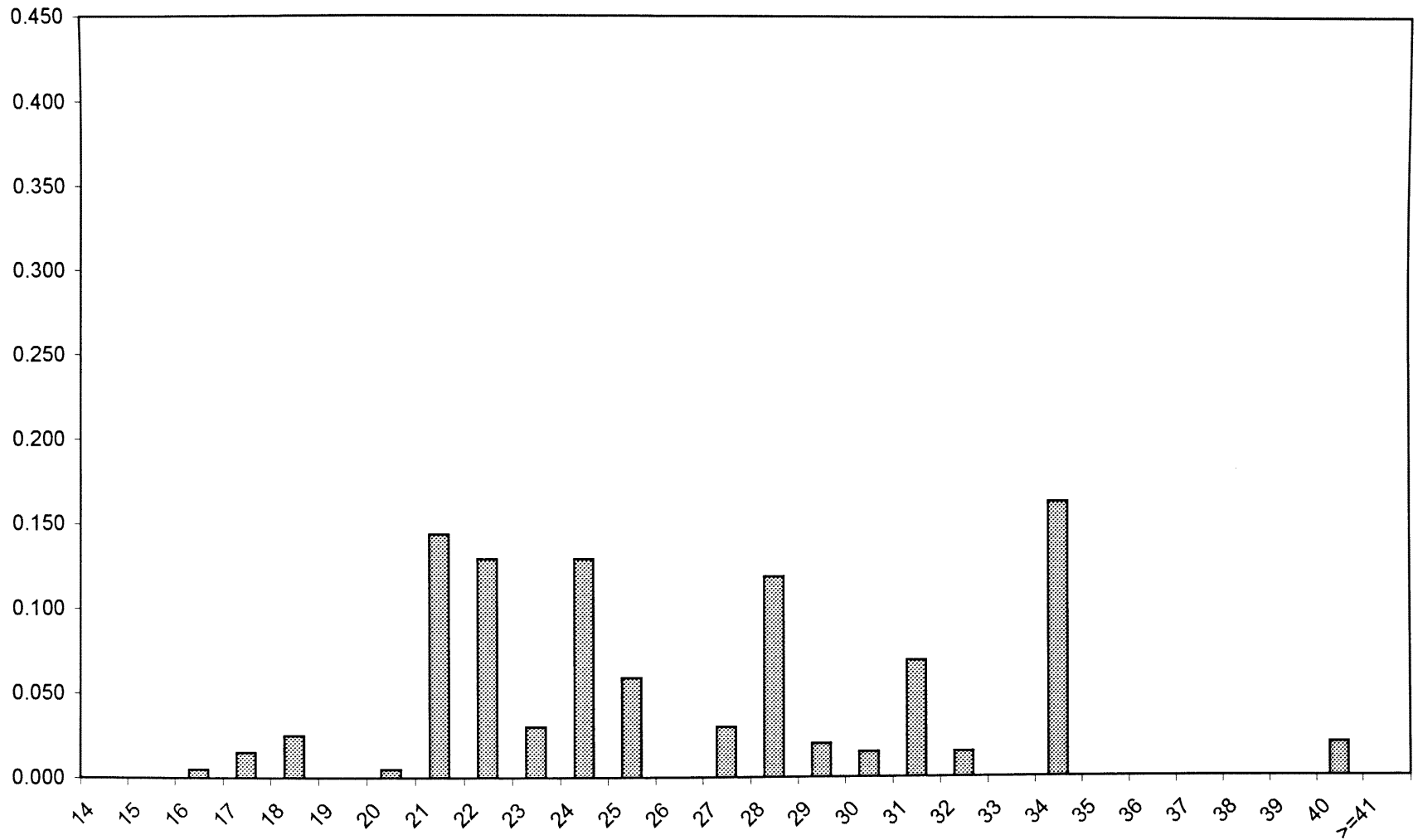


Figure 14. Allelic Distribution Graph of the African American Population [48]

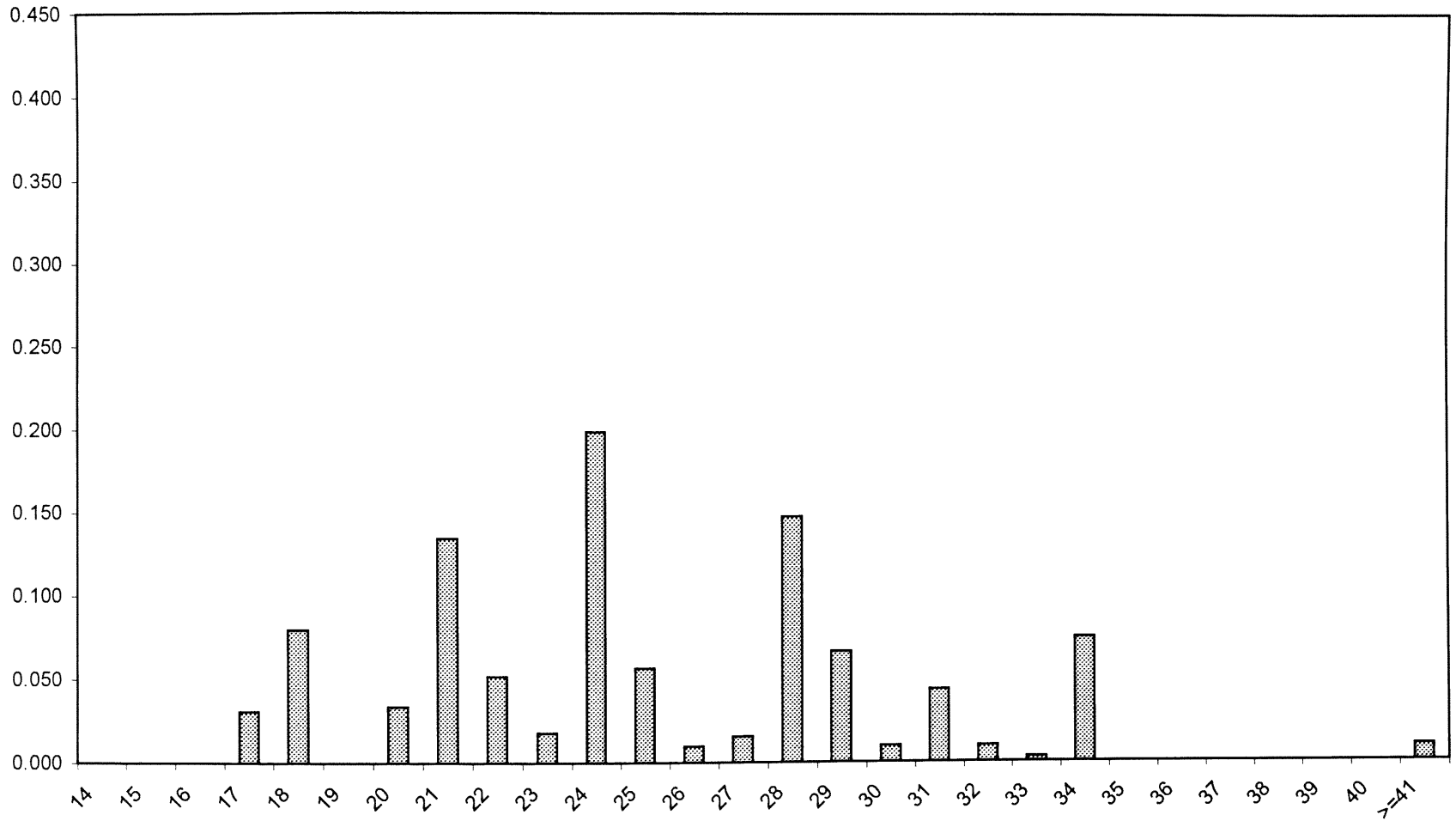
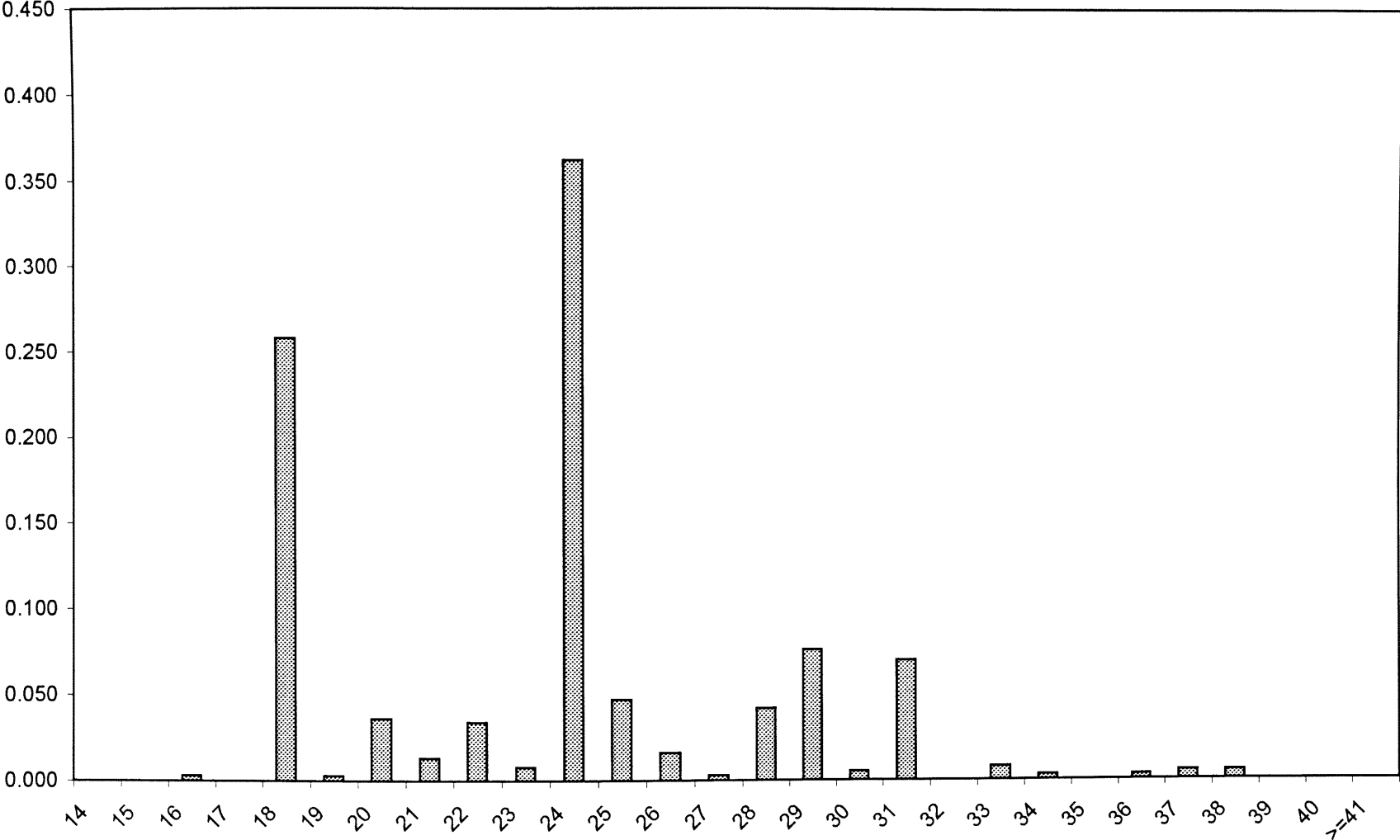


Figure 15. Allelic Distribution Graph of the Caucasian Population [48]



rest of the Caucasians. This is indicative of the variation detected between these groups. The Mapuche sample was more closely grouped to the southwest Asians than to the rest of the Caucasians. The two Alaskan groups show great homogeneity by segregating together. The east Asians are once again separately grouped into northern and southern regions.

Unlike the other two tree building methods, the ML dendrogram represented calculated distances based on the differences between the populations. Within this depiction, the southwest Asian groups (Saudi Arabia, Dubai Arab and Arab Moslem) are again separated from the other Caucasian populations. The east Asians are also shown to be distinct from the Native American groups even though they appear in the same cluster. The east Asians are segregated into northern and southern groups. The two Columbian Indian populations and the two Alaskan populations are shown to be genetically similar.

Throughout the previous analyses, consistent segregations were observed. The five African populations examined in our study were invariably grouped with the sub-Saharan African cluster, with the exception of the one Egyptian sample. The 16 Cameroon samples of our study segregated closely with the previously published Cameroon population [26].

The PC plot yielded a visual representation of allelic variability between the groups. This two-dimensional plot represented 35% of the total variation among the 27 alleles inputted. The x-axis of the PC clearly delineates the sub-Saharan African populations from the other populations. The north African population, Tanta, falls between this group and the Caucasians, nearer to the latter. Along the y-axis, the west Africans and west African derived populations are further segregated within the sub-

Saharan cluster. The y-axis also separates east Asians from the American Indians and middle eastern populations from the rest of the Caucasians. Negroids showed the greatest variability between the alleles with a value of 0.5046. This is indicated in the figure by the sparse arrangement of these populations. The Mongoloids represent the second most diverse group and are also sparsely gathered on the right side of the PC plot. The east Asians segregate together, as do the American Indians. Homogeneity is shown between the two Alaskan populations and the two Columbian populations that segregate near to each other. Similar to the NJ and UPGMA phylogenies, the Mapuche population segregates with the Caucasians, nearer to the southwest Asians. The Caucasian populations are clustered together closely denoting less diversity between the groups. However, segregation is still visible between the southwest Asians and the rest of the Caucasians. The Caucasians showed the least variability between their alleles with a value of 0.2662.

The G-Test displayed homogeneity within the major races. It is in line with the other analyses that we performed in regard to the racial groupings of the populations. This analysis also grouped the Tanta and Mapuche populations with the Caucasians.

In this study we report on D1S80 databases from five African populations. The D1S80 locus provided us with a quick view of the phylogenetic relationships of the 33 populations examined. Although broad categorization of the populations was possible, clearly more polymorphic loci are required to provide for more significant comparisons.

CONCLUSION

Our study has shown that this single locus is highly polymorphic. It was capable of separating 33 different populations into their constituent ethnic groups of Negroid, Mongoloid or Caucasoid. Each of the phylogenetic analyses performed indicated homogeneity between our African samples amongst themselves and in comparison to the few samples that previously existed from Africa.

The allelic distributions in our African populations were similar to that of other African and African derived populations in the literature. The Tanta population, which is believed to have been descendents of Caucasians, showed a similar distribution of alleles as the United States Caucasians [48].

Although D1S80 is an extensively studied locus, the majority of the data is focused on European populations (sparse database information currently exists for the following African nations: Cameroon, Congo, Nigeria and Zimbabwe). This study is the first to include greater than three African populations. A database useful for human differentiation has been created for four additional African populations.

REFERENCES

- [1.] Y. Nakamura, et al., Isolation and mapping of a polymorphic DNA sequence (pMCT118) on chromosome 1p(D1S80). *Nucleic Acids Research*. 16 (1988) 9364.
- [2.] B. Budowle, et al., Multiplex Amplification and Typing Procedure for the Loci D1S80 and Amelogenin. *J. Forensic Sci.* 41 (1996) 660-663.
- [3.] G. Duncan, Microvariation at the Human D1S80 locus. Doctoral Dissertation, FIU, 1996.
- [4.] G. Duncan, et al., Human Phylogenetic Relationships According to the D1S80 locus. *Genetica*. 98 (1996) 277-287.
- [5.] D. Betsch, J. Beard, D1S80 PCR with the \$25 Thermal Cycler. *Biochemical Education*. 27 (1999) 45-47.
- [6.] L.L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes*, Princeton University Press, New Jersey, 1994. pp.73-83, 162-166.
- [7.] G. P. Murdock, *Africa: Its Peoples & their culture history*, McGraw-Hill Book Co, New York, 1959. pp. 111, 127, 392.
- [8.] D. Levinson, *Ethnic Groups Worldwide: A Ready Reference Handbook*, Oryx Press, Phoenix, 1998. pp. 111, 115, 140, 152, 159-160.
- [9.] C. Legum, ed., *Africa Contemporary Record*, Africana Publishing Co., London, 2000. pp. B377.
- [10.] R. M. Maxon, Kenya. Microsoft Encarta Online. 2002.
- [11.] C. C. Li, *First Course in Population Genetics*, Boxwood, Pacific Grove, 1976.
- [12.] S. Cosso, R. Reynolds, Validation of the AmpliFLP D1S80 PCR Amplification Kit for Forensic Casework Analysis According to TWGDAM Guidelines. *J. Forensic Sci.* 40 (1995) 424-434.
- [13.] Perkin-Elmer. *AmpliFLP D1S80 PCR Amplification Kit User Guide*.
- [14.] M. Tahir, et al., Distribution of D1S80 Alleles in the Bahrainian Population. *J. Forensic Sci.* 44 (1999) 1314-1315.

- [15.] H. Rangel-Villalobos, et al., Allele Frequency Distributions of Six Amp-FLPS (D1S80, APO-B, VWA, TH01, CSF1PO and HPRTB) in a Mexican Population. *Forensic Sci. Int.* 105 (1999) 125-129.
- [16.] I. Flores, et al., Population Data for Southern Spain and Canary Islands of HLADQA1, PM and D1S80 Loci. *Forensic Sci. Int.* 119 (2001) 116-118.
- [17.] G. Duncan, et al., Microvariation at the Human D1S80 locus. *Int. J. Legal Med.* 110 (1996) 150-154.
- [18.] I. Eisenbarth, et al., Description of Three New Alleles at the D1S80 (MCT118) Locus. *J. Forensic Sci.* 44 (1999) 841-842.
- [19.] G. Watanabe, K. Shimizu, DNA Sequence Analysis of Long PCR Amplified Products at the D1S80 Locus. *Legal Medicine (In Press)*, (2001).
- [20.] A. M. Gross, et al., Validation Studies for the Genetic Typing of the D1S80 Locus for Implementation into Forensic Casework. *J. Forensic Sci.* 42 (1997) 1140-1146.
- [21.] I. Sanchez-Molina, R. Calvet, Allelic Frequencies for the HLA-DQA1, D1S80, HUMTH01, HUMTPOX, HUMCSF1PO and HUMVWA in Cantabria (Middle North Spain). *J. Forensic Sci.* 45 (2000) 167-169.
- [22.] J. Soares-Vieira, et al., Allele and Genotype Frequencies for D1S80 Locus in a Brazilian Population Sample. *J. Forensic Sci.* 45 (2000) 696-697.
- [23.] R. Deka, et al., Population genetic characteristics of the D1S80 locus in seven human populations. *Human Genetics.* 94 (1994) 252-258.
- [24.] K. Kasai, et al., Amplification of a Variable Number of Tandem Repeats (VNTR) Locus (pMCT118) by the Polymerase Chain Reaction (PCR) and Its Application to Forensic Science. *J. Forensic Sci.* 35 (1990) 1196-2000.
- [25.] A. Buchanan, et al., Long DOP-PCR of rare archival anthropological samples. *Human Biology.* 72 (2000) 911-925.
- [26.] W. Araujo Da Silva Jr., et al., Genetic Diversity of Two African and Sixteen South American Populations Determined on the Basis of Six Hypervariable Loci. *Am. J. Phys. Anthropol.* 109 (1999) 425-437.
- [27.] M. Lorente, et al., Spanish Population Data on Seven Loci: D1S80, D17S5, HUMTH01, HUMVWA, ACTBP2, D21S11 and HLA-DQA1. *Forensic Sci. Int.* 86 (1997) 163-171.

- [28.] B. Peterson, et al., World Population Data for the HLA-DQA1, PM and D1S80 Loci with Least & Most Common Profile Frequencies for Combinations of Loci Estimated Following NRC II Guidelines. *J. Forensic Sci.* 45 (2000) 118-146.
- [29.] S. Gutowski, et al., Statistical Analysis of an Australian Population for the Loci Gc, HLA-DQA1, D1S80 and HUMTH01. *Forensic Sci. Int.* 76 (1995) 1-6.
- [30.] M. Walkinshaw, et al., DNA Profiling in Two Alaskan Native Populations Using HLA-DQA1, PM, and D1S80 Loci. *J. Forensic Sci.* 41 (1996) 478-484.
- [31.] A. Alkhatat, et al., Population Data on the PCR-Based Loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80 from Arabs from Dubai. *Forensic Sci. Int.* 81 (1996) 29-34.
- [32.] M. A. Cariolou, et al., Greek Cypriot Allele and Genotype Frequencies for Amplitype PM-DQA1 and D1S80 Loci. *J. Forensic Sci.* 43 (1998) 661-664.
- [33.] M. Hutz, et al., D1S80 locus variability in South American Indians. *Annals of Human Biology.* 24 (1997) 249-255.
- [34.] S. C. Halos, et al., Philippine Population Database at Nine Microsatellite Loci for Forensic and Paternity Applications. *Forensic Sci. Int.* 101 (1999) 27-32.
- [35.] M. A. Tahir, et al., Distribution of Amplified Fragment Length Polymorphism D1S80 Alleles in a Saudi Arabian Population. *J. Forensic Sci.* 44 (1999) 1159.
- [36.] A. H. Cakir, et al., Distribution of Amplified Fragment Length Polymorphism D1S80 Alleles in Turkish Population. *J. Forensic Sci.* 44 (1999) 1002.
- [37.] J. Felsenstein, PHYLIP (Phylogeny Inference Package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- [38.] F.J. Rohlf, Numerical Taxonomy System (NTSys pc). Version 2.1. Exeter Publishing Ltd., Setauket, NY.
- [39.] G. Carmody, G-Test. Distributed by the author, Department of Biology, Carleton University, Ottawa.
- [40.] S. Sherry, M. Batzer, Modeling Human Evolution – To Tree or Not to Tree? *Genome Research.* 7 (1997) 947-949.
- [41.] P.O. Lewis, D. Zaykin, Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>. (2001)

[42.] J.W. Lee, et al., Evaluation of DNA Match Probability in Criminal Case. *Forensic Sci. Int.* 116 (2001) 139-148.

[43.] K. Monson, B. Budowle, Effect of Reference Database on Frequency Estimates of Polymerase Chain Reaction (PCR)-Based DNA Profiles. *J. Forensic Sci.* 43 (1998) 483-488.

[44.] Population Genetics of Gene Flow. 2002, Wake Forest University. Accessed through www.wfu.edu/users/sipet00g/populationgenetics.html.

[45.] R. Duran, M. Ruiz-Garcia, Genetic Population History Relationships of the Population of Bogota, Columbia, by Using the D1S80, VWA, and TH01 Molecular Markers. *Am. J. Hum. Biol.* 13 (2001) 374-383.

[46.] W. S. Watkins, et al., Patterns of Ancestral Human Diversity: An Analysis of *Alu*-Insertion and Restriction-Site Polymorphisms. *Am. J. Hum. Genet.* 68 (2001) 738-752.

[47.] M. Nei, Genetic Distance Between Populations. *The American Naturalist.* 106 (1972) 283-292.

[48.] B. Budowle, et al., United States Population Data on the Multiplex Short Tandem Repeat Loci – HUMTH01, TPOX, and CSF1PO – and the Variable Number Tandem Repeat Locus D1S80. *J. Forensic Sci.* 42 (1997) 846-849.

[49.] AmpliFLP D1S80 PCR Amplification Kit Product Detail. 2001, Applied Biosystems.

[50.] W. Klitz, et al., Analysis of Genotype Frequencies and Interlocus Association for the PM, DQA1, and D1S80 Loci in Four Populations. *J. Forensic Sci.* 45 (2000) 1009-1015.

[51.] The Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington D.C., 1996. p.5.