African Journal of Biotechnology Vol. 7 (8), pp. 1023-1027, 17 April, 2008
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2008 Academic Journals

Full Length Research Paper

Determination of allele frequencies in nine short tandem repeat loci of five human sub-populations in Botswana

Mpoloka S.W.[1], Kgotlele T.[1] and Wally A.[2].

[1]Department of Biological Sciences, University of Botswana, P/Bag 00704, Gaborone, Botswana.
[2]Botswana Police Forensic Science Laboratory P/Bag 0012, Gaborone, Botswana.

Accepted 15 February, 2008

Allele frequencies for nine short tandem repeat (STR) loci from the AmpFISTR® Profiler Plus™ PCR Amplification Kit were determined in five subpopulations across Botswana and tested for suitability for use in individual identification. DNA was extracted from whole blood samples collected from 150 unrelated individuals. Targeted regions of DNA (vWA, FGA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11 and the sex determining locus Amelogenin) were amplified using multiplexed PCR. The alleles showed a high degree of polymorphism, with heterozygosity above 67%, indicating that there is no significant inbreeding within the subpopulations. Matching probability was below 0.2 hence power of discrimination was high, indicating that the alleles from the five sub-populations can be used in human identifications. The present study is the first reported attempt at determining allele frequencies of subpopulations in Botswana and could possibly be used in developing a national DNA database.

Key words: DNA typing, short tandem repeat (STR), polymerase chain reaction, allele frequency, Botswana.

INTRODUCTION

Humans comprise approximately three billion DNA bases, 2% of which code for proteins while the rest is unflatteringly referred to as “junk” or non-coding DNA (Watson, 2003). This non-coding DNA is made up of apparently functionless stretches of nucleotide sequences of varying lengths that occur repeatedly. Highly repetitive DNA is polymorphic in individuals (Jeffreys et al., 1985) and the sequences vary among individuals at the same site on a chromosome (locus) (Lewis, 2005). That is, any two randomly selected humans are genetically 99.9% identical while the remainder, about 3 million of the 3 billion bases, confers the human uniqueness or individuality (Aaspollu et al., 2000). This polymorphism in repetitive DNA prompted the development of a technique called DNA fingerprinting, alternatively known as “DNA typing” or “DNA profiling”. DNA fingerprinting reconstructs a unique pattern from the DNA of an individual that can be used to distinguish between individuals (Bains, 2004).

DNA markers, including short tandem repeats (STRs), have been developed to take advantage of this variable DNA fraction. Analysis of STRs is the most widely used method of genotyping used in forensic identification of individuals and for testing paternity. This is because STRs consist of short, repetitive sequence elements comprising 3 to 7 base pairs in length distributed throughout the human genome. These loci are a rich source of highly polymorphic markers that may be detected using the polymerase chain reaction (PCR). PCR is a mimic of the normal cellular process of replication of DNA molecules. Each STR is distinguished by the number of times a sequence is repeated, called an allele. Alleles of STR loci are differentiated electrophoretically by the number of copies of the repeat sequence contained within the amplified region. They may be detected from one another using radioactive, ethidium bromide, silver staining or through fluorescence detection. The fact that

*Corresponding author. E-mail: mpoloka@mopipi.ub.bw.

STRs are well characterised and highly polymorphic makes them ideal for use in individual human identification (Lewis, 2005). STRs are commonly used because relatively low DNA quantities are required, and degraded DNA samples can still be used. The STR typing method is also amendable to automation and involves sensitive fluorescent detection that enables rapid collection of genetic data (Butler, 2001) from which the genetic composition of a population can then be described in terms of the frequencies or relative abundances in which alternative alleles are found (Hartl and Jones, 2002).

Several studies (Yasin et al., 2005; Budowle et al., 2005; Linacre, 2001) have shown that there is genetic variation based on allele frequencies between various populations and among different sub-populations. A great deal of genetic variation within and between populations arises from the existence of various alleles at different gene loci. All genetic markers exhibit particular frequencies of occurrence in a population (Inman and Rudin, 1997). Hence it is important to determine the frequency of marker occurrence in order to attach significance to any particular genetic trait. Based on the allele's observed frequency in a particular population, it is assigned the probability of a certain genetic variant (alleles) occuring in two places by chance (Lewis, 2005). The most reliable way to express the overall significance is to multiply the individual frequencies of the different markers (Linacre, 2001). The calculations are only legitimate if the markers pass certain genetic and statistical tests.

In recent years DNA typing has become a primary means of identification when specimens required by conventional means of identification are inadequate or when medical or dental records are unavailable. DNA typing has replaced classical and conventional identification procedures such as visual, x-rays and fingerprint assessment which have been the main methods of identification efforts in mass disasters (Leclair et al., 2000). DNA typing is used in determining perpetrators of violent crimes such as murder and rape, resolving paternity disputes and identifying remains of missing persons or victims of mass destruction and natural disasters. For example, DNA technology was used to identify victims of The World Trade Centre attacks in the USA on September 11, 2001. About 850 of the 1594 victim identifications established were based solely on DNA results (Biesecker et al., 2005). Such identification required reference samples for comparison and these can be obtained from surviving family members or from biological material left on personal items of the deceased such as a tooth brush, comb, or razor (Butler, 2001). In other cases probabilities are used based on observed or predetermined allele frequencies in a particular population (Lewis, 2005).

The pre-determined frequencies are mostly compiled into genetic databases, a collection of DNA profiles obtained from unrelated individuals of a particular group or population (Butler, 2001). Examples of existing genetic databases of human allele frequencies include African Jordanian, Rwandan, Spanish and American populations (Yasin et al., 2005; Budowle et al., 2005; Linacre, 2001). To interpret DNA profiles, allele frequencies derived from population data and genotype frequencies are computed. The result estimates the likelihood that two individuals from a given population share the same genotype for all of the DNA sequences (loci) examined (Lewis, 2005). Two such parameters are matching probability and power of discrimination. Matching probability is the probability of picking someone randomly whose alleles will match with another person in the same group (Butler et al., 2003). Power of discrimination (PD), which is the opposite of matching probability, is the chance of picking someone randomly whose alleles are different from another person in the same group (Butler et al., 2003).

Thus allele frequencies serve as reference material for statistical evaluation in human identification. These frequencies are important tools when traditional identification procedures are uncertain or inapplicable. The aim of the present study was to determine allele frequencies in nine short tandem repeat loci of five subpopulations in Botswana. The study also sought to establish if those allele frequencies can be used to identify individuals within a subpopulation.

MATERIALS AND METHODS

Whole blood was collected from 150 unrelated individuals from five subgroups from five different regions of Botswana (Figure 1) using FTA® cards. The cards were the method of choice because they have preservatives that hinder spoilage and are easily packaged and transported. The subpopulations geographic locations in Botswana are as follows: subgroup 1 (south-east Botswana), subgroup 2 (south-west Botswana), subgroup 3 (north-east Botswana), subgroup 4 (western Botswana) and subgroup 5 (north-west Botswana).

A 1.2 mm punch was made from the bloodstained FTA® card (Whatman Inc. USA). The punch was washed three times with 200 µl FTA purification reagent and rinsed two times with 200 µl TE buffer. The punch or disk was air dried at room temperature. The targeted regions (vWA, FGA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11 and the sex determining locus Amelogenin: see Table 1 for loci details) were amplified on the GeneAmp® PCR System 9700 thermal cycler (Perkin Elmer Corporation) using the AmpFlSTR® Profiler Plus™ PCR Amplification Kit (Promega, USA). Samples were then separated and detected using the ABI 310 Genetic Analyzer (Applied BioSystems) capillary electrophoresis machine.

The ABI GeneScan software, version 1.2.2 was used to interpret the data. Allele frequencies and statistical population parameters were calculated in the "PowerStats" Microsoft Excel Workbook template from Promega (www.promega.com/geneticidtools./)

The different AmpFlSTR® Profiler Plus™ loci used in the study are shown in Table 1. The loci are named as indicated below (with locus D21S11 as an example). D stands for DNA, 21 is chromosome 21, S represents a single copy sequence and 11 is the locus or position described on the chromosome. The chromosomal locations are designated by the letters p and q where p is the short arm of the chromosome and q is the long arm of the chromosome. For example, 4q28 represents band 28 on long arm of chromosome 4. The letters ACGT represents the four DNA bases adenine (A), cytosine (C), guanine (G) and thymine (T).

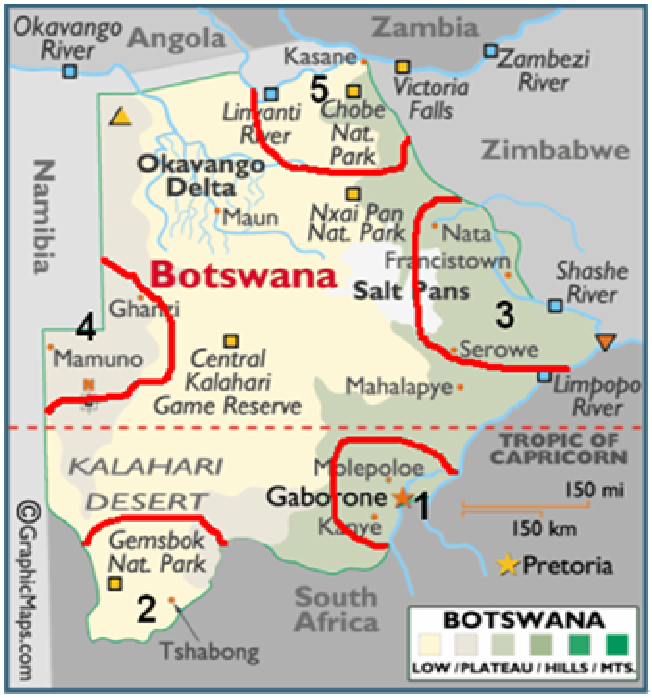

Figure 1. Sample areas from the different regions in Botswana.

RESULTS AND DISCUSSION

The allele frequencies for the nine AmpFISTR® Profiler Plus™ from the five subpopulations are shown in Table 2. The power of discrimination (PD), the probability of paternity exclusion (PE), the typical paternity index (TPI) the polymorphic information content (PIC) and the observed heterozygosity (Ho) were calculated using the PowerStats" Microsoft Excel Workbook template from Promega (www.promega.com/geneticidtools./).

A number of alleles, some of which were previously reported as variants (Butler, 2001; NIST, 2006), were detected in this study. There is a possibility that new alleles that may be unique to the Botswana population and other variant alleles that occur between loci may have been detected. Non-published variant alleles are being observed on a regular basis as STR typing becomes more wide-spread (Higuchi et al., 1988). Some alleles were found to fall within the reported range but have not been reported before, pointing to the possibility of their being variants or new alleles. Examples of these are alleles 13.2 at the D18S51 locus and alleles 24.2, 28.2, 30.2 and 37 in the locus D21S11 (Table 2).

The allele frequencies ranged from 0.004 for the rarest alleles found at most loci studied (except vWA and D13S317) to 0.369 for the most common allele (allele 12 at the D13S317 locus). The level of heterozygosity was found to be high, the lowest observed heterozygotes frequency being 0.0667 at the D13S317 locus and the highest being 0.859 for the FGA locus pointing to a random mating population. The frequency of observed homozygotes was low across all loci studied with locus D18S51 showing the lowest frequency (0.131) and locus D13S317 having the highest frequency of homozygotes (0.333) indicating that inbreeding is insignificant in the study population. Inbreeding reduces the frequency of heterozygous genotypes while increasing the homozygous genotypes as opposed to random mating (Jeffreys et al., 1985). This is because inbreeding results in an excess of homozygotes while random mating leads to an excess of heterozygotes. This is shown in Table 2 where heterozygotes frequencies are above 67%, indicating that mating is random.

Table 1. AmpFISTR® Profiler Plus™ loci used in the study.

Locus	Chromosomal location	Repeat motif
FGA	4q28	CTTT
vWA	12p12-pter	TCTG TCTA
D3S1358	3p	TCTG TCTA
D5S818	5q21-31	AGAT
D7S820	7q11.21-22	GATA
D8S1179	8	TCTA TCTG
D13S317	13q22-31	GATA
D18S51	18q21.3	AGAA
D21S11	21q21	TCTA TCTG
Amelogenin	X:p22.1-22.3 Y:p11.2	Sex determining locus

The matching probability (MP) was found to be low, ranging from 0.026 (locus D21S11) to 0.114 (locus D3S1358). The lower the matching probability the lower the chances of picking someone whose alleles are the same with someone else (Linacre, 2001). The higher the power of discrimination (PD) the higher the chances that someone will be picked randomly whose alleles are different from the first person. The values of PD from Table 2 imply that chances of picking out two people randomly that are genetically different are high. The most polymorphic loci are the most discriminating loci (Leclair et al., 2000; Biesecker et al., 2005).

Only STRs that demonstrate high degrees of variability within the population are selected (Biesecker et al., 2005). The allele frequencies from the study are variable and show that mating is random; therefore they can be used in identification procedures. The power of DNA testing is such that the information provided by several highly variable loci is sufficient to convince us of the source of a sample (Linacre, 2001).

The study showed there is variability between subgroups in Botswana, hence further studies of the other subgroups is needed. This may lead to an establishment of a statistical DNA database of Botswana based on the allele frequencies of the various subgroups. This will provide an alternative method in identifying remains of victims to the current ones and also for academic studies. Further studies need to be conducted to verify the authenticity of the variants and the possibility of newly

Table 2. Allele frequencies for nine STR loci in five subpopulations

	D3S1358	D5S818	D7S820	D8S1179	D13S317	D18S51	vWA		FGA		D21S11
Alleles	**n = 260**	**n = 260**	**n = 246**	**n = 260**	**n = 252**	**n = 248**	**n =260**	***Alleles***	**n = 256**	***Alleles***	**n = 256**
6	-	-	-	-	-	-	-	18	0.02	24.2	0.004
7	-	0.004	0.004	-	-	-	-	19	0.066	25	
8	-	0.108	0.175	-	0.016	-	-	20	0.07	26	0.044
9	-	0.027	0.163	-	0.008	0.004	-	21	0.07	27	0.07
10	-	0.065	0.337	0.004	0.044	0.004	-	22	0.188	28	0.195
10.2	-	-	-	-	-	-	-	23	0.199	28.2	0.004
11	-	0.188	0.187	0.027	0.258	0.008	0.012	24	0.203	29	0.0156
12	0.004	0.323	0.11	0.108	0.393	0.012	-	25	0.07	29.2	
13	-	0.258	0.024	0.212	0.194	0.036	0.012	26	0.082	30	0.137
13.2	-	-	-	-	-	0.004	-	26.2	-	30.2	0.004
14	0.1	0.027	-	0.369	0.079	0.056	0.077	27	0.016	31	0.086
14.2	-	-	-	-	-	0.012	-	28	0.012	31.2	0.059
15	0.365	-	-	0.208	0.008	0.117	0.173	29	0.004	32	0.027
16	0.3	-	-	0.069	-	0.161	0.304	30		32.2	0.078
17	0.181	-	-	-	-	0.149	0.162			33	0.023
18	0.046	-	-	0.004	-	0.169	0.142			33.2	0.035
19	0.004	-	-	-	-	0.173	0.092			34	0.012
20	-	-	-	-	-	0.048	0.027			34.2	0.031
21	-	-	-	-	-	0.024	-			35	0.027
22	-	-	-	-	-	0.012	-			35.2	0.008
25	-	-	-	-	-	0.008	-			36	0.023
										37	0.004
										38	0.012
MP	0.114	0.092	0.087	0.108	0.11	0.034	0.059		0.04		0.026
PD	0.886	0.908	0.913	0.892	0.89	0.966	0.941		0.96		0.974
PIC	0.69	0.74	0.75	0.72	0.69	0.86	0.79		0.84		0.88
PE	0.44	0.543	0.578	0.613	0.379	0.673	0.53		0.713		0.652
TPI	1.71	2.17	2.37	2.6	1.5	3.1	2.1		3.56		2.91
Ho	0.292	0.231	0.211	0.192	0.333	0.131	0.238		0.141		0.172
He	0.708	0.769	0.789	0.808	0.667	0.839	0.762		0.859		0.828

MP: matching probability, **PD**: power of discrimination, **PIC**: polymorphic information content, **PE**: power of exclusion, **TPI**: typical paternity index, **Ho**: observed homozygotes, **He**: observed heterozygotes.

identified alleles in the Botswana population. This can be done by sequencing the variants and new alleles.

REFERENCES

Aaspollu A, Sumeri I, Kelve M (2000). Allele distribution at Nine STR Loci-D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317 and D7S820-in the Estonian Population. In International Congress Series 1193 (Eds GF, Senabaugh P, Lincoln J, Olaisen B) pp. 129-131, Elsevier Science B. V., Amsterdam, The Netherlands.

Bains W (2004). Biotechnology from A to Z. 3rd edition. Oxford University Press Inc., New York.

Biesecker LG, Bailey-Wilson JE, Ballantyre J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B, Eisenberg A, Forman L, Kidd KK, Leclair B, Niezgoda S, Parsons TJ, Shaler R, Sherry ST, Sozer A, Walsh A (2005). DNA Identifications After the 9/11 World Trade Centre Attack. Science, 310(5751): 1122-1123.

Budowle B, Adamowicz M, Aranda XG, Barna C, Chakraborty R, Cheswick D, Dafoe B, Eisenberg A, Frappier R, Gross AM, Ladd C, Lee H-S, Milne SC, Meyers C, Prinz M, Richard ML, Sladahna G, Tierney AA, Viculis L, Krenke BE (2005). Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America. Forensic Sci. Int. 150: 1-15.

Butler JM (2001). Forensic DNA typing: Biology and technology behind STR Markers. Academic Press, London. United Kingdom.

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

Hartl DL, Jones EW (2002). Essential Genetics: A Genomics Perspective. 3rd edition. Jones and Bartlett Inc., Sudbury.

Higuchi R, Von BCH, Sensabaugh GF, Erlich HA (1988). DNA typing from single hairs, Nature, 322: 543-546.

Inman K, Rudin N (1997). An Introduction to Forensic DNA Analysis. CRC Press Inc., Florida. United States of America.

Jeffreys AJ, Wilson V, Thein SL (1985). Individual-specific fingerprints of human DNA. Nature, 316: 76-79

Leclair B, Fregean CJ, Bowen KL, Borys SB, Elliot J, Fourney RM (2000). Enhanced kinship analysis and STR based DNA typing for human identification in mass disasters. In International Congress Series 1193 (Eds Senabaugh GF, Lincoln PJ, Olaisen B), pp. 91-93, Elsevier Science, Amsterdam, The Netherlands.

Lewis R (2005). Human Genetics: Concepts and Applications. Sixth edition. McGraw-Hill, New York.

Linacre AMT, Phillips P, Wally A, Goodwin W (2001). *STR* data for the SGM Plus™ loci from three Indonesian populations. Forensic Sci. Int. 123: 232-234

NIST (2006). The National Institute of Standards and Technology, Biochemical Science Division. Variant Allele Reports. http://www.cstl.nist.gov/biotech/strbase/var_tab.htm.

Watson J (2003). DNA The secret of life. Random House Inc., New York.

Yasin SR, Hamad MM, Elkarmi AZ, Jaran AS (2005). African Jordanian Genetic Database on Fifteen Short Tandem Repeat Genetic Loci. Croatian Med. J. 46: 587-592.