Biochemical Genetics (2019) 57:170–191
https://doi.org/10.1007/s10528-018-9879-0

ORIGINAL ARTICLE

Genetic Variation at 15 Autosomal STR Loci Among Seven Egyptian Populations

Ahmed F. AbdEl-Hafez[1] · Sherif H. El-Alfy[2,3] · Hamdy H. Swelim[2] · Nagwa H. A. Hassan[2]

Received: 29 May 2017 / Accepted: 20 July 2018 / Published online: 3 August 2018

Abstract

Egypt is a transcontinental country containing substantial ethnic, cultural, and linguistic diversity among its people. This study was conducted to investigate the genetic variation at 15 AmpFlSTR Identifiler short tandem repeat (STR) loci, D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA, within and between seven Egyptian populations. Samples of 814 unrelated individuals from Northern Coast, Delta, Greater Cairo, Canal governorates, Northern Upper Egypt, Southern Upper Egypt, and Sinai were investigated. All loci were highly polymorphic in all sample populations. The data were analyzed to give information on allele frequencies and other population statistical parameters. After applying Bonferroni correction, the agreement with Hardy–Weinberg equilibrium (HWE) was confirmed for all loci (exact test), and for all loci with the exception of D3S1358, D19S433, and D18S51 (X^2 test). The levels of genetic differentiation and the genetic relationships among populations were evaluated by coefficient of genetic differentiation (F_{ST}), AMOVA, and genetic distance of Nei. The most differentiated populations were found between Sinai and Southern Upper Egypt. These two populations showed the lowest within-population variation, whereas the population of Greater Cairo showed the highest within-population variation as indicated by the fixation index F_{IS}. The varying levels of genetic relatedness among the populations in relation to their geographical distribution were analyzed using Mantel test. The results demonstrated that the effectiveness of STR markers enhances their value for identifying the genetic variation within and between Egyptian populations.

Keywords DNA typing · Short tandem repeats · Multiplex-PCR · Egyptian populations

✉ Ahmed F. AbdEl-Hafez
ahmed_tasy@hotmail.com

Extended author information available on the last page of the article

Introduction

Egypt is a transcontinental country spanning the Northeast corner of Africa and Southwest corner of Asia via a land bridge formed by the Sinai Peninsula. With over 90 million inhabitants, Egypt is the largest country in Middle East region and Arab World and the third largest in Africa. Egypt has been influenced and shaped by some particular traditional customs through the history (Midant-Reynes 2000; White 2011). The majority of Egyptian people live in the delta and Nile valley from Greater Cairo to south of Aswan where the only arable land is found. Other people live in the Eastern Desert, Western Desert, and Sinai Peninsula. Some Egyptian people are divided into clans and tribes, distributed throughout the country, and practice complementary endogamy. Other people are urban, living mainly in Greater Cairo, Alexandria, and other cities in Egypt, and practice cultural exogamy. They are generally less traditional and conservative than other inhabitants of the more remote southern valley, deserts, and of the physically isolated Sinai Peninsula (Marsot 2007; White 2011). Such customs might contribute, to some extent, to the patterns of genetic variation among Egyptian populations.

Egypt assimilated various foreign influences, including Greek, Persian, Roman, Arab, Ottoman, and European, coupled with demographic and migration events. Egypt contains various ethnic groups, including ethnic Egyptians, the largest ethnic group in the country, Fallahi, Sa'idi, Bedouin, Berber, and Nubian, among others (Shaw 2004; Marsot 2007). In addition, various languages and dialects are spoken in Egypt, including *Egyptian Arabic*, *Sa'idi Arabic*, *Bedawi Arabic*, *Sudanese Arabic*, *Nobiin*, and others. The considerable ethnic and cultural diversity within Egypt make the study of genetic diversity among its populations an interesting effort.

Short tandem repeats (STRs), also known as microsatellites, are short tandemly repeated DNA sequences, with a repeat unit of 2–6 bp. They are polymorphic and the most informative PCR-based genetic markers, whose importance in forensic analysis (Butler 2006; Borosky et al. 2014), paternity testing (Thomson et al. 1999; Mortera et al. 2016), and population studies (Barbarii et al. 2004; Yao et al. 2016) is notable and outstanding. Such markers have now become commonly used in various Arab countries (Halima et al. 2009; Babiker et al. 2011; Khodjet-el-khil et al. 2012; Hadi et al. 2014; Osman et al. 2015). Little is known about the genetic diversity and structure of Egyptian populations. A few previous genetic studies in Egypt have mainly focused on mitochondrial (mt) DNA and the Y chromosome (Krings et al. 1999; Keita and Boyce 2005). In addition, only limited data are available on STR markers in Egyptian populations. Analysis of STR markers that included only a single population from Upper (Southern) Egypt was reported by Omran et al. (2009). Therefore, the present study was conducted in order to comprehensively assess the genetic structure and diversity among seven Egyptian populations. In order to have a better understanding of the molecular variation occurring in the studied populations, AMOVA analysis was carried out. The study compared locus-by-locus the allelic frequencies of Egyptian population with those from other Arab populations. Also, the study was conducted to

investigate whether the geographic distance and distribution likely to influence the patterns of genetic relationships between the populations under study. The study also aimed to establish a forensic database based on 15 autosomal STR markers by way of capillary electrophoresis on 3500 Genetic Analyzer.

Samples and Methods

Samples

Peripheral blood samples (drawn in EDTA) or oral swab samples were obtained from 814 unrelated Egyptian individuals residing in seven main Egyptian regions, Northern Coast, Delta, Greater Cairo, Canal governorates, Northern Upper Egypt, Southern Upper Egypt, and Sinai (Fig. 1). All individuals were adequately informed and an appropriate informed consent was obtained before their participation.

DNA Extraction

Genomic DNA was extracted by an automated extraction and purification procedure using the EZ1 Advanced and EZ1 DNA investigator kit (QIAGEN, California, USA) according to the instructions provided by the manufacturer (QIAGEN, California, USA). In this procedure, genomic DNA was extracted from the individual samples using a digestion buffer and proteinase K (supplied with the Kit) in 2-ml sample tube. After incubation of the tubes at 56 °C for 15 min, DNA from lysate was purified by the automated purification procedure. This silica-based DNA purification method was used to obtain high-quality DNA devoid of PCR

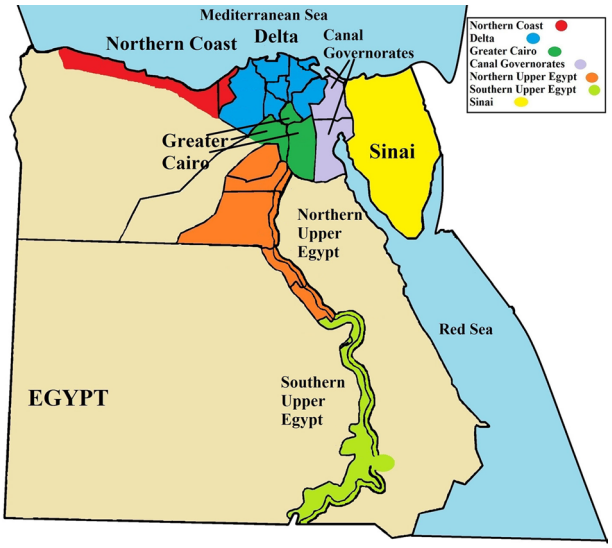

Fig. 1 Map of Egypt representing the geographical location of the seven populations analyzed

inhibitors. This purification method depends on binding of DNA to silica-coated magnetic beads in the presence of chaotropic salts such as guanidine hydrochloride, sodium iodide, and sodium perchlorate. The beads were separated from the lysate using a magnet. The DNA was then washed and eluted in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8). DNA quantity and quality were assessed by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, Delaware, USA), and stored at − 20 °C until analysis.

PCR Amplification

Multiplexed PCR amplification of 15 polymorphic STR loci was performed using the AmpFℓSTR® Identifiler® Plus amplification kit (Applied Biosystems, Foster city, CA, USA). Primer sequences are commercial information and are not available for publication. The characteristics of the 15 STR loci are shown in Table 1. For each locus, one fluorescently labeled primer pair was used [either 6-FAM (blue), VIC (green), NED (yellow) or PET (red)]. PCR was carried out in a total reaction volume of 25 µL consisting of 10 µL of DNA sample with 1–5 ng template DNA, 10 µL of Master Mix buffer and 5 µL of primer set following the procedure recommended by the manufacturer (Applied Biosystems, Foster City, CA, USA). A negative control with no DNA template was used. PCR was performed using GeneAmp-9700 thermal cycler (Applied Biosystems) with the following amplification conditions: an initial denaturation at 95 °C for 11 min, 28 cycles of 94 °C for 20 s, 59 °C for 3 min for both annealing and extension, and final extension of 60 °C for 10 min to enable full A-addition to the PCR products.

Table 1 The characteristics of the 15 STR loci used in the present study

Locus	Chromosome location	Repeat motif	Primer label
D8S1179	8q24.13	TCTA	6-FAM (blue)
D21S11	21q21.1	TCTA	
D7S820	7q21.11	GATA	
CSF1PO	5q33.1	AGAT	
D3S1358	3p21.31	TCTA	VIC (green)
TH01	11p15.5	AATG	
D13S317	13q31.1	TATC	
D16S539	16q24.1	GATA	
D2S1338	2q35-37.1	TGCC	
D19S433	19q12-13.1	AAGG	NED (yellow)
vWA	12p13.31	TCTA	
TPOX	2p25.3	AATG	
D18S51	18q21.33	AGAA	
D5S818	5q23.2	AGAT	PET (red)
FGA	4q31.3	TTTC	

Electrophoresis and Analysis

The amplified fragments were analyzed with the 3500 Genetic Analyzer (8-Capillary) (Applied Biosystems). One microliter of each PCR reaction product was added to 8.7 µL of deionized formamide (Sigma, St Louis, MO, USA) and 0.3 µL of GeneScan 500 LIZ size standard (Applied Biosystems). PCR products were denatured at 95 °C for 3 min and immediately chilled for 3 min. Electrophoresis was performed using the Performance Optimized Polymer pop 4 (Applied Biosystems) with a 36 cm/50 µm capillary at 15–20 kV for 30 min at 60 °C. Data were analyzed using GeneMapper® ID-X Software (Applied Biosystems) to assess the quality of PCR amplification and to assign the specific alleles to each fragment analyzed.

Statistical Analysis

POPGENE software package (Yeh et al. 1999) was used to calculate allele frequencies, observed number of alleles, effective number of alleles (Kimura and Crow 1964), observed (Ho) and expected (He) heterozygosity (Levene 1949) at each locus in the seven populations under study. The agreement of genotype frequencies with Hardy–Weinberg equilibrium (HWE) was determined using the X^2-test based on the number of observed and expected heterozygotes and the exact test based on the number of observed and expected genotypes, as implemented in the GENEPOP software version 4.6 (Raymond and Rousset 1995; Rousset 2008). The Bonferroni correction (Weir 1990) was applied while testing for Hardy–Weinberg equilibrium. Polymorphism Information Content (PIC) developed by Bostein et al. (1980) was calculated for each locus using Arlequin software version 3.1.1 (Excoffier et al. 2005). Arlequin software version 3.1.1 (Excoffier et al. 2005) was also used for the population differentiation tests through locus-by-locus comparisons of the allelic frequencies of Egyptian population with those from other Arab populations with similar loci coverage. These populations included Saudi Arabian (Osman et al. 2015), Sudanese (Babiker et al. 2011), Libyan (Khodjet-el-khil et al. 2012), Moroccan (Bentayebi et al. 2014), and Iraq (Barni et al. 2007).

Using the variance-base method of Weir and Cockerham (1984), population differentiation by Wright's F-statistics (Wright 1951, 1978) was computed using FSTAT version 2.9.3.2 computer program (Goudet 2002). Intrapopulation fixation indices or inbreeding coefficients (F_{IS} or f) were calculated. Interpopulation comparisons were conducted using F_{ST} coefficients of genetic differentiation and analysis of molecular variance (AMOVA) as implemented in the Arlequin software version 3.1.1 (Excoffier et al. 2005). Genetic distances among populations were estimated using standard genetic distance (Ds) of Nei (1972). Genetic relationships between populations were assessed through multidimensional scaling (MDS) plots of the genetic distance matrix, and both genetic distance and geographical distance matrices using SPSS software version 20. Regression analyses

for the correlation between genetic distance matrices and geographical distance matrices were also obtained from Mantel test. Mantel tests were implemented with 10,000 permutations using Arlequin software version 3.1.1 (Excoffier et al. 2005).

Results

Genetic Diversity

The number of alleles identified in the seven populations ranged between 105 and 138 alleles for the 15 STR loci. The highest number of alleles was recorded in Greater Cairo population, while the lowest number of alleles was recorded in Sinai population. For the combination of the seven Egyptian populations, the distribution of observed allele frequencies for the 15 STR loci and other measures of genetic diversity are shown in Table 2. The highest allelic frequency was allele 8 of TPOX (0.485), while the lowest allelic frequency (0.001) was recorded, for example, for allele 34 of D21S11. In addition, the less and the most frequent alleles varied within the examined populations. The most polymorphic markers in this study were D21S11, spanning 18 alleles over all populations, with allele 29 most frequent (0.350) in Canal population and D18S51, spanning 16 alleles, with allele 13 most frequent (0.250) in Sinai population.

The degree of polymorphism at each locus could also be expressed in terms of heterozygosity along with the PIC value. The PIC of the 15 STR loci was greater than 0.507 in all populations examined. The lowest average values were observed for the TPOX locus (Ho = 0.676, PIC = 0.617) and the highest average values were scored for the D18S51 locus (Ho = 0.868, PIC = 0.849) and the D2S1338 locus (Ho = 0.865, PIC = 0.843). Locus-by-locus comparisons with available published data on other Arab populations are shown in Table 3. After applying the Bonferroni correction, population differentiation tests showed that the Egyptian population had significant differences with Moroccan in 12 loci, Iraq in 8 loci, Saudi Arabian in 7 loci, Sudanese in 7 loci, and Libyan in 5 loci out of 15 loci.

Little deviations from Hardy–Weinberg equilibrium, based on the X^2 test, were observed in all populations studied except for Greater Cairo which exhibited agreement with Hardy–Weinberg equilibrium at all studied loci. In the other populations, one to four loci were rejected. However, across all Egyptian populations examined, deviations from Hardy–Weinberg equilibrium were detected for six loci, D8S1179, D3S1358, D13S317, D19S433, TPOX, and D18S51. After applying Bonferroni correction (testing on significance level of 0.003), only departure for D3S1358, D19S433, and D18S51 remained significant. Based on the exact test, no deviations from Hardy–Weinberg equilibrium were observed, with the exception of D8S1179 and D16S539 loci. After using the Bonferroni correction, the agreement with HWE was confirmed for all loci examined. The fixation index F_{IS} (inbreeding coefficient) was considerably different for different loci in all populations examined. It was also different for different populations ranging from 0.012 in Greater Cairo to 0.039 in

Table 2 Allele frequencies distribution for the 15 STR loci and measures of genetic diversity in Egyptian population ($n = 814$)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
5						0.001									
6						0.225						0.010			
7			0.010	0.004		0.212						0.001			
8	0.009		0.176	0.023		0.115	0.121	0.020				0.485		0.021	
9	0.006		0.092	0.014		0.354	0.068	0.150		0.002		0.175		0.069	
9.3						0.058									
10	0.072		0.326	0.297		0.030	0.063	0.078		0.005		0.085	0.001	0.130	
10.2													0.001		
11	0.074		0.250	0.252		0.002	0.256	0.367		0.015		0.221	0.015	0.235	
12	0.112		0.116	0.344	0.002		0.325	0.248		0.125	0.001	0.017	0.180	0.355	
12.2										0.003					
13	0.215		0.024	0.057	0.006		0.116	0.116		0.189	0.002	0.002	0.145	0.172	
13.2										0.020			0.001		
14	0.243		0.002	0.006	0.073		0.043	0.017		0.268	0.089		0.145	0.011	
14.2										0.066					
15	0.206				0.288		0.003	0.001	0.002	0.103	0.143		0.122	0.004	
15.2										0.099					
16	0.053				0.234				0.046	0.037	0.248		0.112		
16.2										0.044					
17	0.009				0.279				0.226	0.008	0.282		0.122		0.003
17.2										0.009					
18	0.001				0.102				0.119		0.148		0.077		0.017
19					0.014				0.146		0.069		0.039		0.059
20									0.129		0.010		0.019		0.093

Table 2 (continued)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
21									0.055		0.001		0.012		0.140
22									0.056		0.002		0.001		0.149
23									0.117				0.002		0.185
24		0.001							0.055						0.171
25									0.040						0.111
26		0.001							0.004						0.042
27		0.019													0.012
28		0.119													0.008
29		0.271													0.004
29.2		0.001													
30		0.257													0.002
30.2		0.022													
31		0.065													
31.2		0.068													
32		0.013													
32.2		0.098													
33		0.001													
33.2		0.047													
34		0.001													
34.2		0.002													
35		0.005													
36		0.006													
No	11	18	8	8	8	8	8	8	12	15	11	8	16	8	14
Ne	5.618	5.524	4.254	3.484	4.142	4.100	4.566	4.062	7.114	6.226	4.861	3.031	7.455	4.131	6.648

Table 2 (continued)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Ho	0.772	0.804	0.742	0.729	0.750	0.752	0.775	0.782	0.865	0.793	0.837	0.676	0.868	0.765	0.838
He	0.831	0.825	0.773	0.719	0.766	0.763	0.788	0.760	0.868	0.844	0.802	0.670	0.874	0.764	0.859
PIC	0.797	0.793	0.730	0.660	0.718	0.716	0.749	0.716	0.843	0.816	0.764	0.617	0.849	0.719	0.832
P (X^2)	0.004	0.539	0.106	0.969	0.001	0.688	0.004	0.103	0.621	0.000	0.713	0.004	0.000	0.604	0.932
P (Exact)	0.017	0.655	0.104	0.831	0.343	0.851	0.486	0.029	0.155	0.103	0.867	0.626	0.845	0.741	0.541

No observed number of alleles, *Ne* effective number of alleles, *Ho* observed heterozygosity, *He* expected heterozygosity, *PIC* Polymorphism Information Content, *P* (X^2) *P* value of X^2 test for HWE, *P (exact) P* value of exact test for HWE

Table 3 Population differentiation tests between Egyptian population and other neighboring/historically-related Arab populations

Population	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Saudi Arabian	0.8873	*0.0387*	*0.0256*	0.5371	**0.0000**	**0.0000**	**0.0000**	0.8120	**0.0000**	**0.0000**	0.2615	*0.0488*	**0.0000**	0.3386	**0.0000**
Sudanese	*0.0372*	**0.0026**	0.1253	0.1786	0.7820	**0.0000**	**0.0000**	*0.0274*	**0.0000**	*0.0250*	0.7648	**0.0000**	**0.0000**	*0.0263*	**0.0000**
Libyan	0.0712	**0.0022**	*0.0427*	*0.0216*	0.1465	*0.0331*	0.6112	**0.0000**	**0.0000**	*0.0274*	0.2895	0.3449	**0.0006**	0.4884	**0.0000**
Moroccan	*0.0262*	**0.0012**	0.6112	**0.0000**	0.3652	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0018**	**0.0000**
Iraq	*0.0263*	**0.0000**	0.8193	0.4174	*0.0080*	**0.0000**	**0.0000**	0.3652	*0.0413*	**0.0000**	**0.0000**	*0.0269*	**0.0028**	**0.0000**	**0.0028**

Significant P values before Bonferroni correction for multiple testing ($P < 0.05$) are shown in italics. Significant P values after Bonferroni correction ($P < 0.0033$) are shown in bold

Table 4 Measures of genetic diversity and fixation index F_{IS} or f values (inbreeding estimates) for the seven populations

	Populations						
	Northern coast	Delta	Greater Cairo	Canal governorates	Northern Upper Egypt	Southern Upper Egypt	Sinai
n	120	120	125	115	120	110	104
No	126	130	138	120	130	118	105
Ho	0.7840	0.7815	0.7860	0.7773	0.7850	0.7770	0.7766
He	0.7952	0.7934	0.7962	0.7944	0.7937	0.7960	0.7892
PIC	0.758	0.759	0.764	0.756	0.761	0.751	0.736
F_{IS}	0.018	0.021	0.012	0.020	0.017	0.032	0.039

n number of individuals, *No* observed number of all alleles, *Ho* observed heterozygosity, *He* expected heterozygosity, *PIC* Polymorphism Information Content

Sinai (Table 4). The fixation index F_{IS} values were 0.039 ($P=0.0028$) and 0.032 ($P=0.0042$) in Sinai and Southern Upper Egypt, respectively.

Genetic Structure

Hierarchical analysis of molecular variance (AMOVA) over 15 STR loci in Egyptian populations is shown in Table 5. The results of AMOVA analysis revealed significant variation among groups (1.24%, $P<0.05$). The most variation (99.71%, $P<0.0001$) was found among individuals of the seven populations when treated as a single group than variation among populations within groups (3.81%, $P<0.001$). The degree of genetic differentiation between all populations was estimated through F_{ST} coefficient of genetic differentiation. Table 6 displays the values of the coefficient of genetic differentiation F_{ST} across the 15 STR loci in the seven population pairs. The estimated F_{ST} values between most population pairs were smaller than 0.01. However, larger estimates of F_{ST} values were observed between population of Sinai and the other six populations, except between Sinai and Delta (0.0099). These values ranged from 0.0114 (between Sinai and Greater Cairo) to 0.0180 (between Sinai and Southern Upper Egypt). Results of pairwise AMOVA analysis revealed significant differences (at the level $P=0.05$) between Sinai population and the other six populations. Moreover, after applying Bonferroni correction, such differences remained significant.

Table 7 displays the genetic relationships among the Egyptian populations examined based on Nei's distance (Nei 1972) and the geographical distances between them. Genetic distance matrix showed that the highest genetic distance was found between Sinai and Southern Upper Egypt populations (0.1530). The lowest value for genetic distance was found between Greater Cairo and Delta populations (0.0199). Multidimensional Scaling (MDS) analyses were constructed to illustrate the genetic distances between the seven Egyptian populations and to display the relationship between the genetic distances and the geographical distribution of these populations

Table 5 Hierarchical analysis of molecular variance (AMOVA) over 15 STR loci in Egyptian samples

Source of variation								
Among groups			Among populations within groups			Within populations		
Percent variance	Covariance estimates	*P*	Percent variance	Covariance estimates	*P*	Percent variance	Covariance estimates	*P*
1.24	0.0120 Va	< 0.05	3.81	0.0189 Vb	< 0.001	99.71	0.5000 Vc	< 0.0001

Table 6 Pairwise analysis of AMOVA (P values), above diagonal, and Wright's F_{ST} estimates among the Egyptian populations (below diagonal)

Population	Northern coast	Delta	Greater Cairo	Canal governorates	Northern Upper Egypt	Southern Upper Egypt	Sinai
Northern coast	–	0.3619	0.5933	0.5857	0.4571	0.4071	0.0023
Delta	−0.0001	–	0.7571	0.8547	0.6333	0.2595	0.0023
Greater Cairo	0.0001	−0.0006	–	0.6428	0.3286	0.3524	0.0024
Canal governorates	0.0010	−0.0000	−0.0002	–	0.6691	0.5309	0.0024
Northern Upper Egypt	0.0011	−0.0002	0.0007	0.0003	–	0.5405	0.0024
Southern Upper Egypt	0.0013	0.0003	−0.0013	−0.0016	−0.0010	–	0.0015
Sinai	0.0160	0.0099	0.0114	0.0139	0.0131	0.0180	–

Table 7 Geographical distances (Km) between the population pairs (above diagonal) and genetic distances, Nei 1972 (below diagonal)

Population	Northern coast	Delta	Greater Cairo	Canal governorates	Northern Upper Egypt	Southern Upper Egypt	Sinai
Northern coast	–	325.64	351.49	429.45	435.45	847.22	570.28
Delta	0.0317	–	82.81	123.26	284.34	665.14	269.79
Greater Cairo	0.0290	0.0199	–	116.44	205.71	585.76	273.53
Canal governorates	0.0419	0.0322	0.0278	–	299.70	628.12	156.29
Northern Upper Egypt	0.0355	0.0239	0.0238	0.0324	–	415.20	452.30
Southern Upper Egypt	0.0620	0.0525	0.0425	0.0514	0.0461	–	689.21
Sinai	0.1237	0.0932	0.0961	0.1155	0.1054	0.1530	–

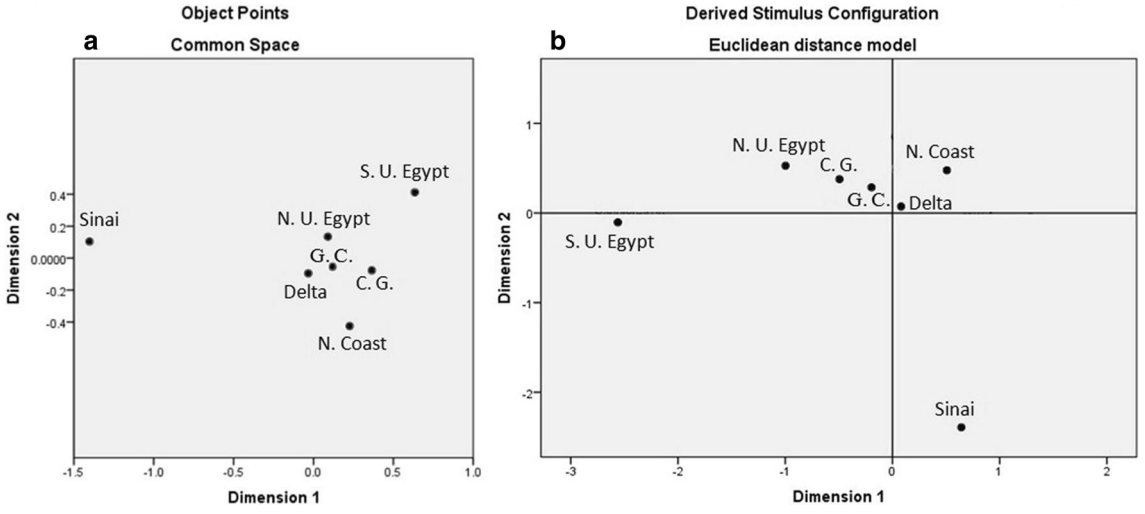

Fig. 2 Multidimensional scaling (MDS) plots based on genetic distances (**a**) and both genetic distances and geographical distances (**b**) for the seven Egyptian populations. The populations are indicated by filled circles. Populations of Sinai and Southern Upper Egypt which show the highest genetic distance are located far away from the other populations. G. C., Greater Cairo; C. G., Canal Governorates; N. Coast, Northern Coast; N. U. Egypt, Northern Upper Egypt; S. U. Egypt, Southern Upper Egypt; Delta; Sinai

(Fig. 2a, b). Plots showed clustering of populations of Greater Cairo, Delta, Canal governorates, and Northern Upper Egypt into one group. Northern Coast population was located slightly away from this group. In contrast, those of Sinai and Southern Upper Egypt were located furthest from the latter group. The correlation between the genetic distances and the geographical distances, or isolation by distance (IBD), of the populations studied was also analyzed using Mantel test (Fig. 3a–c). Mantel test displayed a non-significant tendency of IBD ($r=0.387$, $P=0.091$). Excluding Sinai sample from the population pairs resulted in a significant correlation ($r=0.917$, $P=0.0001$). In addition, all population samples versus Sinai sample showed a significant correlation ($r=0.933$, $P=0.0001$).

Discussion

Genetic Diversity

The present study assessed the genetic structure and diversity among seven Egyptian populations based on 15 STR loci. For the combination of the seven Egyptian populations, the highest allelic frequency was allele 8 of TPOX (0.485), while the lowest allelic frequency (0.001) was recorded, for example, for allele 34 of D21S11. In other studies, the most frequent allele for STR loci can vary depending on the population, for example, 12 (0.4093) at D5S818 in Slovenian population (Drobnic et al. 2005), 15 (0.4143) at D3S1358 in Bolivians (Cifuentes et al. 2008), and 8 (0.4890) at TPOX among Wallachians in South Romania (Stanciu et al. 2009). In studies on the central region of Saudi Arabia (Riyadh Province) by Osman et al. (2015), the highest allelic frequency was allele 10 of TPOX (0.558). In Iraq population, the highest allelic frequency was allele 13 of D5S818, while the lowest allele frequency was allele 13 of D3S1358 (Hadi et al. 2014). The present data showed that D21S11, spanning 18 alleles over all populations studied, and D18S51, spanning 16 alleles, were the most polymorphic markers. Osman et al. (2015) have recorded that the most polymorphic loci observed in the central region of Saudi Arabia were D18S51, FGA, and D21S11 defined by 17, 15, and 15 alleles, respectively.

The PIC of the 15 STR loci was found to be greater than 0.507 in all populations examined indicating that these markers are highly polymorphic. According to the measures of PIC and the observed heterozygosity (Ho), the lowest average values were observed for the TPOX locus (Ho = 0.676, PIC = 0.617) and the highest average values were scored for the D18S51 locus (Ho = 0.868, PIC = 0.849) and the D2S1338 locus (Ho = 0.865, PIC = 0.843). Similarly, according to Jeran et al. (2007), the highest values in Montenegro population were scored for D18S51 (Ho = 0.870, PIC = 0.856) and D2S1338 (Ho = 0.858, PIC = 0.844). Chan et al. (2005) have observed that the heterozygosity in Chinese Hong Kong population ranged from 0.572 (TPOX) to 0.883 (D8S1179). The PIC of the 15 loci in the central region of Saudi Arabia has been reported to be greater than 0.570 by Osman et al. (2015). According to their studies, the highest values were detected in D19S433 locus (Ho = 0.869, PIC = 0.850), while the lowest values were identified in TPOX (Ho = 0.621, PIC = 0.570), in agreement with the present study.

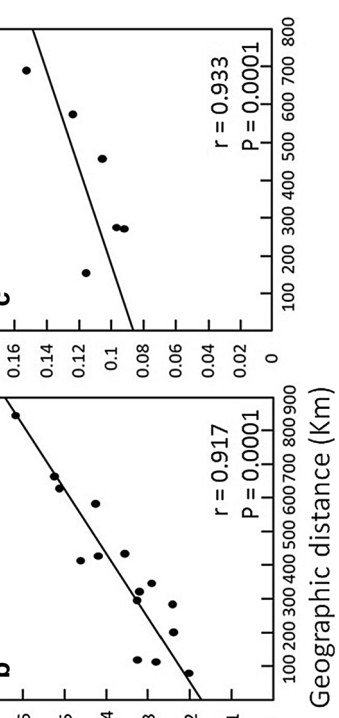

Fig. 3 Isolation by distance patterns for genetic distances between all populations (**a**), between all populations excluding Sinai population (**b**) and between populations versus Sinai population (**c**). Lines represent the least square regression between genetic distance and geographical distance. R statistics and *P* values from Mantel tests are indicated

The population differentiation tests were evaluated between Egyptian population and other Arab populations. The most significant difference was observed between Egyptian population and Moroccan (at 12 loci), followed by Iraq (at 8 loci), Saudi Arabian (at 7 loci), Sudanese (at 7 loci), and Libyan (at 5 loci). In general, an increase in the number of significant differences between the Egyptian dataset and those from the other populations might be closely correlated with an increase in the geographical distance and/or decrease in the amount of historical contact in between.

The agreement with Hardy–Weinberg equilibrium (HWE), based on the X^2-test, was confirmed for all studied loci in Greater Cairo. In the other populations, one to four loci were rejected. Furthermore, across all Egyptian populations examined, departure from HWE was detected for six loci, D8S1179, D3S1358, D13S317, D19S433, TPOX, and D18S51. Based on the exact test, only D8S1179 and D16S539 loci showed departure from HWE. Thus, Egyptian population showed a certain degree of deviation from HWE (six loci based on the X^2-test, and two loci based on the exact test, on the level of $P = 0.05$). This deviation is explained by the excess homozygosity at some loci or due to consanguineous marriage among Egyptian populations (Shawky et al. 2011; Ahmed 2017).

The estimates of fixation index F_{IS} (inbreeding coefficient) were considerably different for different loci in all populations examined pointing to the existence of population subdivisions. They were also different for different populations. The fixation index F_{IS} values in Sinai and Southern Upper Egypt indicate that individuals within each of these populations are more related by consanguinity than in any other Egyptian population investigated. Omran et al. (2009) have reported the presence of insignificant genetic differentiation and homogeneity of Southern Upper Egypt population. Estimates of F_{IS} indicate that the population of Greater Cairo had the highest within-population variation. This suggests random mating and low consanguinity in Greater Cairo population compared with the other studied populations. Also, the noteworthy variation within population of Greater Cairo could be attributed to the continuous movement of people from different governorates in Egypt to this metropolitan area which represents a major economic center in Egypt. Furthermore, this culturally diverse city had witnessed many different historical periods throughout the ages (Shaw 2004; Marsot 2007). Pharaonic, Greek, Roman, Coptic, Islamic, French, and British impact must be considered in this regard.

Genetic Structure

Analysis of molecular variance (AMOVA) in Egyptian populations showed that the most variation was found among individuals of the seven populations when treated as a single group than variation among populations within groups. The degree of genetic differentiation between all populations was estimated through F_{ST} coefficient of genetic differentiation. The estimated F_{ST} values between most population pairs were smaller than the value of 0.01, the conservative value recommended by the National Research Council (Weir 1996). However, larger estimates of F_{ST} values, which indicated more genetically isolated populations, were observed between

population of Sinai and the other six populations, except between Sinai and Delta (0.0099). These values ranged from 0.0114 (between Sinai and Greater Cairo) to 0.0180 (between Sinai and Southern Upper Egypt). These F_{ST} values are consistent with expectations for more isolated groups, such as Sinai population. Results of pairwise AMOVA analysis revealed significant differences between the geographically isolated Sinai population and the other six populations.

The present data demonstrated that the genetic structure observed in the examined Egyptian populations is correlated with the ethnic affiliation and geography. Genetic distance matrix showed that the highest genetic distance was found between Sinai and Southern Upper Egypt populations. The lowest value for genetic distance was found between Greater Cairo and Delta populations. The data indicated that the degree of relatedness between populations of Greater Cairo (mostly urban people) and Delta (Fallahi people), short-distance geographical regions, is high and that between populations of Sinai (Bedouin people) and Southern Upper Egypt (Sa'idi and Nubian people), far distant geographical regions, is low. Multidimensional Scaling (MDS) analyses were constructed to illustrate the genetic distances between the seven Egyptian populations and to display the relationship between the genetic distances and the geographical distribution of these populations. The analyses clustered the populations of Greater Cairo, Delta, Canal governorates, and Northern Upper Egypt into one group. Northern Coast population was located slightly away from this group. In contrast, those of Sinai and Southern Upper Egypt were located furthest from the latter group. The correlation between the genetic distances and the geographical distances, or isolation by distance (IBD), of the populations studied was also demonstrated using Mantel test. Mantel test displayed a non-significant tendency of IBD among populations studied. The non-significance of this correlation was driven by Sinai population. Excluding Sinai sample from the population pairs resulted in a significant correlation. In addition, all population samples versus Sinai sample showed a significant correlation. This indicates that the patterns of genetic variation do not depend only on geographical distances, but also on other factors, such as geographical isolation. The same conclusion was reported by Collin et al. (2013).

The patterns of genetic variation also seem to bear some relationship to the language and dialects. Egypt comprises a variety of the Arabic languages (dialects), which represent a Semitic branch of the Afroasiatic language family (Gary and Gamal-Eldin 1982; Versteegh 2001). The daily language of the Egyptians is the local variety of Egyptian dialects, known as *Egyptian Arabic*, also called *Hadar* (civilized) *dialect* or *Cairene Arabic* or *Masri* (Woidich 1994; Haeri 1997; Bassiouney 2015). Most Egyptians living in Greater Cairo speak *Egyptian Arabic*. Additionally, Egyptians living in Delta have high genetic similarity to those living in Greater Cairo, speak a variety of *Fallahi Arabic* or *Bahrawi*, a dialect that is much like *Egyptian Arabic* (Woidich 1996). Moreover, sizable ethnic groups of Egyptians living in Upper Egypt speak a variety of *Sa'idi Arabic* (Northern Upper Egypt and many regions of Southern Upper Egypt) and *Nubian* language (Lower Nubia in Southern Upper Egypt) (Rouchdy 1991; Nishio 1994). Northern Upper Egypt and Southern Upper Egypt vicinity along with sharing to some extent the *Sa'idi Arabic* dialect is reflected in the genetic similarity between them. Bedouins in Sinai speak

Bedawi Arabic (Jong 2000, 2011). They exhibited high genetic distance with people of Upper Egypt, particularly Southern Upper Egypt. Northern Coast comprises Egyptians who speak *Alexandrian dialect* and *Bedawi Arabic* (Eltouhamy, 2015). The *Alexandrian dialect* is affected by cultural influences from Italians and Greeks, explaining the cultural diversity of Alexandrian people. Population of Northern Coast has its highest genetic similarity with those of Greater Cairo and Delta. People of Northern Coast have various interactions with Greater Cairo and Delta.

The rate of movement of people up and down the Nile corridor, the extent of east–west contact, and the geographical distribution are thus among the mechanisms that have produced different patterns of genetic relationships among Egyptian populations. The study of population admixture levels among Egyptian populations remains an open subject for further investigation.

Conclusion

Based on allelic frequencies and other population statistical analyses of the population genetic structure using AmpF/STR Identifiler detection system, the combination of the 15 STR loci proved to be significant for identifying the genetic variation within and between Egyptian populations. Furthermore, the present study demonstrated the correlation between the genetic relationships of populations and their dialect diversity and geographical distribution.

Acknowledgements The authors gratefully thank the volunteers whose cooperation made this study possible. The authors also would like to thank the Forensic Science Laboratory Unit, Police Academy, Cairo, Egypt, for the support during the investigation, also the Egyptian Ministry of Interior for its cooperation. The authors would like to express their deepest gratitude and thanks to Prof. Mohammed Saber, Cell Biology Department, National Research Center, Egypt, Dr. Sherif Abdel-Ghany, Faculty of Agriculture, Cairo University, Egypt, and Dr. Abdullah Agha, Faculty of Science, Taibah University, Saudi Arabia, for their kind support and suggestions.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interests.

References

Ahmed SM (2017) Consanguineous marriage among Egyptian youth: secondary analysis of survey of young people in Egypt, 2014. EJCM 35:85–94

Babiker HM, Schlebusch CM, Hassan HY, Jakobsson M (2011) Genetic variation and population structure of Sudanese populations as indicated by 15 Identifiler sequence-tagged repeat (STR) loci. Invest Genet 2:1–13

Barbarii LE, Rolf B, Constantinescu C, Hohoff C, Calistru P, Dermengiu D (2004) Allele frequencies of 13 short tandem repeat (STR) loci in the Romanian population. Forensic Sci Int 141:171–174

Barni F, Berti A, Pianese A, Boccellino A, Miller MP, Caperna A, Lago G (2007) Allele frequencies of 15 autosomal STR loci in the Iraq population with comparisons to other populations from the middle-eastern region. Forensic Sci Int 167:87–92

Bassiouney R (2015) Language and identity in modern Egypt. Edinburgh University Press, Edinburgh

Bentayebi K, Abada F, Ihzmad H, Amzazi S (2014) Genetic ancestry of a Moroccan population as inferred from autosomal STRs. Meta Gene 2:427–438
Borosky A, Toscanini U, Gómez AA, Parolín ML, Basso N, Vullo C (2014) Forensic population data for 20 STR loci in Argentina. Forensic Sci Int Genet 13:e20–e21
Bostein D, White RL, Skolnick M, Davis RW (1980) Construction of genetic linkage maps in man using restriction fragment length polymorphism. Am J Hum Genet 32:314–331
Butler JM (2006) Genetics and genomics of core STR loci used in human identity testing. J Forensic Sci 51:253–265
Chan KM, Chiu CT, Tsui P, Wong DM, Fung WK (2005) Population data for the identifier™ 15 STR loci in Hong Kong Chinese. Forensic Sci Int 152:307–309
Cifuentes L, Jorquera H, Acuña M, Ordóñez J, Sierra AL (2008) Allele frequencies for 12 autosomal short tandem repeat loci in two Bolivian populations. Genet Mol Res 7:271–275
Collin H, Burri R, Comtesse F, Fumagalli L (2013) Combining molecular evolution and environmental genomics to unravel adaptive processes of MHC class IIB diversity in European minnows (*Phoxinus phoxinus*). Ecol Evol 3:2568–2585
Drobnic K, Bakal N, Marjanovic D (2005) Allele frequencies for the 15 short tandem repeat loci in Slovenian population. J Forensic Sci 50:1–3
Eltouhamy I (2015) Language attitudes towards dialects of Arabic in Egypt. MA Thesis, The American University in Cairo, Egypt
Excoffier L, Laval G, Schneider S (2005) Arlequin Ver. 3.1.1: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:147–150
Gary JO, Gamal-Eldin S (1982) Cairene Egyptian colloquial Arabic. Lingua Descriptive Studies, Vol. 6, North-Holland, Amsterdam
Goudet J (2002) FSTAT computer package for PCs. Institute of Ecology, UNIL, Lausanne
Hadi I, Abdullah M, Jaber A, Yoke C (2014) Genetic variation of twenty autosomal STR loci and evaluate the importance of these loci for forensic genetic purposes. Afr J Biotechnol 13:1210–1218
Haeri N (1997) The sociolinguistic market of Cairo. Paul Kegan International, London
Halima MSHA, Bernal LP, Sharif FA (2009) Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip. Leg Med (Tokyo) 11:203–204
Jeran N, Havas D, Ivanovic V, Rudan B (2007) Genetic diversity of 15 STR loci in a population of Montenegro. Coll Antropol 3:847–852
Jong R (2000) A grammar of the Bedouin Dialects of the Northern Sinai Littoral: bridging the linguistic gap between the eastern and western Arab world. Brill, Leiden
Jong R (2011) A grammar of the Bedouin Dialects of Central and Southern Sinai. Brill, Leiden
Keita SO, Boyce AJ (2005) Genetics, Egypt, and history: interpreting geographical patterns of Y chromosome variation. Hist Afr 32:221–246
Khodjet-el-khil H, Fadhlaoui-Zid K, Gusmao L, Alves C, Benammar-Elgaaied A, Amorism A (2012) Allele frequencies for 15 autosomal STR markers in the Libyan population. Ann Hum Biol 39:80–83
Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738
Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Paabo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration. Am J Hum Genet 64:1166–1176
Levene H (1949) On a matching problem in genetics. Ann Math Stat 20:91–94
Marsot A (2007) A history of Egypt from the Arab conquest to the present. Cambridge University Press, Cambridge
Midant-Reynes B (2000) The prehistory of Egypt: from the first Egyptians to the first Pharaohs. Blackwell, Oxford
Mortera J, Vecchiotti C, Zoppis S, Merigioli S (2016) Paternity testing that involves a DNA mixture. Forensic Sci Int Genet 23:50–54
Nei M (1972) Genetic distance between populations. Am Nat 106:283–292
Nishio T (1994) The Arabic Dialect of Qifṭ (Upper Egypt): grammar and classified vocabulary. Asian & African Lexicon 27, Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo
Omran GA, Rutty GN, Jobling MA (2009) Genetic variation of 15 autosomal STR loci in Upper (Southern) Egyptians. Forensic Sci Int Genet 3:e39–e44
Osman AE, Alsafar H, Tay GK, Theyab J, Mubasher M, Eltayeb-El Sheikh N, AlHarthi H, Crawford MH, El Ghazali G (2015) Autosomal short tandem repeat (STR) variation based on 15 loci in a population from the central region (Riyadh Province) of Saudi Arabia. J Forensic Res 6:267–271

Raymond M, Rousset F (1995) GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. J Hered 86:248–249
Rouchdy A (1991) Nubians and the Nubian language in contemporary Egypt: a case of cultural and linguistic contacts. Brill, Leiden
Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res 8:103–106
Shaw I (2004) Ancient Egypt: a very short introduction. Oxford University Press, New York
Shawky RM, El-Awady MY, Elsayed SM, Hamadan GE (2011) Consanguineous mating among Egyptian population. EJMHG 12:157–163
Stanciu F, Stoian IM, Popescu OR (2009) Population data for 15 short tandem repeat loci from Wallachia region, South Romania. Croat Med J 50:321–325
Thomson JA, Pilotti V, Stevens P, Ayres KL, Debenham PG (1999) Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. Forensic Sci Int 100:1–16
Versteegh K (2001) The Arabic language. Edinburgh University Press, Edinburgh
Weir BS (1990) Genetic data analysis: methods for discrete population genetic data. Sinauer Associates, Sunderland
Weir BS (1996) The second national research council report on forensic DNA evidence. Am J Hum Genet 59:497–500
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
White JM (2011) Everyday life in ancient Egypt. Dover Publications, New York
Woidich M (1994) Cairo Arabic and the Egyptian dialects. In Caubet D, Vanhove M (eds) Actes des Premières Journees Internationales de Dialectologie Arabe de Paris. Paris, pp 493–507
Woidich M (1996) Rural Dialect of Egyptian Arabic: an overview. Égypte-monde arabe 27–28:325–354
Wright S (1951) The genetic structure of populations. Ann Eugen 15:323–354
Wright S (1978) Evolution and the genetics of populations: variability within and among natural populations, vol 4. University of Chicago Press, Chicago
Yao J, Wang LM, Gui J, Xing JX, Xuan JF, Wang BJ (2016) Population data of 15 autosomal STR loci in Chinese Han population from Liaoning Province, Northeast China. Forensic Sci Int Genet 23:e20–e21
Yeh FC, Boyle T, Rongcai Y, Ye Z, Xian JM (1999) POPGENE, Version 1.31. A Microsoft Window based free ware for population genetic analysis. University of Alberta, Edmonton

Affiliations

Ahmed F. AbdEl-Hafez[1] · Sherif H. El-Alfy[2,3] · Hamdy H. Swelim[2] · Nagwa H. A. Hassan[2]

[1] General Authority of Forensic Evidence, Police Academy, Cairo, Egypt

[2] Department of Zoology, Faculty of Science, Ain Shams University, Cairo, Egypt

[3] Department of Biology, Faculty of Science, Taibah University, Almadinah Almunawarah, Saudi Arabia