Analyses of DNA Short Tandem Repeat Loci Database with Popgene

DNA-STR Lokusları ile Oluşturulan Veri Tabanının Popgene İstatistik Programı ile Analizi

İ. Ömer BARLAS, MD[a]
Ersi ABACI KALFOĞLU, MD[b]
E. Hülya YÜKSELOĞLU,[b]
S. Şebnem ÖZCAN,[b]
Gavril PETRIDIS,[b]

[a]Department of Medical Biology and Genetics,
Mersin University Faculty of Medicine, Mersin
[b]İstanbul University
Institute of Forensic Sciences, İstanbul

Geliş Tarihi/*Received:* 14.08.2009
Kabul Tarihi/*Accepted:* 08.02.2010

Yazışma Adresi/*Correspondence:*
İ. Ömer BARLAS, MD
Mersin University Faculty of Medicine, Department of Medical Biology and Genetics, Mersin, TÜRKİYE/TURKEY
iobarlas@gmail.com

ABSTRACT Objective: Genetic distance is the term which refers to the genetic relationship between populations and subpopulations. Turkey manifests an intricate genetic constitution due to numerous gene flow and genetic admixture between its seven geographical regions. Therefore, a different set of allele frequency database for each region is expected to be addressed in forensic cases studies. Aim: Using POPGENE 1.31, we aimed to analyze short tandem repeat (STR) loci in seven geographical regions of Turkey and to determine the differences between them and the most appropriate frequencies likely to be used in forensic and anthropological studies. **Material and Methods:** The DNA from unrelated healthy individuals from seven geographical regions were typed for the STR loci CSF1PO, D7S820, D13S317, D16S539, F13AO1, FES/FPS, THO1, TPOX, vWA. Data were evaluated with POPGENE, version 1.31 program and compared statistically. **Results:** The obtained data were analyzed in terms of Fst (F-statistics), observed-expected heterozygosities, allele frequencies and whether they meet Hardy-Weinberg equilibrium. The results showed no significant difference in STR loci between geographical areas (FST= 0.0239). To conclude, allele frequencies are similar and genetic distance matrix has a similar typology in seven geographical regions of Turkey. The Turkish population has a heterogeneous genetic profile and a single database sufficiently represents the population in forensic studies.

Key Words: Forensic genetics; microsatellite repeats

ÖZET Genetik uzaklık, populasyonlar ve alt-populasyonlar arasındaki genetik ilişkiyi ortaya çıkaran bir kavramdır. Türkiye, yedi coğrafi bölgesi arasında gen akımları ve genetik karışımların sonucu olarak çok karışık genetik bir yapı sergilemektedir. Bu duruma bağlı olarak, adli amaçlı vaka çalışmalarında, her bölgenin farklı bir allel sıklığı veri tabanıyla temsil edilmesi beklenmektedir. **Amaç:** Bu çalışmada, STR lokuslarını inceleyip POPGENE- version 1.31 programında istatistiksel olarak analiz ederek Türkiye'nin yedi coğrafik bölgesi arasındaki herhangi bir genetik farklılığın olup olmadığını ortaya çıkararak adli bilimler ve antropolojik çalışmalarda kullanılabilecek en uygun frekansları belirlemek istedik. **Gereç ve Yöntemler:** Bu yedi bölgede yaşayan, aralarında akrabalık bağı bulunmayan sağlıklı bireylerden elde edilen DNA'lar CSF1PO, D7S820, D13S317, D16S539, F13AO1, FES/FPS, THO1, TPOX, vWA STR lokusları için tiplendirilmiştir ve istatistiksel olarak karşılaştırılmıştır. **Sonuç:** Elde edilen bilgiler, Fst (F-istatistiği), beklenen-gözlenen heterozigotluklar, allel frekansları ve Hardy-Weinberg'e uyum açısından analiz edilmiştir. Sonuçlar, coğrafik bölgeler arasında STR lokus açısından istatistiksel olarak belirgin bir farklılık olmadığını göstermiştir (FST=0.0239). Sonuç olarak ülkenin yedi coğrafi bölgesinde allel sıklıklarının benzer olduğu ve genetik uzaklık matriksinin benzer topolojiye sahip olduğu görülmüştür. Türk populasyonunun homojen bir genetik profile sahip olduğu ve adli çalışmalarda tek bir veri tabanıyla temsil edilmesinin yeterli olacağı gösterilmiştir.

Anahtar Kelimeler: Adli genetik; mikrosatelit tekrarları

Turkiye Klinikleri J Foren Med 2010;7(1):28-32

Turkey manifests an elaborate genetic constitution reflecting the consequences of numerous gene flow, admixture and local genetic differentiation due to migrations.[1,2]

The question that we ask in this paper is whether the possible degree of fluctuations is proportional to the DNA-level differences and similarities within the Turkish population. Standard genetic distance formula, the calculation of the codon, nucleotide or allele differences, is the most useful one defined by Nei to determine the genetic relationship (distance or similarity) for both the individuals and populations. This approach might be acceptable when the characteristics of the biological units studied are considered since every genetic marker mutates differently and is inherited independently. The genetic distance of an individual to a population was regarded as the average of distances between the individuals and the members of the population sample.[3-5]

The method of inferring relationships between individuals and assigning an individual to the "closest" population in anthropology and forensic genetics is based on the population frequencies of the observed alleles and on the conditional probabilities of the observed genotypes at the microsatellite loci. Using relatively few loci, it is possible to show the variability between ethnic groups and subpopulation structures.[6,7]

Turkey manifests an intricate genetic constitution due to economical, environmental and cultural influences, reflecting the consequences of numerous gene flow and genetic admixture between its seven geographical regions (Figure 1).

Since there is an obvious geographical distance and physical separation between residential parts of the country, the genetic marker frequencies are expected to be different. Therefore, a different set of frequency database for each region is expected to be addressed in case studies. In this study we examined the genetic distances between seven regions of Turkey to specify the most appropriate frequencies to be used in anthropologic and forensic studies.

MATERIALS AND METHODS

The current study was approved by the Local Ethics Committee. In a total of 203 samples from 7 different geographical regions of Anatolia were randomly selected and studied. The samples were collected from randomly selected, unrelated, healthy donors, (interviewed personally), who were the inhabitants of the studied regions over at least 3 generations.

Short tandem repeat (STR) typing (CSF1PO, D7S820, D13S317, D16S539, F13AO1, FES/FPS, THO1, TPOX, vWA) procedures were carried out as described in the literature.[8-13] All allele frequencies were obtained by direct counting. F-statistics, Nei's unbiased genetic identity and genetic distance,[14] expected and observed heterozygosities/homozygosities, allele frequencies and their conformity to Hardy-Weinberg equilibrium were estimated on PopGene, Version 1.31 program.[15]

FIGURE 1: Map of geographical regions of Turkey (left) and Turkey between Europe and Asia (right)

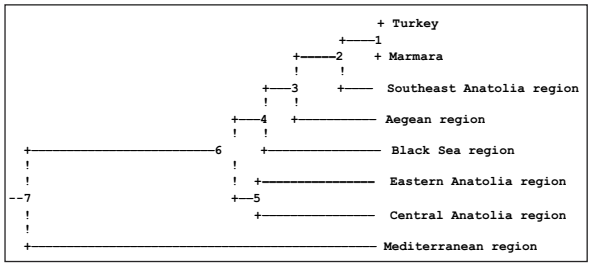

FIGURE 2: The dendogram based on data generated in the present study. Nei's genetic distance was used in the construction of the dendrogram.

The dendrogram was constructed with Nei's genetic distance measure.[3,14]

RESULTS AND DISCUSSION

The genotype and allele frequency distributions for all of the loci do not significantly deviate from the HWE expectations. That shows negligible inbreeding effects. There was no significant difference in allele frequency values between all regions. In the search of the relationships among the regions, dendrogram was constructed (Figure 2). Using the allele frequency data of the nine STR loci, the distance matrix gave a similar topology. Thus, all regions forming the Turkish population are genetically related. The genetic distance measures were used to assess the relatedness of the regions.

The number of the individuals in some regions was lower since the samples were selected randomly. However, this disadvantage was minimized using Nei's unbiased estimator.[4,16] The expected heterozygosity ratio of Turkey was 0.7482 whereas the mean heterozygosity ratio of some regions was over 0.72 except the Mediterranean region. This might have been due to the geographic localizations of the cities in this area. The Mediterranean region is a costal zone, where the residential districts are far from each other, restricting gene flow between them. Depending on the heterozygosities, the lower homozygosity ratios pointed out that there was no accumulation in any loci and also no significant genetic difference between the regions due to the admixture of genetic markers between the districts. These outcomes were confirmed with the F-statistics analysis (Table 1), a main indicator of a genetic differentiation levels in a population.[3,4] The mean F_{ST} value was found to be 0.0239, showing no significant fixation in any of the loci due to subpopulation formation.[17]

According to the dendrogram analysis (Figure 2), Marmara region was in the same cluster (cluster1) as Turkey. For these two populations Nei's genetic similarity was 1.000 and genetic distance was 0.0000 (Table 2). These values were derived from the similarities of the allele frequencies, showing a randomized distribution of the genetic markers. This finding was not surprising because the Marmara region is the most crowded district of Turkey since it has experienced migration from every parts of the country.

Southeast Anatolia is in the same cluster (2) with cluster 1 (namely Marmara). This clustering can be attributed to the fact that the social and eco-

TABLE 1: Nei's unbiased genetic distance measures for seven regions of Turkey

Population	Turkey	Marmara	Mediterranean	Eastern A.	Aegean	Southeast A	Central A.	Black Sea
Turkey	****	1.0000	0.9485	0.9860	0.9876	0.9908	0.9934	0.9929
Marmara	0.0000	****	0.9307	0.9813	0.9833	0.9932	0.9876	0.9830
Mediterranean	0.0528	0.0718	****	0.9382	0.9230	0.9082	0.9302	0.9461
Eastern A.	0.0141	0.0189	0.0638	****	0.9584	0.9606	0.9754	0.9542
Aegean	0.0125	0.0168	0.0801	0.0425	****	0.9776	0.9493	0.9702
Southeast A	0.0092	0.0069	0.0963	0.0402	0.0226	****	0.9642	0.9623
Central A	0.0066	0.0124	0.0724	0.0249	0.0520	0.0365	****	0.9739
Black Sea	0.0071	0.0171	0.0554	0.0469	0.0303	0.0384	0.0264	****

Upper diagonal: Nei's genetic identity
Lower diagonal: Nei's genetic distance

TABLE 2: FST estimates for the 7 region population samples

Locus	Fst
CSF1PO	0.0320
D7S820	0.0232
D13S317	0.0150
D16S539	0.0138
F13A01	0.0369
FES/FPS	0.0372
THO1	0.0203
TPOX	0.0205
vWA	0.0174
Fst over all loci	0.0239

* FST (F statistics) estimated according to Hartl and Clarck [28]

nomical disadvantages in the Southeast Anatolia region forcing precincts people to migrate and settle to the most abundant residential area of Turkey for years.

Aegean region together with cluster 2 constitutes the 3rd cluster. The genetic distance between Aegean and Southeast Anatolia regions (0.0226) (Table 2) was one of the closest identities, denoting that the Aegean region is the second biggest pool influenced by the gene flow from southeast of Turkey after Marmara. Considering the topologies of the clusters 2 and 3 (Figure 2) and genetic measure (Table 2) together, one can see that these districts are genetically similar. This similarity is originated not only from migration from Southeast Anatolia, but also slow and gradual population (cultural/educational) exchange between two well developed regions, Marmara and Aegean, for a long period of time.

Black Sea region constitutes the 4th cluster (fig 2) together with the 3rd cluster (Marmara, Aegean, and Southeast Anatolia), the genetic similarity of the regions represented by 0.0171, 0.303, 0.384 respectively (Table 2). This similarity is due to the geographic localization of Black Sea region. The high mountains, lying parallel to the Black sea, constrain the populations in a narrow coastline on the north side of Turkey. These mountains, while restricting gene flow from North to South, only allow migration from Eastern Anatolia to Black Sea region and from there to the Marmara.

Cluster 5 (Eastern Anatolia and Central Anatolia) can be analyzed separately (Figure 2). Eastern Anatolia is the most underdeveloped region of Turkey due to its hard climate and geographic features. This situation forced precinct people to migrate to the modest areas following the only overland road across from Central Anatolia and from there to Marmara. This drift causes a genetic admixture of two neighboring districts, lowering the genetic distance to 0.0249; Table 2) and let them share the same cluster. The 5th and the 4th clusters constitute the 6th cluster. As shown in Table 2, the ratios of genetic similarities are high, which represents a steady amalgamation of the genetic markers.

The Mediterranean region may be considered different from the other regions (6th cluster) (Figure 2). However, the F_{ST} value (0.0239) and allelic frequency counts did not show any fixation. This situation may result from unspoiled genetic constitution of the region by migration effects.

The geographical and demographical circumstances can determine the genetic constitutions of countries, by means of migration, genetic drift and isolation. These genetic profiles peculiar to populations are significant not only in the anthropology, but also in clinics and forensics. Considering their post histories (isolation of populations due to the geographic barriers and conserved life styles in Eastern Anatolia and Black Sea region, endogamies, originated from tribal traditions in Southeast Anatolia, sparse population distribution in Mediterranean coastline), if the seven regions of Turkey is taken in to account particularly, it is expected that regions will constitute their own genetic profiles. However, the socio-economic and ecological circumstances forced precinct people to migrate to other regions throughout the years. This flow (still continuing) caused a slow but dynamic exchange of genes across districts, resulting in a randomized distribution of genetic traits. Correlated with the demographic features, HWE equilibrium, Fst values, heterozygosity ratios and allele frequencies showed that Turkey has a heterogeneous genetic profile representing all the regions.

The obtained results indicated that Turkish population had a heterogeneous genetic variety; therefore, the allele frequencies in the gene pool can be useful and utilized in all forensic analyses. It is not necessary to form a separate database for each region as there was not a significant genetic differentiation between the regions due to the STR loci studied in the current study.

Acknowledgement

This study was supported by the Research Fund of Istanbul University T-1160/18062001

REFERENCES

1. Cavalli-Sforza LL, Menozzi P, Piazza A. Prehistory and history in West Asia. In: Cavalli-Sforza LL, Menozzi P, Piazza A, eds. The history and geography of human genes. 1st ed. Princeton: Princeton University Press; 1994. p.214.
2. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O , Cavalli-Sforza LL, Underhill PA. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet. 2004; 114(2) : 127–148.
3. Nei M. Genetic distance between populations. Amer Natur 1972; 106:283-292.
4. Nei M. Frontiers of Biology. In: Nei M eds. Molecular population genetics and evolution. 1st ed. Amsterdam:North-Holland Pub. Co. New York : American Elsevier Pub. Co; 1975. p.288.
5. Nei M, Roychoudhruy AK. Sampling variaences of heterozygosity and genetic distance, Genetics 1974; 76 (2):379-390.
6. Calafell F, Shuster A, Speed WC, Kidd J, Kidd KK. Short tandem repeat polymorphism evolution in humans. Eur. J. Hum. Genet. 1998; 6 (1):38-49.
7. Jin L, Baskett ML, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA. Microsatellite evolution in modern humans: a comparision of two data sets from the same populations. Ann Hum Genet. 2000; 64 (2):117-134.
8. Anker R, Steinbrueck T, Donis-Keller H. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum Mol Genet 1992;1(2):137.
9. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet. 1991;49 (4):746-756.
10. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994; 55 (1):175-189.
11. Jin L, Underhill PA, Buoncristiani M, Robertson JM. Defining microsatellite alleles by genotyping global indigenous human populations and non-human primates. J For Sci 1997; 42(3):496-499.
12. Selim AG, Ryan A, El-Ayat G, Wells CA. Loss of heterozygosity and allelic imbalance in apocrine metaplasia of the brest: microdissection microsatellite analysis. J.Pathol 2002; 196 (3): 287-291.
13. Urquhart A, Oldroyd NJ, Kimpton CP, Gill P. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. Biotehcniques 1995; 18(1):116-121.
14. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals Genetics 1978; 89(3):583-590.
15. Yeh, FC, Boyle, TJB. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 1997;129: 157.
16. Nei M. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci 1973; 70 (12):3321-3323.
17. Hartl DL, Clark AG. Principles of population genetics. In. Hartl DL, Clark AG.eds. Principles of population genetics.2nd ed. Sunderland, Massachusetts: Sinauer Associates; 1989. P.1-682.