

EFFECTS OF POPULATION DATA UPDATE ON GENETIC VARIANCE OF 15 POLYMORPHIC AUTOSOMAL STR LOCI TESTED IN TRANSYLVANIA

J. Sz. Demeter¹, Beatrice Kelemen¹, Gyöngyi Székely², O. Popescu^{1,3}

¹ MOLECULAR BIOLOGY CENTER, INTERDISCIPLINARY RESEARCH INSTITUTE ON BIO-NANO-SCIENCES, ² EXPERIMENTAL BIOLOGY DEPARTMENT, FACULTY OF BIOLOGY AND GEOLOGY, BABEȘ-BOLYAI UNIV., CLUJ-NAPOCA, ROMANIA
³ INSTITUTE OF BIOLOGY, THE ROMANIAN ACADEMY; BUCHAREST, ROMANIA

Summary

Since genomic DNA typing projects were started in Romania, five studies representing six populations have been carried out in Transylvania. All these results were cumulated and an updated Transylvanian database (Trs-6) was created. A comparison of the initial and updated Transylvanian databases was made in order to check consequences of the update. The comparison performed between the two databases revealed significant differences at 4 loci. When compared to five other databases (Moldavia, Dobruja, Wallachia, Romania and FBI), significant differences between the original and the updated databases were reported for 10 of the 15 loci analyzed. Therefore, the use of the updated database is recommended any time population data comparison with Transylvanian database has to be done.

Keywords: updated Transylvanian database; population comparison tests.

dszabolcs@yahoo.com

Introduction

Technological and scientific innovations have an impact on human life in many ways. Criminal acts, forensic cases, terrorist attacks or human mass disasters are just a few of the undesirable, unpredictable, and inevitable situations relying on human identification (Budowle *et al.* 2005).

Short Tandem Repeats (STRs) are among the most polymorphic DNA markers suitable in human identification (Butler 2005). Since genomic DNA typing projects were started in Romania, five studies representing six populations have been carried out in Transylvania. The initial Transylvanian STR database (Trs) contained 1977 individuals. Although for one (Zalău) out of the sixteen Transylvanian counties there is still no population data available, we consider that by adding data characterizing the population of Western Romania (West),

Szeklers from Harghita (HR-Sze), Csángós from Harghita (HR-Csn), Szeklers from Covasna (CV-Sze) and Hungarians from Cluj (CJ-Hu) to the initial Transylvanian database (Trs) we contribute to the completeness of the Transylvanian database. All these results were cumulated and an updated Transylvanian database (Trs-6; contains data for six populations) was created (Demeter 2010).

Materials and methods

Populations: The following populations were included in the updated Transylvanian Database (Trs-6; n=3097): non-Szekler Hungarians from Cluj county (CJ-Hu; n=146) (Demeter *et al.* 2010), Szeklers from Covasna county (CV-Sze; n=278) (Demeter *et al.* 2010), Szeklers from Harghita county (HR-Sze; n=257) (Egyed *et al.* 2005), Csángós from Harghita county (HR-Csn; n=220) (Egyed *et al.* 2005), initial Transylvanian STR Database (Trs; n=1977) (Stanciu *et al.* 2009a) and

general population of western part of Transylvania (West; n=219) (Marian *et al.* 2007).

Analysis of data: Data comparison was performed using the Arlequin software (Excoffier *et al.* 2005). Pairwise population comparison tests (F_{st}) were performed with 1023 permutations. Exact population differentiation tests (exact test) were performed using 100,000 Markov-chain and 10,000 dememorization steps. The significance level was set to $P < 0.05$.

Results

Performing the pairwise population comparison test (F_{st}), significant difference between the initial and the updated Transylvanian databases was identified at locus D21S11. Under exact population differentiation test, significant differences were observed at 4 loci: D21S11, D13S317, D2S1338 and vWA (Table 1). When the two databases were compared to general population data of Moldavia (n=1321) (Stanciu *et al.* 2009b), Dobruja (n=569) (Stanciu *et al.* 2009c), Wallachia (n=1910) (Stanciu *et al.* 2009d), Romania (n=6897) (Demeter 2010), and FBI (n=195-203) (Budowle *et al.* 2001), three cases concerning significant differences were observed. The first situation identified refers to a given significance level, which in both databases is $P < 0.05$: exact differentiation test P value for D21S11 and D13S317 loci, when compared to entire

Romanian database; F_{st} and exact test P values for D5S818 and D16S539 loci, when compared to FBI and Wallachia databases. In the second situation significant differences of the initial database lost their significance after the population data update: F_{st} P value for the D19S433 locus when compared to Moldavia; exact test P value for the FGA locus when compared to FBI database; F_{st} and exact test P value for the D3S1358 locus when compared to the FBI database. The third and the most frequent situation identified refers to changes of initially insignificant values to significant ones: F_{st} and exact test P value for the D21S11 locus, when compared to Moldavia and Wallachia; exact test P value for the vWA locus, when compared to Moldavia, Dobruja, Wallachia, FBI and Romanian databases; exact test P value for the D13S317 locus, when compared to Moldavia, Dobruja, Wallachia. Trs and Trs-6 were matched to the general Romanian population database. Exact tests reveal significant differences at 2 loci (D21S11 and D13S317) and 4 loci (D21S11, D13S317, vWA and D18S51) for Trs and Trs-6, respectively (Table 2). Within a comparison to the FBI database, the F_{st} comparison for Trs shows significant differences at one locus (D3S1358) and none for Trs-6. Exact differentiation tests reveal differences for D3S1358 and FGA loci versus Trs, and D21S11, vWA and D5S818 versus Trs-6 (Table 2).

Test performed	Population	Trs-6 n=3097	Moldavia n=1321	Dobruja n=569	Wallachia n=1910	Romania n=6897	FBI n=195-203	Total number of significantly different loci*
Fst	Trs	D21S11	D19S433	0	D16S539	0	D3S1358	3
	Trs-6		D21S11	0	D21S11 D13S317 D16S539	0	0	4
Exact test	Trs	D21S11 D13S317 D2S1338 vWA	0	0	D16S539	D21S11 D13S317	D3S1358 FGA	5
	Trs-6		D21S11 D13S317 D2S1338 vWA	D21S11 D13S317 vWA	D21S11 D13S317 D16S539 D2S1338 vWA	D21S11 D13S317 vWA D18S51	D21S11 vWA D5S818	19

Table 1: Summary of significantly different loci identified during population comparison tests;

Abbreviations: exact test (exact population differentiation test); Fst (pairwise population comparison test); n (number of samples); Trs (initial Transylvanian database); Trs-6 (updated Transylvanian database); vs (versus); * significant values for Trs versus Trs-6 were not counted.

Table 2: Comparison of the initial and updated Transylvanian databases with both pairwise population comparison (F_{st}) and exact population differentiation tests P values.

Marker	Test	Population	Trs	Moldavia	Dobruja	Wallachia	Romania	FBI
			n=1977	n=1321	n=569	n=1910	n=6897	n=195-203
D8S1179	Fst	Trs-6 vs	0.496	0.801	0.597	0.450	0.917	0.423
	Fst	Trs vs		0.434	0.954	0.960	0.753	0.199
	Exact test	Trs-6 vs	0.768	0.593	0.594	0.647	0.989	0.380
	Exact test	Trs vs		0.295	0.927	0.901	0.864	0.148
D21S11	Fst	Trs-6 vs	0.002*	0.007	0.115	<0.001	0.055	0.245
	Fst	Trs vs		0.579	0.945	0.228	0.170	0.203
	Exact test	Trs-6 vs	<0.001*	<0.001	<0.001	<0.001	<0.001	0.013
	Exact test	Trs vs		0.251	0.726	0.631	<0.001	0.079
D7S820	Fst	Trs-6 vs	0.592	0.869	0.930	0.055	0.708	0.792
	Fst	Trs vs		0.796	0.994	0.533	0.867	0.400
	Exact test	Trs-6 vs	0.935	0.978	0.972	0.137	0.970	0.194
	Exact test	Trs vs		0.894	0.995	0.387	0.868	0.156
CSFIPO	Fst	Trs-6 vs	0.596	0.531	0.499	0.627	0.993	0.922
	Fst	Trs vs		0.423	0.329	0.686	0.585	0.922
	Exact test	Trs-6 vs	0.807	0.541	0.641	0.457	0.931	0.118
	Exact test	Trs vs		0.693	0.438	0.737	0.646	0.161
D3S1358	Fst	Trs-6 vs	0.297	0.640	0.938	0.256	0.959	0.161
	Fst	Trs vs		0.163	0.304	0.210	0.140	0.043 [†]
	Exact test	Trs-6 vs	0.649	0.775	0.790	0.331	0.997	0.144
	Exact test	Trs vs		0.410	0.523	0.586	0.364	0.041 [†]
TH01	Fst	Trs-6 vs	0.789	0.896	0.485	0.146	0.733	0.477
	Fst	Trs vs		0.680	0.865	0.449	0.850	0.383
	Exact test	Trs-6 vs	0.925	0.886	0.204	0.290	0.941	0.177
	Exact test	Trs vs		0.734	0.425	0.564	0.895	0.201
D13S317	Fst	Trs-6 vs	0.074	0.111	0.449	0.043	0.379	0.208
	Fst	Trs vs		0.628	0.760	0.767	0.501	0.208
	Exact test	Trs-6 vs	<0.001*	<0.001	<0.001	<0.001	<0.001	0.187
	Exact test	Trs vs		0.753	0.677	0.904	<0.001	0.342
D16S539	Fst	Trs-6 vs	0.757	0.462	0.878	<0.001	0.394	0.505
	Fst	Trs vs		0.472	0.983	0.034	0.816	0.417
	Exact test	Trs-6 vs	0.945	0.655	0.926	0.005	0.795	0.404
	Exact test	Trs vs		0.694	0.985	0.012	0.958	0.368
D2S1338	Fst	Trs-6 vs	0.478	0.720	0.865	0.658	0.974	n/a
	Fst	Trs vs		0.938	0.706	0.960	0.767	n/a
	Exact test	Trs-6 vs	<0.001*	0.005	0.334	<0.001	0.265	n/a
	Exact test	Trs vs		0.972	0.586	0.971	0.151	n/a

Marker	Test	Population	Trs	Moldavia	Dobruja	Wallachia	Romania	FBI
			n=1977	n=1321	n=569	n=1910	n=6897	n=195-203
D19S433	Fst	Trs-6 vs	0.319	0.313	0.347	0.797	0.985	n/a
	Fst	Trs vs		0.016 [†]	0.515	0.915	0.192	n/a
	Exact test	Trs-6 vs	0.949	0.835	0.410	0.865	0.999	n/a
	Exact test	Trs vs		0.290	0.788	0.780	0.873	n/a
vWA	Fst	Trs-6 vs	0.332	0.306	0.782	0.107	0.319	0.851
	Fst	Trs vs		0.682	0.982	0.883	0.959	0.574
	Exact test	Trs-6 vs	<0.001*	<0.001	<0.001	<0.001	<0.001	0.035
	Exact test	Trs vs		0.691	0.948	0.616	0.869	0.694
TPOX	Fst	Trs-6 vs	0.909	0.433	0.498	0.546	0.952	0.290
	Fst	Trs vs		0.255	0.733	0.901	0.918	0.247
	Exact test	Trs-6 vs	0.986	0.527	0.453	0.967	0.983	0.405
	Exact test	Trs vs		0.451	0.333	0.940	0.888	0.214
D18S51	Fst	Trs-6 vs	0.846	0.952	0.207	0.080	0.584	0.442
	Fst	Trs vs		0.938	0.604	0.582	0.917	0.312
	Exact test	Trs-6 vs	0.994	0.606	0.560	0.139	<0.001	0.715
	Exact test	Trs vs		0.757	0.846	0.729	0.064	0.635
D5S818	Fst	Trs-6 vs	0.662	0.457	0.753	0.589	0.894	<0.001
	Fst	Trs vs		0.134	0.943	0.522	0.410	0.003
	Exact test	Trs-6 vs	0.977	0.663	0.816	0.926	0.996	0.016
	Exact test	Trs vs		0.610	0.758	0.925	0.970	0.002
FGA	Fst	Trs-6 vs	0.746	0.992	0.828	0.682	0.987	0.554
	Fst	Trs vs		0.976	0.964	0.631	0.799	0.218
	Exact test	Trs-6 vs	0.943	0.989	0.735	0.680	0.998	0.193
	Exact test	Trs vs		0.956	0.925	0.627	0.814	0.027 [†]

Abbreviations: exact test (exact population differentiation test); Fst (pairwise population comparison test); n (number of samples); vs (versus); Trs (initial Transylvanian database); Trs-6 (updated Transylvanian database); FBI (US Caucasian population); n/a (not available);

Note: values marked with “*” symbol show significant differences between Trs-6 and Trs; values in italics show no changes of significant values when both Trs and Trs-6 were compared to other populations; values marked with “†” symbol show how significant value of the initial database loses its significance after the population data update; values in bold prove how initially insignificant value changes to significant one after update; Fst was done with 1000 permutations; exact test settings: 100000 Markov-chain and 10000 dememorization steps; significance level $P < 0.05$

Discussion

We performed additional comparison tests with a modified Romanian database (Hungarian populations were excluded). As a result, the number of significantly different loci decreased from three to zero versus Wallachia and from two to zero versus Moldavia. The change raised the number of significantly different loci from three to four versus Trs-6 but did not influence significantly the match to Dobruja. Although no tests have been performed to explain the given effect of the Transylvanian Hungarian communities upon the Romanian DNA database, this study indicates that ethnical origin should be also taken into consideration when comprehensive DNA databases are created.

Conclusions

Statistical comparison tests performed in order to check the consequences of population data update made to the Transylvanian STR database resulted in 4 significantly different loci when Trs was matched to Trs-6. Also, 10 of the 15 loci reveal differences when both the initial and the updated databases were matched to data of other historical Romanian regions, general Romanian database and FBI. Therefore, we recommend the usage of the updated database any time when population data comparison with Transylvanian database has to be done.

References

Budowle B., Bieber F., Eisenberg A.: Forensic aspects of mass disasters: strategic considerations for DNA-based human identification. *Leg. Med.* **7**, 230-243, 2005

Budowle B., Shea B., Niezgodá S., Chakraborty R.: CODIS STR Loci Data from 41 Sample Populations. *J. Forensic Sciences*, **46**, 453-489, 2001

Butler J. M.: *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. Second Ed. Academic Press. 2005

Demeter J. Sz., Kelemen B., Székely Gy., Popescu O.: Genetic Variation at 15 Polymorphic Autosomal Short Tandem

Repeat (STR) Loci of Two Hungarian Populations in Transylvania, Romania. *Croat. Med. J.*, In press, 2010

Demeter J. Sz.: Allelic frequency of STR loci in some populations from Transylvania. PhD thesis, Babeş-Bolyai University, Cluj-Napoca, 2010

Egyed B., Füredi S., Pádár Z.: Population genetic study in two Transylvanian populations using forensically informative autosomal and Y-chromosomal STR markers. *Forensic Sci. Int.* November 25, PMID: 16314060 (Epub ahead of print), 2005

Excoffier L., Laval G., Schneider S.: Arlequin ver. 3.5: an integrated software package for population genetics data analysis, *Evolutionary Bioinformatics*, 47–50, 2005

Marian C., Anghel A., Bel S. M., Ferencz B. K., Ursoniu S., Dressler M., Popescu O., Budowle B.: STR data for the 15 AmpFISTR identifier loci in the Western Romanian population. *Forensic Sci. Int.* **170**, 73-75, 2007

Stanciu F., Stoian I. M., Popescu O. R., Comprehensive STR data for the AmpFISTR Identifier from Transylvania (NW Romania). *Legal Med.* **11**, 48–50, 2009a

Stanciu F., Popescu O. R., Stoian I. M.: Allele frequencies of 15 STR loci in Moldavia region (NE Romania). *Forensic Sci. Int.* **4**, 39–40, 2009b

Stanciu F., Stoian I. M., Popescu O. R., STR data for the AmpFISTR Identifier from Dobruja region (SE Romania), *Forensic Sci. Int.* **3**, 146–147, 2009c

Stanciu F., Stoian I. M., Popescu O. R.: Population Data for 15 Short Tandem Repeat Loci from Wallachia Region, South Romania. *Croat. Med. J.* **50**, 321-325, 2009d