Avraham Amar,[1,2] Ph.D.; Chaim Brautbar,[1,2] Ph.D.; Uzi Motro,[3] Ph.D.; Tovi Fisher,[2] M.Sc.; Batsheva Bonne-Tamir,[4] Ph.D.; and Shoshana Israel,[1,2] Ph.D.

Genetic Variation of Three Tetrameric Tandem Repeats in Four Distinct Israeli Ethnic Groups

REFERENCE: Amar A, Brautbar C, Motro U, Fisher T, Bonne-Tamir B, Israel S. Genetic variation of three tetrameric tandem repeats in four distinct Israeli ethnic groups. J Forensic Sci 1999;44(5):983– 986.

ABSTRACT: The allele frequency distributions of three STR loci amplified by PCR have been studied in four Israeli communities: Ashkenazi Jews and three non-Ashkenazi groups, namely Moroccan, Yemenite, and Ethiopian Jews. The loci analyzed were CSF1PO, TPOX, and HUMTHO1. The typing was performed in sequencing polyacrylamide gels under denaturing conditions that could separate alleles with differences of a single base. The population data were analyzed with respect to Hardy-Weinberg (H-W) equilibrium and found that all loci meet the H-W expectations. Noticeable differences were encountered between the four Jewish ethnic groups studied hereby indicating the importance of establishing a local database to be used in human identity testing in these different Israeli Jewish groups.

KEYWORDS: forensic science, short tandem repeats, ethnic groups, allele frequencies, DNA typing, database, population genetics, THO1, TPOX, CSF1PO, Israel

The human genome contains tandemly repeated sequence elements (1), that can be loci of highly polymorphic variable number of tandem repeats (VNTR). A subgroup of the VNTR consists of short tandem repeat (STR) loci (2,3), that contain short repetitive sequence elements of 3 to 7 base pairs in length (1–3). These abundant repeats are distributed throughout the human genome and are a rich source of highly polymorphic markers. STR alleles are small in size, generally <350 base pairs, and thus may be detected using the polymerase chain reaction (PCR) method (4–6), using unique flanking sequence primers. Puers et al. (7) described an analytical system comprised of three STR loci, HUMTHO1 (1,2,8), TPOX (9), and CSF1PO (10) that can be amplified simultaneously by PCR.

It is essential to establish a database for each genetically distinct population of the genetic markers in use. The Israeli Jewish population consists of many ethnic groups, but is generally divided into two main backgrounds: Ashkenazim, originating in Central and Eastern Europe, and non-Ashkenazim, originating from Middle Eastern, and North African countries. Studies on the different ethnic groups, using blood groups and markers of Y-chromosome and mitochondria, have revealed the existence of genetic distances among these groups (11).

In the present study, we present allele frequency data for the three tetrameric STR loci that were typed by using multiple PCR and subsequent electrophoresis in DNA sequencing gels with silver staining detection. We have included four different communities, namely: Ashkenazi Jews and Jews originating from Morocco, Yemen, and Ethiopia.

Materials and Methods

Samples

DNA was extracted from unrelated donors from four different Israeli Jewish ethnic groups. The DNA was extracted using the salting out method as reported by Miller et al. (12).

PCR Amplification

All the amplifications were performed in a final volume of 12.5 μL in a MJR thermocycler (Watertown, Massachusetts). Each reaction contained 1-25 ng genomic DNA. The coamplification of HUMTHO1, TPOX, and CSF1PO was performed using the GenePrint Kit (Promega Corporation). The reactions were subjected to an initial denaturation at 96°C for 2 min and then to 10 cycles of denaturation at 94°C for 1 min; primer annealing at 64°C for 1 min; and primer extension at 70°C for 1.5 min. The PCR was then continued for 20 cycles of denaturation at 90°C for 1 min; primer annealing at 64°C for 1 min; and primer extension at 70°C for 1.5 min. PCR products were subjected to 2% agarose gel electrophoresis to confirm the success and approximate yield of the amplification reaction.

In order to assign for the different alleles in the three distinct STR, 3 μL of loading buffer (10 mM NaOH, 95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol FF) were mixed with 5 μL of PCR product. The samples were denatured for 5 min, cooled in ice until used and 6 μL were loaded onto a polyacrylamide gel. The gels of 4% acrylamide were prepared from a stock solution of 40% acrylamide: bis (19:1), they contained 7 M urea and 0.5 X Tris-Borate-EDTA buffer and were 30 cm long and 0.4 mm thick. Electrophoresis was carried out on an SA 32 electrophoresis apparatus (Gibco-BRL). The conditions for electrophoresis were set at a constant power of 40W at ambient temperature. Electrophoresis was stopped when the xylene cyanol dye migrated 3 cm from the anode (~75 min). Allelic designations were determined by comparison of the sample fragments with those of the allelic ladder supplied by Promega Corporation.

[1] Tissue Typing Unit, Hadasah Medical Organization, Jerusalem, Israel.
[2] Lautenberg Center for General and Tumor Immunology, Hadassah–Hebrew University Medical Organization, Jerusalem, Israel.
[3] Department of Statistics, the Hebrew University of Jerusalem, Jerusalem, Israel.
[4] Department of Human Genetics, Sackler School of Medecine, Tel Aviv University, Tel Aviv, Israel.

Received 14 July 1998; and in revised form 31 Dec. 1998; accepted 25 Jan. 1999.

Statistical Analysis—Hardy-Weinberg Equilibrium

We tested compatibility of each of the three STR markers with the Hardy-Weinberg (H-W) equilibrium by a chi-square test for goodness of fit. Since these tests involved some rather small expected frequencies, the chi-square distribution cannot be a reliable approximation. The P values (i.e., the probabilities of rejecting the null hypothesis that the relevant population is in Hardy-Weinberg equilibrium with respect to the marker under consideration) were thus estimated by computer simulations. For each marker, 1000 simulated samples were drawn under the assumption of H-W equilibrium. The proportion of samples which had a chi-square statistic larger than the observed chi-square was taken as an estimate of the real P value.

Comparing Allele Distribution among Communities

For each marker, we compared the allele distribution among the four communities (Ashkenazi, Moroccan, Yemenite, and Ethiopian Jews), using a chi-square test for independence. The *P* values (i.e., the probabilities of rejecting the null hypothesis that the allele distribution is the same in all four communities) were estimated by computer simulations. For each marker, 1000 simulated samples were drawn under the null hypothesis assumption, and the proportion of samples which had a chi-square statistic larger than the observed chi-square was taken as an estimate of the real *P* value.

Pairwise genetic distances between the different communities were determined in the following way: First, for each marker, the distance between any two communities was calculated using the approach suggested by Nei (13, page 177). Thus, for marker j (j = 1,2,3), the distance between communities *A* and *B*, say, is

$$d_j(A,B) = 1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}},$$

(where x_i and y_i are the frequencies of allele i in each community). The overall distance $D(A,B)$ between communities *A* and *B* was then taken as the Euclidean measure of the three marker distances

$$D(A,B) = \sqrt{d_1(A,B)^2 + d_2(A,B)^2 + d_3(A,B)^2}.$$

These distances served to construct a population tree, using the unweighted pair group clustering method (14).

Results and Discussion

All samples from the four different ethnic groups considered in the present study (Ashkenazi, Moroccan, Yemenite, and Ethiopian Jews) were typed successfully for the three STR loci using the protocol supplied with the GenePrint Kit (Promega Corporation). The distribution of observed allelic frequencies for CSF1PO, TPOX, and THO1 are shown in Tables 1–3. All alleles differed in size by one repeat unit (4 bp) for all loci, except for the THO1 allele 9.3. The 9.3 allele is 1 bp smaller in size than the 10 allele (15). The ability to type unequivocally the 9.3 allele demonstrates the resolving capacity and utility of the electrophoretic system used in this study.

Among the 10 known alleles in the CSF1PO locus (10), three of them (alleles 10, 11, and 12) are relatively common, and quite evenly distributed among the four different groups (Table 1). When all the four groups were compared with respect to allele distribution in the CSF1PO locus, a significant difference emerged ($\chi^2_{21} = 34.02$, $P = 0.024$). Among the 8 alleles identified in locus TPOX (9), three alleles (8,9,11) were found to be relatively common in the 4 ethnic groups studied hereby, with significant differences between the groups ($\chi^2_{18} = 44.26$, $p < 0.001$) (Table 2). The THO1 locus is the most variable locus in the groups studied. Among the 8 alleles identified in this locus (8), five were relatively common in the groups studied by us. The differences in allele distribution between the communities studied were significant ($\chi^2_{15} = 64.93$, $p < 0.001$) (Table 3).

When we explored the allele distributions between any two distinct ethnic groups we sometimes found significant differences for each marker (Table 4).

TABLE 1—*CSF1PO allele frequencies in four different ethnic groups in Israel.*

	Ashkenazi (N = 128)		Moroccan (N = 133)		Yemenite (N = 67)		Ethiopian (N = 105)	
Allele	N*	P†	N	P	N	P	N	P
7	0	0	2	0.75	0	0	3	1.43
8	3	1.17	2	0.75	0	0	4	1.90
9	4	1.56	5	1.88	1	0.75	9	4.29
10	60	23.44	68	25.56	35	26.12	38	18.10
11	88	34.38	110	41.35	43	32.09	70	33.33
12	88	34.38	69	25.94	45	33.58	71	33.80
13	13	5.08	10	3.76	8	5.97	12	5.71
14	0	0	0	0	2	1.49	3	1.43

* N = Number.
† P = Proportion.

TABLE 2—*TPOX allele frequencies in four different ethnic groups in Israel.*

	Ashkenazi (N = 128)		Moroccan (N = 133)		Yemenite (N = 67)		Ethiopian (N = 105)	
Allele	N*	P†	N	P	N	P	N	P
7	0	0	1	0.38	0	0	1	0.48
8	140	54.69	130	48.87	74	55.22	88	41.90
9	35	13.67	36	13.53	27	20.15	60	28.57
10	20	7.81	14	5.26	12	8.96	18	8.57
11	58	22.66	76	28.57	19	14.28	36	17.14
12	3	1.27	8	3.0	2	1.49	7	3.33
13	0	0	1	0.38	0	0	0	0

* N = Number.
† P = Proportion.

TABLE 3—*THO1 allele frequencies in four different ethnic groups in Israel.*

	Ashkenazi (N = 128)		Moroccan (N = 133)		Yemenite (N = 67)		Ethiopians (N = 105)	
Allele	N*	P†	N	P	N	P	N	P
6	54	21.09	53	19.92	37	27.60	51	24.29
7	32	12.50	40	15.04	23	17.16	45	21.43
8	31	12.11	29	10.90	13	9.70	49	23.33
9	66	25.78	96	36.09	37	27.60	47	22.38
9.3	62	24.22	36	13.53	22	16.42	12	5.71
10	11	4.30	12	4.51	2	1.49	6	2.86

* N = Number.
† P = Proportion.

TABLE 4—*Pairwise comparisons of allele distributions between the four Jewish ethnic groups.*

CSF1PO	Ashkenazi	Moroccan	Yemenite	Ethiopian
Ashkenazi	—	$\chi^2_6 = 7.692$ P value = 0.242	$\chi^2_6 = 6.634$ P value = 0.391	$\chi^2_7 = 12.445$ P value = 0.080
Moroccan		—	$\chi^2_7 = 11.899$ P value = 0.099	$\chi^2_7 = 16.236$ P value = 0.015
Yemenite			—	$\chi^2_7 = 10.673$ P value = 0.151
Ethiopian				—
TPOX	**Ashkenazi**	**Moroccan**	**Yemenite**	**Ethiopian**
Ashkenazi	—	$\chi^2_6 = 7.490$ P value = 0.251	$\chi^2_4 = 5.848$ P value = 0.229	$\chi^2_5 = 21.837$ P value < 0.001
Moroccan		—	$\chi^2_6 = 14.647$ P value = 0.013	$\chi^2_6 = 23.684$ P value < 0.001
Yemenite			—	$\chi^2_5 = 7.537$ P value = 0.183
Ethiopian				—
THO1	**Ashkenazi**	**Moroccan**	**Yemenite**	**Ethiopian**
Ashkenazi	—	$\chi^2_5 = 13.596$ P value = 0.028	$\chi^2_5 = 8.329$ P value = 0.136	$\chi^2_5 = 40.623$ P value < 0.001
Moroccan		—	$\chi^2_5 = 7.476$ P value = 0.198	$\chi^2_5 = 30.079$ P value < 0.001
Yemenite			—	$\chi^2_5 = 20.594$ P value = 0.001
Ethiopian				—

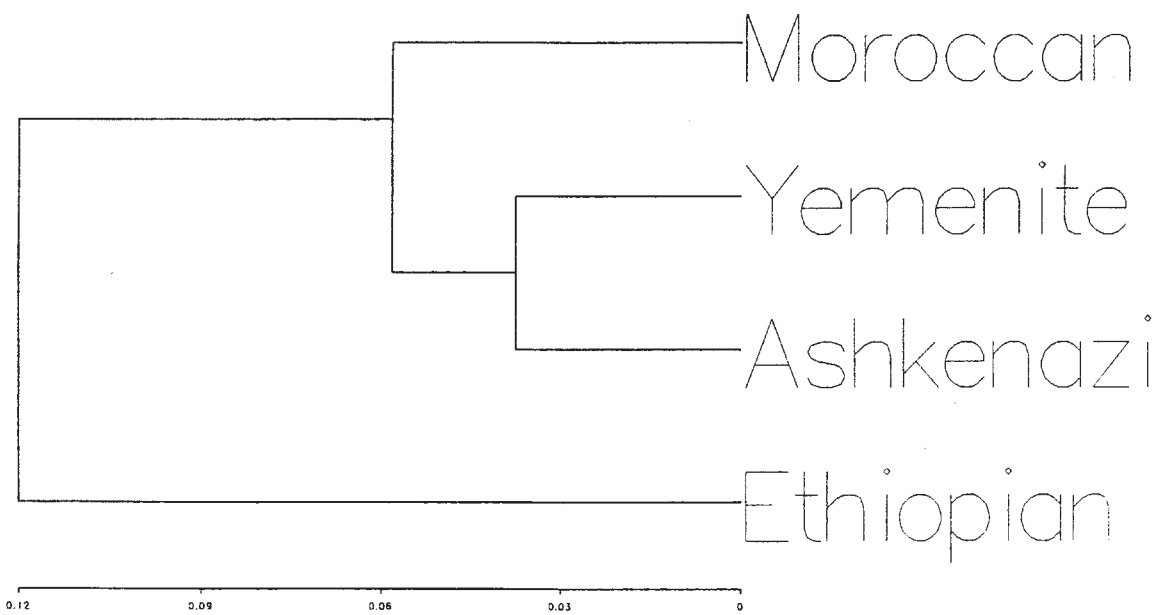

FIG. 1—*A population tree, based on the genetic distances among the four Jewish communities in the three STR markers.*

Any genetic marker needs to be tested for deviation from Hardy-Weinberg equilibrium before its application for forensic cases. No significant deviations from the Hardy-Weinberg expected values were found for the three STR loci used in this study (Table 5).

A population tree, based on a cluster analysis of the genetic distances in the three markers, is presented in Fig. 1. The dendrogram, constructed by the unweighted pair group method, indicates that the Ethiopian Jewish community is most distantly related to the other Jewish communities in our study. These results are in agreement with the results previously obtained using markers of the Y-chromosome and mitochondrial DNA performed on six Jewish communities, where the Ethiopians Jews were found to be significantly different from the other communities (11). Adding data for three American communities (African-Americans, Caucasian-Americans, and Hispanic-Americans) reported by Creacy et al. (16) to our study, we obtain the population tree in Fig. 2, which

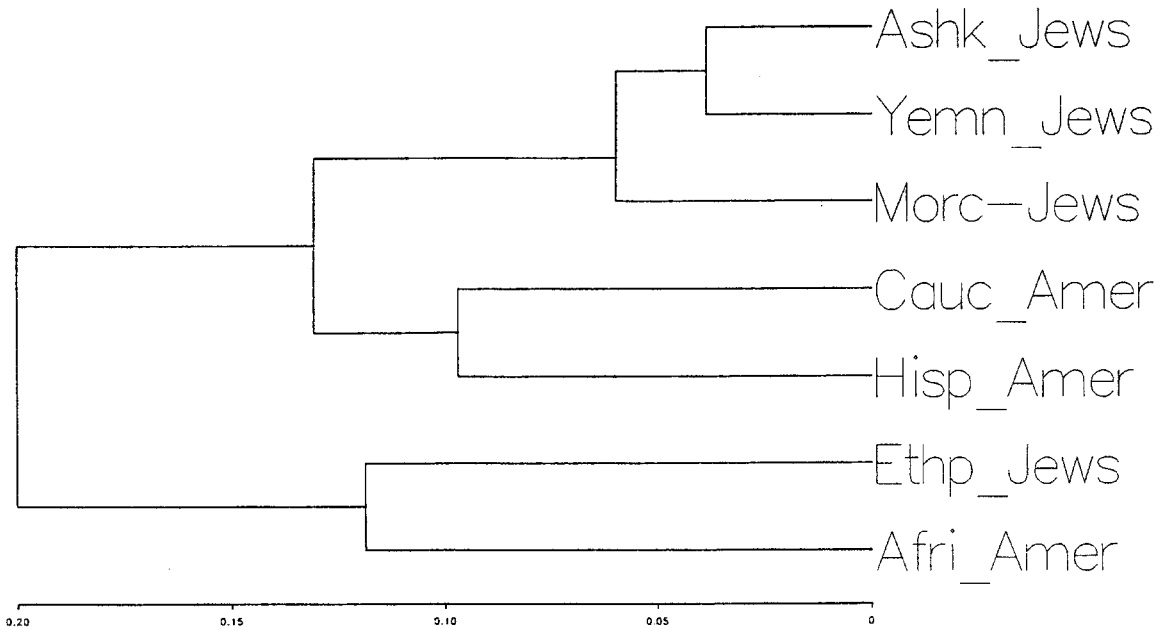

FIG. 2—*A population tree, based on the genetic distances among the four Jewish and the three American communities in the three STR markers.*

TABLE 5—*The goodness of fit tests for compatibility with Hardy-Weinberg equilibrium.*

Population	CSF1PO	TPOX	THO1
Ashkenazi	$\chi^2_{15} = 22.52$ P value = 0.090	$\chi^2_{10} = 6.88$ P value = 0.610	$\chi^2_{15} = 7.30$ P value = 0.952
Moroccan	$\chi^2_{21} = 11.87$ P value = 0.705	$\chi^2_{10} = 16.50$ P value = 0.072	$\chi^2_{15} = 16.33$ P value = 0.356
Yemenite	$\chi^2_{10} = 9.58$ P value = 0.400	$\chi^2_{10} = 8.13$ P value = 0.498	$\chi^2_{10} = 5.37$ P value = 0.886
Ethiopian	$\chi^2_{28} = 11.39$ P value = 0.946	$\chi^2_{10} = 4.18$ P value = 0.927	$\chi^2_{15} = 8.39$ P value = 0.893

demonstrates that the three Jewish communities, Ashkenazi, Moroccan, and Yemenite Jews, form a clearly visible cluster, distinct from the cluster of the Caucasian and Hispanic Americans. The Ethiopians Jews and the African Americans form another cluster.

Conclusion

In conclusion, an Israeli population database has been established for CSF1PO, TPOX, and THO1. The differences observed among the four Jewish ethnic groups indicate the importance of establishing a local database to be used in human identification in a society consisting of many different ethnic groups.

References

1. Edwards A, Civitello A, Hammond HA, Caskey T. DNA typing and genetic mapping with thimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
2. Edwards A, Hammond HA, Jin L, Caskey T, Chakraborty R. Genetic variation at five thimeric and tetrameric tandem repeats loci in four human population groups. Genomics 1992;12:241–53.
3. Warne D, Watkins C, Bodfish P, Nyberg K, Spurr NK. Tetranucleotide repeat polymorphism at the human beta-actin related fosendegene 2 (ACTBP2) detected using the polymerase chain reaction. Nucleic Acids Res 1991;19:69–80.
4. Saiki RK, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. Science 1985;230:1350–4.
5. Erlich HA, editor. PCR technology: Principles and applications for DNA amplification, New York, Stockton Press 1989.
6. Innis MA, Gelfand H, Sninsky JJ, White TJ, editors. PCR protocols: A guide to methods and applications, San Diego, Academic Press 1990.
7. Puers C, Lins AM, Sprecher CJ, Brinkmann B, Schumm JW. Analysis of polymorphic STR loci using well-characterized allelic ladders. In: Proc. 4th Int. Symp. Human Identification 1993. Promega Corporation, 1994;161–72.
8. Polymerofeules MH, Xiao H, Rath DS, Merill CR. Tetranucleotide repeat polymorphism at the human tyrosine hydrolase gene (TH). Nucleic Acids Res 1991;19:37–53.
9. Anker R, Steinbrueck T, Donis-Keller M, Steinbrueck T, Donis-Keller H. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (RTPO) locus. Hum Mol Genet 1992;1:137.
10. Hammond HA, Jin L, Shong Y, Caskey T, Chakraborty R. Evaluation of 13 STR loci for use in personal identification applications. Am J Hum Genet 1994;55:175.
11. Ritte U, et al. The differences among Jewish communities—maternal and paternal contributions. J Med Evol 1993;37:435–40.
12. Miller SA, Dykes DD, Polesky HF. A simple salting-out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1998;16:12–5.
13. Nei M. Molecular population genetics and evolution. Amsterdam, North-Holland 1975.
14. Abbott LA, Bisby FA, Rogers DJ. Taxonomic analysis in biology. New York, Columbia University Press 1985.
15. Puers C, Hammond HA, Jin L, Caskey T, Schumm JW. Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUTHO1 (AATG)n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–8.
16. Creacy S, Bever RA, Sprecher CJ, Schumm JW. In: Promega technical manual Gene Print STR systems, 1996.

Additional information and reprint requests:
Dr. Avraham Amar, Tissue Typing Unit
Hadassah Medical Organization
P.O. Box 12000, Jerusalem 91120, Israel
FAX: 972-2-6422403