

Genetic Structure of Seven Mexican Indigenous Populations Based on Five Polymarker Loci

LEONORA BUENTELLO-MALO,¹ ROSENDA I. PEÑALOZA-ESPINOSA,² FRANCISCO LOEZA,³ FABIO SALAMANCA-GOMEZ,² AND RICARDO M. CERDA-FLORES^{4,5*}

¹Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México, México D.F.

²Unidad de Investigación Médica en Genética Humana, Centro Médico Nacional, Siglo XXI, Instituto Mexicano del Seguro Social (IMSS), México D.F.

³Universidad Michoacana de San Nicolás de Hidalgo, Michoacan, México

⁴Departamento de Genética de Poblaciones, Centro de Investigación Biomédica del Noreste, IMSS, Nuevo León, México

⁵Facultad de Enfermería, Universidad Autónoma de Nuevo León, México

ABSTRACT This descriptive study investigates the genetic structure of seven Mexican indigenous populations (Mixteca Alta, Mixteca Baja, Otomies, Purepecha, Nahuas-Guerrero, Nahuas-Xochimilco, and Tzeltales) on the basis of five PCR-based polymorphic DNA loci: LDLR, GYPA, HBG, D7S8, and GC. Genetic distance and diversity analyses indicate that these Mexican indigenous are similar and that more than 96% of the total gene diversity (H_T) can be attributed to individual variation within populations. Mixteca-Alta, Mixteca-Baja, and Nahuas-Xochimilco show indications of higher admixture with European-derived persons. The demonstration of a relative genetic homogeneity of Mexican Indians for the markers studied suggests that this population is suitable for studying disease-marker associations in the search for candidate genes of complex diseases. *Am. J. Hum. Biol.* 15:23–28, 2003. © 2002 Wiley-Liss, Inc.

Mexican mestizos represent the largest Spanish-speaking population in Mexico. They were generated by an admixture of Europeans and Africans with natives (Serrano, 1995). Presently, the proportion of Mexican indigenous populations in regard to the total Mexican population (97,843,412) is 10% and, considering language as selection criteria, this proportion exhibits a diminishing tendency (Census, 2000; Fernandez and Serrano, 1996).

Genetic variability detected by variable number of tandem repeat loci (VNTRs) has been shown to be highly informative for gene mapping, forensic identification of individuals, determination of relatedness of individuals, and for evolutionary studies of closely related populations or species.

Some Native American databases have been established for the polymorphic loci LDLR, GYPA, HBG, D7S8, and GC, but for Mexican Indians the information is scarce. This descriptive study explores the genetic structure of seven Mexican indigenous populations (Fig. 1) on the basis of five PCR-based polymorphic DNA loci.

Mexican Indians [57 Mixteca-Alta (MA), 34 Mixteca-Baja (MB), 50 Otomies (OT), 50 Purepecha (PU), 40 Nahuas-Guerrero (NG), 46 Nahuas-Xochimilco (NX), and 53 Tzeltales (TZ)]. A detailed description of the genetic structure (using blood groups and protein loci), geographic location, number of individuals, and language of these seven populations and others were published previously by Lisker (1981). Genomic DNA was extracted by the standard phenol-chloroform method (Sambrook et al., 1989).

PCR amplification and typing

Amplification and typing for the systems LDLR, GYPA, HBG, D7S8, and GC were carried out using the AmpliType PM PCR Amplification and Typing Kit (Perkin Elmer, Norwalk, CT) according to the manufacturer's protocol. Amplification was carried out in a Perkin-Elmer DNA thermal cyclor 9600.

MATERIALS AND METHODS

Sample preparation

Whole-blood samples were collected in EDTA by venipuncture from 330 unrelated

*Correspondence to: Dr. Ricardo M. Cerda-Flores, División de Genética, CIBIN, IMSS, Administración de Correos #4, Apdo. Postal #20, 64720 Monterrey, Nuevo León, México. E-mail: ricardocerda_mx@yahoo.com.mx

Received 3 April 2002; Revision received 21 June 2002; Accepted 24 June 2002

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ajhb.10116



Fig. 1. Location of the seven Mexican indigenous populations.

Statistical analysis

The statistical analysis was conducted in six steps. First, allele frequencies were calculated by the gene-counting method (Li, 1976). Second, Hardy-Weinberg Equilibrium (HWE) was tested by the likelihood ratio (Weir, 1990), exact (Guo and Thompson, 1992), and χ^2 for total heterozygosity tests. The significance level was 0.05. These calculations were carried out using a permutation program described by Chakraborty et al. (1991, 1993). The Bonferroni correction (Weir, 1990) for multiple comparisons was used. Third, genetic distances were computed considering Nei's standard genetic distances (Nei, 1972), and their standard errors (SE) calculated as described in Nei and Roychoudhury (1974). Significance of the genetic distances was assessed by the χ^2 statistic (Nei and Roychoudhury, 1974). For dendrogram construction, we used Saitou and Nei's (1987) neighbor-joining method, in which the significance of the branches was evaluated by bootstrapping. Gene frequency data for comparison were obtained from the compilation of Scholl et al. (1995) for three Native

American populations (Navajo, Pueblo, and Sioux) and at <http://www.uni-duesseldorf.de/WWW/MedFak/Serology/polymarker.html> for Spain (pooled without Canary Isles), Afro-Americans (pooled), and USA Hispanics (pooled). Fourth, the extent of genetic variation within and between the populations was assessed using the nested gene diversity computer program (NEGST) developed by Chakraborty et al. (1982). Fifth, computation for nonrandom association of alleles among different genetic loci was conducted according to the methods of Brown et al. (1980) and Chakraborty (1981, 1984). Finally, levels of polymorphism and their utility through the estimation of locus-specific heterozygosities (Nei and Roychoudhury, 1974), power of exclusion (for paternity testing, Ohno et al., 1982), and power of discrimination (Jones, 1972) were computed.

RESULTS AND DISCUSSION

Genotype and allele frequencies are shown in Tables 1 and 2. With a few exceptions, genotype frequency distributions do not deviate from HWE.

TABLE 1. Observed genotype frequency distributions (%) for five DNA polymorphic loci

Genotype		MA n = 57	MB n = 34	OT n = 50	PU n = 50	NG n = 40	NX n = 46	TZ n = 53
LDLR	AA	45.61	41.18	28.00	70.00	32.50	34.78	37.74
	AB	24.56	20.59	46.00	26.00	32.50	50.00	39.62
	BB	29.82	38.24	26.00	4.00	35.00	15.22	22.64
GYPA	AA	35.09	44.12	59.18	68.00	95.00	39.13	75.47
	AB	49.12	29.41	30.61	30.00	2.50	45.65	24.53
	BB	15.79	26.47	10.20	2.00	2.50	15.22	0.00
HBGG	AA	15.79	23.53	14.00	8.00	7.50	4.35	11.32
	AB	42.11	41.18	38.00	30.00	32.50	78.26	37.74
	BB	35.09	20.59	46.00	58.00	60.00	17.39	50.94
	AC	7.02	5.88	2.00	0.00	0.00	0.00	0.00
	BC	0.00	2.94	0.00	4.00	0.00	0.00	0.00
CC	0.00	5.88	0.00	0.00	0.00	0.00	0.00	
D7S8	AA	28.07	41.18	51.02	46.00	40.00	28.26	43.40
	AB	49.12	44.12	28.57	36.00	42.50	58.70	45.28
	BB	22.81	14.71	20.41	18.00	17.50	13.04	11.32
GC	AA	3.51	11.76	2.04	0.00	0.00	2.17	3.77
	AB	14.04	8.82	6.12	14.00	2.50	19.57	15.09
	BB	21.05	8.82	6.12	8.00	10.00	4.35	13.21
	AC	14.04	20.59	20.41	18.00	15.00	39.13	16.98
	BC	29.82	29.41	34.69	36.00	50.00	21.74	28.30
CC	17.54	20.59	30.61	24.00	22.50	13.04	22.64	

MA: Mixteca-Alta; MB: Mixteca-Baja; OT: Otomies; PU: Purepecha; NG: Nahuas-Guerrero; NX: Nahuas-Xochimilco; TZ: Tzeltales.

TABLE 2. Observed allele frequency distributions (%)

Allele		MA n = 57	MB n = 34	OT n = 50	PU n = 50	NG n = 40	NX n = 46	TZ n = 53
LDLR	A	57.90	51.47	51.00	83.00	48.75	59.78	57.55
	B	42.10	48.53	49.00	17.00	51.25	40.22	42.45
GYPA	A	59.65	58.82	74.49	83.00	96.25	61.96	87.74
	B	40.35	41.18	25.51	17.00	3.75	38.04	12.26
HBGG	A	40.35	47.06	34.00	23.00	23.75	43.48	30.19
	B	56.14	42.65	65.00	75.00	76.25	56.52	69.81
	C	3.51	10.29	1.00	2.00	0.00	0.00	0.00
D7S8	A	52.63	63.23	65.31	64.00	61.25	57.61	66.04
	B	47.37	36.77	34.69	36.00	38.75	42.39	33.96
GC	A	17.54	26.47	15.31	16.00	8.75	39.52	19.81
	B	42.98	27.94	26.53	33.00	36.25	25.00	34.91
	C	39.48	45.59	58.16	51.00	55.00	43.48	45.28

MA: Mixteca-Alta; MB: Mixteca-Baja; OT: Otomies; PU: Purepecha; NG: Nahuas-Guerrero; NX: Nahuas-Xochimilco; TZ: Tzeltales.

TABLE 3. Standard genetic distances (10^{-3}) among Mexican indigenous populations

	MA	MB	OT	PU	NG	NX
MB	6.20 ± 6.43					
OT	20.94 ± 9.93	14.16 ± 11.81				
PU	52.86 ± 14.27	79.32 ± 33.16	37.14 ± 33.28			
NG	61.49 ± 35.39	77.85 ± 42.03	15.06 ± 12.35	40.03 ± 39.27		
NX	2.28 ± 10.40	-2.68 ± 5.32	12.59 ± 6.46	45.64 ± 12.88	63.34 ± 29.33	
TZ	34.55 ± 23.49	43.37 ± 28.44	5.59 ± 6.05	18.23 ± 20.84	2.42 ± 1.94	28.34 ± 19.20

MA: Mixteca-Alta; MB: Mixteca-Baja; OT: Otomies; PU: Purepecha; NG: Nahuas-Guerrero; NX: Nahuas-Xochimilco; TZ: Tzeltales.

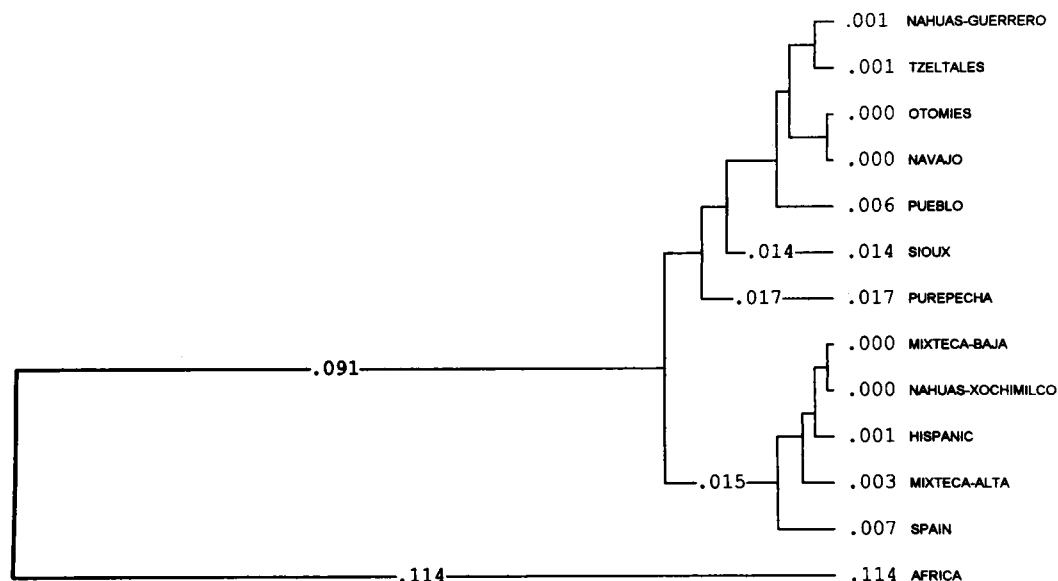


Fig. 2. Neighbor-joining tree showing the genetic relationships among the seven Mexican populations studied here and six others for comparison.

Genetic distances are presented in Table 3. Differences among them are not statistically significant. The dendrogram obtained from the genetic distances obtained using the data obtained here and those of six other populations are shown in Figure 2. Four of the seven Mexican populations (Nahuas-Guerrero, Tzeltales, Otomies, and Purepecha) and the three Native American groups cluster together, while the Mixteca-Baja and Nahuas-Xochimilco populations cluster with USA Hispanics and the Spanish sample. All are clearly distinct from Africans.

Table 4 presents the hierarchical gene diversity analysis. The total average gene

diversity (H_T) is 48.9%. Over 96% of H_T can be attributed to individual variation within the population.

From the available genotype data on each individual, we defined the multilocus genotype for each individual for the five loci. The number of loci with respect to which individual was heterozygous was determined. This generated an observed distribution on the number of heterozygous loci across 330 individuals. Chakraborty (1981) provided a numerical algorithm to compute the expected distribution for such observations, assuming random association of alleles at the different loci. In general, the observed distribution agrees with the expected one

TABLE 4. Gene diversity analysis of the allele frequency data

Locus	G_{ST}^*		H_T^{**}
	Within population	Between populations	Total gene diversity
LDLR	95.27	4.73	0.486
GYP A	89.89	10.11	0.380
HBGG	95.59	4.41	0.484
D7S8	99.13	0.87	0.473
GC	98.07	1.93	0.625
Mean	95.96	4.04	0.489
s.e.	1.40	1.40	0.039

*Expressed as percentage of total.

**Absolute total gene diversity in the entire sample.

TABLE 5. Summary statistics of levels of polymorphism in 330 Mexican Indians

Loci	LDLR	GYP A	HBGG	D7S8	GC	Combined
Power of exclusion	0.183	0.154	0.201	0.181	0.337	0.700
Power of discrimination	0.650	0.545	0.628	0.626	0.780	0.995
Observed heterozygosity	0.484	0.379	0.477	0.475	0.625	
Expected heterozygosity	0.484	0.379	0.478	0.475	0.626	

($\chi^2 = 9.60$, $P = 0.10$). The mean number of heterozygous loci is 2.20 and the variance is 1.26, while their expected values (under the random association model) are 2.44 and 1.22, respectively. Clearly, these values provide no evidence of nonrandom association of these alleles in the total indigenous populations.

Table 5 presents locus-specific values of observed and expected (unbiased) heterozygosity, exclusion probability (PE), and power of discrimination (PD), based on the allele frequency data of Table 2. The combined average PE and PD were 0.700 and 0.995 (1 in 200) for the total sample. Thus, for every 200 comparisons between two people chosen at random, about 199 pairs would have different types. These five loci should be a useful battery of markers for DNA testing.

These findings have a number of important implications in relation to the utility of these populations for anthropogenetic and epidemiological studies. The demonstration of a relative genetic homogeneity of Mexican Indians for the markers studied suggests that this population is suitable for studying disease-marker associations in the search for candidate genes of complex diseases (Chakraborty and Weiss, 1988). On the other hand, their multilocus genotypic distribution satisfies the premises of random segregation of unlinked loci. Therefore, the probability of finding a specific multiple-locus genotype in them can be determined by the product rule of locus-specific genotype probabilities. Finally, Mixteca-Alta, Mixteca-Baja, and Nahuas-Xochimilco show a higher degree of intermixture with European-derived persons than the other four groups.

ACKNOWLEDGMENTS

The authors are grateful to the study participants. Institutional Review Board approval was obtained for all sample collections.

LITERATURE CITED

- Brown AHD, Feldman MW, Nevo E. 1980. Multilocus structure of natural populations of *Hortedum spontaneum*. *Genetics* 98:523-536.
- Census. 2000. Instituto Nacional de Estadística, Geografía e Informática, 2000. XII Censo General de Población y Vivienda. México, D.F.: Instituto Nacional de Estadística, Geografía e Informática (<http://www.INEGI.gob.mx>).
- Chakraborty R. 1981. The distribution of the number of heterozygous loci in natural populations. *Genetics* 98:461-466.
- Chakraborty R. 1984. Detection of non-random association of alleles from the distribution of the number of heterozygous loci in a sample. *Genetics* 108:719-731.
- Chakraborty R, Weiss KM. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 85:9119-9123.
- Chakraborty R, Haag M, Ryman N, Stahl G. 1982. Hierarchical gene diversity analysis and its implication to brown trout population data. *Hereditas* 97:17-21.
- Chakraborty R, Fornage M, Guegue R, Boerwinkle E. 1991. Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: Burkle T, Dolf G, Jeffreys AJ, Wolff R, editors. *Fingerprinting: approaches and applications*. Basel: Birkhauser. p 127.
- Chakraborty R, Jin L, Zhong Y, Budowle B. 1993. Nondetectability of restriction fragments and tests of independence of alleles within and between VNTR loci scored by RFLP analysis. *Am J Hum Genet* 53(Suppl):785.
- Fernandez P, Serrano E. 1996. La población indígena de México en los recuentos censales de 1990 y 1995 (CONAPO/DAF-INAH). Evaluation of the population counting 1995. México, D.F.: Mexican Demography Society (Meeting).
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361-372.
- Jones DA. 1972. Blood samples: probability of discrimination. *J Forensic Sci Soc* 19:355-359.
- Li CC. 1976. *First course in population genetics*. Pacific Grove, CA: Boxwood Press.
- Lisker R. 1981. *Estructura genética de la población Mexicana: Aspectos médicos y antropológicos*. Mexico: Salvat Mexicana de Ediciones.
- Nei M. 1972. Genetic distance between populations. *Am Nat* 106:283-292.
- Nei M, Roychoudhury AK. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76:379-390.

- Ohno Y, Sebetan IM, Akaishiy S. 1982. A simple method for calculating the probability of excluding paternity with any number of codominant alleles. *Forensic Sci Int* 19:93-98.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Scholl S, Budowle B, Radecki K, Salvo M. 1996. Navajo, Pueblo, and Sioux population data on the loci HLA-DQA1, LDLR, GYPA, HBG, D7S8, Gc, and D1S80. *J Forensic Sci* 41:47-51.
- Serrano C. 1995. 500 Años de historia: la conquista y el mestizaje biológico en México. In: Ochoa L, editor. *Conquista, transculturación y mestizaje: raíz y origen de México*. México, DF: Instituto de Investigaciones Antropológicas (UNAM). p 39.
- Weir BS. 1990. Multiple tests. In: *Genetic data analysis*. Sunderland, MA: Sinauer. p 109-110.