

Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping

C. Bonilla^{1,*}, E. J. Parra², C. L. Pfaff³, S. Dios⁴, J. A. Marshall⁵, R. F. Hamman⁵, R. E. Ferrell⁶, C. L. Hoggart⁷, P. M. McKeigue⁷ and M. D. Shriver⁴

¹National Human Genome Center, Howard University, Washington, DC 20060, USA

²Department of Anthropology, University of Toronto at Mississauga, Mississauga, Ontario, Canada

³Human Genetics Department, The University of Michigan, Ann Arbor, MI 48109-0618, USA

⁴Department of Anthropology, The Pennsylvania State University, University Park, PA 16802, USA

⁵Department of Preventive Medicine and Biometrics, University of Colorado School of Medicine, Denver, CO 80262, USA

⁶Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

⁷Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E7HT, UK

Summary

Hispanic populations are a valuable resource that can and should facilitate the identification of complex trait genes by means of admixture mapping (AM). In this paper we focus on a particular Hispanic population living in the San Luis Valley (SLV) in Southern Colorado. We used a set of 22 Ancestry Informative Markers (AIMs) to describe the admixture process and dynamics in this population. AIMs are defined as genetic markers that exhibit allele frequency differences between parental populations $\geq 30\%$, and are more informative for studying admixed populations than random markers. The ancestral proportions of the SLV Hispanic population are estimated as $62.7 \pm 2.1\%$ European, $34.1 \pm 1.9\%$ Native American and $3.2 \pm 1.5\%$ West African. We also estimated the ancestral proportions of individuals using these AIMs. Population structure was demonstrated by the excess association of unlinked markers, the correlation between estimates of admixture based on unlinked marker sets, and by a highly significant correlation between individual Native American ancestry and skin pigmentation ($R^2 = 0.082$, $p < 0.001$). We discuss the implications of these findings in disease gene mapping efforts.

Introduction

Admixed populations facilitate mapping complex trait genes due to the linkage disequilibrium (LD) that is created when two or more previously separated populations intermix (Nei & Li, 1973; Chakraborty & Weiss, 1988). If the founding populations differ in trait prevalence due to differences in frequency of high risk genotypes, nearby markers that also differ in allele frequency between the parental populations will enable the local-

ization and identification of the genetic variants that are responsible for the increased disease risk. This general approach has been referred to as Mapping by Admixture Linkage Disequilibrium (MALD; Stephens *et al.* 1994; Briscoe *et al.* 1994) or simply Admixture Mapping (AM; Zheng & Elston, 1999; McKeigue *et al.* 2000). In the New World a number of admixed populations arose as a result of a history of conquest and colonization, followed by genetic exchange among the members of previously isolated parental populations that came into contact during the past five centuries. Hispanic and African-American populations are examples of such groups that can be found in the U.S. Recently, it has been shown that LD in African-American popula-

*Corresponding author: Carolina Bonilla, National Human Genome Center, Howard University, Washington, DC 20060. Phone: (202) 806 9436. Fax: (202) 986 3972. E-mail: cbonilla@genomecenter.howard.edu

tions may extend over distances considerably larger than 1 cM: 19 cM in the 13p13–p20 region (Rybicki *et al.* 2002); and 30 cM around the FY locus on chromosome 1q22–q23 (Parra *et al.* 1998; Lautenberger *et al.* 2000; Pfaff *et al.* 2001). Given these long distances over which LD extends, admixed populations may be uniquely useful for whole genome association scans.

The U.S. Census Bureau applies the term Hispanic or Latino to all people who classify themselves as Cuban, Puerto Rican, Mexican, Mexican American, Chicano, or other Spanish/Hispanic/Latino. The latter category includes individuals from the Dominican Republic, the Spanish-speaking countries of Central and South America, Spain, and individuals who self-identify as Spanish Americans (US Census Bureau 2000). Such a broad definition clusters populations of diverse cultural features and genetic backgrounds. In addition, each population has its own patterns and histories of immigration and settlement, marriage practices, as well as specific fertility and mortality rates. Thus, “Hispanic” is a broad term referring to individuals that can trace their ancestry, to varying degrees, to several continents. A good example of this can be found in a paper by Hanis *et al.* (1991). These authors estimated the European, Native American and African genetic contribution to Mexican, Puerto Rican and Cuban Hispanic samples, based on serum protein and blood group markers. They described a wide variation in the relative contribution from the three parental groups to each sample, with the European ancestry ranging between 45% and 62%, the Native American ancestry between 18% and 31%, and the African ancestry between 8% and 37%. Hanis *et al.*'s results were based on a few classical markers of low information content with respect to ancestry. Now, it is possible to use highly informative markers, carefully selected on the basis of a large frequency differential between parental groups, to provide a more accurate picture of the admixture proportions and dynamics in the population under study. In view of the cultural and genetic heterogeneity observed among Hispanic populations it is, therefore, critical to examine each population independently to understand their genetic histories and to study how these populations might best be used for the identification of disease genes using AM.

Hispanics from a region known geographically as the San Luis Valley (SLV), that extends through southern Colorado and northern New Mexico, as well as Hispanics from surrounding areas, commonly self-identify as Spanish Americans, and generally emphasize their Spanish ancestry over their Mexican and/or Native American roots (Swadesh, 1974). The situation of Spanish Americans differs from that of most other Hispanic groups in that they became U.S. citizens following the annexation of present-day New Mexico and parts of Colorado to the U.S. in 1848, and are settled in a relatively compact group centered around Santa Fe, NM, which allowed them to preserve their culture (Zeleny, 1974). As a result of its geographical location, a valley 7100 feet above sea level and surrounded on three sides by mountain ranges, the SLV has remained fairly isolated. Furthermore, in the early days of the colonization of the area the presence of several Native American tribes in the SLV prevented the Spaniards from successfully occupying this territory (McConnell Simmons, 1999). It was not until the 1830s that the first Spanish Americans migrated from Taos and Abiquiu in New Mexico to settle in the valley. The SLV, unlike Texas and California, received little Mexican immigration until the late 20th century (Hamman *et al.* 1989).

We have studied the admixture pattern, dynamics, and extent of LD, of the Hispanic population living in the SLV in southern Colorado, using a panel of Ancestry Informative Markers (AIMs). We have also analyzed the relationship between individual ancestry and a complex phenotype, skin pigmentation, which varies between parental groups. Skin pigmentation is an ideal model phenotype, since it varies widely between the parental populations, has little environmental variability and is easily measured. Several complex diseases, e.g. type 2 diabetes, obesity, and gallbladder disease, are thought to show increased prevalence in the various Hispanic populations, in part because there are risk alleles at high frequencies in the Native American parental populations. These diseases and the genes that modify their risk may have similar properties to the pigmentation phenotypes and the genes determining population differences in skin, hair and eye color.

Subjects and Methods

Characteristics of the SLV Hispanic sample. The sample was selected from participants of the San Luis Valley Diabetes Study (SLVDS), a geographically based study of the natural history, incidence and risk factors for type 2 diabetes conducted in the counties of Alamosa and Conejos in southern Colorado. These counties are 47.7% Hispanic (US Census, 2000). Informed consent from all participants and approval by the IRB of the University of Colorado were obtained prior to data collection. Additional approval of this work was obtained through the Penn State University IRB (ORC# 00M0453). Ethnicity, based on the 1980 US census question, was self-reported as Mexican, Mexican American or Chicano (N = 118), Cuban or Puerto Rican (N = 0) and other Spanish/Hispanic (N = 292). The procedure for selecting the SLVDS study subjects has been described in detail by Hamman *et al.* (1989). In summary, diabetics in the study area were identified through all health care facilities and through advertisements in local newspapers, presentations to local organizations, and local radio programmes. Eligible subjects with a medical diagnosis of diabetes were 20–74 years of age, residents of the study area, mentally competent and spoke either Spanish or English. Eighty-two percent of eligible subjects (N = 440) attended the baseline data collection clinic (1984–1988). Controls were selected using a two-stage sampling method. First, 57% of all occupied structures in the two county area were sampled and enumerated. Enumerated persons of 20–74 years of age were the sampling frame for the second stage of control selection, where subjects were randomly selected within age, sex, ethnic group and country strata to match the age and sex distribution of diabetics. Sixty-seven percent of eligible controls (N = 1351) attended the baseline clinic. Fifty-nine percent of the Hispanic SLVDS participants with a history of diabetes were confirmed by oral glucose tolerance test, and controls with confirmed normal glucose tolerance test were selected for these admixture studies. Overall, a total of 184 individuals with diabetes and 260 individuals with normal glucose tolerance at the SLVDS baseline visit were identified and genotyped for this study.

Parental population samples. In order to estimate admixture proportions in the population of the SLV, we genotyped samples representing three parental populations. These samples consisted of European Spanish from Valencia (N = 72), Native Americans including Mayas from the Yucatan Peninsula (N = 96) and Pima (N = 33), Cheyenne (N = 31) and Pueblo (N = 24) from Southwestern U.S., and Africans from Nigeria (N = 134), Sierra Leone (N = 93) and the Central African Republic (N = 52). Parental groups were tested for the presence of association between unlinked AIMs. Using the program STRUCTURE (Pritchard *et al.* 2000) we were able to distinguish several recently admixed individuals in the parental populations and eliminate them from the analysis. Altogether we eliminated 8 Mayan subjects and 15 Southwestern Native American subjects because of evidence of recent admixture. In this way we reduced the structure created by recent admixture in the parental populations. It is important to point out that we removed only those individuals showing strong evidence (>10%) of non-Native American ancestry. Other, less admixed, subjects were retained as part of the sample.

Ancestry Informative Markers (AIMs). Accurate estimates of admixture require a panel of unlinked markers with a large frequency difference (δ) in alleles between the parental populations. For dimorphic markers, $\delta = |p_1 - p_2|$, where p_1 and p_2 are the frequencies of one of the alleles in parental populations 1 and 2, respectively. Given the limited available allele frequency information on Native Americans and Hispanics, we established the threshold δ as a frequency differential of 30% in this study. We selected Single Nucleotide and Deletion/Insertion polymorphisms (SNPs and DIPs) from on-line databases (Marshfield, The SNP Consortium, dbSNP, Whitehead Institute, ALFRED) and published reports of allele frequencies in diverse populations (Bieber *et al.* 1996; Parra *et al.* 1998; Collins-Schramm *et al.* 2002), and tested our set of parental populations to corroborate the reported δ levels. In this work we have used a panel of 22 autosomal markers, informative for Native American/European, African/European, and Native American/African comparisons. Table 1 lists all the markers used in this study, their allelic

Table 1 AIM allele frequencies in the parental populations and the SLV Hispanic population and corresponding δ levels

| AIMs | Spanish | Native Americans | | Southwestern populations | Africans (average) | δ | δ | δ | SLV Hispanics |
|----------------------|---------|------------------|--------------------|--------------------------|--------------------|-------------------------|-----------------|-------------------------|---------------|
| | | (average) | Mayas | | | Spanish/Native American | Spanish/African | Native American/African | |
| TSC-1102055*T | 0.921 | 0.137 | 0.138 | 0.136 | 0.487 | 0.784 | 0.434 | 0.350 | 0.598 |
| MID-93*ins | 0.220 | 0.919 | 0.917 | 0.921 | 0.739 | 0.699 | 0.519 | 0.180 | 0.442 |
| MID-52*ins | 0.918 | 0.237 | 0.180 | 0.293* | 0.737 | 0.681 | 0.181 | 0.500 | 0.686 |
| WI-11909*G | 0.845 | 0.181 | 0.143 | 0.220 | 0.805 | 0.664 | 0.040 | 0.624 | 0.517 |
| CKM*T | 0.257 | 0.904 | 0.925 | 0.882 | 0.164 | 0.647 | 0.093 | 0.740 | 0.479 |
| WI-11153*G | 0.172 | 0.819 | 0.779 | 0.858 | 0.785 | 0.647 | 0.613 | 0.034 | 0.394 |
| PV 92*ins | 0.171 | 0.792 | 0.799 | 0.786 | 0.225 | 0.621 | 0.054 | 0.567 | 0.420 |
| DRD2-Taq D*T | 0.630 | 0.045 | 0.052 | 0.038 | 0.135 | 0.585 | 0.495 | 0.090 | 0.427 |
| MID-575*ins | 0.993 | 0.416 | 0.420 | 0.412 | 0.876 | 0.577 | 0.117 | 0.460 | 0.831 |
| WI-14319*C | 0.142 | 0.716 | 0.651 | 0.781* | 0.386 | 0.574 | 0.244 | 0.330 | 0.295 |
| DRD2-Bcl I*C | 0.135 | 0.665 | 0.653 | 0.677 | 0.063 | 0.530 | 0.072 | 0.602 | 0.277 |
| WI-17163*G | 0.197 | 0.690 | 0.721 | 0.659 | 0.054 | 0.493 | 0.143 | 0.636 | 0.391 |
| WI-7423*T | 0.517 | 0.058 | 0.088 | 0.028* | 0.000 | 0.459 | 0.517 | 0.058 | 0.303 |
| CYP19-E2*T | 0.287 | 0.741 | 0.733 | 0.750 | 0.332 | 0.454 | 0.045 | 0.409 | 0.377 |
| TYR 192*A | 0.485 | 0.034 | 0.012 | 0.057* | 0.005 | 0.451 | 0.480 | 0.029 | 0.274 |
| MID-161*ins | 0.508 | 0.109 | 0.064 | 0.154* | 0.637 | 0.399 | 0.129 | 0.528 | 0.391 |
| SGC-30610*T | 0.300 | 0.699 | 0.793 | 0.604** | 0.401 | 0.399 | 0.101 | 0.298 | 0.442 |
| WI-4019*A | 0.295 | 0.618 | 0.489 | 0.746** | 0.430 | 0.323 | 0.135 | 0.188 | 0.419 |
| D11S429*T | 0.440 | 0.119 | 0.109 | 0.128 | 0.087 | 0.321 | 0.353 | 0.032 | 0.407 |
| DYS199*T | 0.000 | 0.605 | 0.710 ^a | 0.500 ^b | 0.000 | 0.605 | 0.000 | 0.605 | 0.038 |
| FY-null*T | 0.993 | 1.000 | 1.000 | 1.000 | 0.001 | 0.007 | 0.992 | 0.999 | 0.951 |
| F13B*G | 0.104 | 0.018 | 0.035 | 0.000* | 0.704 | 0.086 | 0.600 | 0.686 | 0.063 |
| GNB3*T | 0.414 | 0.364 | 0.402 | 0.326 | 0.795 | 0.050 | 0.381 | 0.431 | 0.371 |

NOTE. - Useful Spanish/Native American, Spanish/African, and Native American/African δ s are shown in bold. Asterisks denote significant differences between Native American populations: * $p < 0.05$, ** $p < 0.001$. DYS199 allelic frequencies for parental groups were obtained from the literature (Underhill *et al.* 1996; Lell *et al.* 1997). ^aincludes data from: Mayas, Mixe, Mixtecs and Zapotecs. ^bincludes data from: Seminoles and Navajo.

frequencies in the parental groups and the SLV population, and the observed δ between Spanish and Native Americans, Spanish and Africans, and Africans and Native Americans. Additionally, we typed DYS199, a Y-chromosome polymorphism that has been identified as a Native American-specific AIM (Underhill *et al.* 1996; Lell *et al.* 1997). Detailed information regarding these markers is available at dbSNP, under PSU-ANTH as the submitter handle. Most AIMs were screened with the McSNP method as described by Akey *et al.* (2001) and Ye *et al.* (2002). PV92, an *Alu* insertion, was assayed using conventional agarose gel electrophoresis and ethidium bromide staining. PCR was performed in 96-well OmniUltra plates with a PCR MBS thermal cycler, both obtained from ThermoHybaid (Oxford, UK). Amplifications consisted of buffer (50 mM KCl + 10 mM Tris-HCl pH 8.9), MgCl₂ 1.0–2.5 mM, deoxynucleotide triphosphates (dNTPs)

200 μ M each, primers 0.5 mM each, *Taq* polymerase 1 U, genomic template 20 ng, and sterile filtered water (Sigma) up to 25 ml. PCR conditions were as follows: 5 min at 94°C; 35 cycles at 94°C/specific annealing temperature/72°C, 30 seconds each; and 5 min at 72°C. Restriction analysis was performed as recommended by the supplier. Table 2 shows PCR conditions and restriction enzymes used for typing each marker.

Skin pigmentation measurement. The methodology used is explained in detail in Iyengar (1988). Briefly, skin reflectance was measured on the upper inner arm, using a Photovolt model 575 spectrophotometer (Photovolt Instruments Inc., Minneapolis, MN) with a Y-sensor unit and three glass filters: tri-amber (650 nm), tri-green (550 nm) and tri-blue (450 nm). The raw reflectance values of the tri-green filter were used to calculate L* (Lightness),

Table 2 AIMS used for estimating Native American, African and European ancestry in the Hispanic population of the San Luis Valley

| AIMs | alleles | chromosomal location | 5'-3' forward/reverse primers | annealing temperature (°C) | MgCl ₂ (mM) | restriction enzyme |
|--------------------|-----------|----------------------|--|----------------------------|------------------------|--------------------|
| MID-575 | TTC/— | 1p34.3 | F-CAAAATCTGCTCCATGTCCA R-CAAGGGTTAGGGAGGTTGGT | 56 | 1.5 | Ear I |
| FY-null | C/T | 1q23.2 | F-GAACCTGATGGCCCTCATT R-TCAGCGCCTGTGCTTCCATG | 55 | 2 | Nco I |
| F13B | A/G | 1q31.3 | F-CCTGAGTAATGGTTACATCTCTGA R-CCCTCCAGTGGTTTTGTACC | 58 | 1.5 | Nsi I |
| TSC-1102055 | C/T | 1q32.1 | F-GCCTCTCGATGAGTAAATATG R-TTTGATGATACCTACGCATAGTCTG | 63 | 2.5 | Hae III |
| WI-11153 | G/C | 3p12.1 | F-CTTCAAATTGCTTTAAGTAC R-ATCCAACAGTCAAGGTCTaC | 55 | 1.8 | Bsa A I |
| MID-52 | TTG/— | 4q24 | F-AGAGGCAGCTAATAATTgAC R-GATGGTTTTGTGTAGTgAC | 51 | 2.5 | Ahd I |
| SGC-30610 | A/T | 5q11.2 | F-GCCTGAGAATCCCACACATT R-CCTCTGATGGGAGATGGAAA | 60 | 2 | Xmn I |
| WI-17163 | A/G | 5q33.1 | F-CCATTTCTTTGTAAATAACAATAACtTT R-CTTTTCCAATGTTGGTTTACAGAATC | 60 | 2 | Dra I |
| WI-4019 | A/G | 7q21.3 | F-CAGGCCAAGAGCGTCCTA R-TGCCACTCTGTGAACAGCAA | 55 | 1.5 | Nla III |
| WI-11909 | A/G | 9q21.31 | F-ATTTGTGTTGGGTGGTcAG R-GTCCTCCTCTGAGATTTTCTG | 52 | 1.5 | Xba I |
| D11S429 | C/T | 11q13.3 | F-GGATTCCTCCCTTTGTAGG R-CAGAGACAGCGGCTAGAGA | 60 | 2 | Sac I |
| TYR 192 | A/C | 11q14.3 | F-TTATGTGTCAATGGATGCAC R-GCTTCATGGGCAAATCAAT | 62.5 | 2.5 | Mbo I |
| DRD2-Bcl I | C/T | 11q23.2 | F-TGCTGTCAGAATCACCTATTCAA R-TAAGGGCAGCAGGAACCAC | 55 | 1 | Bcl I |
| DRD2-Taq D | C/T | 11q23.2 | F-TGGGGGTGTGAAGAAAAGAG R-TTTAGTAGCAGAGGAAGGAGTGG | 61 | 1.5 | aTaq I |
| GNB3 | C/T | 12p13.31 | F-CATCATCTGCGGCATCAC R-AATAGTAGGCGGCCACTGAG | 58 | 1.5 | BseD I |
| WI-14319 | C/T | 15q14 | F-CATCTGAGTGCAAGATAAAAAGGA R-CCCACCCCAAATCATCTAT | 55 | 1 | Rsa I |
| CYP19-E2 | G/T | 15q21.2 | F-GCATACTCCTATGGGTTgTC R-TGTGAACAGGAGCAGATGGC | 58 | 1.5 | Hae III |
| PV92 | Alu indel | 16q23.3 | F-AACTGGGAAAATTTGAAGAGAAAGT R-TGAGTTCTCAACTCTTGTGTGTTAG | 55 | 2.5 | N/A |
| WI-7423 | C/T | 17p13.1 | F-CTCCTTGGCAGGGATTTGT R-CTGGTATCCACGGTGCAAG | 62 | 2 | Sma I |
| CKM | C/T | 19q13.2 | F- AGCTCATGGTGGAAATGGAG R-GCAGGCGCCTACTTCTGG | 55 | 1 | aTaq I |
| MID-161 | ATC/— | 20q11.21 | F-GTACTTCCACGGCACAATcC R-GCCTAGAGTACAGGGTGAGCA | 58 | 2.5 | Bsl I |
| MID-93 | AGA/— | 22q13.2 | F-GGGGTGTTTCATGAGCCATAG R-CGTTGTGTTTTATTTGTGCAGTC | 59 | 2.5 | Mbo II |
| DYS 199 | C/T | Yq11.221 | F-CCTGACAATGGGTCACCTCT R-TTTAGGTACCAGCTCTTCCcAAT | 56 | 1.5 | Mfe I |

NOTE.- Chromosomal location as given in the UC-Santa Cruz database. Restriction sites were engineered for those markers lacking a natural site; the modified base is shown in lower case.

a color parameter that represents the white-black axis, as $116 * (y/100)^{1/3} - 16$, where y is the average of the tri-green filter measurements taken on the upper inner aspect of both arms (Billmeyer & Saltzman, 1981).

Data analysis. Group admixture estimates were calculated using Long's (1991) weighted least squares (WLS) method with the program ADMIX. Individual admixture was estimated by the maximum-likelihood (ML)

method used by Hanis *et al.* (1986) and Chakraborty *et al.* (1986), using the program IBGA (Individual Bio-Geographical Ancestry) written by CLP and MDS. This program also estimates the support intervals on the maximum likelihood estimates as the points at which M (admixture proportion) equals the maximum minus one log unit. Additionally, we tested for admixed individuals in the parental samples, and for structure in the study population, with the program STRUCTURE 2.0 (Pritchard *et al.* 2000). This program uses Markov chain simulation to generate the posterior distribution of individual admixture proportions given the observed marker genotypes, based on a hierarchical model for population admixture, individual admixture and locus ancestry. Ancestry-specific allele frequencies (the frequencies of each allele given the subpopulation of ancestry of the gene copy) are estimated from the data. Haplotype frequencies were estimated by an expectation maximization algorithm using the program 3LOCUS (Long *et al.* 1995). To test for allelic association between pairs of loci, a likelihood ratio statistic (G) was calculated as implemented in 3LOCUS. To detect the presence of genetic structure in the sample, we compared the proportion of significant associations between unlinked markers with the expected values at the 5% significance level. We performed two additional tests for detecting population structure. Since the SLV Hispanic population has originated mainly from the admixture of Native American and European populations, we analyzed the ratio D_t/D_0 as described by Pfaff *et al.* (2001) using the dihybrid model of admixture. Initial disequilibrium levels (D_0) were calculated as $m(1-m)\delta_A\delta_B$, where m and $1-m$ correspond to the admixture contribution from each parental group, and δ_A and δ_B are the frequency differentials between these populations at loci A and B, respectively, and plotted these against observed disequilibrium levels (D_t). It should be noted that when calculating D_0 , δ is defined as (p_1-p_2) , and so can be either positive or negative. We also tested for correlation between individual admixture estimates obtained with different sets of markers. The whole panel of markers was randomly split into two subsets, and we estimated the correlation of individual ancestry estimates generated by each subset of markers. The fit of the genotype frequencies to the Hardy-Weinberg equilibrium was checked with the program HW implemented in the com-

puter package GENEPOP 2.0 (Raymond & Rousset, 1995).

We used linear regression models implemented in the program SPSS (version 10) to test for an association of ancestry with skin pigmentation (L^*) as a dependent variable.

Results

Admixture Proportions in the San Luis Valley

Based on our set of 22 autosomal AIMs and a trihybrid model of admixture, we used the WLS method to estimate the composition of the total sample as $62.7 \pm 2.1\%$ European, $34.1 \pm 1.9\%$ Native American, and $3.2 \pm 1.5\%$ African. The allele frequencies for each parental population and the admixed population are shown in Table 1, together with the European/Native American, the European/African, and the African/Native American δ levels. When the sample was subdivided according to disease status we observed no significant difference in admixture proportions between diabetics and controls (Table 3). Therefore we used the combined sample for subsequent analyses. However, it is important to note that we have not tested the hypothesis of whether there is an effect of admixture on the risk of type 2 diabetes. Doing so would require a careful consideration of known environmental covariates (e.g. socioeconomic status; Parra *et al.* submitted).

We calculated admixture for each individual in the sample using a ML method implemented in the program IBGA, which also gives one-log support intervals for the estimates. The average individual estimates using the ML method are very similar to the estimates obtained with the WLS method. The average European, Native American and African components were estimated as 61.6%, 32.8%, and 5.6%, respectively. The distribution of individual admixture values can be plotted as a triangle, where each vertex represents the contribution of one of the parental populations. Figure 1 shows the triangle plot corresponding to the individual admixture estimates of the SLV Hispanic sample. Admixture estimates in SLV Hispanics encompass the whole range of Native American and European contributions. A lower African contribution is also evident in this graphical representation.

Table 3 Admixture proportions in SLV Hispanics by disease status, sex, ethnic self-identification, and mention of having Native American ancestry (mean \pm s.e.)

| | N | Spanish contribution (%) | Native American contribution (%) | African contribution (%) |
|---|-----|--------------------------|----------------------------------|--------------------------|
| Diabetics | 168 | 60.9 \pm 2.2 | 35.1 \pm 2.0 | 4.0 \pm 1.7 |
| Controls | 230 | 64.0 \pm 2.3 | 33.4 \pm 2.2 | 2.6 \pm 1.6 |
| Males | 179 | 63.0 \pm 2.0 | 34.2 \pm 1.8 | 2.7 \pm 1.4 |
| Females | 231 | 62.7 \pm 2.3 | 33.7 \pm 2.1 | 3.7 \pm 1.7 |
| Mexican Americans/Chicanos | 118 | 60.4 \pm 2.1 | 36.9 \pm 1.9 | 2.7 \pm 1.4 |
| Spanish Americans | 292 | 63.8 \pm 2.3 | 32.7 \pm 2.1 | 3.5 \pm 1.7 |
| Known Native American ancestry | 91 | 61.5 \pm 2.4 | 35.2 \pm 2.2 | 3.3 \pm 1.8 |
| Unknown Native American ancestry | 319 | 63.2 \pm 2.2 | 33.6 \pm 2.0 | 3.2 \pm 1.6 |
| Parental: Mayas | 405 | 63.0 \pm 2.1 | 33.7 \pm 1.9 | 3.3 \pm 1.6 |
| Parental: Southwestern tribes | 405 | 62.8 \pm 2.3 | 34.0 \pm 2.1 | 3.2 \pm 1.6 |

NOTE. - The last two rows are the estimates obtained when only one Native American population is used as parental.

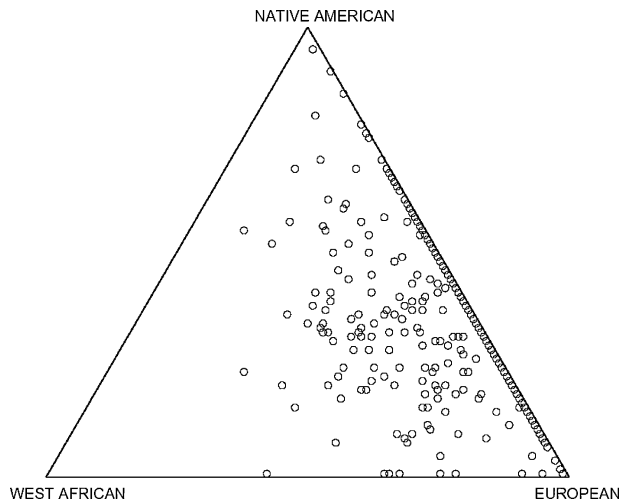


Figure 1 Distribution of individual admixture estimates for the SLV Hispanic population. Each vertex of the triangle represents a parental population. Open circles depict individuals, and their relative position in the diagram corresponds to each individual's ancestry proportions.

We used two Native American samples to estimate the Native American ancestral frequencies, Mayas and Southwestern American Indians (including Pima, Cheyenne and Pueblo). We observed significant differences between the two samples in eight of the markers in Table 1 (WI-4019, $p = 0.00002$; SGC-30610, $p = 0.002$; MID-161, $p = 0.01$; WI-14319, $p = 0.024$; MID-52, $p = 0.025$; WI-7423, $p = 0.032$; F13B, $p = 0.033$; TYR-192, $p = 0.049$). Nonetheless, if admixture was calculated using either the Mayas or the Southwestern tribes as the Native American parental

population instead of the average allele frequencies of both, no significant difference was found between the estimates (Table 3). The fact that both Native American groups may be used individually or combined to generate the parental frequencies is also apparent from the graphs in Figure 2. As suggested by Long (1991), we plotted the frequency difference between the admixed and one of the parental populations ($p_h - p_2$) against the frequency difference between both parental populations ($p_1 - p_2$) for each marker. Figure 2 shows the results of these plots using either of the two Native American parental populations, the Mayas (panel 2A), the Southwestern Pueblo, Pima and Cheyenne (panel 2B), or the average Native American frequencies (panel 2C). The goodness of fit of each of these three options is remarkable (Mayas, $R^2 = 0.9859$; Southwestern groups, $R^2 = 0.9830$; Native American average, $R^2 = 0.9866$). The slope of the trend lines represents an estimate of European ancestry (Long, 1991).

When Native American ancestry is calculated using the Y-chromosome AIM (DYS199), the estimate drops considerably to 3.8%, indicating that the male Native American contribution to this population has been very small, and that there has been sex-biased gene flow in this population.

Genotype frequencies for 17 markers did not deviate from Hardy-Weinberg expectations. Five AIMS, on the other hand, showed a significant deviation from Hardy-Weinberg equilibrium: FY-null ($p = 0.002$), MID-93 ($p = 0.021$), TSC1102055 ($p = 0.031$), DRD2-Bcl

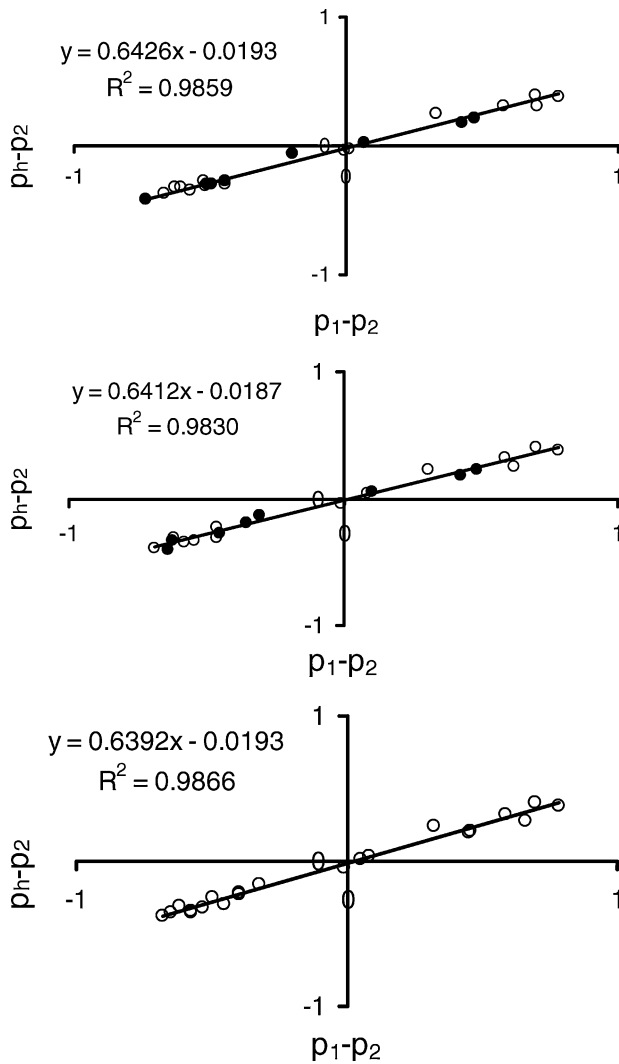


Figure 2 The difference between the allele frequencies in the admixed population and each (panels 2A and 2B) or both (panel 2C) of the Native American parental populations is plotted against the allele frequency difference between the parental populations, for every marker. ● = Markers that show significant differences in allele frequency between Mayas and Southwestern Native Americans (WI-4019, SGC-30610, MID-161, WI-14319, MID-52, WI-7423, F13B, TYR-192).

I ($p = 0.033$), and GNB3 ($p = 0.039$). This deviation was due either to excess of homozygotes (TSC1102055, MID-93 and FY-null) or heterozygotes (DRD2-Bcl I and GNB3).

We subdivided the sample according to sex (males = 179, females = 231), ethnic self-identification (Spanish American = 292, Mexican American/Chicano = 118), and whether the subject stated had any Native American ancestry (yes = 91, no = 319), and calculated the ad-

mixture proportions for each subgroup. Results of this analysis are shown in Table 3. Basically, there are no differences between the groups for most of the above categories, except for ethnic affiliation, where Mexican Americans show more Native American admixture than Spanish Americans.

Admixture Dynamics and Population Structure in the San Luis Valley

Since population structure has been shown to be a source of confounding in association studies performed in admixed populations (Kittles *et al.* 2002) we tested for stratification in the SLV population in several ways. In the SLV sample, 5.9% of unlinked marker pairs exhibited gametic disequilibrium. This is slightly higher than what is expected by chance assuming a 5% alpha level. As anticipated, high G values were consistently observed between two markers located within the *DRD2* gene ($G = 126.2$), closely linked at 4.7 kb on chromosome 11q. Other AIMs in this region that were linked at a distance of 39 cM (D11S429/*DRD2*), 22 cM (TYR 192/D11S429) and 17 cM (TYR192/*DRD2*) did not show significant association. Markers linked at 20 cM on chromosome 15q (WI-14319/*CYP19-E2*) showed significant LD ($G = 8.17$, $p < 0.01$).

The plot of the observed disequilibrium (D_t)/initial disequilibrium (D_0) ratio as formulated by Pfaff *et al.* (2001) for SLV Hispanics is presented in Figure 3. There is a significant correlation ($R^2 = 0.0948$, $p = 2.9^*$

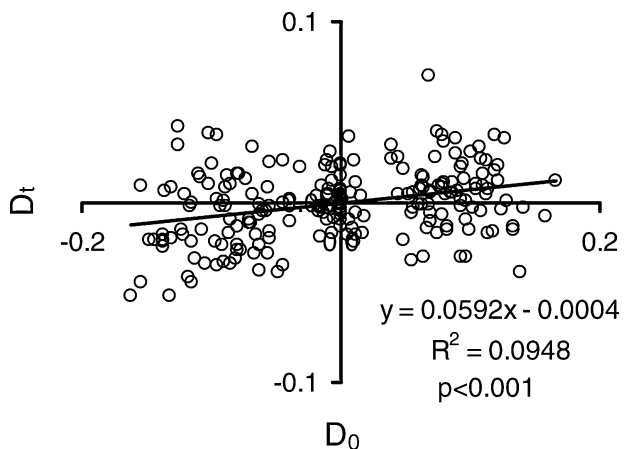


Figure 3 Observed LD (D_t) versus initial LD due to admixture (D_0), between pairs of markers.

10E-6) between these variables. Pfaff *et al.* (2001) have simulated the relationship between D_t and D_0 for two models of admixture dynamics. In the Hybrid Isolation (HI) model, admixture takes place in a single generation, with the hybrid population subsequently evolving by drift and recombination alone and receiving no further gene flow from the parental populations. Under this model there is no significant correlation between the estimated initial disequilibrium due to admixture, and the observed disequilibrium. In the alternative model of admixture dynamics, which we have termed the Continuous Gene Flow (CGF) model, there is a constant rate of gene flow in each generation between one of the parental populations and the admixed population. Under this model, a positive and significant correlation would be expected between D_t and D_0 .

In a population with no genetic structure we do not expect individual ancestry estimates obtained with different sets of markers to be correlated. Evidence for this comes from the analysis of a simulated population with the same admixture proportions and size as the SLV Hispanics under both models of admixture dynamics. In the simulations of the CGF model there was a low but significant correlation between independent estimates ($R^2 = 0.0242$, $p = 9.8 \times 10^{-4}$). In contrast, in the HI model estimates were not correlated ($R^2 = 1.3 \times 10^{-4}$, $p = 0.814$). When analyzing the SLV population our results indicated structure in this group ($R^2 = 0.0266$, $p = 5.4 \times 10^{-4}$).

Finally, we used the STRUCTURE program to evaluate the existence of hidden structure in the sample. We ran the program with $K = 1$, $K = 2$, and $K = 3$ as the predefined settings for the number of populations, using 30,000 iterations for the burn-in period and 70,000 additional iterations to obtain parameter estimates. We used several of the available options of the admixture and the linkage models implemented in the STRUCTURE 2.0 program, such as independent allele frequencies, correlated allele frequencies, single alpha, or independent alphas. When the admixture model was used no population structure was detected (probability of $K = 1$ was highest). However, under the linkage model $K = 2$ was most probable. In summary, the program STRUCTURE, the D_t/D_0 plot, and the positive correlation of admixture estimates using marker subsets,

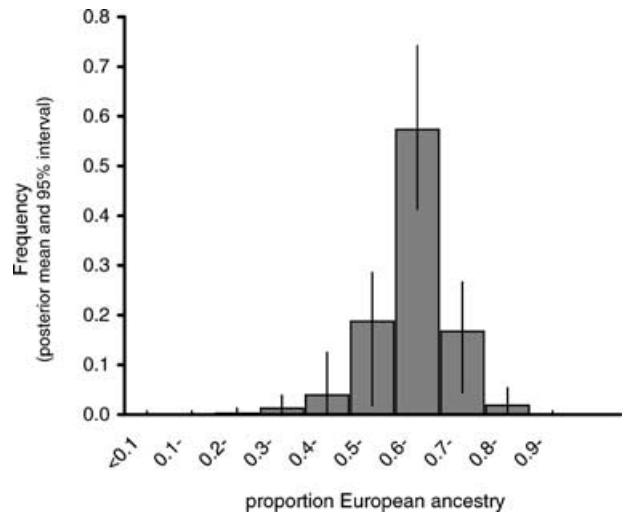


Figure 4 Estimated distribution of individual admixture in the SLV Hispanic population obtained with the Bayesian model implemented in the program ADMIXMAP.

indicate that there is some structure due to admixture in the SLV sample.

The presence of genetic structure is also evident in the distribution of individual admixture in the population. Figure 4 shows the posterior means and 95% posterior intervals for the proportions of the sample falling into each 10% interval of European admixture from 0 to 1. The estimates of individual admixture were obtained with the program ADMIXMAP. More than 90% of the subjects have between 50% and 80% European admixture. The average crossover rate between European and non-European ancestry was estimated as 3.2 per 100 cM (95% posterior interval 2.0 to 5.0).

Native American Ancestry and Skin Pigmentation

We compared individual ancestry as estimated with our molecular markers to skin pigmentation, a complex phenotype that differs between the parental populations. In a population with no genetic structure no correlation is expected between these variables, unless a marker is linked to a gene influencing pigmentation and admixture is recent or continuous. Figure 5 shows the relationship between individual admixture estimates and skin reflectance represented as L^* . The age-adjusted slope of the regression of reflectance on European admixture was estimated as 4.9 units ($p < 0.001$) in the classical analysis,

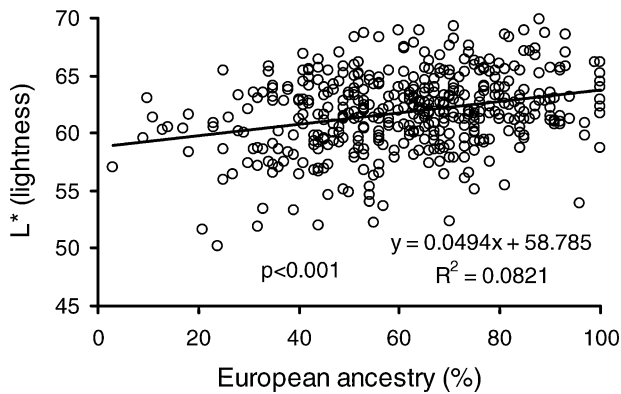


Figure 5 Correlation between skin pigmentation measured as L^* and percentage of European ancestry for each subject in the SLV Hispanic population.

and as 29.6 units (95% posterior interval 29.0 to 30.2) in the Bayesian analysis, which allows for uncertainty in the estimation of admixture. The European contribution is greater in subjects with a higher L^* value and, hence a lighter skin. This relationship is only expected in the presence of population structure.

Discussion

We have compiled a set of genetic markers that are useful for estimation of ancestral proportions in admixed populations and individuals. Since most of these markers show no significant differences between two discrete Native American populations, Southwestern American Indians and Mayans, these markers may be well suited to studying a wide range of Hispanic populations either in the U.S. or in Latin America. Clearly, population history must be taken into consideration for each particular population, and specific markers added that represent other populations involved in the admixture process. Polymorphic markers situated throughout the genome, with large frequency differences between parental populations, like those presented here, will facilitate genome-wide AM in admixed populations.

The estimates of admixture proportions in the Hispanic population of the SLV show a gene pool derived primarily from Europeans (63%) and Native Americans (34%), with a minor but definite African contribution of approximately 3%. Native American tribes that populated the area initially comprised the Pueblo Indians, and the Utes, who entered western Colorado

by A.D 1000–1200. After this time, Navajos, Apaches, Comanches, Arapahos, Cheyennes and Kiowas also migrated through the valley. In 1598 Don Juan de Oñate established the first Spanish settlement in the Southwest at San Gabriel, New Mexico (McConnell Simmons, 1999). Despite several incursions by the Spanish, the Utes successfully discouraged the colonization of the SLV region until the 1830s. Mexico became independent of the Spanish crown in 1821 and soon encouraged the settlement of regions north of New Mexico to control increasing U.S. interest in the area. To accomplish this goal the Mexican government made several land grants to families living in New Mexico. According to historical accounts, these first settlers were Spaniards, some were mestizos (individuals of mixed Spanish and Native American ancestry), and some were Indians who joined the villagers, embracing their culture, taking up their language and even their names (Zeleny, 1974). Besides New Mexican indigenous peoples, Mexican Indians and Mexican mestizos might also have contributed to the SLV population. Spanish colonization parties frequently contained Mexican Indian soldiers, probably Tlaxcalans and natives from Mexico's northern states. It was not until the early 1850s that the Anglo-Americans first entered southern Colorado (Zeleny, 1974). A reliable estimate of admixture proportions depends largely on how closely related to the real ancestral frequencies are those found in populations selected as putative parentals (Chakraborty, 1986). In light of the historical records, we chose as parentals a Mexican native population: Mayas from the Yucatan Peninsula, and a group of Native American populations from the Southwest: Pueblo, Pima and Cheyenne. Although there were slight differences in allele frequencies at some loci, ancestry estimates did not vary significantly whether obtained with one, the other or the average of both Native American parental choices. Similarly, Collins-Schramm *et al.* (2002) found small intra-ethnic differences for their microsatellite and insertion/deletion polymorphisms between Pima and Yavapai, two Arizona tribes belonging to different linguistic groups. Since we have searched for markers that maximize the difference between Native Americans and Europeans it is likely that the intrapopulation variability in each ancestral group is generally reduced. This may be one reason why our marker set is apparently robust with respect to the

parental populations used to determine the AIM allele frequencies.

Other Hispanic populations of Mexican origin show similar levels of Native American ancestry: 29% in Arizona (Long *et al.* 1991), 30–31% in Starr County, Texas (Hanis *et al.* 1986, 1991; Cerda-Flores *et al.* 1992), 32% in Oakland, California (Reed, 1974), 40% in Colorado (Gottlieb & Kimberling, 1979), 18–46% in San Antonio, Texas (Relethford *et al.* 1983), and 45–50% in northern California (Collins-Schramm *et al.* 2002). Previous Native American admixture estimates in the SLV population using serological and skin reflectance data were 46.0% and 33.5%, respectively, in a sample of 392 Hispanics, most of which are included in our sample set (Iyengar, 1988). An increase in the number of serological markers used (from eight to eleven) reduced the estimate to 33.2% in the same sample, although only the RH*cde, FY*A and *B, and F13B*1 and *3 had allele frequency differences between founder populations greater than 30% (Merriwether *et al.* 1997). Thus, our marker set provides an estimate that coincides with previously reported findings, and is also more reliable because of the large δs displayed by the AIMs we used.

Merriwether *et al.* (1997) described the existence of directional gene flow in this population, as mtDNA admixture estimates indicated a 2.5-fold higher Native American contribution than nuclear estimates (85.11% vs 33.15%). We have tested the DYS199*T allele, which is restricted to the indigenous populations of the Americas. The DYS199*T has a frequency of 2.25% in the Hispanics of the SLV, giving an admixture estimate of 3.8% Native American male ancestry. Both studies have revealed a pattern of directional mating in this population, an asymmetric interaction between Spanish males and Native American females, much like in other Hispanic populations of Latin America (Green *et al.* 2000; Carvajal-Carmona *et al.* 2000, 2003; Rodriguez-Delfin *et al.* 2001). During the conquest and colonization of America the immigration of women from the Iberian Peninsula was significantly lower than that of men, so European males frequently took native women as wives or partners (Mörner, 1967). After the initial directional contact between European and Native American populations it seems likely that the admixed group became mostly endogamic, which would explain the

high levels of Native American mtDNA (Merriwether *et al.* 1997).

It is possible that the African component detected in this sample ($3.2 \pm 1.5\%$) was already present in the Mexican mestizos that migrated north. Mexico experienced a massive slave trade during the 16th and 17th centuries, at which point the black population was larger than the white population (Aguirre-Beltrán, 1946). Palmer (1976) reported that at least 200,000 enslaved Africans were brought to Mexico during the colonial period, although 500,000 might be a more accurate figure due to unlicensed importation or smuggling (Aguirre-Beltrán, 1946). During Mexico's last century of Spanish rule, slave trade had diminished substantially, and by the beginning of the 19th century the majority of Afro-Mexicans were admixed. A study of African admixture in current Mexican populations found it to range from 2.7% in Oaxaca to 40.5% on the east coast (Lisker *et al.* 1996). A recently published article by Cerda-Flores *et al.* (2002), where they analyzed three Mexican mestizo populations from the states of Nuevo León, Jalisco, and the Federal District for two nuclear DNA polymorphisms (D1S80 and HLA-DQA1), reveals similar admixture proportions in these populations to our findings in the SLV Hispanics (50–60% European, 37–49% Native American, 1–3% African). Remarkably, the northeastern population of Nuevo León presents the largest African contribution (3%). Historical records also refer to individuals of African ancestry who accompanied Spanish explorers in their expeditions throughout the Southwest (Taylor, 1998). Although gene flow from African Americans in the U.S. cannot be ruled out, a very small number of blacks were reported in New Mexico (85, 0.09% of the population) and Colorado (46, 0.13% of the population) in the census of 1860 (Taylor, 1998).

It is interesting to note that Spanish Americans are to some extent less admixed ($32.7 \pm 2.1\%$) with Native Americans than Mexican Americans/Chicanos ($36.9 \pm 1.9\%$), but self-identification as one or the other has, nevertheless, more cultural than biological meaning. Spanish Americans and Mexican Americans in Colorado revealed a degree of Native American admixture comparable to that of other Mexican American populations nationwide, but the latter may not exhibit similar

admixture dynamics. This is something that should be explored particularly for each Hispanic population.

Even though the valley has been relatively isolated since the arrival of the Anglo-Americans in the mid-1800s, there is still genetic structure present in the Hispanic population of the SLV. The significant correlation between individual ancestry and skin pigmentation, and the correlations observed between individual admixture estimates based on different subsets of markers, and between the current disequilibrium levels and the estimated initial disequilibrium due to admixture, give support to the notion that this population is structured. Probable causes of structure in this population might be assortative mating, and a pattern of admixture dynamics that combines continuous gene flow from one of the parental populations for a few generations with subsequent hybrid isolation. Devor (1980) reported a marked departure from random mating in the rural Hispanic population of the village of Abiquiu in northern New Mexico. The number of potential mates appeared to be restricted by culture, religion, and geographic proximity. While these conclusions may not be valid in the case of the SLV, it should be noted that the population we studied is basically rural, with farming and tourism as the primary industries (Hamman *et al.* 1989). However, the level of structure due to admixture seems to be lower than that observed in other admixed populations, such as African Americans from Mississippi and South Carolina. In these samples, 37% and 20%, respectively, of unlinked markers showed a significant association (Pfaff *et al.* 2001). Although there is very little excess association between unlinked markers, the correlation of skin pigmentation with estimated admixture clearly indicate that proportionate admixture varies between individuals, generating cryptic stratification. As mentioned before hidden structure could potentially generate false positive results in association studies that involve admixed populations (Kittles *et al.* 2002). As a result, several methods have been devised that account for the presence of population structure (Ewens & Spielman, 1995; Devlin & Roeder, 1999; Pritchard & Rosenberg, 1999; McKeigue *et al.* 2000; Pritchard *et al.* 2000; Reich & Goldstein, 2001; Satten *et al.* 2001; Pfaff *et al.* 2002; Hoggart *et al.* 2003), one of which has been used effectively with the SLV population (Hoggart *et al.* 2003).

What then are the implications for using this population for AM of complex phenotypes? Having almost 35% of Native American ancestry makes it well suited for applying AM to a disease such as type 2 diabetes, as statistical power to detect genes that underlie ethnic differences in disease risk is maximal when the proportion of admixture from the high-risk population is about 35% (McKeigue, 1998). Even though we detected moderate levels of population structure that could eventually represent a problem, ways of controlling for it are readily available (Hoggart *et al.* 2003). The estimated ancestry crossover rate of 3.2 per 100 cM between European and non-European ancestry is consistent with a hybrid isolation model, in which the average time back to the most recent unadmixed ancestor is about 7 generations. The ancestry crossover rate determines the density of markers required for AM to search a genomic region, and the mapping resolution of such studies. From previous simulations (McKeigue, 1998), we can estimate that AM in a population with ancestry crossover rate of 3.2 would require average marker spacing of 2–3 cM to extract at least 70% of information in the initial genome search. Mapping resolution, however, would be higher than in a population where admixture has been more recent. As the marker set contains only one pair and two triples of linked markers, estimates of the average crossover rate are only approximate at this stage.

Conclusions

This work represents an initial step in the systematic study of Hispanic populations as admixed populations that might be useful for AM. Admixture between populations with widely differing disease rates will produce LD between genes responsible for the disease and genetic markers with different frequencies in the parental populations (McKeigue, 1997). AM is a promising approach that takes advantage of this fact to locate genes for complex diseases using admixed populations. Hispanics in the U.S. experience a high prevalence of chronic disorders, such as diabetes mellitus type 2, obesity, and gall-bladder disease, compared to the general population. It is well known that, due to the demographics of the U.S. Hispanic populations, basically relative youth, high fertility and high immigration rates, their disease burden will only rise with time. The rates of these

particular diseases in Hispanics are intermediate between those of Native Americans and Europeans, and covary with the distribution of genetic admixture (Weiss, 1993). Yet, before the application of AM, an extensive characterization of the admixed population is required, as we have done with SLV Hispanics. Using a Bayesian hierarchical model that deals with the uncertainty of the individual ancestry estimates, we were able to directly and accurately estimate the distribution of admixture proportions in the population. Our study also provides a set of polymorphic AIMs with large δ between parental groups, and quite homogeneous frequencies among different Native American populations that could be used to detect Native American ancestry in a wide array of Hispanic populations. Finally, it contributes to a better understanding of the way admixture has taken place in a particular Hispanic population, by relating the genetic analysis to available historical information.

Acknowledgements

We thank the participants in the San Luis Valley Diabetes Study for their cooperation. We are grateful to Drs. Kenneth Weiss and Anne Buchanan for providing us with the Maya samples, and Dr. Jeffrey Long for the Southwestern Native American samples. We also thank Drs. Kenneth Weiss and Rick Kittles for helpful comments on earlier drafts of this manuscript. This work was supported in part by grants from NIH/NIDDK (DK53958) and NIH/NHGRI (HG002154) to M.D.S. and from NIH/NIMH (MH60343) to P.M.M., and a Wenner-Gren Foundation Developing Countries Training Fellowship to C.B.

Electronic Database Information

Accession numbers and URLs for data in this article are as follows:

ALFRED, <http://alfred.med.yale.edu/alfred>

dbSNP, <http://www.ncbi.nlm.nih.gov/SNP>

Center for Medical Genetics, Marshfield Medical Research Foundation, <http://research.marshfieldclinic.org/genetics>

The SNP Consortium (TSC), <http://snp.cshl.org>

University of California, Santa Cruz, Human Genome

Project Working Draft, <http://genome.ucsc.edu>

US Census Bureau, <http://www.census.gov>

Whitehead Institute, <http://www-genome.wi.mit.edu>

References

- Aguirre-Beltrán, G. (1946) *La población negra de México, 1519–1810: Estudio etnohistórico*. Mexico: Fuente Cultural.
- Akey, J. M., Sosnoski, D., Parra, E., Dios, S., Heister, K., Su, B., Bonilla, C., Jin, L. & Shriver, M. D. (2001) Melting Curve Analysis of SNP's (McSNP): A Simple Gel-Free Low-Cost Approach to SNP Genotyping and DNA Fragment Analysis. *BioTechniques* **30**, 358–367.
- Bieber, H., Bieber, S. W., Rodewald, A. & Barrantes, R. (1996) Microevolution and genetic affinities among six Amerindian tribes of lower Central America: comparative genetic study of serum proteins. *Hum Biol* **68**, 929–953.
- Billmeyer Jr, F. W. & Saltzman, M. (1981) *Principles of color technology*, 2nd edition. New York: John Wiley and Sons.
- Briscoe, D., Stephens, J. C. & O'Brien, S. J. (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. *J Hered* **85**, 59–63.
- Carvajal-Carmona, L. G., Ophoff, R., Service, S., Hartiola, J., Molina, J., Leon, J. P., Ospina, J., Bedoya, G., Fremer, N. & Ruiz-Linares, A. (2003) Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* **112**, 534–541.
- Carvajal-Carmona, L. G., Soto, I. D., Pineda, N., Ortiz-Barrientos, D., Duque, C., Ospina-Duque, J., McCarthy, M., Montoya, P., Alvarez, V. M., Bedoya, G. & Ruiz-Linares, A. (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* **67**, 1287–1295.
- Cerda-Flores, R. M., Kshatriya, G. K., Bertin, T. K., Hewett-Emmett, D., Hanis, C. L. & Chakraborty, R. (1992) Gene diversity and estimation of genetic admixture among Mexican-Americans of Starr County, Texas. *Ann Hum Biol* **19**, 347–360.
- Cerda-Flores, R. M., Villalobos-Torres, M. C., Barrera-Saldaña, H. A., Cortés-Prieto, L. M., Barajas, L. O., Rivas, E., Carracedo, A., Zhong, Y., Barton, S. A. & Chakraborty, R. (2002) Genetic admixture in three Mexican mestizo populations based on D1S80 and HLA-DQA1 loci. *Am J Hum Biol* **14**, 257–263.
- Chakraborty, R. (1986) Gene admixture in human populations: models and predictions. *Yearbook of Physical Anthropology* **29**, 1–43.
- Chakraborty, R., Ferrell, R. E., Stern, M. P., Haffner, S. M., Hazuda, H. P. & Rosenthal, M. (1986) Relationship of prevalence of non-insulin dependent diabetes mellitus to Amerindian admixture in the Mexican Americans of San Antonio, Texas. *Genet Epidemiol* **3**, 435–454.

- Chakraborty, R. & Weiss, K. M. (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic associations between loci. *Proc Natl Acad Sci USA* **85**, 9119–9123.
- Collins-Schramm, H. E., Phillips, C. M., Operario, D. J., Lee, J. S., Weber, J. L., Hanson, R. L., Knowler, W. C., Cooper, R., Li, H. & Seldin, M. F. (2002) Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* **70**, 737–750.
- Devlin, B. & Roeder, K. (1999) Genomic control for association studies. *Biometrics* **55**, 997–1004.
- Devor, E. J. (1980) Marital structure and genetic isolation in a rural Hispanic population in northern New Mexico. *Am J Phys Anthropol* **53**, 257–265.
- Ewens, W. J. & Spielman, R. S. (1995) The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* **57**, 455–464.
- Gottlieb, K. & Kimberling, W. J. (1979) Admixture estimates for the gene pool of Mexican Americans in Colorado. *Am J Phys Anthropol* **50**, 444 (abstract).
- Green, L. D., Derr, J. N. & Knight, A. (2000) mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* **66**, 989–998.
- Hamman, R. F., Marshall, J. A., Baxter, J., Kahn, L. R., Mayer, E. J., Orleans, M., Murphy, J. R. & Lezotte, D. C. (1989) Methods and prevalence of Non-Insulin Dependent Diabetes Mellitus in a biethnic Colorado population. The San Luis Valley Diabetes Study. *Am J Epidemiol* **129**, 295–311.
- Hanis, C. L., Chakraborty, R., Ferrell, R. E. & Schull, W. J. (1986) Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. *Am J Phys Anthropol* **70**, 433–441.
- Hanis, C. L., Hewett-Emmett, D., Bertin, T. K. & Schull, W. J. (1991) Origins of US Hispanics. Implications for diabetes. *Diabetes Care* **14**, 618–627.
- Hoggart, C. J., Parra E. J., Shriver, M. D., Bonilla, C., Kittles R. A., Clayton D. G. & McKeigue P. M. (2003) Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* **72**, 1492–1504.
- Iyengar, S. (1988) Estimation of admixture and association with a number of markers in a population with non-insulin dependent diabetes mellitus: The San Luis Valley Diabetes Study. Master thesis, University of Pittsburgh.
- Kittles, R. A., Chen, W., Panguluri, R. K., Ahaghotu C., Jackson, A., Adebamowo, C. A., Griffin, R., Williams, T., Ukoli, F., Adams-Campbell L., Kwagyan, J., Isaacs, W., Freeman, V. & Dunston, G. M. (2002) CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet* **110**, 553–560.
- Lautenberger, J. A., Stephens, J. C., O'Brien, S. J. & Smith, M. W. (2000) Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am J Hum Genet* **66**, 969–978.
- Lell, J. T., Brown, M. D., Schurr, T. G., Sukernik, R. I., Starikovskaya, Y. B., Torroni, A., Moore, L. G., Troup, G. M. & Wallace, D. C. (1997) Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. *Hum Genet* **100**, 536–543.
- Lisker, R., Ramirez, E. & Babinsky, V. Genetic structure of autochthonous populations of Meso-America: Mexico. (1996) *Hum Biol* **68**, 395–404.
- Long, J. C. (1991) The genetic structure of admixed populations. *Genetics* **127**, 417–428.
- Long, J. C., Williams, R. C., McAuley, J. E., Medis, R., Partel, R., Tregellas, W. M., South, S. F., Rea, A. E., McCormick, S. B. & Iwaniec, U. (1991) Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. *Am J Phys Anthropol* **84**, 141–157.
- Long, J. C., Williams, R. C. & Urbanek, M. (1995) An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am J Hum Genet* **56**, 799–810.
- McConnell Simmons, V. (1999) *The San Luis Valley, Land of the Six-Armed Cross*, 2nd edition. Niwot, Colorado: University Press of Colorado.
- McKeigue, P. M. (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am J Hum Genet* **60**, 188–196.
- McKeigue, P. M. (1998) Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet* **63**, 241–251.
- McKeigue, P. M., Carpenter, J. R., Parra, E. J. & Shriver, M. D. (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet* **64**, 171–186.
- Merriwether, D. A., Huston, S., Iyengar, S., Hamman, R., Norris, J. M., Shetterly, S. M., Kamboh, M. I. & Ferrell, R. E. (1997) Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* **102**, 153–159.
- Mörner, M. (1967) *Race Mixture in the History of Latin America*. Boston: Little, Brown and Company.
- Nei, M. & Li, W. H. (1973) Linkage disequilibrium in subdivided populations. *Genetics* **75**, 213–219.
- Palmer, C. (1976) *Slaves of the White God: Blacks in Mexico, 1570–1650*. Cambridge: Harvard University Press.
- Parra, E. J., Marcini, A., Akey, J., Martinson, J., Batzer, M. A., Cooper, R., Forrester, T., Allison, D. B., Deka, R., Ferrell, R. E. & Shriver, M. D. (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* **63**, 1839–1851.

- Pfaff, C. L. (2001) Estimating admixture dynamics: implications for mapping genes. PhD thesis, The Pennsylvania State University.
- Pfaff, C. L., Parra, E. J., Bonilla, C., Heister, K., McKeigue, P. M., Kambou, M. I., Hutchinson, R. G., Ferrell, R. E., Boerwinkle, E. & Shriver, M. D. (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* **68**, 198–207.
- Pfaff, C. L., Kittles, R. A. & Shriver, M. D. (2002) Adjusting for population structure in admixed populations. *Genet Epidemiol* **22**, 196–201.
- Pritchard, J. K. & Rosenberg, N. A. (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* **65**, 220–228.
- Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P. (2000) Association mapping in structured populations. *Am J Hum Genet* **67**, 170–181.
- Raymond, M. & Rousset, F. (1995) Genepop, version 2.0.
- Reed, T. E. (1974) Ethnic classification of Mexican-Americans. *Science* **185**, 283.
- Reich, D. E. & Goldstein, D. B. (2001) Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* **20**, 4–16.
- Relethford, J. H., Stern, M. P., Gaskill, S. P. & Hazuda, H. P. (1983) Social class, admixture, and skin color variation in Mexican-Americans and Anglo-Americans living in San Antonio, Texas. *Am J Phys Anthropol* **61**, 97–102.
- Rybicki, B. A., Iyengar, S. K., Harris, T., Liptak, R., Elston, R. C., Maliarik, M. J. & Iannuzzi, M. C. (2002) Prospects of admixture linkage disequilibrium mapping in the African-American genome. *Cytometry* **47**, 63–65.
- Rodriguez-Delfin, L. A., Rubin-de-Celis, V. E. & Zago, M. A. (2001) Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. *Hum Hered* **51**, 97–106.
- Satten, G. A., Flanders, W. D. & Yang, Q. (2001) Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. *Am J Hum Genet* **68**, 466–477.
- Stephens, J. C., Briscoe, D. & O'Brien, S. J. (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am J Hum Genet* **55**, 809–824.
- Swadesh, F. L. (1974) *Los primeros pobladores. Hispanic Americans of the Ute frontier*. Notre Dame, Indiana: University of Notre Dame Press.
- Taylor, Q. (1998) *In search of the racial frontier: African Americans in the American West, 1528–1990*. New York: W. W. Norton & Company.
- Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. & Cavalli-Sforza, L. L. (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* **93**, 196–200.
- Weiss, K. M. (1993) *Genetic variation and human disease. Principles and evolutionary approaches*. Cambridge: Cambridge University Press.
- Ye, J., Parra, E. J., Sosnoski, D. M., Heister, K., Underhill, P. A. & Shriver, M. D. (2002) Melting curve SNP (McSNP) genotyping: a useful approach for diallelic genotyping in forensic science. *J Forensic Sci.* **47**, 593–600.
- Zeleny, C. (1974) *Relations between the Spanish-Americans and Anglo-Americans in New Mexico*. New York: Arno Press.
- Zheng, C. & Elston, R. C. (1999) Multipoint linkage disequilibrium mapping with particular reference to the African-American population. *Genet Epidemiol* **17**, 79–101.

Received: 13 June 2003

Accepted: 17 November 2003