



Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors.

Emily A Cameron, BSc¹, Veronica L Martinez-Marignac, PhD¹, Adan Valladares, PhD², Andrea Chan, BSc¹, Arjuna Perera, BSc¹, Rachel Globus-Goldberg, BSc¹, Niels Wachter, MD, M.Sc³, Jesus Kumate, MD, PhD⁴, Paul M McKeigue, PhD⁵, David O'Donnell, PhD⁵, Mark D Shriver, PhD⁶, Miguel Cruz, PhD² and Esteban J Parra, PhD¹.

¹Department of Anthropology, University of Toronto, Mississauga, Ontario, Canada, L5L 1C6; ²Biochemistry Research Unit and ³Clinical Epidemiology Research Unit, Hospital de Especialidades, Centro Médico "Siglo XXI", IMSS, Mexico; ⁴Fundación IMSS, Mexico, Mexico; ⁵Conway Institute, University College Dublin, Dublin, Ireland and ⁶Department of Anthropology, Penn State University, USA.



Introduction

✓The majority of the contemporary Mexican population consists of mestizos, an admixed group with a genetic background derived from the original Native American inhabitants, the European settlers (primarily from Spain), and to a lesser extent, West Africans. There has been an increased interest in the biomedical field in admixed populations because it is possible to make use of recent admixture to map genes underlying ethnic variation in disease risk. This approach is known as admixture mapping (AM) (1-2).

✓Because risk of type 2 diabetes (T2D) varies greatly among metapopulation groups (3), it can be studied using AM (4).

✓Our main goal was to characterize the admixture proportions and dynamics in a sample of T2D patients and controls from Mexico City and to define the design requirements to carry out admixture mapping in this population.

Subjects, Materials and Methods

✓We studied a sample of 286 T2D patients (198 females, 88 males) and 275 controls (86 females, 189 males)¹. Information on sex, age, body mass index (BMI) and education was also available. DNA was isolated from whole blood using the QIAamp DNA Blood Maxi Kit.

✓Ancestry informative markers (AIMs) are genetic markers that display large frequency differences between the parental populations. We genotyped up to 69 AIMs to estimate relative parental contributions at the population and individual level.

✓To estimate maternal/paternal contributions to our admixed sample, 6 mtDNA haplogroups and 2 Y-chromosome markers were typed for a subset of the sample (due to limited DNA availability).

✓Genetic markers were genotyped either by McSNP (5-6) or restriction enzyme assays in the University of Toronto molecular anthropology laboratory, real time PCR and FRET by KBiosciences (Herts, UK), or a modified allele-specific PCR with Universal energy-transfer-labeled primers by Prevention Genetics (Wisconsin, USA)².

✓The admixture proportions and the average number of generations since the admixture event (τ - sum of intensities parameter) were estimated using the software ADMIXMAP v3.2. We also used ADMIXMAP to test for stratification and to build a logistic regression model with diabetes as an outcome, controlling for covariates such as sex, age, BMI and education. ADMIXMAP is a general purpose program for modeling population admixture³.

✓Departure of genotype frequencies from Hardy-Weinberg proportions were tested using an exact test (7).

¹Samples collected by the Biochemistry and Clinical Epidemiology Research Units of the Centro Médico "Siglo XXI".
²Concordance rates between 99.4% and 100%.
³Versions of ADMIXMAP for Windows and Linux and a manual are available at <http://www.ucld.edu/genepi/software.html>.

Results and Discussion

Admixture proportions and T2D in Mexico City

Admixture	Median	Mean	Pct 2.5	Pct 97.5
Sum Intensities	6.7	6.7	5.7	8.0
Prop Nam	0.646	0.646	0.629	0.662
Prop Eur	0.304	0.304	0.287	0.319
Prop Afr	0.050	0.050	0.043	0.057
Odds ratios	Median	Mean	Pct2.5	Pct 97.5
Sex ¹	0.21	0.21	0.16	0.28
Age	1.17	1.17	1.15	1.19
BMI	1.05	1.05	1.02	1.09
Education ²	0.57	0.57	0.50	0.66
Prop Nam	1.62	1.62	0.63	4.35

Table 1. Analysis of admixture and estimated odds ratios in logistic regression models with T2D as an outcome variable. Significant covariates are indicated in red. ¹Female=1, Male=2. ²Primary=1, Secondary=2, Preparatory=3, University and/or postgraduate=4.

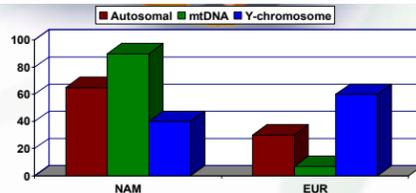


Figure 1. Triangle plot showing the distribution of individual admixture estimates obtained using 69 AIMs.

Figure 2. Evidence of sex-biased gene flow: Percentage of autosomal, mtDNA and Y-chromosome contributions from Native American (NAM) and European (EUR) parental populations.

- ✓The parental contributions to the Mexico City sample are:
 - 65% Native American, with individual ancestry ranging from 22-90%
 - 30% European, with individual ancestry ranging from 8-75%
 - 5% West African, with individual ancestry ranging from 1-20%
- ✓There is strong evidence of sex biased gene flow with a European maternal contribution of around 7%, and a paternal contribution of 60%
- ✓The average number of generations since the admixture event (τ) was estimated at 7 generations (~175 years, 95% CI: 150-200 yrs).
- ✓Age, female sex, BMI and low education levels were significantly associated with T2D.
- ✓The odds ratio for T2D associated with unit change in Native American admixture proportion (from 0 to 1) was estimated as 1.6, but the confidence interval overlaps 1.

Population Stratification

% of associations between unlinked markers showing significance

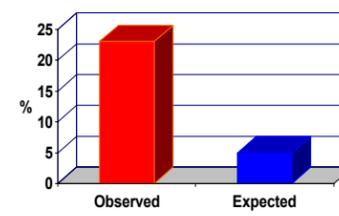


Figure 3. Strong evidence of genetic structure: The number of unlinked AIMs showing significant associations is much higher than expected.

Upper bound - Lower bound - Average

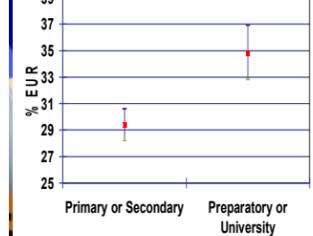


Figure 4. Strong evidence of socioeconomic stratification: The percentage of European ancestry is significantly higher in individuals with preparatory or University education.

- ✓There is strong evidence for the presence of genetic structure in Mexico City
 - ADMIXMAP was used to test for genetic stratification, using models with 1, 2 or 3 subpopulations: A model with 2 populations is required to fit the data
 - 23% of unlinked markers show significant associations (expected: 5%)
- ✓We found strong evidence of socioeconomic stratification: Not everyone in Mexico City has the same access to education. Individuals with 100% European ancestry are 9.4 times more likely to have higher education than those with no European ancestry.
- ✓Mating is probably not random with respect to socioeconomic status, and socioeconomic status is associated with ancestry. This finding is likely a major factor explaining the presence of genetic structure in this population. The presence of genetic structure could be also due in part to continuous gene flow.
- ✓These results emphasize the importance of controlling for population stratification as a possible confounder in genetic association studies in the Mexican population (8-9). Potential differences in socioeconomic status between cases and controls could result in false positive results

Implications for the application of AM to identify T2D genes

✓Two of the most important factors to consider for the application of AM are the distribution of individual ancestry and τ , the average number of generations since the admixture event.

•In our Mexican sample, most individuals are within the ideal admixture proportions of 10% to 90%.

• τ determines the density of genetic markers required for AM and the mapping resolution, and was estimated as 7 generations. Based on this estimate, an AM study in the Mexican mestizo population will require around 1,400 AIMs spanning the genome.

•Sample sizes of around 2,000 will be required to detect any locus that contributes an ancestry risk ratio of about 1.5.

✓The development of a genome-wide panel of AIMs for AM applications in Hispanics is well within reach. Using the newly developed 500K microarray from Affymetrix, we have identified a preliminary panel with more than 3,000 AIMs spanning the whole genome.

Number of AIMs	Average inter-marker physical distance
3,519	777 Kb
Initial <i>f</i> cutoff value	Average <i>f</i>
0.25	0.42
Average δ	
0.62	

Table 2. A preliminary AM genome-wide panel for Hispanics. δ is defined as the absolute frequency difference between Native American and European populations. *f* is the Fisher information content for ancestry. Markers were selected using the newly developed Affymetrix 500K microarray.

Conclusions

Our results indicate that the Mexican population is suitable for AM applications to identify genes involved in diseases showing prevalence differences between continental populations, such as T2D. The availability of a genome-wide AM panel with Native American/European AIMs will facilitate the application of this novel gene mapping method in Hispanic populations.

References

- Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. *Am J Hum Genet* 74(5):965-978, 2004.
- Patterson N, Hallgrímsson B, Lane B et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 74(5):979-1000, 2004.
- American Diabetes Association. <http://www.diabetes.org/home.asp>.
- Haldar I, Shriver MD. Measuring and using admixture to study the genetics of complex diseases. *Hum Genomics* 1(1):52-62, 2003.
- Akey JM, Sosnoski D, Parra E et al. Melting Curve Analysis of SNPs (McSNP): A Simple Gel-Free Low-Cost Approach to SNP Genotyping and DNA Fragment Analysis. *BioTechniques* 30: 358-367, 2001.
- Ye J, Parra EJ, Sosnoski DM et al. Melting curve SNP (McSNP) genotyping: a useful approach for diallelic genotyping in forensic science. *J Forensic Sci* 47: 593-600, 2002.
- Hoggart CJ, Parra EJ, Shriver MD et al. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 72(6):1402-1504, 2003.
- Parra EJ, C. J Hoggart, C Bonilla et al. Relation of type 2 diabetes to individual admixture and candidate gene polymorphisms in the Hispanic American population of San Luis Valley, Colorado. *J Med Genet* 41:116, 2004.

Acknowledgements. We would like to thank all the individuals from Mexico City for their participation in this study. We are also grateful to the personnel at the National Blood Bank, Medical Center Siglo XXI (IMSS) in Mexico. This work was supported by the Fondo para el Fomento de Investigación (FOFI) and the Instituto Mexicano del Seguro Social (IMSS) in Mexico. EJP was supported by a Pilot grant from the Banting and Best Diabetes Centre at the University of Toronto and by the Canadian Institutes of Health Research (CIHR), Institute of Genetics. MDS, DO and PM are supported by a grant from the NIH/NHGRI (HG02154). We would like to thank David Ramos, Alfonso Alcantara and Olga Gaja for their excellent technical support. Finally, the preliminary genome-wide AM panel has been developed in collaboration with the company Affymetrix (Santa Clara, CA). Background image courtesy of the U.S. Department of Energy Genomics:GTL program, <http://www.ornl.gov/hgmis>