Supporting Information

Silva-Zolezzi et al. 10.1073/pnas.0903045106

SI Materials and Methods

Samples. All participants declared to have 4 grandparents not self-recognized as recent immigrants and born in the state in which recruitment was performed. All Zapotecos were individuals with no first-degree familial relationship.

Community and Individual Engagement. The ad hoc process for community consultation and engagement included state government officials, university and health authorities, and members of the local community. The strategy previous to sample collection included (*a*) distribution of a brochure explaining the project using simple language, (*b*) display of a poster reproducing the informed consent form, (*c*) delivery of 4–6 public information sessions, and (*d*) communication via television, radio, and printed press. In all participant states, samples were collected during a public academic event. All participants provided consent in the presence of 2 local witnesses. For the Zapotecos, informed consent was translated into their native language and all parts of the process took place in the presence of a bilingual translator. All Mestizo participants were mostly, but not exclusively, members of the local state university community.

DNA Extraction and Genomewide SNP Genotyping. DNA extraction, genotyping, and data quality control were performed at INMEGEN (Mexico). The average call rate using the BRLMM algorithm was 99.45%. Only SNPs with a missingness $<20\%$, Hardy–Weinberg equilibrium P value <0.0001, and HET $<1\%$ in chromosome X of males were included in our analysis. The phased haplotypes for the 100K genotypes of the HapMap samples were obtained from Phase II HapMap (www.hapmap.org).

Statistical Methods. All principal components analyses (PCA) were performed without outlier removal iteration. Ancestry differences were assessed with a Mann–Whitney U test. Pearson correlation coefficients between F_{ST} and average ancestral contribution differences were calculated (StatPlus:mac 2008, AnalystSoft) to evaluate the contribution of ancestry differences to genetic distance. Box-plot distributions of ancestral components and their coefficient of variation, defined as the ratio of the standard deviation σ to the mean μ, $CV = \sigma/\mu$, were calculated with R.

Ancestry Analysis. Ancestry informative markers (AIMs) were SNPs with allele frequency differences (δ) ≥ 0.4 for pairwise comparisons between HapMap groups and ZAP. To minimize background, LD markers at <500 kb from any other marker were eliminated. These AIMs were used to run STRUCTURE using the linkage model, allowing for admixture with the number of parental populations between $K = 3$ and $K = 7$. Ten repeats were performed for every K, with 10,000 burn-in cycles and 10,000 replicates without prior population information.

LD and Haplotype Sharing (HS) Analyses. LHRD was calculated dividing the genome into windows of 35 markers on average, spanning $\approx$1 Mb. Haplotype diversity was inferred by comparing the average number of haplotypes across these regions in each population. HS was assessed by comparing haplotype frequencies composed of 5 SNPs spanning $\approx$100 kb. Haplotypes with frequencies $>1\%$ and 5% were compared across populations.

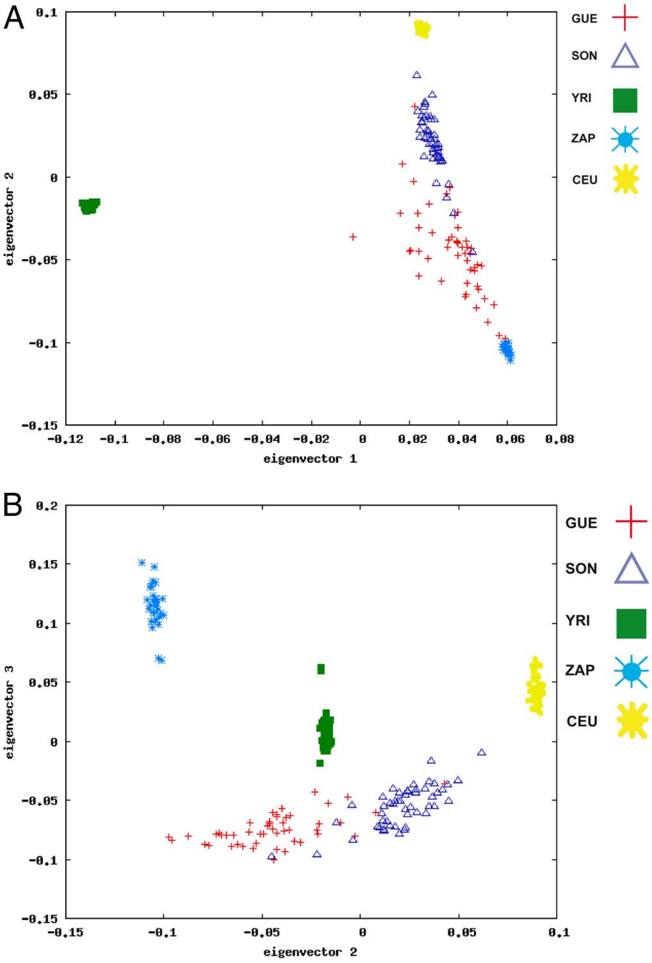

Fig. S1. Principal components analysis. The 4 most informative eigenvectors were plotted for a data set including Mexican Mestizo subpopulations showing the largest difference in HET (SON and GUE), ZAP, CEU and YRI. (*A*) First and second eigenvectors; (*B*) second and third eigenvectors.

Fig. S2. Frequency distribution of SNPs with the highest information content between Mexican Mestizo subpopulations. The frequency of the 14 SNPs with the highest information content is represented in the geographical context of the 6 Mexican Mestizo populations analyzed. The darker shading indicates the region with the highest frequency, and the lighter shading the region with the lowest frequency for a particular marker.

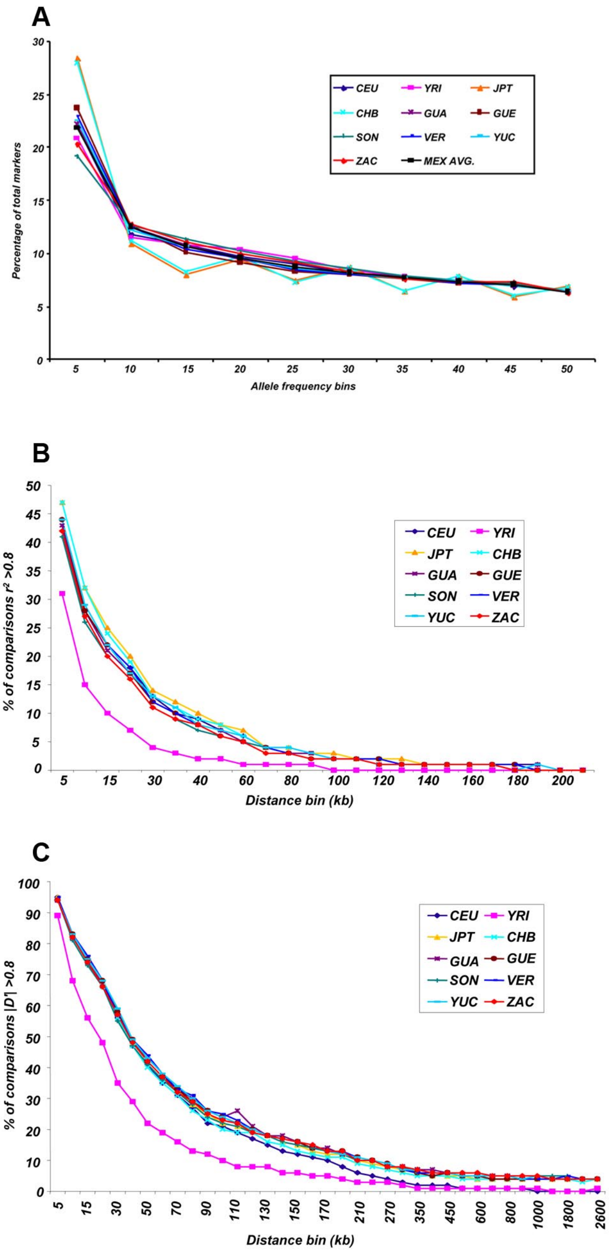

Fig. S3. (*A*) Allele frequency distribution of the 99,953 SNPs typed for the 4 HapMap, 6 Mexican Mestizo, and Amerindian populations is shown. (*B* and *C*) LD decay over distance. LD was measured as a pairwise comparison of markers with a MAF ≥ 15% that fell in the same distance bin. LD decay over distance is represented by the percentage of pairwise comparison scores equal to 0.8 using r^2 (*B*) or *D*′ (*C*).

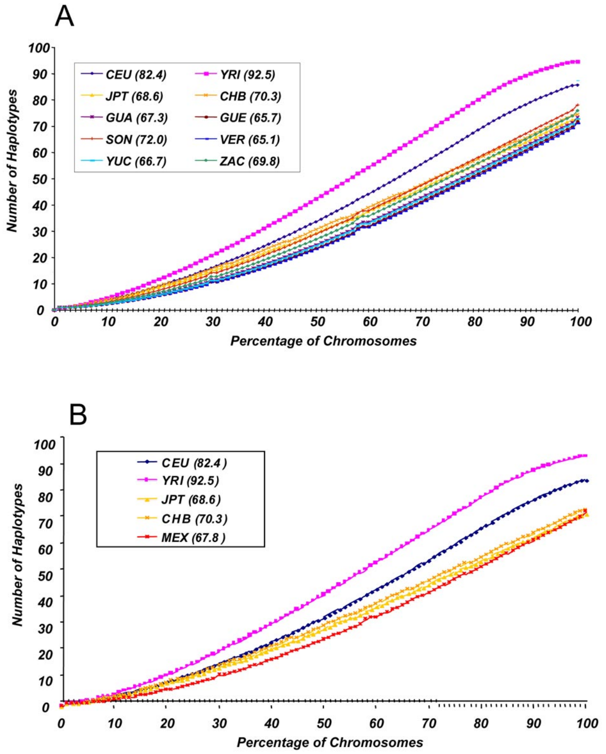

Fig. S4. Long-range haplotype diversity. After phasing, ≈1-Mb genomic windows of ≈35 markers were generated, and the frequencies of the derived haplotypes were averaged and compared to the percentage of chromosomes represented by those haplotypes. (*A*) Mexican subpopulations vs. HapMap populations. (*B*) Haplotype diversity average in Mexican and HapMap populations.

Table S1. Average ancestry proportions of 6 Mexican Mestizo subpopulations, Amerindian Zapotecos from Oaxaca, and all HapMap populations

	STRUCTURE, unsupervised clustering, $K = 4$, $\delta \geq 0.4$ ($n = 1{,}814$)			
Population	EUR	AMI	AFR	EA
CEU	0.956 ± 0.024	0.038 ± 0.024	0.002 ± 0.002	0.004 ± 0.005
YRI	0.009 ± 0.007	0.008 ± 0.006	0.981 ± 0.011	0.002 ± 0.003
JPT + CHB	0.017 ± 0.011	0.024 ± 0.020	0.005 ± 0.008	0.954 ± 0.025
ZAP	0.006 ± 0.004	0.992 ± 0.005	0.001 ± 0.001	0.001 ± 0.001
GUA	0.399 ± 0.100	0.576 ± 0.096	0.011 ± 0.018	0.013 ± 0.021
GUE	**0.285 ± 0.120**	**0.660 ± 0.138**	**0.041 ± 0.061**	**0.014 ± 0.021**
SON	**0.616 ± 0.085**	**0.362 ± 0.089**	0.012 ± 0.017	**0.010 ± 0.012**
VER	0.356 ± 0.130	0.613 ± 0.141	0.020 ± 0.042	0.011 ± 0.016
YUC	0.392 ± 0.162	0.588 ± 0.161	**0.008 ± 0.012**	0.012 ± 0.020
ZAC	0.457 ± 0.084	0.511 ± 0.770	0.018 ± 0.023	0.013 ± 0.018
Mexican mean	0.418 ± 0.155	0.552 ± 0.154	0.018 ± 0.035	0.012 ± 0.018

Ancestry proportions were calculated using a set of 1,814 ancestry-informative markers with $\delta \geq 0.4$ after pairwise comparisons of 4 parental populations (CEU-JPT + CHB = 234, CEU-YRI = 308, CEU-ZAP = 314, JPT + CHB-YRI = 311, JPT + CHB-ZAP = 262, YRI-ZAP = 385). The Mexican Mestizo subpopulations demonstrating the highest and lowest proportion of each ancestral component are in boldface type. European (EUR), Amerindian (AMI), African (AFR), and East Asian (EA) ancestry proportions were estimated with STRUCTURE in 6 Mexican subpopulations, Guanajuato (GUA), Guerrero (GUE), Sonora (SON), Veracruz (VER), Yucatan (YUC), and Zacatecas (ZAC); in 1 Mexican Amerindian group, Zapotecas (ZAP); and in all HapMap populations, Caucasians (CEU), African (YRI), and Asians (JPT + CHB). Results from STRUCTURE are reported for the best model when using AIMs with δ cutoffs ≥ 0.4 ($n = 1{,}814$).

Table S2. Pairwise comparisons of individual ancestry estimates (Mann–Whitney *U* test) for 6 Mexican Mestizo subpopulations

	GUA	GUE	SON	VER	YUC
			Amerindian ancestry		
GUE	0.997				
SON	$<1 \times 10^{-4}$	$<1 \times 10^{-4}$			
VER	0.107	0.088	$<1 \times 10^{-4}$		
YUC	0.526	3.1×10^{-2}	$<1 \times 10^{-4}$	0.506	
ZAC	1×10^{-4}	$<1 \times 10^{-4}$	$<1 \times 10^{-4}$	$<1 \times 10^{-4}$	5.7×10^{-3}
			European ancestry		
GUE	$<1 \times 10^{-4}$				
SON	$<1 \times 10^{-4}$	$<1 \times 10^{-4}$			
VER	6.7×10^{-2}	2.50×10^{-3}	$<1 \times 10^{-4}$		
YUC	4.80×10^{-1}	6.00×10^{-4}	$<1 \times 10^{-4}$	3.48×10^{-1}	
ZAC	1.3×10^{-3}	$<1 \times 10^{-4}$	$<1 \times 10^{-4}$	$<1 \times 10^{-4}$	1.35×10^{-2}
			African ancestry		
GUE	3.83×10^{-1}				
SON	2.44×10^{-1}	9.50×10^{-1}			
VER	3.29×10^{-1}	1.45×10^{-1}	3.04×10^{-2}		
YUC	2.61×10^{-1}	9.12×10^{-2}	2.44×10^{-2}	9.64×10^{-1}	
ZAC	9.06×10^{-2}	8.98×10^{-1}	3.98×10^{-1}	2.74×10^{-2}	5.9×10^{-3}
			Asian ancestry		
GUE	5.50×10^{-1}				
SON	9.97×10^{-1}	4.65×10^{-1}			
VER	8.20×10^{-1}	3.66×10^{-1}	8.63×10^{-1}		
YUC	5.37×10^{-1}	2.40×10^{-1}	5.06×10^{-1}	9.54×10^{-1}	
ZAC	4.52×10^{-1}	9.42×10^{-1}	3.63×10^{-1}	2.82×10^{-1}	1.74×10^{-1}

Statistically significant differences (*P*-value $\leq 5 \times 10^{-2}$) are indicated for differences between ancestral contributions in each Mexican Mestizo subpopulation.

Table S3. Genetic distance between Mestizo subpopulations related to differences in main continental ancestral contributions (AMI and EUR)

Pop1-Pop2	Δ_{EUR}	$F_{ST_{Pop1\text{-}Pop2}}$	Genetic distance by ancestry	Genetic distance by ancestry/$F_{ST_{Pop1\text{-}Pop2}}$
GUA-GUE	0.114	2.00*E*-03	2.00*E*-03	1.00
GUE-ZAC	−0.172	5.00*E*-03	4.56*E*-03	0.91
GUE-SON	−0.331	1.90*E*-02	1.69*E*-02	0.89
SON-VER	0.260	1.30*E*-02	1.04*E*-02	0.80
VER-ZAC	−0.101	2.00*E*-03	1.57*E*-03	0.79
GUE-VER	−0.071	1.00*E*-03	7.76*E*-04	0.78
GUA-SON	−0.217	1.10*E*-02	7.25*E*-03	0.66
SON-ZAC	0.159	6.00*E*-03	3.89*E*-03	0.65
SON-YUC	0.224	1.20*E*-02	7.73*E*-03	0.64
GUA-ZAC	−0.058	1.00*E*-03	5.18*E*-04	0.52
GUE-YUC	−0.107	4.00*E*-03	1.76*E*-03	0.44
GUA-VER	0.043	1.00*E*-03	2.85*E*-04	0.28
YUC-ZAC	−0.065	3.00*E*-03	6.51*E*-04	0.22
VER-YUC	−0.036	2.00*E*-03	2.00*E*-04	0.10
GUA-YUC	0.007	3.00*E*-03	7.55*E*-06	0.00

Average European (EUR) ancestry proportions were estimated with STRUCTURE for 6 Mexican subpopulations: Guanajuato (GUA), Guerrero (GUE), Sonora (SON), Veracruz (VER), Yucatan (YUC), and Zacatecas (ZAC) (Table S1). Δ_{EUR}, EUR average ancestry proportion of Pop1 − EUR average ancestry proportion of Pop2; genetic distance between subpopulations, $F_{ST_{Pop1\text{-}Pop2}}$ (Table 1). Genetic distance by ancestry was calculated as follows: $(\Delta_{EUR})^2 \times F_{ST_{AMI\text{-}EUR}}$; $F_{ST_{AMI\text{-}EUR}} = 0.154$ is the genetic distance between Amerindian Zapotecas (ZAP) and Caucasians (CEU) from the HapMap (Table 1).

Table S4. Median and coefficient of variation (CV) for individual ancestry estimates in Mexican Mestizo subpopulations

	EUR		AMI		AFR		EA	
Population	Median	CV	Median	CV	Median	CV	Median	CV
GUA	0.378	0.250	0.599	0.167	0.005	1.583	0.005	1.546
GUE	0.289	0.421	0.668	0.210	0.005	1.501	0.006	1.488
SON	0.626	0.139	0.347	0.246	0.005	1.391	0.006	1.264
VER	0.354	0.366	0.608	0.230	0.003	2.096	0.005	1.542
YUC	0.391	0.414	0.595	0.273	0.004	1.527	0.005	1.625
ZAC	0.461	0.184	0.510	0.151	0.009	1.236	0.005	1.322

CVs reflect normalized variance between individuals in each group. The highest CV values were present for East Asian and African distributions, indicating greater variance in these contributions within each subpopulation. Ancestry proportions: European (EUR), Amerindian (AMI), African (AFR), and East Asian (EA). Mexican Mestizo subpopulations: Guanajuato (GUA), Guerrero (GUE), Sonora (SON), Veracruz (VER), Yucatan (YUC), and Zacatecas (ZAC); Mexican Amerindian group, Zapotecas (ZAP)

Table S5. Frequency of the 14 SNPs with the highest information content ($I_n > 0.04$) to differentiate between Mexican Mestizo groups, as calculated by the informativeness for assignment statistic

Chr	SNP ID	Position	Gene	CEU	JPT+CHB	YRI	ZAP	GUA	GUE	SON	VER	YUC	ZAC
1	rs4528122	149680414	*POGZ*	0.142	0.856	0.692	0.933	0.650	0.790	0.350	0.690	0.630	0.630
1	rs986690	238518042	*FMN2*	0.250	0.534	0.158	0.983	0.590	0.790	0.360	0.690	0.640	0.650
12	rs6487927	30717602	*IPO8*	0.475	0.466	0.575	0.033	0.270	0.210	0.540	0.130	0.340	0.300
13	rs2147155	92878308	*GPC6*	0.500	0.029	0.025	0.000	0.130	0.092	0.360	0.070	0.080	0.170
4	rs10516422	98483983	—	0.017	0.103	0.183	0.283	0.150	0.350	0.080	0.220	0.337	0.100
5	rs10515716	154822132	—	0.208	0.399	0.133	0.733	0.590	0.650	0.320	0.720	0.470	0.640
6	rs1878071	93533402	—	0.217	0.815	0.008	0.683	0.420	0.670	0.220	0.480	0.480	0.400
9	rs4084051	828022	—	0.175	0.371	0.483	0.750	0.430	0.690	0.270	0.570	0.440	0.420
9	rs7853112	8322353	*PTPRD*	0.350	0.478	0.875	0.750	0.520	0.750	0.310	0.640	0.540	0.510
9	rs10511491	8335923	*PTPRD*	0.392	0.472	0.925	0.750	0.530	0.740	0.320	0.620	0.550	0.520
9	rs1039336	8366287	*PTPRD*	0.242	0.676	0.658	0.867	0.630	0.830	0.430	0.690	0.700	0.510
9	rs10116714	12397578	—	0.050	0.522	0.333	0.817	0.520	0.650	0.229	0.590	0.580	0.460
9	rs1980888	9109037	—	0.100	0.404	0.200	0.966	0.590	0.830	0.360	0.550	0.673	0.470
9	rs4743556	97924766	—	0.167	0.315	0.292	0.828	0.460	0.710	0.306	0.470	0.540	0.450

Table S6. Percentages of common haplotypes shared between Mexican subpopulations

Population	GUA GUE	GUA SON	GUA VER	GUA YUC	GUA ZAC	GUE SON	GUE VER	GUE YUC	GUE ZAC	SON VER	SON YUC	SON ZAC	VER YUC	VER ZAC	YUC ZAC
GUA						97	96	96	96	96	97	97	96	96	96
GUE		96	96	96	96					96	96	96	96	96	96
SON	93		93	94	94		92	93	93				93	94	94
VER	96	97		96	97	97		96	96		96	97			97
YUC	95	96	96		96	96	95		96	96		96		96	
ZAC	95	96	95	95		96	95	95		96	96		95		
MEX Av	95	96	95	95	96	97	95	95	95	96	96	97	95	96	96

Percentages of shared common haplotypes (>5% frequency) using all possible pairs of Mexican subpopulations as the reference group are shown.