

Simulation study: We performed computer simulations of the admixture process under several theoretical models: the main differences between the models are the presence or absence of admixture stratification (e.g. variation of ancestry proportions in the individuals of the admixed sample), and the presence or absence of pigmentation differences between the parental populations. Briefly, we simulated the following scenarios:

Admixed samples

1-Absence of admixture stratification. We simulated an admixed sample in which all the individuals of the sample have similar admixture proportions using the program SIMSAMPLE. This program creates genotype data for individuals with a history of admixture that can be specified up to any number of generations ago. Any number of individuals can be created with any number of marker loci. The parental frequencies are specified in an input file. We simulated genotype data for 35 and 100 AIMs, which were used to estimate individual ancestry. AIM frequencies correspond to real data extracted from an analysis of the Affymetrix 10K Mapping Array in different world populations (Shriver et al., unpublished data). The differences in frequency between the parental populations range between 92.4% and 59.9%. In this simulation, we constructed genotype data for 200 individuals resulting from an admixture process between 2 populations that took place 10 generations ago. The proportion of ancestors from population 1 was defined as 40% and the proportion of ancestors from population 2 as 60%. We also simulated genotype data for up to 15 pigmentation genes (2, 5, 10 and 15), which determine melanin content in each individual following a simple polygenic additive model (more details described below). All the genetic markers (both AIMs and pigmentation genes) are unlinked.

2- Presence of admixture stratification. Using the program SIMSAMPLE, we simulated an admixed sample of 200 individuals in which there is variation in ancestry proportions. This sample comprises individuals showing different contributions from the two parental populations. The sample was modeled to closely resemble the admixture proportions found in the African American sample from Washington DC. In this simulated sample, the relative contribution of population 2 was 0% in 20% of the individuals, 6.25% in 20% of the individuals, 12.5% in 20% of the individuals, 25% in 20% of the individuals, 37.5% in 10% of the individuals, 50% in 5% of the individuals and 62.5% in 5% of the

individuals. Genotypes were simulated for 35 or 100 AIMs, and up to 15 pigmentation genes, as described above.

Pigmentation phenotypes

Constitutive pigmentation was simulated using a simple polygenic additive model in which one allele of each pigmentation locus specifies twice the amount of melanin as the alternative allele. We considered four alternative scenarios in which pigmentation was determined by the action of 2, 5, 10 and 15 loci, respectively. We defined the effect of each allele to be dependent on the number of genes determining pigmentation. In the case where only two genes influence constitutive pigmentation, the effect of the “light” allele was 10 melanin units, and the effect of the “dark” allele 20 melanin units. Therefore, melanin content would range between 40 (4 “light” alleles) and 80 (4 “dark” alleles). For 5 genes, the effect of the “light” allele was defined as 4 and the effect of the “dark” allele 8 (melanin content between 40 and 80). For 10 genes, the effect of the “light” allele was 2 and the effect of the “dark” allele 4 (melanin content between 40 and 80) and for 15 genes, the effects of the “light” and “dark” alleles were 1.5 and 3, respectively (melanin content 45 and 90). Additionally, we considered two alternative models regarding pigmentation differences between the parental populations.

1-Pigmentation differences between the parental populations are large, due to divergent allele frequencies in the pigmentation genes. One parental population (population 1) has a higher constitutive pigmentation than the other parental population (population 2), due to the higher frequency of “dark” alleles. The differences in allele frequency between the parental populations for the pigmentation loci range between 30% and 63%.

2-Pigmentation differences between the parental populations are small, with similar allele frequencies in the pigmentation genes. Differences in allele frequency between the parental populations do not exceed 10%.

Overall, in the simulation study, we considered three hypothetical scenarios: 1- Absence of admixture stratification in the admixed sample and large pigmentation differences between the parental populations due to genetic factors. 2- Presence of admixture stratification in the admixed sample (e.g. variation in admixture proportions between individuals) and large pigmentation differences between the parental populations due to genetic factors. 3- Presence of admixture stratification in the admixed sample and small pigmentation differences between the parental populations.

Results

The results of these simulations are depicted in Table 2. The correlations are not significant under models 1 and 3, irrespective of the number of AIMs genotyped and the number of pigmentation genes influencing melanin content. All the correlations estimated under model 2 were significant ($p < 0.001$). The extent of the correlation increases only slightly when individual ancestry is estimated with 100 AIMs, instead of 35 AIMs. The number of pigmentation genes influencing melanin content has a stronger effect on the correlation between pigmentation and ancestry (see Table 1).

Table 1. Relationship of melanin content and individual ancestry in simulated admixed samples (N=200) under three alternative models: 1- Absence of structure in the admixed sample and large pigmentation differences between the parental populations due to allele frequency differences in pigmentation genes. 2- Presence of structure in the admixed sample (e.g. variation in admixture proportions between individuals) and large pigmentation differences between the parental populations. 3- Presence of structure in the admixed sample and small pigmentation differences between the parental populations (e.g. small allele frequency differences in pigmentation genes). Individual ancestry was estimated with 35, or alternatively 100 AIMs. Melanin content was modeled as an additive polygenic trait influenced by 2, 5, 10 or 15 pigmentation genes. Significant correlations are indicated in bold.

Simulated Sample	35 AIMs		100 AIMs	
	Spearman's rho	p value	Spearman's rho	p value
Model 1				
2 Pigmentation genes	R=-0.007	p=0.919	R=-0.017	p=0.806

5 Pigmentation genes	R=0.088	p=0.216	R=0.086	p=0.227
10 Pigmentation genes	R=0.105	p=0.140	R=0.075	p=0.294
15 pigmentation genes	R=0.038	p=0.589	R=0.001	p=0.993
Model 2	Spearman's rho	p value	Spearman's rho	p value
2 Pigmentation genes	R=-0.444	p<0.001	R=-0.462	p<0.001
5 Pigmentation genes	R=-0.494	p<0.001	R=-0.503	p<0.001
10 Pigmentation genes	R=-0.630	p<0.001	R=-0.638	p<0.001
15 pigmentation genes	R=-0.640	p<0.001	R=-0.654	p<0.001
Model 3	Spearman's rho	p value	Spearman's rho	p value
2 Pigmentation genes	R=-0.075	p=0.290	R=-0.119	p=0.094
5 Pigmentation genes	R=-0.107	p=0.133	R=-0.112	p=0.113
10 Pigmentation genes	R=-0.083	p=0.245	R=-0.108	p=0.128
15 pigmentation genes	R=-0.089	p=0.211	R=-0.076	p=0.284

Comparison of Correlation Coefficients

There were significant differences in the correlations observed in the samples that were measured with the same instrument (DermaSpectrometer). The correlation observed in Puerto Rico differs significantly from the values observed in Mexico ($p<0.001$) and African Caribbeans ($p=0.019$), and the difference with the African-American sample is of borderline significance ($p=0.059$). There are also differences between the correlations observed in the Mexican sample with respect to the value observed in African Americans ($p=0.014$). No significant differences were observed between the African Caribbean sample and the Mexican ($p>0.05$) or the African American samples ($p>0.05$).