LETTERS

On the use of Chinese population as a proxy of Amerindian ancestors in genetic admixture studies with Latin American populations

European Journal of Human Genetics (2016) **24**, 326–327;
doi:10.1038/ejhg.2015.184; published online 2 September 2015

Dear Editor,

We read the recent article of Magalhães da Silva *et al*¹ reporting the correlation between biogeographic ancestries, estimated using 30 ancestry informative markers (AIMs), and self-reported skin color in two different Brazilian Northeastern populations (Fortaleza and Salvador, capitals of the states of Ceará and Bahia, respectively). The authors observed that African ancestry is more correlated in the sample from Salvador than in the one from Fortaleza and that the use of different African populations as proxies of the Brazilian's African ancestors may influence the results.

One unusual point of this study was taking Han Chinese from Beijing (CHB) as pseudo-ancestors for Amerindians as there is no Native American population included in the HapMap Project (<http://hapmap.ncbi.nlm.nih.gov/>) and, in addition, because CHB population has been shown to have allele frequencies similar to those of Native Amerindians.¹

It is well known that most Latin American countries are inhabited by tri-hybrid populations derived from African, Amerindian and European roots, in which their proportions show considerable variability.² As the vast majority of published studies aimed at inferring the admixture proportions of Latin American populations have not used CHB population as a proxy for Amerindian ancestors, we compared the admixture inference of some Latin American populations using Chinese or Native American ancestors as proxies.

In order to estimate admixture proportions, we used the following Latin American populations from 1000 Genomes Project Phase III³: Colombians from Medellín, Colombia (CLM); Mexican Ancestry from Los Angeles, USA (MXL); Peruvians from Lima, Peru (PEL); and Puerto Ricans from Puerto Rico (PUR). Besides the CHB population, we also used as proxies for African and European ancestors the Utah Residents with Northern and Western European ancestry (CEU) and Yoruba in Ibadan, Nigeria (YRI), respectively, also obtained from 1000 Genomes Project Phase III database. For Amerindian ancestors, a combination of Mayans, Quechuans and Nahua Natives (we call this group AMI) were used as pseudo-ancestors for Indigenous Americans as described by Kosoy *et al*.⁴

One hundred and twenty-seven SNPs were used as AIMs from a set of 128 SNPs validated by Kosoy *et al*⁴ (the rs10954737 SNP is not

present in the 1000 Genomes database). After merging 1000 Genomes with AMI data using PLINK version 1.90,⁵ we employed STRUSTRUCTURE software v. 2.3.4 (Pritchard *et al*)⁶ running into ParallelStructure R package.^{7,8} The parameters applied were 10 independent runs with 100 000 burn-in steps and 100 000 Markov chain Monte Carlo replicates assuming three ancestral populations ($K=3$) in admixture model, allele frequencies correlated and the parameter USEPO-PINFO=1. CLUMPAK Server allowed us to generate bar plots, referring to individual and population ancestry proportions.⁹ We made two separated analyses: one using CEU, YRI and AMI as pseudoancestors and another using CEU, YRI and CHB.

The bar plot with the admixture estimate is shown in Figure 1 and the ancestry proportions for the CLM, MXL, PEL and PUR populations are displayed in Table 1. It is possible to observe a tendency towards increasing of Amerindian ancestry and a decreasing of European ancestry when the CHB population was used as proxy for Native American ancestors.

In general, the results obtained using AMI pseudo-ancestors are more similar to those found in the literature, also when using AMI pseudo-ancestors in studies concerning populations from the same city such as CLM, MXL, PEL and PUR (Table 1). The results published by Magalhães da Silva *et al*¹ predicted 54.7, 12.3 and 33% for EUR, AFR and CHB contributions, respectively, in a self-declared 'white' individuals from Fortaleza. Whereas in a study of Pena *et al*,¹⁰ using 40 validated AIMs, the authors found 75.8, 13.3 and 10.9% in a group of 'white' individuals from the same city. The same tendency for increased AMR contribution, using CHB as pseudo-ancestors, occurred when analyzing 'brown' individuals. Curiously, this tendency was not confirmed when the results of the population from Salvador are compared with the findings from the study of Pena *et al*.¹⁰

We also evaluated whether the allele frequency of ancestry markers were different between AMI and CHB populations: in our set of 127 AIMs, 100 showed significant differences between allele frequencies (P -value from χ^2 test <0.05). Moreover, we observed a weak correlation between the allele frequencies of AMI and CHB populations ($r=0.33$). Finally, average pairwise F_{ST} values were higher when comparing AMI with CEU or YRI (0.30 and 0.35, respectively) than when comparing CHB with CEU and YRI (0.16 and 0.25, in that order). As higher divergence corresponds to better Informativeness,¹¹ we could hypothesize that the use of AMI individuals as pseudo-ancestors of Native Americans should be preferred to that of CHB individuals.

In light of this, we believe that the results of Magalhães da Silva *et al*¹ may be biased owing to the use of CHB population instead AMI as reference samples of Amerindian ancestors. Some genotyping data from AMI populations such as those from Kosoy *et al*⁴ are publically available as well as those from Human Genetic Diversity Project (<http://www.hagsc.org/hgdp/files.html>) and may serve as source of AIMs where a subset of the panel found by Galanter *et al*¹² can be downloaded.

In future studies dealing with Latin American tri-hybrid admixtures, our suggestion would be to consider genotyping data from AMI populations as first choice, eventually comparing the results with those obtained by using CHB as proxy as carried out by Magalhães da Silva *et al*.¹ Their comparison will allow a definitive choice based on the better representatives of Amerindian ancestry.

Figure 1 Bar plot with the admixture estimate of Latin American populations using AMI (top) and CHB (bottom) populations as proxies for the Amerindian ancestors. CEU=Utah Residents with Northern and Western European ancestry; YRI=Yoruba in Ibadan, Nigeria; AMI=combination of Mayans, Quechuans Amerindians and Nahua Native Americans; CHB=Han Chinese from Beijing; CLM=Colombians from Medellin, Colombia; MXL=Mexican Ancestry from Los Angeles, USA; PEL=Peruvians from Lima, Peru; PUR=Puerto Ricans from Puerto Rico.

Table 1 Proportions of European (EUR), African (AFR) and Native American (AMR) ancestors in Latin American populations

Population	Using AMI			Using CHB			Literature data			Authors
	EUR	AFR	AMR	EUR	AFR	AMR	EUR	AFR	AMR	
CLM	0.6770	0.0610	0.2620	0.5904	0.0390	0.3706	0.6000	0.1200	0.2800	Wang <i>et al</i> ¹³
MXL	0.4769	0.0384	0.4848	0.3361	0.0150	0.6489	0.4580	0.1100	0.4290	Qu <i>et al</i> ¹⁴
PEL	0.2020	0.0342	0.7638	0.0705	0.0138	0.9157	0.1400	0.0200	0.8400	Sandoval <i>et al</i> ¹⁵
PUR	0.7422	0.1127	0.1450	0.7006	0.0890	0.2104	0.6400	0.2100	0.1500	Via <i>et al</i> ¹⁶

Abbreviations: CLM=Colombians from Medellin, Colombia; MXL=Mexican Ancestry from Los Angeles, USA; PEL=Peruvians from Lima, Peru; PUR=Puerto Ricans from Puerto Rico. The results using Amerindian (AMI) and Chinese individuals (CHB) are shown. Some findings retrieved from literature for similar populations are also displayed.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Ronald R de Moura^{*,1,2}, Valdir de Queiroz Balbino¹, Sergio Crovella^{1,2}
and Lucas AC Brandão^{1,2,3}

¹Department of Genetics, Federal University of Pernambuco
(UFPE), Recife, Brazil;

²Laboratory of Immunopathology Keizo Asami (LIKA/UFPE),
Recife, Brazil;

³Department of Pathology, Federal University of Pernambuco
(UFPE), Recife, Brazil

E-mail: ronaldmoura1989@gmail.com

- Magalhães da Silva T, Sandhya Rani M R, Nunes de Oliveira Costa G *et al*: The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *Eur J Hum Genet* 2014; **23**: 984–989.
- Salzano FM, Sans M: Interethnic admixture and the evolution of Latin American populations. *Genet Mol Biol* 2014; **37**: 151–170.
- 1000 Genomes Project Consortium: An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **135**: 0–9.
- Kosoy R, Nassir R, Tian C *et al*: Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat* 2009; **30**: 69–78.

- Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945–959.
- Besnier F, Glover KA: ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One* 2013; **8**: 1–5.
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria, 2015. Available at <http://www.r-project.org/>.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I: Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 2015; **15**: 1179–1191.
- Pena SDJ, Di Pietro G, Fuchshuber-Moraes M *et al*: The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 2011; **6**: e17063.
- Ding L, Wiener H, Abebe T *et al*: Comparison of measures of marker informativeness for ancestry and admixture mapping. *BMC Genomics* 2011; **12**: 622–640.
- Galanter JM, Fernandez-Lopez JC, Gignoux CR *et al*: Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet* 2012; **8**: e1002554.
- Wang S, Ray N, Rojas W *et al*: Geographic patterns of genome admixture in Latin American mestizos. *PLoS Genet* 2008; **4**: 1–9.
- Qu HQ, Li Q, Lu Y *et al*: Ancestral effect on HOMA-IR levels quantitated in an American population of Mexican origin. *Diabetes Care* 2012; **35**: 2591–2593.
- Sandoval JR, Salazar-Granara A, Acosta O *et al*: Tracing the genomic ancestry of Peruvians reveals a major legacy of pre-Columbian ancestors. *J Hum Genet* 2013; **58**: 627–634.
- Via M, Gignoux C, Roth LA *et al*: History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. *PLoS One* 2011; **6**: e16513.