Forensic Science International: Genetics Supplement Series 8 (2022) 91–96

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

Sample collection strategies when building mitochondrial DNA forensic databases

Filipa Simão [a,*], Adriana Castillo [b], Germán Burgos [c,d,e], Leonor Gusmão [a]

[a] DNA Diagnostic Laboratory, State University of Rio de Janeiro, Brazil
[b] Genetics Laboratory, Industrial University of Santander, Bucaramanga, Colombia
[c] Escuela de Medicina, Facultad de Ciencias de la Salud, Universidad de Las Américas, Quito, Ecuador
[d] Grupo de Medicina Xenómica. Instituto de Ciencias Forenses, Universidad de Santiago de Compostela, Spain
[e] One Health Research Group, Facultad de Medicina, Universidad de Las Américas, Quito, Ecuador

ARTICLE INFO

Keywords:
MtDNA
Admixture
South America
Frequency databases

ABSTRACT

For establishing databases that capture the existing diversity in populations, the sample collection strategy is a determining factor and caution must be taken when choosing the suitable approach. Many researchers choose to restrict the sampling to individuals with inheritance for three generations in a specific geographic location. However, the appropriate database in a forensic context is the one representing the current population. We analyzed mtDNA composition across generations in populations from Colombia, Ecuador, and Paraguay. An overall genetic homogeneity was detected, with statistically significant differences on macrohaplogroup frequencies for few department/regions.

1. Introduction

In a forensic context, the statistical weight of a match between two identical mtDNA profiles depends on the frequency of the haplotype in a database. The high diversity of mtDNA haplotypes reported worldwide highlights the importance of large representative databases that capture the existing variability and allow population substructure to be evaluated [1]. In these circumstances, the sample collection strategy is a crucial factor. Many authors choose to restrict the sample collection to individuals with residence and proven inheritance for, at least, three generations in a specific geographic location, which not necessarily represent the reference database in most forensic scenarios, namely in populations subject to recent migrations.

The current genetic diversity in South American populations is mainly attributed to admixture events during the colonial period. More recent immigration from Europe and Asia are also influencing the genetic composition of these populations, as well as the continuous movement of individuals between and within countries. It is known that the admixture processes between individuals from different continental backgrounds have been happening distinctively across South America and resulted in patterns of admixture that vary throughout the subcontinent [e.g. [2,3]]. Considering this, the aim of this work was to evaluate to what extent sample strategies impact on capturing the existing diversity in South America. Accordingly, the maternal genetic background of admixed populations was studied, to evaluate if differences exist on the genetic composition of populations over close generations.

2. Materials and methods

Detailed information on the samples used are described on Table 1.

MtDNA haplotypes from Paraguay were retrieved from Simão et al. [4] and haplotypes from Ecuador and Colombia were obtained with the same methodologies as in Simão et al. [4]. Values of haplotype diversity (H) and Analysis of molecular variance (AMOVA) were obtained with Arlequin [5]. The exclusion power (mtCE) was calculated according to Simão et al. [6]. Analyses were performed after discarding indels at homopolymeric tracts.

3. Results

The values of H and mtCE obtained for each department/region did not change significantly over the generations established (Table 2). Nonetheless, in some departments/regions it is possible to detect a loss of diversity for the individuals with three generations at a specific birthplace, when compared with the values obtained considering

* Corresponding author.
E-mail address: f.simao@campus.fct.unl.pt (F. Simão).

https://doi.org/10.1016/j.fsigss.2022.09.033
Received 13 September 2022; Accepted 29 September 2022
Available online 1 October 2022

Table 1
Origin and number of the samples used in the present study.

	Colombia[a]									Paraguay[b]
Region/ department	Antioquia	Boyacá	Caldas	Cundinamarca	Huíla	Norte Santander	Risaralda	Santander	Tolima	Alto Paraná
Living place	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	215
Birthplace	42	49	37	56	47	49	40	203	52	112
Mother birthplace	33	38	28	40	41	24	32	141	49	27
Grandmother brithplace	30	43	33	27	39	22	23	138	53	Ø
All generations*	26	33	18	21	37	18	16	126	36	25

Legend: Ø Information not available, *individuals with inheritance in that geographic location for three generations, [a]unplublished, [b]Simão et al. [4]
Note$_1$: The total number of samples for each country vary throughout generations because information was not available for all individuals.
Note$_2$: Only datasets with more than 15 samples were used in the analyses.

Paraguay[b]										Ecuador[a]		
Caaguazú	Caazapá	Capital	Central	Concepción	Cordillera	Guaíra	Itapúa	Misiones	Paraguarí	Central	North	South
2	0	30	61	0	3	111	76	1	37	Ø	Ø	Ø
23	15	91	31	7	22	105	67	9	39	68	70	42
30	30	52	33	16	42	113	64	20	72	22	28	20
Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	31	17	28
1	0	13	14	0	2	80	48	1	27	16	16	15

Table 2
Values of haplotype diversity (H), exclusion power (mtCE) and AMOVA obtained for three generations in Paraguay, Colombia, and Ecuador.

Country	Region/ Department	Haplotype diversity (H)						Exclusion power (mtCE)						AMOVA			
		Living place	Birthplace	Mother birthplace	Grandmother birhtplace	All generations	highest minus lowest	Living place	Birthplace	Mother birthplace	Grandmother birhtplace	All generations	highest minus lowest	Among populations	Within populations	F_{ST}	P-value
Colombia	Antioquia		0,9849	0,9811	0,9747	0,9846	0010		0,9512	0,9545	0,9333	0,9600	0018	-2,69	102,69	-0,02691	1,00000 ± 0,00000
	Boyacá		0,9651	0,9744	0,9767	0,9754	0012		0,9048	0,9317	0,9358	0,9318	0031	-1,82	101,82	-0,01816	1,00000 ± 0,00000
	Caldas		0,9685	0,9735	0,9394	0,9542	0034		0,9174	0,9048	0,8769	0,8889	0041	-1,36	101,36	-0,01357	0.86911 ± 0.00320
	Cundinamarca		0,9864	0,9859	0,9858	0,9857	0001		0,9316	0,9436	0,9316	0,9286	0012	-1,78	101,78	-0,01776	0.99980 ± 0.00014
	Huíla		0,9833	0,9805	0,9798	0,9775	0003		0,9636	0,9634	0,9636	0,9595	0,000	-2,22	102,22	-0,02223	1,00000 ± 000000
	Norte Santander		0,9872	0,9891	0,9870	0,9935	0000		0,9307	0,9457	0,9307	0,9739	0015	-2,25	102,25	-0,0225	0.99624 ± 0.00059
	Risaralda		0,9359	0,9435	0,9407	0,9333	0005		0,8090	0,8306	0,8142	0,8167	0022	-2,42	102,42	-0,02425	0.99901 ± 0.00030
	Santander		0,9773	0,9786	0,9780	0,9777	0001		0,9258	0,9338	0,9364	0,9389	0011	-0,49	100,49	-0,00493	1,00000 ± 0,00000
	Tolima		0,9842	0,9847	0,9877	0,9857	0004		0,9637	0,9541	0,9637	0,9635	0010	-1,18	101,18	-0,01176	0.99158 ± 0.00094
Paraguay	Alto Paraná	0,9927	0,9907	0,9858		0,9867	0007	0,9839	0,9904	0,9829		0,9833	0008	-0,65	100,65	-0,00645	0.99970 ± 0.00017
	Caaguazú		0,9921	0,9839			0008		0,9960	0,9747			0021	-2,22	102,22	-0,02217	0.95772 ± 0.00210
	Caazapá		0,9905	0,9816			0009		0,9714	0,9724			0001	-2,9	102,9	-0,02905	0.94386 ± 0.00252
	Capital	0,9931	0,9956	0,9932			0003	0,9425	0,9812	0,9857		0,9615	0043	0,78	99,22	0,00778	0.12139 ± 0.00366
	Central	0,9945	0,9978	0,9905			0007	0,9869	0,9892	0,9886		0,9890	0002	-1,7	101,7	-0,01703	1,00000 ± 0,00000
	Concepción Cordillera		1,0000	0,9917 0,9884			0012		0,9481	0,9833 0,9733			0025	-1,94	101,94	-0,01943	0.95059 ± 0.00219
	Guaíra	0,9758	0,9762	0,9817		0,9753	0006	0,9613	0,9599	0,9663		0,9557	0006	-0,74	100,74	-0,00743	1,00000 ± 0,00000

(*continued on next page*)

Table 2 (*continued*)

Country	Region/ Department	Haplotype diversity (H)					Exclusion power (mtCE)					AMOVA				
		Living place	Birthplace	Mother birthplace	Grandmother birhtplace	All generations	highest minus lowest	Living place	Birthplace	Mother birthplace	Grandmother birhtplace	All generations	highest minus lowest	Among populations	Within populations	F_{ST}	P-value
	Itapúa	0,9965	0,9946	0,9906		0,9929	0006	0,9877	0,9869	0,9767		0,9805	0011	-1,33	101,33	-0,01328	1,00000 ± 0,00000
	Misiones Paraguarí	0,9700	0,9811	0,9895 0,9922		0,9772	0022	0,9580	0,9703	0,9789 0,9832		0,9658	0025	-1,28	101,28	-0,01281	0,99505 ± 0.00067
Ecuador	Central		0,9934	0,9957	0,9935	1,0000	0002		0,9781	0,9957	0,9914	1,0000	0018	-1,45	101,45	-0,01452	0,98455 ± 0.00118
	North		0,9983	1,0000	1,0000	1,0000	0002		0,9963	0,9947	1,0000	1,0000	0005	-1,25	101,25	-0,01248	0.92554 ± 0.00256
	South		0,9930	0,9947	0,9921	0,9905	0003		0,9907	0,9842	0,9894	0,9810	0006	-1,59	101,59	-0,01595	0.96535 ± 0.00202

Note$_1$: AMOVA results obtained after 10100 permutations.
Note$_2$: highest vs. lowest columns refers to the different between the highest and lowest value of diversities calculated, independent of the generations.

individuals' birthplace (Table 2).

The AMOVA performed for each department/region (after grouping samples according to the three generations established) showed no statistically significant differentiation among generations (data not shown), in both haplotype and haplogroup composition.

A Fisher test was performed to assess the presence of statistically significant differences in macro-haplogroup proportions across generations (after grouping samples into A, B, C, D, Eurasian and African). Haplogroup frequencies were constant over the generations, with some exceptions. The frequency of haplogroup B was higher in the group of individuals living in Capital (Paraguay) (57%) than in the subsets of individuals and mothers born in the region (30% and 35%, respectively). Statistically significant differences ($p < 0.05$) were also detected in haplogroup C from Capital, and European lineages from Alto Paraná and South Ecuador (data not shown).

4. Discussion

An overall genetic homogeneity was detected, although residual differentiation seems to exist in some department/regions. If changes occurred on the maternal composition over recent generations, the high mtDNA diversity in South America [e.g. [4,7]] may have hampered the detection of differences among the subsets established. Based on genealogical data, these results can also be explained due to geneflow among populations that are not significantly different. For example, in the case of Paraguay, for the last three generations there was a high migration among departments with similar mtDNA genetic background [4].

5. Conclusion

Haplotype frequency databases built to disclose population history are often used for forensic purposes. However, such databases do not always represent the current genetic diversity of the population, but the fraction of those individuals with three-generation heritage in a geographic region. Therefore, to assess if databases built with different purposes can be interchangeable, it is crucial to consider demographic and historical data that may point to genetic differences between generations. This aspect is particularly important in the construction of databases of South American populations, which have complex population dynamics, with different levels of recent immigration and of isolation.

Financial support

L.G. and F.S. were supported by FAPERJ, Brazil (CNE-2022 and E-26/202.275/2019). L.G. was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico–CNPq, Brazil (ref. 306342/2019-7). AC was supported by Vicerectoría de Investigación y Extensión, Universidad Industrial de Santander (internal financing code 2488 of 2019). G.B was supported by MED.GBF.20.07 funded by DGIV from Universidad de Las Américas; Quito, Ecuador.

Conflict of interest

None.

References

[1] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing, Forensic Sci. Int. Genet. 13 (2014) 134–142.

[2] F. Saloum de Neves Manta, R. Pereira, R. Vianna, A. Rodolfo Beuttenmüller de Araújo, D. Leite Góes Gitaí, D. Aparecida da Silva, E. de Vargas Wolfgramm, I. da Mota Pontes, J. Ivan Aguiar, M. Ozório Moraes, E. Fagundes de Carvalho, L. Gusmão, Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels, PLoS One 8 (2013), e75145.

[3] D. Corach, O. Lao, C. Bobillo, K. van Der Gaag, S. Zuniga, M. Vermeulen, K. van Duijn, M. Goedbloed, P.M. Vallone, W. Parson, P. De Knijff, M. Kayser, Inferring continental ancestry of argentineans from autosomal, Y-chromosomal and mitochondrial DNA, Ann. Hum. Genet 74 (2010) 65–76.
[4] F. Simão, J. Ribeiro, C. Vullo, L. Catelli, V. Gomes, C. Xavier, G. Huber, M. Bodner, A. Quiroz, A.P. Ferreira, E.F. Carvalho, W. Parson, L. Gusmão, The ancestry of eastern paraguay: a typical south american profile with a unique pattern of admixture, Genes 12 (2021).
[5] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0); an integrated software package for population genetics data analysis, Evol. Bionform. Online 1 (2005) 47–50.
[6] F. Simão, A.P. Ferreira, E.F. de Carvalho, W. Parson, L. Gusmão, Defining mtDNA origins and population stratification in Rio de Janeiro, Forensic Sci. Int. Genet. 34 (2018) 97–104.
[7] A. Gómez-Carballa, F. Moreno, V. Álvarez-Iglesias, F. Martinón-Torres, M. García-Magariños, J.A. Pantoja-Astudillo, E. Aguirre-Morales, P. Bustos, A. Salas, Revealing latitudinal patterns of mitochondrial DNA diversity in Chileans, Forensic Sci. Int. Genet. 20 (2016) 81–88.