



# Accumulation of population polymorphisms in the mitonuclear genome with probable adaptive effect to extreme environments as altitude for forensic purposes

G.C. Iannacone<sup>a,b,c,\*</sup>, P. Ramírez<sup>b</sup>

<sup>a</sup> Institute of Legal Medicine and Forensic Science, Lima, Peru

<sup>b</sup> Laboratory of Molecular Microbiology and Biotechnology, Universidad Nacional Mayor de San Marcos, Lima, Peru

<sup>c</sup> Universidad Peruana de Ciencias Aplicadas, Lima, Peru

## ARTICLE INFO

### Keywords:

Mitonuclear interaction  
Population genomic  
Genome  
DNA polymorphism  
Forensic DNA  
Extreme environmental

## ABSTRACT

We analyzed the accumulation of population polymorphism in 2504 individuals - nuclear genomes (nDNA) of 26 populations (81 genes associated to extreme environments) and 3295 mitochondrial genomes (mtDNA) of 47 populations with the aim to found mitonuclear relationship associated an extremes environment as altitude. For that, we use an algorithm developed by us to determine the accumulation of polymorphisms by segments in the genome and thus be able to perform the multivariate analysis to found SNPs differences and similarities among populations. The results showed in Peruvian population a statistically significant mitonuclear relationship for 113/293970 nDNA SNPs in 16/81 genes. In the case of the mtDNA, we found a statistically significant mitonuclear relationship for 6/22 mtDNA positions – Gene. Additionally for the Peruvian population, the MRPP3 had the greatest polymorphism contribution with respect to other populations. Then, these nDNA and mtDNA SNPs in genetically close populations to Peru can be applied to forensic genomic phenotyping to identify groups likely adapted to extreme conditions (such as altitude) or make individualization between low and high altitude populations.

## 1. Introduction

Studies at the genomic level are generating a large amount of information with application in forensic genetics, that require tools capable of scanning the genome to identify the degree of accumulation of DNA polymorphisms in segments throughout populations (population genomics). Thus, these polymorphism segments (Multiple SNPs) in the genome, allow to increase the power of discrimination among individuals and some of them allows to determine individualizing phenotypic characteristics at the level of population subgroups [1], such as particular geographic adaptations accumulated in the time; and jointly with other molecular markers, contribute to the resolution of forensic cases. In this context, we proceeded to determine population polymorphisms using an algorithm development by us with the aim to determinate the mitonuclear relationship with a probable adaptive effect (genes associated with altitude) to characterize Peruvian population with respect to other populations.

## 2. Material and methods

We analyzed nuclear genome sequences of 2504 individuals, which includes Peruvian population (approved by the Research Ethics Committee, OI-003-11 and OI-087-13, of the National Institute of Health of Peru (Population data and coding at <https://www.adnsolucion.com/ndna>). These sequences correspond to 81 nuclear (nDNA) genes associated to joint function with mitochondrial (mtDNA) and in the same time are genes associated to extreme environmental metabolic conditions adaptation such as altitude. Likewise, 3295 individuals were used for the mitochondrial genome analysis distributed in 47 populations where the nuclear genome populations selected in this study are contained (Population data and coding at <https://www.adnsolucion.com/mtdna-population>). In this sense, the aim was to determine the differential distribution of mitonuclear polymorphisms in the Peruvian population with respect to close and distant populations in relation to ancestral migratory origin.

The genomic data for both genomes was aligned, transformed and

\* Correspondence to: Av. Los Conquistadores 575, San Isidro, Lima, Peru.

E-mail address: [giannacone@mpfn.gob.pe](mailto:giannacone@mpfn.gob.pe) (G.C. Iannacone).

<https://doi.org/10.1016/j.fsigs.2022.09.027>

Received 5 September 2022; Accepted 28 September 2022

Available online 4 October 2022

1875-1768/© 2022 Elsevier B.V. All rights reserved.

analyzed to determine the Combined Segment Index CSI [2], algorithm formula at <https://www.adnsolucion.com/algorithm-formula>. Once the CSI values were obtained in both genomes, the segments were classified as probably selective (S), Slightly selective (SS), Slightly neutral (SN) and Neutral (N). With these results a ranking was made for the S, SS and SN with respective CSI values and we introduce them in a matrix of nucleotide positions range by populations versus Factors (type of segment CSI, Chromosomes, Genes, Gene Families) for multivariate analysis with PRIMER v7.0.21. Likewise, with the results obtained for the values of S, SS and SN, a matrix of nucleotide positions versus populations were constructed in order to analyze the significance of Z distribution for the hypotheses of differential distribution of polymorphisms for the Peruvian population at the interpopulation level for both genomes (Values < 0.05), as well as to determine the similar interpopulation distribution of mitonuclear relationship for the Peruvian population (Values > 0.05) and then this last matrix was analyzed also with the PRIMER v7.0.21 program.

### 3. Results and discussion

In relation to the population clustering with the CSI values for segments S, SS and SN in the nDNA and mtDNA, the results show for both genomes in relation to Perú (PEL) with respect to the other populations, a congruence with the population distributions (phylogenetic tree) obtained with the traditional genetic distances methodologies [3], where the migratory origins of human populations range from Africa to America. In relation to the distribution of CSI values for nDNA and mtDNA, the Peruvian population shows a differential population pattern with respect to the other populations and is located in the tree next to the Latin American population. In the case of nDNA, a total of 14772/25389 segments for Perú were obtained for a normalized CSI for S, SS and SN, corresponding to 51/81 nDNA genes. In relation to the type of the segment in the nDNA, the MTPAP and EGLN1 genes are only the type S, the HIF1A is exclusively the type SS and the 28/51 genes are exclusively type SN. In the case of the remaining 20/51 nDNA genes, they have segments of type S, SS and SN within the same gene. Likewise, it was observed higher CSI values in the chromosome 14, followed by the chromosome X, 11 and 21 and it correspond to segments of greater accumulation of polymorphism, (<https://www.adnsolucion.com/figure-1>).

In the case of the heat map plotting by genes and chromosomes for the nDNA segments S, SS and SN showed differential variability in the case of the Peruvian population for the gene: DSCAM (Chromosome 21), MRPP3 (Chromosome 14) and FLT1 (Chromosome 13). Likewise, it was found that the highest normalized CSI values were for the BDKRB2, CYP17A1 and MRPP3 genes, followed by PKIA, TEK, SNPs rs1692120, PPP3CA, VDR and HIF1A, (<https://www.adnsolucion.com/figure-2>).

In relation to the statistical significance of the distribution of nDNA SNPs in the Peruvian population of characteristic SNPs of Peru (52873/293870 SNP) with respect to the populations of South America showed a 497/52873 SNPs, which correspond to 16/81 nDNA genes analyzed and corresponding to ADRB3, DLC1, COL5A1, DSCAM, COL24A1, MRPP3, EPAS1, FLT1, PPARGC1A, PPP3CA, ANKS1B, TENM2, BACT1 and the environment of the SNPs rs1692120, rs1372635, rs3564453. The MRPP3 showed the highest normalized CSI values of 9/16 nDNA genes in the analysis of heat map plotting with respect to other populations (<https://www.adnsolucion.com/figure-2>). Additionally, a greater contribution of polymorphisms from Latin America and/or Peru to the world was observed in the case of the nDNA SNPs of genes PPARGC1A, PPP3CA, ANKS1B, TENM2, FLT1, MRPP3, HIF1A, SNP rs1692120 and SNP rs1372635 (Data not show).

To determine the probable mito-nuclear relationship, we analyzed the significant distribution of mtDNA SNPs using the same procedure performed for nDNA. The results show accumulation of segments type S, SS and SN characteristic for the Peruvian population and its correspond to 22 SNPs/16569nt of mtDNA, of which they have statistical

significance in 12/22 mtDNA SNPs with respect to Latin American populations such as the Mexican population, which has similar mtDNA haplotypic diversity with the Peruvian population. The positions of these mtDNA SNPs with statistically significant value (position – Gene): 11177nt - ND4, 6473nt - COX1, 6755nt - COX1, 663nt - 12s, 1736nt - 16s, 4248nt - ND1, 12007nt - ND4, 3547nt - ND1, 827nt - 12S, 9950nt - COX3, 15535nt - CYTB and 14053nt - ND5.

When analyzing the distribution of both nDNA and mtDNA SNPs under the hypothesis of equal mitonuclear distribution, only 6/22 mtDNA SNPs (11177nt - ND4, 6473nt - COX1, 6755nt - COX1, 3547nt - ND1, 827nt - 12S and 9950nt - COX3) show similar distribution (statistically not different) with the 16/81 significant nuclear SNPs genes for the Peruvian population (113/497 significant nDNA SNPs). Additionally, we also observed distributions similar to “haplotypes” (transmission related between nDNA and mtDNA SNPs variants) in a heat map plot with greater values of similar significance. We found 104 mitonuclear “haplotype” (The definition of haplotypes can be seen at <http://www.adnsolucion.com/mitohaplotipos>) that include mtDNA SNPs with significant distribution in the mitonuclear relationship for Peruvian population, mtDNA SNPs 6473 - COX1, 11177 - ND4, 827-12S, 15535 - CYTB as mitonuclear “haplotype” 3, 2, 4, 31,30, 29, 28; mtDNA SNPs 3547 - ND1, 9950 - COX3 as mitonuclear “haplotype” 4, 31, 30, 29, 28, 27, 25; mtDNA SNPs 6755 COX1 as mitonuclear “haplotype” 42, 48, 47, 45, 43, 40, 39, 38; (<https://www.adnsolucion.com/figure-3>). The distribution of the 113 nDNA SNPs with respect to the mitonuclear “haplotypes” is listed at <https://www.adnsolucion.com/mitonuclear-a-sociation>.

### 4. Conclusion

With the developed algorithm, the population genomic polymorphism can be quickly scanned through a methodology based on segmenting the genome and quantifying each of these segments. Thus, we found in the Peruvian population 16/81 nDNA genes associated in previous studies to extreme conditions (as altitude) and that in turn these have a mitonuclear relationship with respect to 6/22 genes of mtDNA in the Peruvian population. In this sense, the 113 nDNA SNPs and the 6 mtDNA SNPs can be used as a characterization pattern for forensic individualization purposes of the Peruvian population and at the same time it allows to determine differences between genetically close (regional) populations that have been exposed in generations time to extreme conditions (such as altitude) compared to those that have not been. Without forgetting, the adaptation history in the same extreme condition could include some different SNPs genes in genetically distant populations, as has been observed in some altitude adaptations studies between Andean and Tibetan populations. Likewise, by having greater specificity of location of the DNA segments using this specific SNPs information, we will be able to design specific methodologies for new generation sequencing platforms [4] with the aim to increase the capacity of samples per run and also increase the sequencing depth that in conditions of degraded DNA or low DNA quantity would increase the possibility to obtaining more genetic information in relation to quality, quantity and marker type.

### Conflict of interest statement

None.

### References

- [1] P.M. Schneider, B. Prainsack, M. Kayser, The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry, *Dtsch. Arzteblatt Int.* 51–52 (2019) 873–880.

- [2] G.C. Iannacone, Population genomics of the mitochondrial genome segments and the prediction of neutral and selective trends for identification and association studies, *Forensic Sci. Int.: Genet. Suppl. Ser.* 7 (1) (2019) 826–828.
- [3] G.C. Iannacone, R.C. Parra, Genetic structure and kinship analysis from the Peruvian Andean area, *Forensic Sci. Humanit. Action* (2020) 473–489.
- [4] A. Serrano, Forensic DNA phenotyping: a promising tool to aid forensic investigation. Current situation, *Span. J. Leg. Med.* 46 (2020) 183–190.