Mitochondrion 12 (2012) 237–241

Contents lists available at SciVerse ScienceDirect

Mitochondrion

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

Mitogroup: Continent-specific clusters of mitochondrial OXPHOS complexes based on nuclear non-synonymous polymorphisms

Denis Pierron [a], Thierry Letellier [b], Lawrence I. Grossman [a,*]

[a] Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI 48201, USA
[b] Laboratoire de Physiopathologie Mitochondriale, INSERM, Université Victor Segalen Bordeaux 2, 146, rue Léo Saignat 33076, Bordeaux, France

ARTICLE INFO

Article history:
Received 19 July 2011
received in revised form 5 September 2011
accepted 16 September 2011
Available online 24 September 2011

Keywords:
Haplogroup
OXPHOS
Polymorphism
PCA
Pathogenicity score
Respiratory chain

ABSTRACT

OXPHOS polymorphisms are known to be population specific and to influence disease. Previous studies have focused on mtDNA polymorphisms. Based on a world sampling of 629 unrelated individuals, we have now studied the polymorphisms of the 80 genes encoding OXPHOS nuclear subunits. We have shown that (i) amino-acid replacement frequencies are significantly correlated with their pathogenicity probability, and (ii) populations can be distinguished based only on amino-acid replacements in nuclear encoded OXPHOS subunits. These results are congruent with the major mtDNA haplogroups, which suggests that OXPHOS complexes are different across the populations in both nuclear and in mitochondrial encoded subunits.

1. Introduction

Mitochondrial DNA (mtDNA) has played a substantial role in human population genetics for twenty years owing to its absence of recombination and its high mutation rate (Cann et al., 1987). Human mtDNA phylogeny was characterized by the designation of large lineages defined by a few specific positions, i.e., haplogroups (Torroni et al., 1993). Extensive phylogeographic studies have shown astonishing correlations between individuals' geographic origin and mitochondrial haplogroup, revealing the migration and settlement history of human species. For example, the finding that most of the oldest mitochondrial lineages (L0–6) are only present in sub-Saharan Africa and only two lineages (M and N, derived from African lineage L3) are present outside of Africa has significantly substantiated the theory of a recent African origin of *Homo sapiens* (Behar et al., 2008; Cann, 1992; Quintana-Murci et al., 1999).

Recent studies have suggested that the particular distribution of human haplogroups is not only due to demographic events, in which largely neutral changes are isolated by events such as migrations, but also due to some selective strengths (Mishmar et al., 2003; Ruiz-Pesini and Wallace, 2006). Indeed, mtDNA encodes thirteen genes of mitochondrial oxidative phosphorylation (OXPHOS), which is responsible for the trade off between food consumption and production of heat and energy. It has been proposed that some polymorphisms, which would induce higher heat production, have been beneficially selected in cold environments (Balloux et al., 2009; Ruiz-Pesini et al., 2004). The presence or absence of any "positive selection signal" on the human mtDNA phylogeny is highly debated in the human population genetics field (Kivisild et al., 2006; Pierron et al., 2011; Sun et al., 2007). Also, despite suggestions from numerous genetic epidemiology studies that haplogroups influence patient phenotypes (Pierron et al., 2008; Torroni et al., 1997), results from functional biochemical studies are contradictory (Elson et al., 2007; Ghelli et al., 2009; Martinez-Redondo et al., 2010; Tranah et al., 2011).

We propose that nuclear genetic background could be a missing variable to understand the discrepancies in relating mtDNA polymorphisms to phenotype. Indeed, we think that mutations on nuclear encoded OXPHOS subunit nuclear genes could influence the phenotypic expression of mitochondrial genotype. Supporting this idea, our group has demonstrated the existence of co-evolution between mitochondrial and nuclear subunits of complex IV on anthropoid primate lineages (Doan et al., 2004). Other studies have shown that human cells lacking mtDNA could not be fully rescued by chimpanzee and gorilla mtDNA (complex I deficiency) and not at all by mtDNA from other close relatives of humans such as orangutan or, in fact, any other primate (complex IV assembly deficiency), which suggest a tight cooperation between mitochondrial and nuclear genetic background (Barrientos et al., 1998).

In the present report we have investigated whether it is possible to distinguish OXPHOS nuclear genetic background across human populations and particularly between African and non-African

* Corresponding author. Tel.: +1 313 577 5326; fax: +1 313 577 5218.
E-mail address: l.grossman@wayne.edu (L.I. Grossman).

1567-7249/$ – see front matter
doi:10.1016/j.mito.2011.09.005

populations. To do so, we have listed the polymorphisms present on OXPHOS nuclear genes on European, East-Asian, and African population using the 1000 genomes dataset (Durbin et al., 2010). Based on these polymorphisms, we have performed principal component analysis (PCA) in order to identify any genotype cluster and then performed F-statistics to identify discriminating polymorphisms. Finally, we have screened the potential effects of positive and negative selection on theses polymorphisms.

2. Methods

2.1. List of nucleotides encoding OXPHOS complexes

Based on the Kegg database (Kanehisa, 2002), we have listed the genes encoding OXPHOS complexes I, III, IV, and V. Given our hypothesis suggesting coevolution of mitochondrial and nuclear genetic background, we have focused only on complexes containing both nuclear and mitochondrial genes (Table S1). Then, using the high quality CCDS database, we have listed on the reference grh37 human genome all the positions belonging to any coding region of any splicing variants (Pruitt et al., 2009). For every position, we have listed both reference nucleotide and ancestral nucleotides. The ancestral nucleotide came from the ENSEMBL database, which has inferred ancestral sequences from primate EPO multiple alignments using Ortheus (Flicek et al., 2011).

2.2. Polymorphisms on nucleotides encoding OXPHOS complexes

Based on the genomic sequence of 629 unrelated individuals from the 1000 genome database (November 2010 release) and Tabix using a python script, we have listed and characterized as synonymous or non-synonymous all the polymorphisms present at the positions described above (Durbin et al., 2010; Li, 2011). We then calculated the frequency of each polymorphism in each population and constructed a genotype of each individual containing only the polymorphic position coding for any OXPHOS gene (Table S2). We also generated an ancestor's genotype corresponding to the genotype build from the ENSEMBL ancestral sequence. We have generated three pools of genotypes: set (i) based on all polymorphic positions, set (ii) based only on positions where polymorphisms generate non-synonymous mutation, and set (iii) based only on positions where polymorphisms generate synonymous mutation. Then, for each genotype pool, we performed smartPCA using EIGENSOFT 4.0 software (Price et al., 2006).

2.3. Population stratification

In order to know whether or not the OXPHOS nuclear genetic background is different across populations, we have used EIGENSTRAT to explicitly test the population stratification (Price et al., 2006). For this statistical purpose, we have grouped each individual, based on origin, into one of the four large groups of populations: (i) "African" regrouping: African ancestry in Southwest (USA), Luhya in Webuye (Kenya), Yoruba in Ibadan (Nigeria); (ii) "East Asian" regrouping: Han Chinese in Beijing (China), Han Chinese in south (China), Japanese in Tokyo (Japan); (iii) "European" regrouping: European ancestry from Utah (USA), Finnish in Finland, British in England and Scotland, Toscans in Italy; and (iv) "American" regrouping: Mexican ancestry in Los Angeles (USA) and Puerto Rican in Puerto Rico (USA) (Table S3). The number (K) of axes of variation considered and set equal to the number of statistically significant PCA vectors (i.e., 3 axes for non-synonymous set) followed the authors' recommendation (Price et al., 2006).

In order to test for the existence of correlation between mtDNA and nuclear polymorphisms, we performed a new PCA on the 163 individuals for which haplogroups are thus far available (HapMap_Consortium, 2005). We have grouped individuals into four major haplogroups according to their continental distribution: L1–2–3 (African haplogroups), R (European haplogroup), M and N-nonR (Asiatic haplogroup).

2.4. Non-synonymous polymorphisms

In order to gain insight into these population-specific clusters of mitochondrial OXPHOS complexes (mitogroups), we have focused on the identification of the non-synonymous polymorphism differences between the populations. For each polymorphism we have calculated the F_{ST} value (Wright, 1950) between the populations (Table S2). Then we have evaluated the pathogenicity predicted for each amino acid substitution by MutPred (Li et al., 2009) (Table S2). Because this software does not evaluate nonsense mutations, we give to them the maximum pathogenic value (1.0).

2.5. Positive selection

In order to detect positive selection we have listed the empirical p-value for selection on each gene of the OXPHOS complex reported on HAPLOTTER and based on HAPMAP phase 2 database (HapMap_Consortium, 2005; Voight et al., 2006). This p-value is based on the number of SNPs with an extreme iHS score (Integrated Haplotype Score), which has high statistical power to study selection on non-fixed polymorphisms.

3. Results and discussion

3.1. Polymorphisms on nucleotides encoding OXPHOS complexes

Based on the Kegg and CCDS databases, we report altogether 51,228 bp, belonging to 96 CDS from 80 genes coding for OXPHOS nuclear encoded subunits (Table S1). The analysis of genomic sequence of 629 unrelated individuals shows the presence of 112 synonymous polymorphisms and 136 non-synonymous polymorphisms on 67 genes (Table S2).

3.2. Population stratification

PCA analysis of these identified SNPs shows that the polymorphisms of nuclear encoded OXPHOS genes significantly structure the population independent of whether we have used all polymorphisms (10 vectors significant), only synonymous polymorphisms (6 vectors significant), or only non-synonymous polymorphisms (3 vectors significant).

In order to know whether or not the OXPHOS nuclear genetic background is different across populations, we have tested the population stratification on the PCA results (Price et al., 2006). The three population groups, "African," "East Asian," and "European," appear to be very significantly different (p-value$<10^{-100}$) whichever set of data is used. The "American" population is less significantly different from the others (p-value$<10^{-10}$) and seems to be close to the European population. When all polymorphisms are used for the PCA there is virtually no overlap between African and non-African populations (Fig. S4).

Fig. 1 shows the PCA results based only on non-synonymous polymorphisms; the most important vector (eigenvalue = 22.48) reflects genetic variation between African and non-African populations. It follows that the second vector (eigenvalue = 16.27) separates the non-African populations. Because these results are based only on non-synonymous polymorphisms, they represent structural differences in the OXPHOS complexes clustering "African," "Asian," and "European" populations. Combined with previous studies on mtDNA haplogroups, these results suggest that OXPHOS complexes have significant differences between populations in *both* nuclear encoded and mitochondrial encoded subunits.