Inter and intra species comparative analyses of RHNumtS sequences

Calabrese FM[1], Lang M[1], Simone D[1], Mineccia G[1], Piredda R[1,3], Gasparre G[1,2], Attimonelli M

Motivation

Eukaryotic nuclear genomes contain fragments of mitochondrial genomes: the Nuclear mitochondrial Sequences, NumtS [1]. They are not equally abundant in all species, they are redundant and polymorphic in terms of copy number, thus they can be considered Copy Number Variants (CNVs) and hence candidate markers for population based association studies. NumtS can also be used as outgroups in phylogenetic studies. Although NumtS generation has been dated back to a primordial endosymbiontic event [2], the process is still ongoing. Thus, it is of paramount importance to have a complete overview of NumtS location within a genome, to analyze their population variability and to dissect the mechanisms leading to their generation. To optimize quantification and mapping of NumtS, our group has produced a consensus compilation of human NumtS both by applying various in silico approaches that exploit database similarity searching methods (Blast and BLAT) and by comparing the obtained data with pre-existing compilations. The resulting compilation has been named RHNumtS [3] and has been in part validated by direct sequencing. Recently a revised version of RHNumtS has been produced by applying the above methods with less stringent conditions. Starting with the RHNumtS compilation, we have carried out several analyses in order to date human NumtS and classify them as ancestral or human specific. Within these two classes, on the basis of the locus they map to, the subset of NumtS derived from genome rearrangements has been identified. If NumtS insertions can be dated back to the endosymbiotic event, we hypothesize that the NumtS' flanking region could contain traces of genomic regions of the ancestral bacteria from which the mitochondrion originated. Studies of these flanking regions will give us additional information on NumtS evolution.

Methods

Two different approaches have been implemented to fulfill the aims of our study, one based on comparative genomics and the other based on a protocol that exploits protein-protein database similarity searching with the aim to locate protein relicts on the human genome. The comparative genomics approach utilizes net

[1] Dipartimento di Biochimica e Biologia molecolare "Ernesto Quagliariello", University of Bari, Italy [2] Unità di Genetica Medica, Policlinico Universitario S. Orsola-Malpighi, Bologna, Italy. [3] Dipartimento Ambiente e Foreste (D.A.F.), University of Tuscia, 01100 Viterbo, Italy

tracks available at the UCSC Genome Browser site [4] (http://genome.ucsc.edu/). When comparing two genomes, the net track graphically displays the best alignment between them. This tool allows us to identify and unwrap syntenic NumtS regions of organisms whose genomes have been completely or almost completely sequenced. A syntenic region containing the NumtS can be classified either as human specific or ancestral, and thus orthologous, depending on the presence/absence of a gap in the alignment [5]. The analysis has been performed on Chimp, Macaca, Orangutan and Marmoset. Available UCSC tables have been used to extract the data. To verify our hypothesis of the occurrence of fossil mtDNA regions within the nuclear genome, we have applied Blastx program on each NumtS 5' and 3' flanking region versus the human proteome. We define flanking regions as the genomic region covering 1000 nucleotides on 3' and 5' of a NumtS.

Results

So far we have performed comparative genomics analyses of the NumtS reported in the RHNumtS compilation (first issue [BMC Genomics. 2008 Jun 3;9:267]). No more than ten NumtS displayed a gap when compared with the selected primates indicating that the majority of human NumtS are fixed in the evolution and certainly they have been inserted before Macaca origin (27 Mya). Moreover, primers outside the NumtS sequence, used to amplify and sequence human-specific NumtS, have been tested on Macaca DNA. As expected, sequencing of the PCR products obtained indicates absence of the NumtS, thus confirming the in silico prediction. The application of Blastx on each flanking region versus the human proteome suggests the presence of protein fragments, here classified in three categories: i) proteins whose gene is coded by mtDNA, ii) fragmented proteins located on repeat elements, iii) other fragments of proteins whose origin requires further studies. The first category (i) allowed us to identify extensions of NumtS that had been unseen with Blastn; category (ii) allowed us to reconstruct the duplication path of the NumtS. Preliminary results will be reported.

Contact e-mail

franccalabrese@libero.it

Supplementary information

References

[1] Bensasson D, Zhang D, Hartl DL, Hewitt GM: Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 2001, 16 (6) :314-321 [2] Hartman H. The origin of the eukaryotic cell. Speculations Sci Technol. 1984;7 (2) :77-81. [3] Lascaro D, Castellana, S, Gasparre, G, Romeo G, Saccone, C, Attimonelli, M: The RHNumtS compilation: features and bioinformatics approaches to locate and quantify Human NumtS, BMC Genomics. 2008 Jun 3;9:267. [4] UCSC Genome Browser [http://genome.ucsc.edu/] [5] Einat Hazkani-Covo and Dan Graur: A Comparative Analysis of numt Evolution in Human and Chimpanzee. Mol Biol Evol. 2007 Jan;24 (1) :13-8