NumtS footsteps from the past: an interbreed between eukaryotic nucleus and the mitochondrion

F.M. Calabrese✉, D.L. Balacco, D. Simone, M. Attimonelli

Department of Biosciences, Biotechnologies and Pharmacological Sciences, University of Bari, Bari, Italy

Motivations

NumtS, the Nuclear sequences of mitochondrial origin, populates eukaryotic genomes more or less abundantly [1,2]. Mitochondrion gene order varies between lineages and its conservation has been observed among neighbour clades. The availability of complete and draft eukaryotic nuclear genomes encourages us to produce NumtS tracks (loadable on UCSC genome browser), a suitable tool for the comparison of NumtS pattern among species. Analyses have been performed on 21 different species ascribable to any of the clades reported in the UCSC gateway. The inter-species comparisons led to species-specific and shared NumtS sets. In order to infer where a shared NumtS locus has been fixed during evolution and to test if a NumtS set definition may be influenced by the mitochondrial genome used in the application of the in silico hybridisation nuclear DNA vs mitochondrial DNA, we have carried out crossed analyses of each nuclear genome versus any of the other mitochondrion genomes within our species dataset.

Methods

Blastn software [3] has been used to obtain the NumtS compilations then implemented, using python scripts, as tracks at the UCSC Genome browser. Galaxy suite has allowed us to intersect and extract data. SAMtools [4] have been used for BAM tracks and in manipulating alignments in the SAM format. The alignments and the mitochondrial gene orders have been checked by using Mauve software and by manual inspection.

The matrix reporting the content of the NumtS dataset resulting from crossed analyses has displayed differences in NumtS content depending on the mitochondrion genome used. Overall, among monotremes, platypus mitochondrion versus other nuclear genomes (Ornithoryncus anatinus) has shown the widest and the most recurrent NumtS number with the smallest standard deviation. Otherwise, Caenorhabditis elegans mitochondrion has led to a higher NumtS number among some mammals despite low values in the other species (itself included).

Results

Phylogenetic inferences will be reported based on inter species gene order and gene sequence conservation, nuclear genome size and other elements that will contribute to infer when and how NumtS have reached the nuclear genome.

References

1. Calabrese FM, Simone D, Attimonelli M. Primates and Mouse NumtS in the UCSC Genome Browser BMC bioinformatics (in press)
2. Simone D, Calabrese FM, Lang M, Gasparre G, Attimonelli M (2011) The reference human nuclear mitochondrial sequences compilation validated and implemented on the UCSC genome browser. BMC Genomics 12, 517.
3. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W (2003) Human-Mouse Alignments with BLASTZ. Genome Res. 13(1), 103-7.
4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) 1000 Genome Project Data Processing Subgroup: The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078-9.