Molecular Evolution

Towards an integrated resource for the study of population and disease associated variability of the human mitochondrial genome

M.A. Diroma✉, M. Attimonelli

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

Motivations

Thanks to the development of valid DNA sequencing technologies and statistical methods of analysis, nowadays we have a great amount of human mtDNA data. The mitochondrion genome presents a high rate of variability: the turning point of a polymorphism into a mutation is linked to heteroplasmy. A variability analysis of sequenced human mtDNAs has been performed by comparing published sequences with the reference sequence, rCRS, to detect and to characterize polymorphic positions in order to recognize relationships between i. variability values and haplogroups patterns, and ii. variability and pathogenicity.

Methods

Different web resources, such as Phylotree [1] and MITOMAP [2] besides GenBank and Pubmed, have allowed a variability analysis that has been carried out on mitochondrion genomes of healthy and pathologic individuals within HmtDB [3] where an evaluation of site-specific nucleotide and aminoacid variability is implemented, SiteVar [4]. Pathogenicity analysis has been possible by applying i. Polyphen2 [5] which compares wild type and variant alleles based on the aminoacidic conservation observed in proteins multialignment, and ii. SNPs&GO [6], a method based on support vector machines, for the prediction of functional effects of human non synonymous SNPs on mitochondrial proteins.

Results

The usage of SiteVar on continent-specific datasets has supported the knowledge of specific variability values for each ethnic group leading to the classification of haplogroups and to the characterization of potentially pathogenic mutations, even if there is no completely agreement among methods of prediction of pathogenicity yet. Among 5902 variable sites on the entire mito-

chondrial genome a half fits with those reported in Phylotree and 67 show the highest variability values. Patients present a considerable increase of low variability positions which could be specific mutations associated to a specific pathology while a lot of somatic mutations corresponds to polymorphisms in healthy individuals defining a specific haplogroup, suggesting that these mutations may be not necessarily associated to a pathology. Conclusion: The available results suggest that the integration of the abovementioned web resources, Phylotree, MITOMAP and HmtDB, could surely add values to the knowledge concerning human mitochondrial DNA, population histories and mitochondrion associated diseases. This is a message to the Bioinformatics community to activate interdisciplinary collaborations.

References

1. van Oven, M. and Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat, 30(2): E386-E394.

2. Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P. and Wallace, D.C. (2007) An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Research, 35 (Database issue):D823-D828.

3. Rubino, F., Piredda, R., Calabrese, F.M., Simone, D., Lang, M., Calabrese, C., Petruzzella, V., Tommaseo-Ponzetta, M., Gasparre, G. and Attimonelli, M. (2012) HmtDB, a genomic resource for mitochondrion-based human variability studies. Nucleic Acids Res., 40(D1), D1150-D1159.

4. Pesole, G. And Saccone, C. (2001) A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences. Genetics, 157, 859-865.

5. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010), Nat. Methods, 7(4), 248-249.

6. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P.L. and Casadio, R. (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Human Mutation, 30(8), 1237-1244.