# HMDB, the Human Mitochondrial Data Base, a genomic resource supporting population genetics studies and biomedical research on mitochondrial diseases

Marcella Attimonelli [(1)], Matteo Accetturo [(2)], Gaetano Scioscia[(2)], Caterina Marinelli [(2)], Piero Leo[(2)], Monica Santamaria[(1)], Stefano Mona[(3)], Daniela Lascaro[(1)], Imma Cascione[(1)] and M.Tommaseo-Ponzetta[(3)]

[(1)] Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Via E.Orabona 4, 70126 Bari – Italy
m.attimonelli@biologia.uniba.it
[(2)] Java Technology Center - IBM SEMEA Sud, Via Tridente 10, 70126 Bari – Italy
.JTC Emerging Technology W3 Home: http://ispc1.italy.ibm.com
[(3)] Dipartimento di Zoologia, Università di Bari, Via E.Orabona 4, 70126 Bari – Italy
m.tommaseo@biologia.uniba.it

**Keywords.** Genomics, Mitochondrion, Human, Database, Variability

## Introduction

Population genetics studies based on the analysis of mtDNA and mitochondrial disease studies have produced a huge quantity of sequence data and related information. These data, classified as RFLPs, mtDNA SNPs, pathogenic mutations, HVS1 and HVS2 sequences, and complete mtDNA sequences, are at present distributed worldwide in differently organised databases and web sites, not well integrated among them.
Several mitochondrial specialised databases and databases related with variability data have been designed and implemented, but generally they are structured as simple databases where data are stored, without the possibility to perform any analysis.
Moreover it is not generally possible for the user to submit and contemporarily analyse its own data comparing them with the content of a given database and this is valid both for population genetics data, and for mitochondrial disease data.
As far as population genetics data, for example, the problem of sequence classification in haplogroups is becoming more and more important as the improvement of sequencing technologies is increasing the availability of new complete mitochondrial genomes. Indeed up to now the only way to establish the haplogroup paternity of a given mitochondrial sequence is to manually observe its variant sites respect to a reference sequence, referring to literature in order to define its haplogroup-specific polymorphisms.
Also as far as mitochondrial disease data, despite the large number of disease-associated mutations already discovered in the last few years, the sequencing of the complete human mt genome [1] is allowing the discovery of new pathogenic mutations. Indeed, up to now, the pathogenicity of mtDNA mutations has been, in most cases, prevalently validated by their segregation with the disease and their consequent loss of function when the mutation involves a structural gene. However, no systematic statistical analysis of the mtDNA SNPs has been performed until now.
Here we present the design of a Human Mitochondrial genome DataBase (HMDB) that will collect the complete human mitochondrial genomes publicly available interfaced to analysis programs, allowing the classification of newly sequenced human mitochondrial genomes, and the prediction, through site-specific nucleotidic and aminoacidic analysis[2][3], of the pathogenic potential of mitochondrial polymorphisms.

## HMDB Design

HMDB will allow the user to perform retrieval of data from the database, to analyse the query resulting genomes, to compare a newly sequenced genome with the stored data in order to classify it according to the haplogroup classification. Moreover this resource could contribute to estimate the pathogenetic proneness of a newly detected mutation.
The project is in progress.
The database and the whole resource have been completely designed (Fig. 1 shows some of the slides describing the whole project.
The variability study of the available complete human mt genomes has been performed. Refinement of the resulting data and implementation of the resource will be performed in the present year. We hope to release the first version of the whole package at the end of 20004.

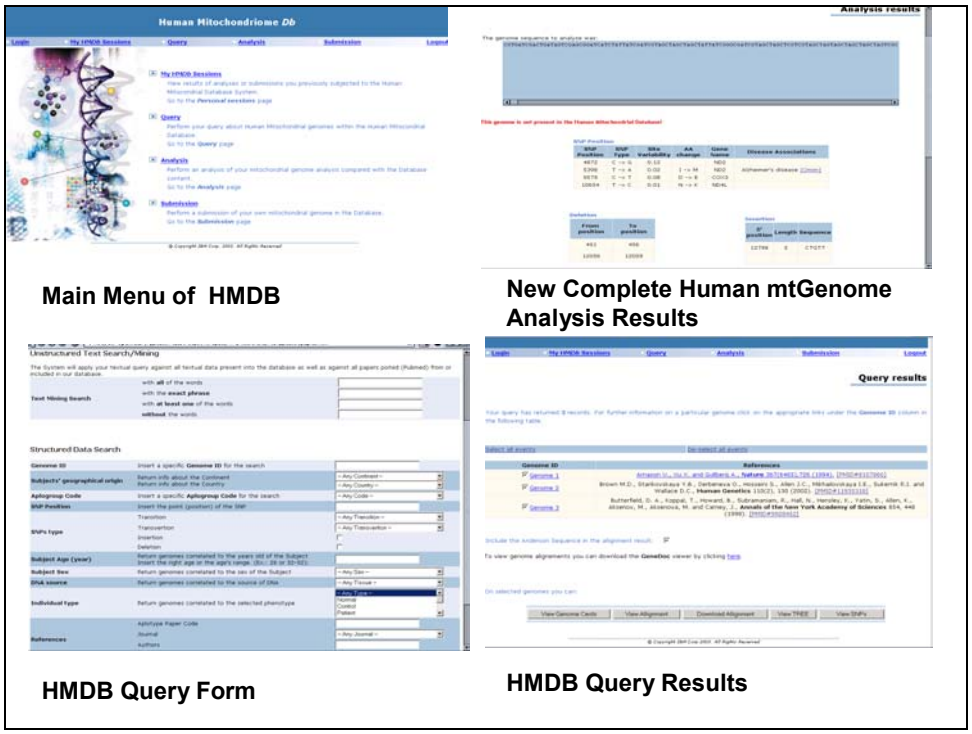


**Fig. 1** HMDB Resource Main Pages

## Acknowledgements

This work is supported by MIUR

## References

[1] M. Crimi, M.Sciacco, S. Galbiati, A. Bordoni, G. Malferrari, R. Del Bo, I. Biunno, N. Bresolin, G.P. Comi, A collection of 33 novel human mtDNA homoplasmic variants. *Hum Mutat.,* 20: 409, 2002

[2] G. Pesole, C. Saccone, A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences. *Genetic*s, 157: 859–865, 2001

[3] D. S. Horner, G. Pesole, The estimation of relative site variability among aligned homologous protein sequences. *Bioinformatics* 19: 600–606, 2003