

THE UMD SOFTWARE TO CREATE LOCUS SPECIFIC DATABASES

Christophe Béroud¹, Gwenaëlle Béroud¹, Catherine Boileau² and Mireille Claustres¹

¹*Laboratoire de Génétique Moléculaire, CHU de Montpellier, France*

²*INSERM U383, Hôpital Necker, Paris, France*

In addition to general mutation databases, locus specific mutation databases are needed to confront clinical and molecular data. The UMD (*Universal Mutation Database*) generic software was created to develop laptop and more recently Internet accessible **L**ocus **S**pecific **D**ata**B**ases (LSDBs). This project was initiated 10 years ago and has now come to maturity with the release of the 2003 version of the UMD software. This version goes beyond the usual LSDB including both coding and non-coding sequences for a specific locus and information about SNPs. A user friendly interface allows one to import data directly from the human genome project and the NCBI website. In addition, a graphical distribution of these SNPs along the locus is available with various zoom in options. The availability of intronic sequences lead us to develop various input validation routines and new processes such as the evaluation of the effect of mutations on splicing using the consensus values described by Cartegni. This is particularly useful to evaluate potential intronic mutations that may result in cryptic exon activation.

Furthermore, many features previously developed for specific LSDBs have also been added to the generic software. Thus the novel database structure includes new tables: pedigree, pictures, epitopes, polymorphic markers and modules. The automatic computation of the impact of the mutation has been improved and displays new information on: (i) the reading frame, (ii) the consensus value of splice sites, (iii) the theoretical size of transcripts and translational products, (iv) the loss of epitopes, polymorphic markers and critical protein domains. In addition, it computes all possible exon-skipping patterns capable of restoring the reading frame. Concomitantly, new graphical displays have been added: mutation map, deletion map and polymorphism map. Furthermore, as some genes contain repetitive modules, we have designed a new set of functions to define these modules and display the distribution of mutations among the consensus sequence.

UMD-LSDBs are not inert repositories but interactive tools using the power of computer analysis to answer complex queries for phenotypic heterogeneity and genotype/phenotype correlations. They will help to pinpoint discrepant cases in which the clinical expression differs from that expected. They should help elaborate allele-specific gene-based therapeutic strategies.

One drawback of LSDBs is that one has to query various databases in case of genetic heterogeneity (one disease - several alternative genes), before knowing which gene is associated with a specific phenotype (for example a myopathy associated with a mental retardation). To solve this problem, we have developed the UMD-central system, which is able to query all UMD-LSDBs and give an overview of their data.

The UMD software is freely available and can be used on laptop computers (Mac and PCs). Furthermore we offer hosting facilities for Internet access to UMD-LSDBs, local servers can also be used to set-up an Internet accessible UMD-LSDB.