

IMGT/HMMs: IG and TR protein database of alignments and Hidden Markov Models

Wafae EL ALAOUI, Hervé PHILIPPE and Marie-Paule LEFRANC

IMGT® the international ImMunoGenetics Information system® <http://imgt.cines.fr>, LIGM Institut de Génétique Humaine IGH CNRS UPR1142 141 rue de la Cardonille, 34396 Montpellier Cedex 05, France



Following recent sequencing of new genomes, the requirement in annotation has become significantly more important. The development of bioinformatics methods and software allows to analyze nucleotide and amino acid sequences in an automatic way. However the generalist approaches do not take into account the characteristics of the immunoglobulin (IG) and T cell receptor (TR) sequences. In order to overcome that problem, we will set up a new system based on the Hidden Markov Models (HMMs). We will create specific models of HMMs which will contain profiles of IG and TR coding regions for each gene and allele based on the IMGT unique numbering for V-DOMAIN and C-DOMAIN. This will be extended to the V-LIKE and C-LIKE domains of the proteins that belong to the immunoglobulin superfamily IgSF. The HMMs, together with phylogenetic analysis, will be used for the identification and annotation of newly sequenced IgSF genes and proteins, and in the study of multigene family evolution.

Alignment of alleles: Sheep (*Ovis aries*) TRGC1

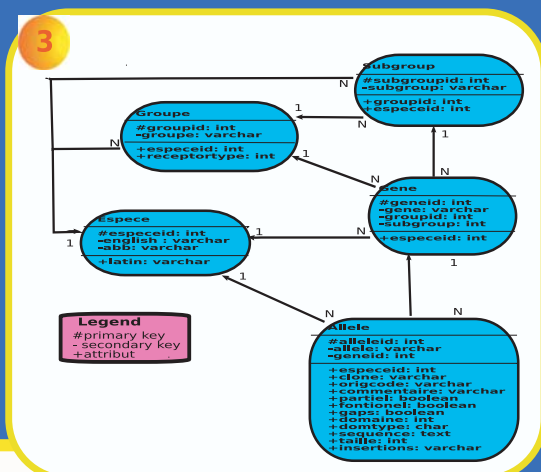
Alignment of alleles: Mouse (*Mus musculus*) IGLC1

Alignment of alleles: Human TRBV2

Protein display: African clawed frog (*Xenopus laevis*) TRBJ

Protein display: Zebrafish (*Danio rerio*) IGHJ

Protein display: Human IGH V-REGIONs



- 1 Aligned sequences in IMGT Protein display
- 2 Aligned sequences in IMGT Alignment of alleles
- 3 UML database for all available alignments
- 4 Database of HMMs profiles for all sequences (Generated by the Hmmer software)
- 5 Web interface for analysis of unknown sequences against HMMs profiles
- 6 Unknown sequences classified and aligned against the appropriate group or subgroup
- 7 Construction of a new phylogeny for a complete analysis of the sequences
- 8 IMGT Colliers de Perles available for the new sequences

Analyse a query sequence by searching IMGT/HMMs

Cut and paste your sequence here.

FASTA format or raw sequence are acceptable:

New unknown sequences

Or

Select the query sequence file you wish to use:

HMMs PROFILES

6

```

YYADYV...GFTTIRGMANKSLYIQNSLAEQ
YYAVSV...GFTTIRGDSKSLAYIQNSLAEQ
YYAVSV...GFTTIRGNSKNTLYIQNSLAEQAVYV
YYAVSV...GFTTIRGNSKNTLYIQNSLAEQAVYV
YYAVSV...GFTTIRGNSKNTLYIQNSLAEQAVYV
EYAVSV...GFTTIRGNSKNTLYIQNSLAEQAVYV
AYAVSV...GFTTIRGNSKNTLYIQNSLAEQAVYV
SYAVSV...GFTTIRGNSKNTLYIQNSLAEQAVYV
YYAVSR...GFTTIRGNSKNTLYIQNSLAEQAVYV
    
```

IMGT Collier de Perles for V-DOMAIN [8.8.20]

Homo sapiens VH (IGHV1-3 - IGHV1* - IGHV6) from b12 (PDB and IMGT/3Dstructure-DB: 1hah_H)

IMGT Collier de Perles for C-DOMAIN [10.4.11]

Homo sapiens CH2 (IGHG1) from b12 (PDB and IMGT/3Dstructure-DB: 1hah_H)

Phylogeny example of CH1 C-DOMAIN

7