IMGT/LIGMotif

A tool for immunoglobulin and T cell receptor gene identification and description in large genomic sequences

Jérôme Lane, Patrice Duroux and Marie-Paule Lefranc
IMGT®, the international ImMunoGenetics information system®, Université Montpellier 2, Laboratoire d'ImmunoGénétique Moléculaire, LIGM, UPR CNRS1142, Institut de Génétique Humaine IGH, 141 rue de la Cardonille, 34396, Montpellier cedex 5, France
{jerome.lane, patrice.duroux, marie-paule.lefranc}@igh.cnrs.fr

Im Muno Gene Tics Information system®

http://www.imgt.org

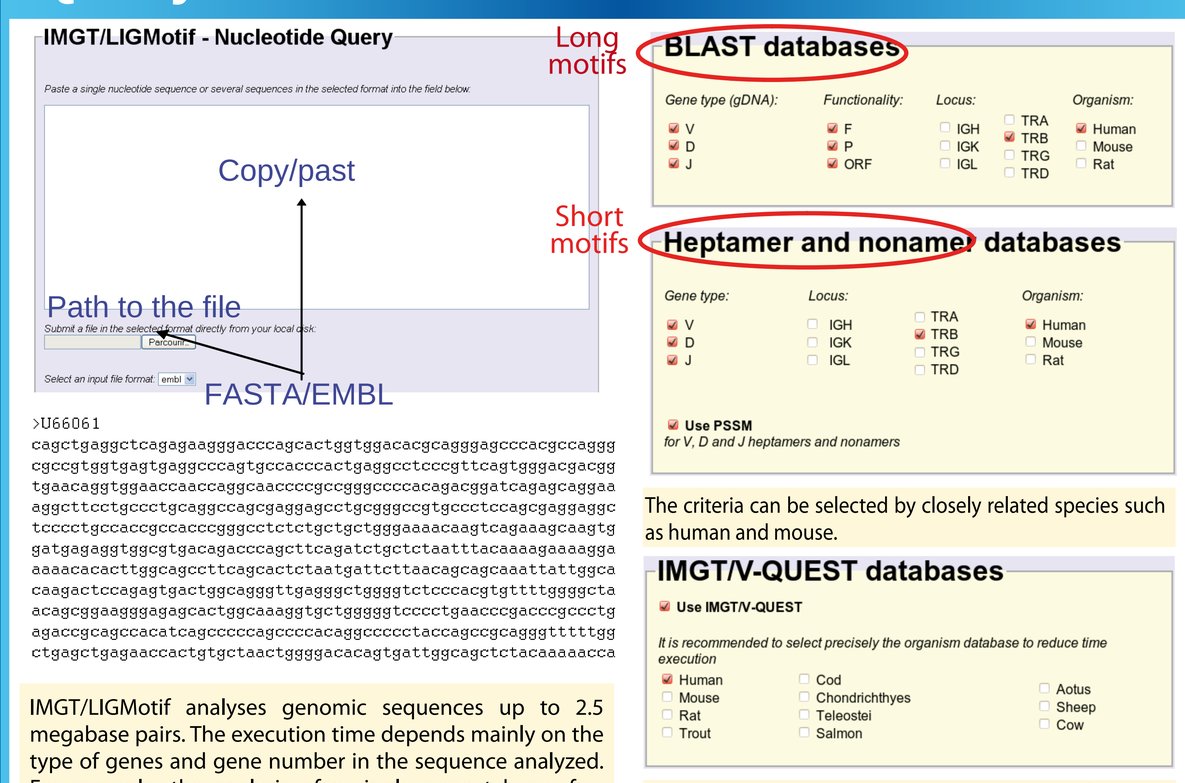
Abstract

IMGT/LIGMotif is a tool for immunoglobulin (IG) and T cell receptor (TR) gene annotation developed by IMGT®, the international ImMunoGeneTics information system®. The annotation process includes 4 modules that allow (i) the gene identification of variable V, diversity D and joining J genes by assigning gene type, localization and orientation in the analysed sequence, (ii) the gene description by assigning labels, (iii) the functionality identification and (iv) the gene delimitation and cluster assembly. The standardized rules used in IMGT/LIGMotif are based on IMGT-ONTOLOGY, the first ontology in immunogenetics and immunoinformatics.

J. Lane, P. Duroux and M.-P. Lefranc BMC Bioinformatics, 11:223 (2010)
M.-P. Lefranc, V. Giudicelli, C. Ginestoux et al Nucleic Acids Res., 37,1006-1012 (2009)
V. Giudicelli and M.-P. Lefranc Bioinformatics, 15:1047-1054 (1999)
P. Duroux, Q. Kaas, X. Brochet et al Biochimie, 90:570-583 (2008)

IMGT/LIGMotif Welcome page

Query

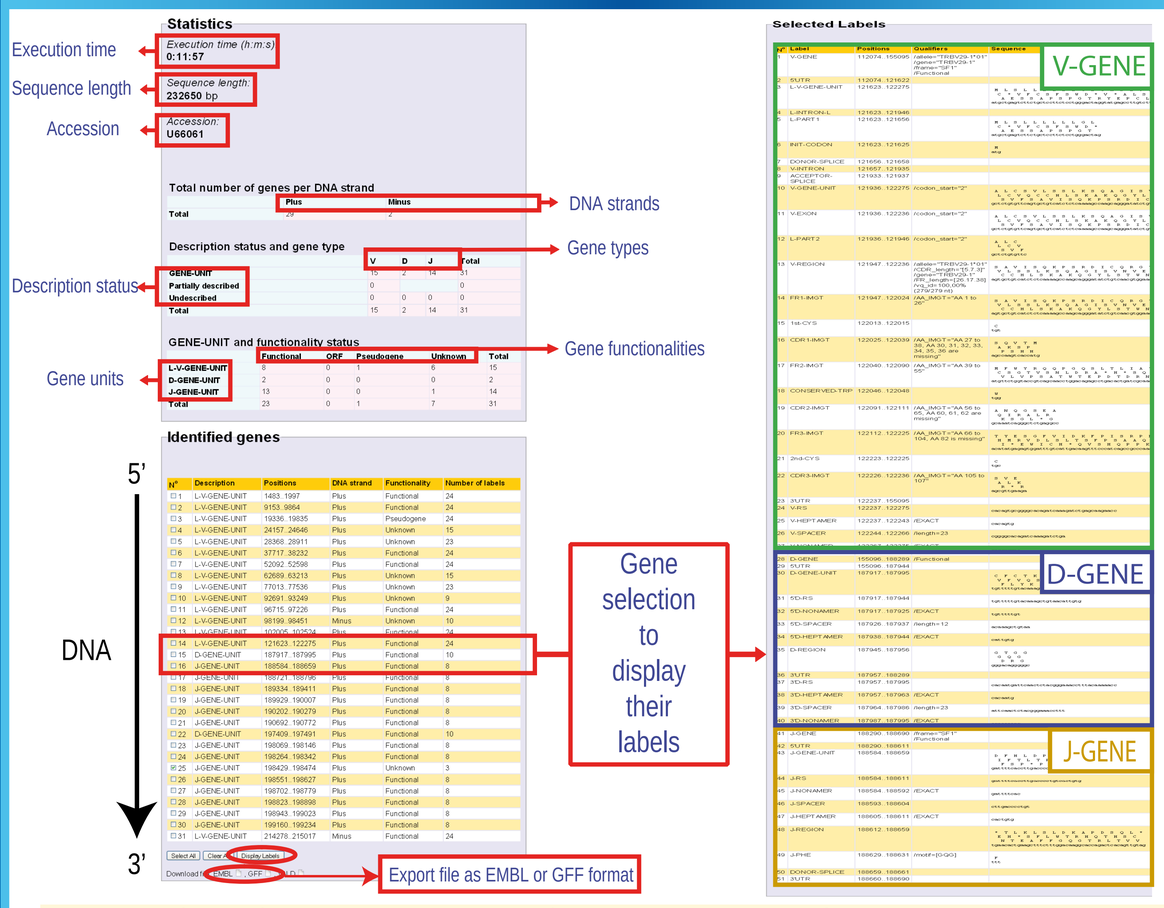


IMGT/LIGMotif analyses genomic sequences up to 2.5 megabase pairs. The execution time depends mainly on the type of genes and gene number in the sequence analyzed. For example, the analysis of a single gene takes a few seconds while the full analysis of a locus containing more than 100 genes takes about one hour using standard parameters.

The criteria can be selected by closely related species such as human and mouse.

IMGT/V-QUEST is integrated to IMGT/LIGMotif. The user can invalidate IMGT/V-QUEST by unchecking «use IMGT/V-QUEST» box. The V genes identified will be less well described but the program will run faster.

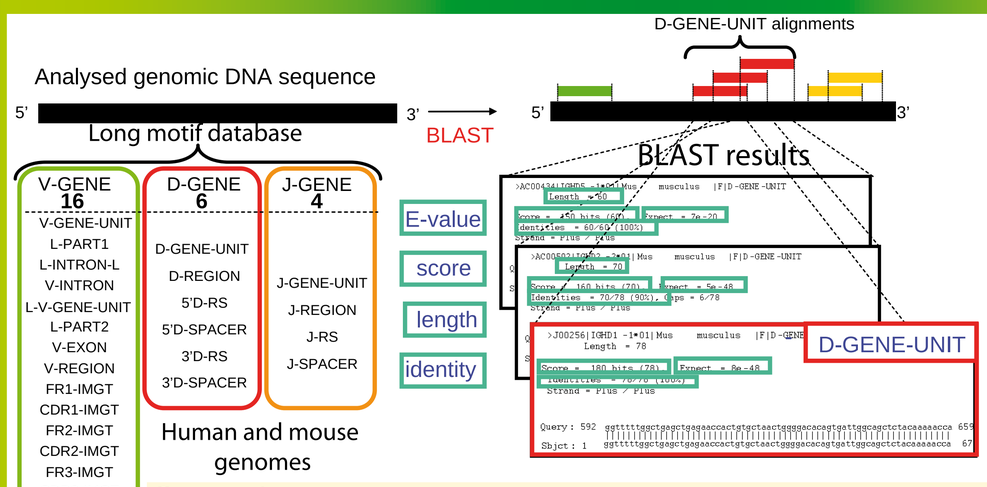
Results



IMGT/LIGMotif results for the 684,973 bp sequence of the human TRB locus identified 83 gene units (67 V, 2 D and 14 J). The screen shots show results of a portion of the TRB locus analysis (U66061).

Algorithm

1. Gene identification



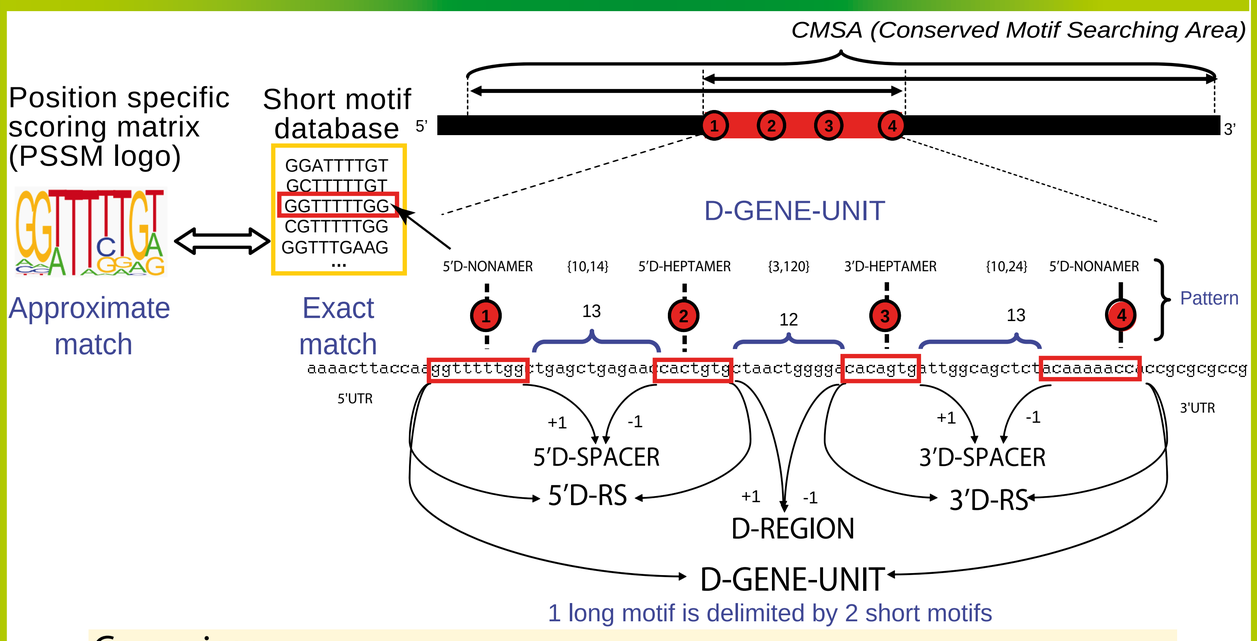
Comprises:

1- Search of long motifs by alignment using BLAST

2- Alignment selection

Alignments are selected according to BLAST parameters such as E-Value, score, length and identity. This allows to identify the gene type, its localization and orientation in the DNA sequence.

2. Gene description



Comprises:

1- Delimitation of Conserved Motif Search Area delimitation (CMSA)

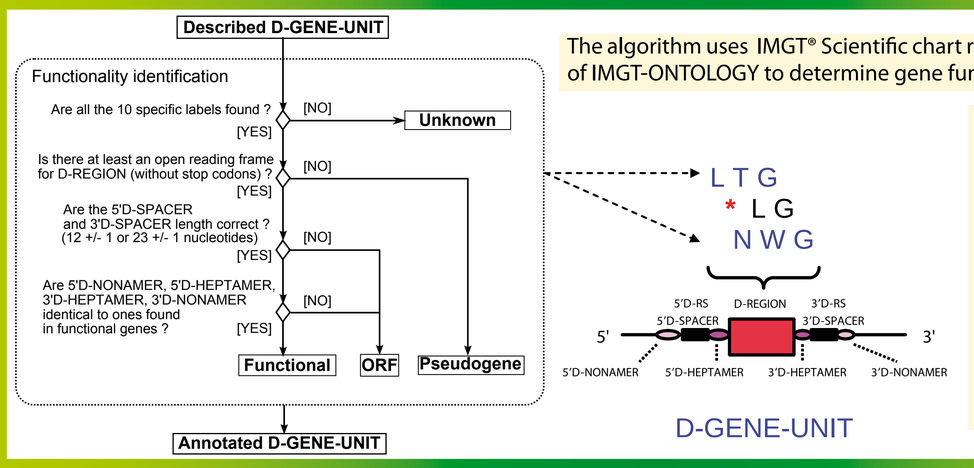
2- Search and association of short motifs

The short motifs from V, D and J gene databases are aligned in their search areas. The alignments obtained are associated to each other by comparing the pattern of the gene identified.

3- Long motif delimitation using the short motifs as anchors

Exception: V-REGION is delimited by the software IMGT/V-QUEST.

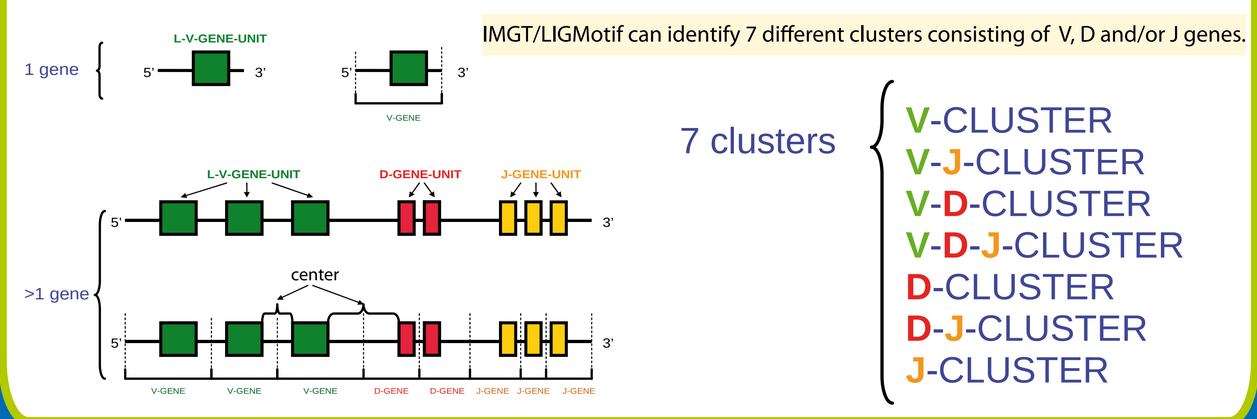
3. Functionality identification



The algorithm uses IMGT® Scientific chart rules based on the concepts of IMGT-ONTOLOGY to determine gene functionality.

Functionality «ORF», «Pseudogene» and «Functional» of genes are identified essentially by the number of labels, the presence of stop codons, frameshift and the recombination signal quality. A functionality is considered to be «Unknown» if a short motif is missing.

4. Gene delimitation and cluster assembly



IMGT/LIGMotif can identify 7 different clusters consisting of V, D and/or J genes.

7 clusters:

- V-CLUSTER
- V-J-CLUSTER
- V-D-CLUSTER
- V-D-J-CLUSTER
- D-CLUSTER
- D-J-CLUSTER
- J-CLUSTER

IMGT® founder and director: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr)
Bioinformatics manager: Véronique Giudicelli (Veronique.Giudicelli@igh.cnrs.fr)
Computer manager: Patrice Duroux (Patrice.Duroux@igh.cnrs.fr)
Webmaster: Chantal Ginestoux (Chantal.Ginestoux@igh.cnrs.fr)

© Copyright 1995-2010 IMGT®, the international ImMunoGeneTics information system®

©2010 Jérôme Lane