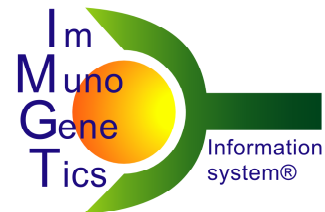


IMGT/LIGMotif : A tool to predict and describe germline vertebrate V-, D- and J-GENEs of immunoglobulins and T cell receptors



Jérôme Lane, Marie-Paule Lefranc and Patrice Duroux

IMGT® the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR 1142, Institut de Génétique Humaine, IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 05, France
Jerome.LANE@igh.cnrs.fr

<http://imgt.cines.fr>

Abstract

The aim of this work is to predict and describe the immunoglobulin (IG) and T cell receptor (TR) genes in genomic DNA of vertebrates. Indeed, owing to their DNA rearrangements the IG and TR variable, diversity and joining gene (V-, D- and J-GENE, respectively) have an unusual structure. Conventional bioinformatic software based on standard gene structure cannot identify them precisely. To solve this problem, a new tool named IMGT/LIGMotif, specific for IG and TR gene prediction and description has been developed. This tool allows genomic annotation at IMGT®, the international ImMunoGeneTics information system® [1]. The processing of the annotation is based on the DESCRIPTION concept established in IMGT-ONTOLOGY [2], the first ontology in the domain of immunogenetics and immunoinformatics. The implemented algorithm combines a similarity search (BLASTN [3]) with the matching of V-, D- and J-GENE patterns.

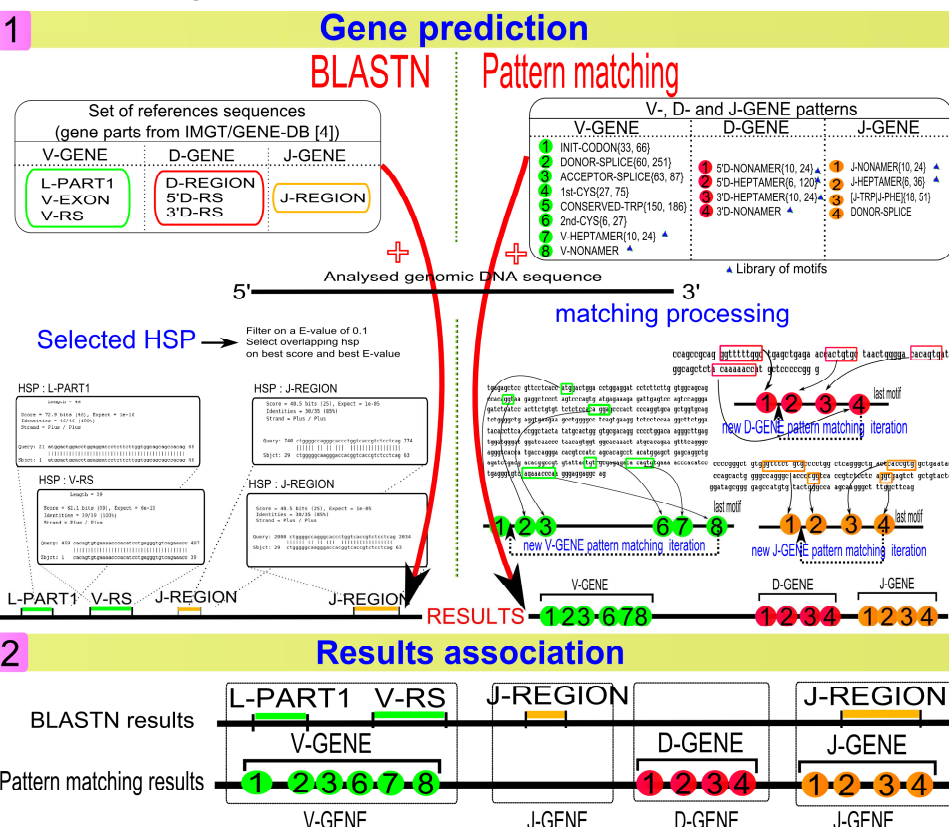
Objectives

- 1- To predict efficiently IG and TR germline vertebrate genes
- 2- To describe precisely genomic IG and TR
- 3- To facilitate genomic annotation

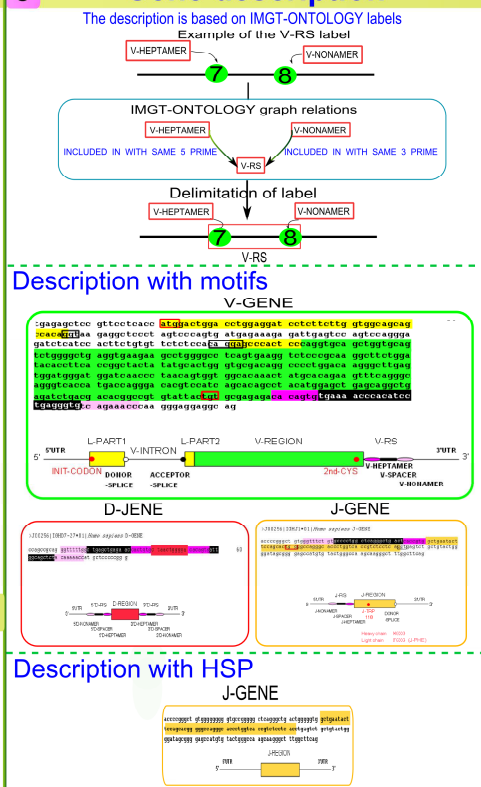
Approach

Combines a similarity search (BLASTN) for a rapid prediction with the matching of V-, D- and J-GENE patterns for validation and complete description of V-, D- and J-GENE.

Annotation algorithm



Gene description



Results

	IMGT/LIGMotif			GeneScanW			GeneMark		
	-	∅	+	-	∅	+	-	∅	+
V-GENE	2	0	11	0	3	7	0	1	8
D-GENE	0	5	23	0	28	0	0	28	0
J-GENE	1	2	7	0	18	0	0	17	1
Other	0	0	0	9	0	0	13	0	0
TOTAL	3	7	41	9	49	7	13	46	9

- : false predicted gene, ∅: false negative, + : true predicted gene

Results show that conventional gene finding software (e.g. GeneScanW[5] and GeneMark[6]) predict reasonably well V-GENE. The reason is that the structure of a V-GENE is the closest of a conventional gene. Indeed, V-GENE contains a common recognizable peptide signal and splicing sites. The major differences between conventional genes and V-, D- and J-GENEs is the presence of recombination signals and the shortness of the D- and J-REGIONS. IMGT/LIGMotif takes in account these additional characteristics which explains higher performance in gene prediction. Moreover IMGT/LIGMotif provides detailed standardized gene annotations.

Conclusion and perspectives

The combination of BLASTN and pattern matching in IMGT/LIGMotif allows the prediction of IG and TR genes in genomic sequences and the standardized description. This facilitates the annotation process by providing the delimitation of IMGT labels. The tool works on IG and TR loci of human and mouse. Preliminary work shows that the software works for IG and TR loci of other vertebrates species. Sets of patterns will be added accordingly to new genomic loci.

References

- [1] M.-P. Lefranc, IMGT, the international ImMunoGeneTics information system®: a standardized approach for immunogenetics and immunoinformatics, *Immunome Res.*, 1-3, (2005).
- [2] V. Giudicelli and M.-P. Lefranc, Ontology for immunogenetics: the IMGT-ONTOLOGY, *Bioinformatics* 15, 1047-1054, 1999.
- [3] S.F. Altschul, et al., Basic local alignment search tool, *J. Mol. Biol.*, 215:403-410, (1990).
- [4] V. Giudicelli, D. Chaume and M.-P. Lefranc, IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes *Nucl. Acids Res.*, 33, D256-D261, (2005).
- [5] C. Burge and S. Karlin, Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.* 288, 78-94, (1997).
- [6] A. Lomsadze et al., Gene identification in novel eukaryotic genomes by self-training algorithm, *Nucl. Acids Res.*, 33, 6949-6956, (2005).

©2007 J. Lane

