BIOINFORMATICS

Vol. 20 Suppl. 1 2004, pages i379–i385
DOI: 10.1093/bioinformatics/bth945

IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V–J and V–D–J JUNCTIONs

Mehdi Yousfi Monod[1], Véronique Giudicelli[1], Denys Chaume[1] and Marie-Paule Lefranc[1,2,]*

[1]Laboratoire d'ImmunoGénétique Moléculaire LIGM, Université Montpellier II, UPR CNRS 1142, Institut de Génétique Humaine IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France and [2]Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005 Paris, France

Received on January 15, 2004; accepted on March 1, 2004

ABSTRACT

Motivation: To create the enormous diversity of 10^{12} immunoglobulins (IG) and T cell receptors (TR) per individual, very complex mechanisms occur at the DNA level: the combinatorial diversity results from the junction of the variable (V), diversity (D) and joining (J) genes; the N-diversity represents the addition at random of nucleotides not encoded in the genome; and somatic hypermutations occur in IG rearranged sequences. The accurate annotation of the junction between V, D, J genes in rearranged IG and TR sequences represents therefore a huge challenge by its uniqueness and complexity. We developed IMGT/JunctionAnalysis to analyse automatically in detail the IG and TR junctions, according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY concepts.

Results: IMGT/JunctionAnalysis is the first tool for the detailed analysis of the IG and TR complex V–J and V–D–J JUNCTION(s). It delimits, at the nucleotide level, the genes resulting from the combinatorial diversity. It identifies accurately the D genes in the junctions of IG heavy (IGH), TR beta (TRB) and delta (TRD) chains. It delimits the palindromic P-REGION(s) and the N-REGION(s) resulting from the N-diversity. It evaluates the number of somatic hypermutations for each gene, within the JUNCTION. IMGT/JunctionAnalysis is capable of analysing, in a single run, an unlimited number of junctions from the same species (currently human or mouse) and from the same locus.

Availability: IMGT/JunctionAnalysis is available from the IMGT Home page at http://imgt.cines.fr

Contact: lefranc@ligm.igh.cnrs.fr

1 INTRODUCTION

The molecular synthesis of the immunoglobulins (IG) and T cell receptors (TR) is particularly complex and unique since it generates an important diversity within the IG and TR sequences. This sequence diversity determines the specificity of these antigen receptors. The repertoire of each individual is estimated to comprise 10^{12} different IG and TR, making possible to recognize any potential antigen. At least, three main mechanisms of sequence diversity have been identified during the IG and TR synthesis (for a review see Lefranc and Lefranc, 2001a,b):

The combinatorial diversity is created by the DNA rearrangements of the variable (V), diversity (D) and joining (J) genes for the IG heavy (IGH) chains and TR beta (TRB) and delta (TRD) chains, and of the V and J genes for the IG kappa (IGK) and lambda (IGL) light chains and TR alpha (TRA) and gamma (TRG) chains (Sakano *et al*., 1979).

The N-diversity (N for nucleotide) results from the addition of nucleotides (designated as N-REGION) at random by the terminal deoxynucleotidyl transferase (TdT, a template-independent DNA polymerase), at the junction of the V, D, J or V, J genes during the rearrangements (Landau *et al*., 1984). This addition is frequently preceded by the deletion of nucleotides at the 3′ end of the V gene, at the 5′ end of the J gene, and at both ends of the D genes which recombine. If there is no nucleotide deletion, very short inverted sequences, called P-REGION (for palindromic) can be observed at the V–(D)–J junctions (Lafaille *et al*., 1989). These nucleotides are adjacent to and complementary to intact coding ends of the rearranged genes.

The somatic hypermutations occur during the IG (but not TR) synthesis and specifically affect the rearranged V–J and V–D–J genes (Gearhart *et al*., 1981).

Therefore, the accurate analysis of the V–D–J and V–J gene junctions represents a difficult challenge and has major biological consequences. Indeed, these junctions correspond to the third hypervariable regions of the IG or TR variable domains (or complementarity determining region 3, or CDR3). Each antigen binding site is formed by two variable

*To whom correspondence should be addressed.

domains (from IGH and IGK or IGL chains, or TRA and TRB chains, or TRG and TRD chains). Structurally, the CDR3 of each V domain form a loop in close contact with the antigen, and, by their diversity, determine the IG and TR specificity. To our knowledge, automatic tools have never been designed to analyse in detail the IG or TR junctions. IMGT, the international ImMunoGeneTics information system® (Lefranc, 2003a,b; Lefranc *et al*., 2003a), created in 1989, by the Laboratoire d'ImmunoGénétique Moléculaire (LIGM) (Université Montpellier II and CNRS) at Montpellier, France, is a high-quality integrated knowledge resource specialized in IG, TR, major histocompatibility complex (MHC) and related proteins of the immune systems (RPI) of human and other vertebrate species (Giudicelli *et al*., 1997; Lefranc, 2001, 2003a,b, 2004; Lefranc *et al*., 1998, 1999; Ruiz *et al*., 2000). IMGT/LIGM-DB, the IMGT nucleotide sequence database, integrates and manages all the IG and TR nucleotide sequences from human and other vertebrate species that have been published in the generalist databases DDBJ/EMBL/GenBank (Benson *et al*., 2003; Miyazaki *et al*., 2003; Stoesser *et al*., 2003) (more than 81 000 entries in March 2004). Most of these IG and TR sequences (73%) are rearranged (genomic DNA or cDNA). IMGT/LIGM-DB provides high-quality annotations that follow the IMGT Scientific chart rules and are based on the IMGT-ONTOLOGY concepts (Giudicelli and Lefranc, 1999; Lefranc, 2004; Lefranc *et al*., 2003a). To deal with this high number of sequences, we developed an automatic tool, IMGT/JunctionAnalysis, which analyses the junction area between the variable, the diversity (for IGH chains and TRB and TRD chains), and the joining genes. This area is defined and labelled as 'JUNCTION' in IMGT. IMGT/JunctionAnalysis characterizes and describes the biological results of each diversity mechanism at the nucleotide level, according to the IMGT annotation rules. IMGT/JunctionAnalysis performs the following tasks: it delimits precisely the 3′ end of the V gene, the 5′ end of the J gene and, for the IGH, TRB and TRD genes, both ends of the D genes, taking into account the potential nucleotide deletions at their extremities. It identifies accurately the D genes and alleles in the IGH, TRB and TRD junctions. It delimits the palindromic P-REGION(s) (if they exist), and the N-REGION(s) resulting from the N-diversity mechanism. It evaluates the number of somatic hypermutations which occurred for the V, (D) and J genes, within the JUNCTION. IMGT/JunctionAnalysis is available on the Web from the IMGT Home page (http://imgt.cines.fr). IMGT/JunctionAnalysis is the first tool that analyses in detail the IG and TR V–J and V–D–J JUNCTIONs.

2 METHODS

2.1 IMGT definition of a JUNCTION

'JUNCTION' belongs to the 'DESCRIPTION' concept of IMGT-ONTOLOGY (Giudicelli and Lefranc, 1999). It describes, in IG and TR rearranged sequences, the feature that includes the 3′ end of a V-REGION (core coding region of a V gene), the D-REGION(s) (core coding region of a D gene), the P-REGION(s), the N-REGION(s), and the 5′ end of a J-REGION (core coding region of a J gene) (Fig. 1).

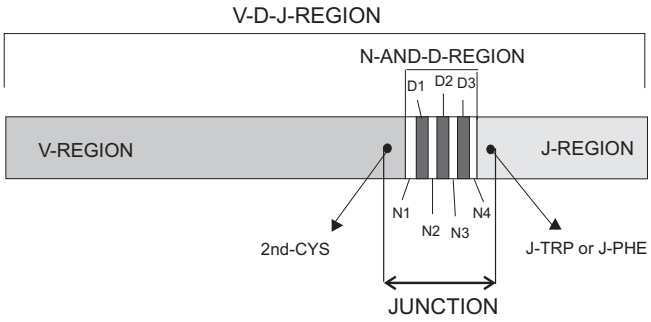

Fig. 1. Schematic representation of an IG V–D–J-REGION with its JUNCTION. A JUNCTION may be particularly complex with up to nine regions: 3′ end of V-REGION, D-REGIONs (D1, D2 or D3, if several), N-REGIONs (N1, N2, N3 or N4, if several), and 5′ end of J-REGION, to which can be added P-REGIONs.

A JUNCTION begins at the second conserved cysteine of the V-REGION (2nd-CYS), at position 104 according to the IMGT unique numbering (Lefranc *et al*., 2003b). A JUNCTION ends with the conserved tryptophan (J-TRP, for the IGH chains) or the conserved phenylalanine (J-PHE, for the IG light chains and the TR chains) at position 118 according to the IMGT unique numbering (Lefranc *et al*., 2003b).

2.2 Principles of search by IMGT/ JunctionAnalysis

IMGT/JunctionAnalysis performs a sequential analysis of a JUNCTION provided by the user. As the nucleotide sequences of the 3′ V-REGION and 5′ J-REGION are too short to be identified by the tool, IMGT/JunctionAnalysis requires the V and J gene and allele names. The gene and allele names need to be provided according to the IMGT nomenclature, which has been officially accepted by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) (Lefranc and Lefranc, 2001a,b). IMGT/JunctionAnalysis searches the constitutive regions of the JUNCTION by comparing the user sequence with the IMGT reference directory. The IMGT directory contains, for each functional and open reading frame (ORF) alleles of the IG and TR genes, the coding region, in nucleotides, of the corresponding reference sequences (V-REGIONs, D-REGIONs, J-REGIONs).

Order of the search for each region. Each nucleotide of the JUNCTION has to be precisely assigned to the most probable region (up to nine regions as shown in Fig. 1, to which can be added P-REGIONs). In a few cases, nucleotides at the limit of two contiguous regions can be assigned indifferently to one or the other region (the downstream end of the 5′ region,

or the upstream end of the 3′ region). In order to assign a nucleotide to the most probable region, a priority order was set up, based on the biological events, for the search of each region. The order for the search, from the first region to the last one, is the following: V-REGION, J-REGION, D-REGIONs, P-REGIONs and N-REGIONs. Once a region is delimited, the following search is carried out, excluding this region, on the remaining sequence.

Sequence comparison and allowed mutations. During the search of a given region, the user sequence is compared, nucleotide by nucleotide, with the sequence of the corresponding region in the IMGT reference directory. Because somatic mutations are known to occur in the IG rearranged V–J and V–D–J genes, we must accept differences between the user sequence and the IMGT reference sequence. The maximum number of allowed differences in the IG V-REGION, D-REGION or J-REGION was defined following an extensive analysis of human and mouse rearranged sequences available in IMGT/LIGM-DB. These numbers, indicated below for each region, were shown to provide, in most cases, the most probable interpretation of the biological event. The additional nucleotide difference found beyond the selected limit, determines the end of the regions. By contrast, no differences are allowed in the TR V-REGION, D-REGION or J-REGION since somatic mutations are not found in the TR rearranged sequences, and the first nucleotide difference found by IMGT/JunctionAnalysis determines the end of the regions. In the Web release, users are allowed to modify the default numbers of accepted mutations.

3 ALGORITHM

3.1 Search for the V-REGION in the user sequence: determination of the 3′ end

The names of the variable gene and allele of the user sequence are known (an IMGT/JunctionAnalysis prerequisite, see above), and the sequence is provided from the 2nd-CYS at position 104 (see above 'IMGT definition of a JUNCTION' section) (Lefranc *et al*., 2003b). The sequence is compared nucleotide by nucleotide, in a 5′ to 3′ orientation, to the V-REGION reference sequence, starting from 2nd-CYS. IMGT/JunctionAnalysis accepts up to two mutations for IGH, and four for IGK and IGL V-REGIONs. The next nucleotide difference found, or the end of the reference sequence, determines the 3′ end of the V-REGION in the user sequence (V-REGION end position). If this delimited V-REGION ends with consecutively mutated nucleotides, these are excluded. No mutations are accepted in the TR sequences, and the V-REGION stops at the first nucleotide difference found.

3.2 Search for the J-REGION in the user sequence: determination of the 5′ end

Once the V-REGION is delimited, IMGT/JunctionAnalysis searches the J-REGION 5′ end. The names of the joining gene and allele of the user sequence are known (an IMGT/JunctionAnalysis prerequisite, see above), and the sequence is provided up to J-TRP or J-PHE at position 118 (see above 'IMGT definition of a JUNCTION' section) (Lefranc *et al*., 2003b). The sequence is compared nucleotide by nucleotide, in a 3′ to 5′ orientation, to the J-REGION reference sequence, starting from J-TRP or J-PHE. As for V-REGION, IMGT/JunctionAnalysis accepts up to two mutations for IGH, and four for IGK and IGL J-REGIONs. The next nucleotide difference found, or the end of the reference sequence, determines the 5′ end of the J-REGION in the user sequence (J-REGION start position). No mutations are accepted in TR sequences, and the J-REGION stops at the first nucleotide difference found.

3.3 Search for D-REGION(s) in the user sequence: identification and delimitation

The identification, the number, the position and the extremities of the D-REGION(s), which are present in the user sequence, are unknown. Therefore, this third step represents the biggest challenge of IMGT/JunctionAnalysis, knowing that D-REGION(s) can be considerably shortened by deletions at both ends and, for IG sequences, heavily modified by somatic mutations. According to the species and the locus, the sequence must be compared to all possible D-REGION(s) from the IMGT reference directory. For each D-REGION, the comparison is carried out for all the nucleotides between the V-REGION and the J-REGION limits that were defined in the two previous steps (Fig. 2). These nucleotides occupy the positions between v and j where:

v = V-REGION end position +1;
j = J-REGION start position −1;

The method for the alignment of a given IMGT reference directory D-REGION with the user sequence is described in Figure 2, where sp is the start position of the search (nucleotide number of the user sequence to which the first 5′ nucleotide of the D-REGIONs from the IMGT reference directory should be aligned, and ep is the end position of the search (nucleotide number of the user sequence to which the last 3′ nucleotide of the D-REGIONs from the IMGT reference directory should be aligned).

Start and end positions of the search go over into the V-REGION and the J-REGION sequences respectively, as limited in the IMGT directory, since the V-REGION and the J-REGION of the user sequence may be deleted and the corresponding positions occupied by D nucleotides.
sp and ep are determined as follows:

$$\text{sp} = v - \text{dl} + 2; \quad \text{ep} = j + \text{dl} - 2,$$

where dl is the D-REGION reference sequence length.

The number 2 represents the minimum number of D-REGION nucleotides that should be compared with the user sequence (Fig. 2).

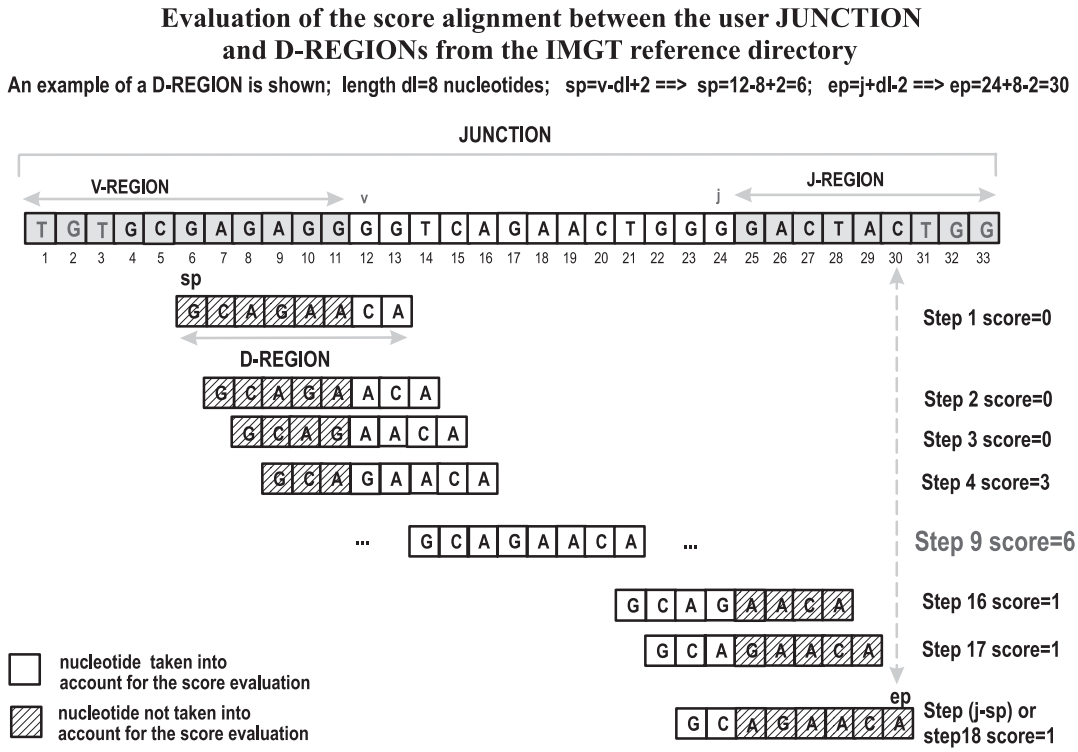

Fig. 2. Method for the alignment of a given D-REGION from the IMGT reference directory with the user JUNCTION, and score evaluation of the shown D-REGION. The JUNCTION extends from 2nd-CYS (encoded by TGT, left end) to J-TRP (encoded by TGG, right end). In the example, the best score is that of Step 9 with 6 nt being aligned between the D-REGION and the user JUNCTION. Each D-REGION of the IMGT reference directory is checked in the same way, and the D-REGION giving the highest score is kept.

For instance, in the example shown in Figure 2, the search will go from sp = 6 to ep = 30.

The D-REGION giving the highest score is kept and the positions giving that score are used to delimit the D-REGION in the user JUNCTION. It should be noted that successive ordered rearrangements of D genes may occur within the genome and the user sequences may therefore include two or three D genes (Alt *et al.*, 1984). Once a first D gene is found, the algorithm looks for a second one, and then a third one in the user sequence. These searches take into account data on the genomic D gene organization within the loci, and the potential chronological D–D rearrangements. Thus, each new search corresponds to genomic D genes located in a more downstream position within the locus.

3.4 Search for P-REGION(s) in the user sequence: identification and delimitation

For each identified and delimited V, D, and J gene in the user sequence, the algorithm looks for short palindromic P-REGION(s), which result from the opening of DNA hairpins at the extremity of intact V-REGION, D-REGION or J-REGION nucleotide sequences. The comparison is made between the germline sequences of the IMGT reference file and the user sequence. The analysis starts from the last nucleotide of an intact V-REGION, from the first or last nucleotide of an intact D-REGION and from the first nucleotide of an intact J-REGION and identifies if the neighbour nucleotide in the JUNCTION is complementary. If so, it identifies if the nucleotide before the last one in the V-REGION, D-REGION or J-REGION is complementary to the second nucleotide in the JUNCTION, and so on. The P-REGIONs are usually very short (1–3 nt).

3.5 Search for N-REGION(s) in the user sequence: identification and delimitation

N-REGIONs are the parts of the sequences, which remain after the delimitations of the V-, D-, J-REGIONs, and the identification of the P-REGIONs. When several N-REGIONs are found, they are numbered N1-REGION, N2-REGION, N3-REGION, N4-REGION (Fig. 1).

3.6 Translation of JUNCTIONs and amino acid numbering

The resulting Web page of IMGT/JunctionAnalysis displays the translation of the JUNCTION into amino acids. The amino acids are numbered according to the IMGT unique numbering for V-DOMAIN (Lefranc *et al.*, 2003b). If many JUNCTIONs have been submitted to IMGT/JunctionAnalysis in a

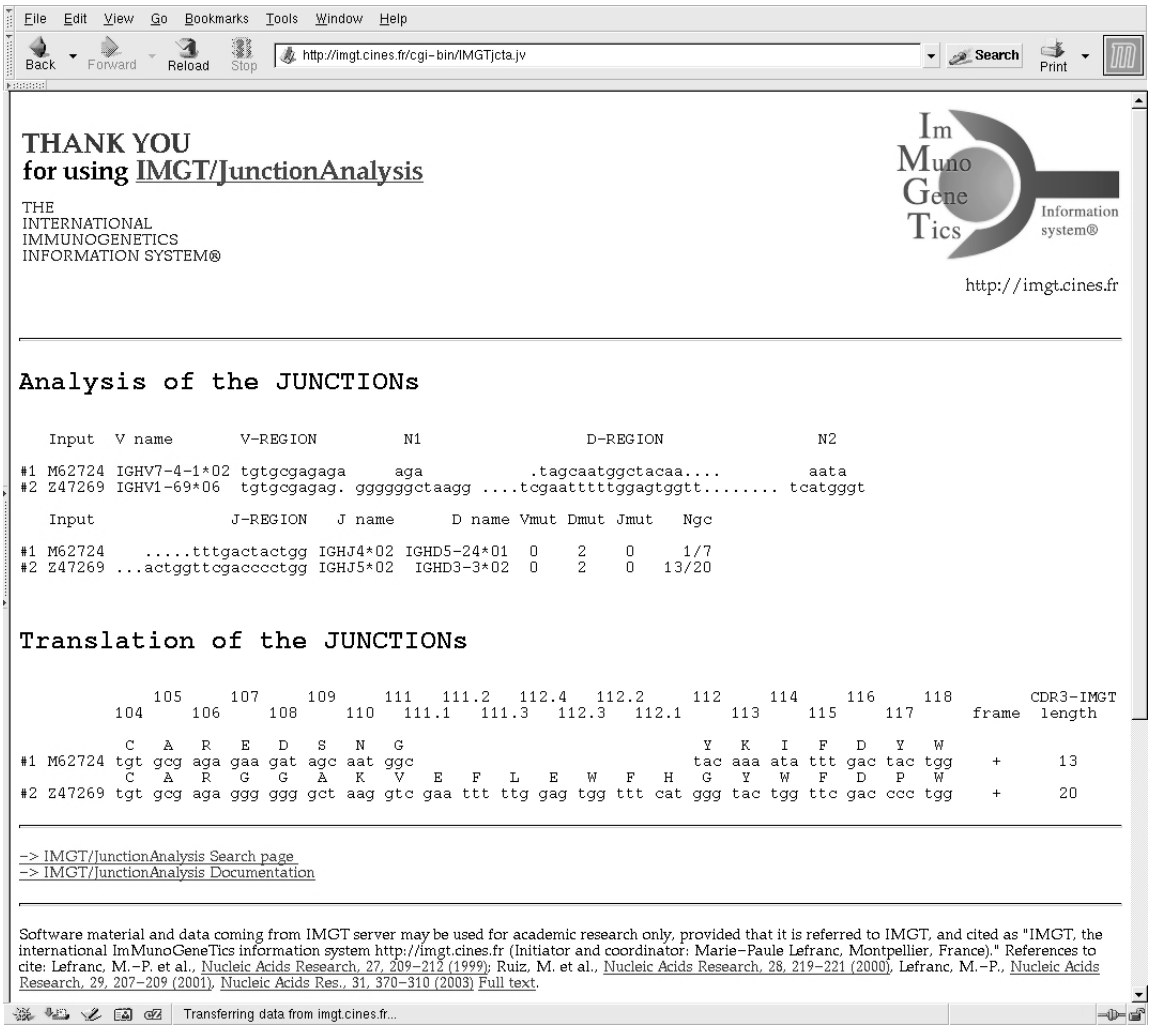

Fig. 3. An example of IMGT/JunctionAnalysis results. The two human IGH JUNCTIONs were analysed in a single run.

single run, their translation may have a different length: the translation of a given JUNCTION is adjusted to the longest one by inserting gaps according to the IMGT unique numbering for V-DOMAIN (Lefranc *et al*., 2003b): the IMGT unique numbering is based on a JUNCTION length of 15 amino acids (from 2nd-CYS 104 to J-PHE or J-TRP 118). For a JUNCTION longer than 15, amino acid position numbers are added between 111 and 112, starting by 112.1, then 111.1, 112.2, 111.2, and so on. For a JUNCTION shorter than 15 amino acids, numbers are deleted, starting by 111, then 112, 110, 113, and so on (Lefranc *et al*., 2003b).

4 RESULTS

IMGT/JunctionAnalysis is the first tool that analyses in detail the IG and TR V–J and V–D–J JUNCTIONs. IMGT/JunctionAnalysis was developed using Java 2 and cgi-bin programs. It is available on the IMGT server at http://imgt.cines.fr. IMGT/JunctionAnalysis is capable of analysing, in a single run, an unlimited number of junctions from a same species and from a same locus. The current Web release deals with human and mouse JUNCTIONs. Because the nucleotide sequences of the JUNCTIONs are relatively short, IMGT/JunctionAnalysis requires the variable and the joining gene and allele names. The program displays the results of the following tasks: it identifies accurately D-GENE and ALLELE in the IGH, TRB and TRD junctions. It delimits precisely the different regions of the junctions: V-REGION, D-REGION(s), J-REGION, as well as N-REGION(s) and P-REGION(s). It determines the number of mutations in the 3′ V-REGION, D-REGION and 5′ J-REGION of the IG JUNCTIONs and calculates the ‘gc’ content of the N-REGIONs of the IG and TR JUNCTIONs. The results produced by IMGT/JunctionAnalysis can be either displayed or downloaded into a local file. The user can also select the maximum number of characters per line to be displayed. Figure 3 shows the results following the analysis of two JUNCTIONs. In ‘Analysis of the JUNCTIONs’, nucleotides of each region identified in a JUNCTION are

displayed. Dots in V-REGION, D-REGION and J-REGION indicate nucleotides that are deleted in the user sequences, by comparison to the corresponding germline reference sequence of the IMGT directory. N1 and N2 indicate N-REGIONs.

The information provided by the user comprises the input (here an accession number), V name (gene and allele) and J gene (gene and allele). The results of IMGT/JunctionAnalysis comprise the following: D name, name of the D-GENE and ALLELE; Vmut, number of mutations in the V-REGION of the input sequence, from 2nd-CYS to the 3′ end, by comparison to the corresponding IMGT reference allele sequence; Dmut, number of mutations in the D-REGION sequence identified by the IMGT/JunctionAnalysis tool, by comparison to the corresponding IMGT reference allele sequence; Jmut, number of mutations in the input sequence from the J-REGION 5′ end to J-PHE or J-TRP, compared with the corresponding IMGT reference allele sequence; and Ngc, ratio of the number of g + c nucleotides to the total number of N region nucleotides.

In 'Translation of the JUNCTIONs' (Fig. 3), each junction nucleotide sequence is translated into amino acid sequences. In the case of frameshifts, gaps, indicated by one or two dots, are inserted to maintain the J-REGION reading frame and to facilitate sequence comparison. Codons and amino acids are numbered according to the IMGT unique numbering for the V-DOMAINs (Lefranc *et al*., 2003b). If several JUNCTIONs are analysed in a single run, the order of the translation results can be displayed, depending from the user choice, either according to the input order of the sequences or based on the CDR3-IMGT length. Standardized criteria for the statistical analysis of immunoglobulin and T cell receptor V-REGION amino acid properties were recently defined by IMGT (Pommié *et al*., 2004). These criteria can be applied easily to the study of the translated JUNCTIONs.

5 DISCUSSION

IMGT/JunctionAnalysis is the first tool that automatically analyses in detail the V–J and V–D–J JUNCTIONs from IG and TR nucleotide rearranged sequences. This represents a huge challenge since JUNCTIONs result from very complex mechanisms, which generate the antigen receptor diversity.

IMGT, the international ImMunoGeneTics information system®, has developed IMGT/V-QUEST, a Web tool that analyses whole IG or TR rearranged sequences and identifies the V, D and J genes (Lefranc, 2003c,d, 2004; Giudicelli *et al*., 2004). We compared the IMGT/JunctionAnalysis results with those provided by IMGT/V-QUEST.

The IMGT/JunctionAnalysis is by far a more accurate tool for the D gene and allele identification and delimitation. However, IMGT/V-QUEST has the advantage of proposing several solutions, which can be useful in some cases.

The way IMGT/V-QUEST and IMGT/JunctionAnalysis identify the D genes is not identical, therefore the scores can be compared for a given tool, but score differences may be observed between the tools.

For two D genes with an identical score in the IMGT/V-QUEST results, IMGT/JunctionAnalysis, in its default configuration, selects the solution which gives the smallest N regions, or, in other terms, prefers a longer D (accepting nucleotide differences) to a shorter D (without nucleotide differences).

IMGT/JunctionAnalysis is currently integrated in IMGT/V-QUEST Web release, as an input option. It allows the user to get in a single tool the identification of the V and J gene and allele names of IG and TR rearranged sequence, and the detailed analysis of the JUNCTION according to IMGT/JunctionAnalysis. Owing to the accuracy of the results obtained by the combination of both tools, we also combined them to automate the annotation of the IG and TR cDNA sequences of IMGT/LIGM-DB (Giudicelli *et al*., 2003).

IMGT/JunctionAnalysis is currently available for the analysis of human and mouse JUNCTIONs since their respective IG and TR loci have been extensively studied in IMGT. Other species will be added when their genomic IG and TR repertoire will be sequenced and characterized. However, in the Web release of IMGT/JunctionAnalysis, users can modify the default values for the numbers of accepted mutations in the user sequences. This allows the biologists to use IMGT/JunctionAnalysis for the JUNCTION analysis of other species: satisfying results have been obtained, in particular, for primates (data not shown).

To our knowledge, IMGT/JunctionAnalysis is the first tool, available on the Web that is capable of displaying such detailed and accurate analysis of IG and TR junctions. The standalone release of IMGT/JunctionAnalysis is available at http://imgt.cines.fr.

ACKNOWLEDGEMENTS

We thank Gérard Lefranc for helpful discussion. We are deeply grateful to the IMGT team for its expertise and constant motivation. IMGT is funded by the European Union's 5th PCRDT programme (QLG2-2000-01287), the Centre National de la Recherche Scientifique (CNRS) and the Ministère de la Recherche et de l'Education Nationale.

REFERENCES

Alt,F.W., Yancopoulos,G.D., Blackwell,T.K., Wood,C., Thomas,E., Boss,M., Coffman,R., Rosenberg,N., Tonegawa,S. and Baltimore,D. (1984) Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.*, **3**, 1209–1219.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2003) GenBank. *Nucleic Acids Res.*, **31**, 23–27.

Gearhart,P.J., Johnson,N.D., Douglas,R. and Hood,L. (1981) IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature*, **291**, 29–34.

Giudicelli,V., Chaume,D., Bodmer,J., Muller,W., Busin,C., Marsh,S., Bontrop,R., Lemaitre,M., Malik,A. and Lefranc, M.-P.

(1997) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, **25**, 206–211.

Giudicelli,V. and Lefranc,M.-P. (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics*, **15**, 1047–1054.

Giudicelli,V., Protat,C. and Lefranc,M.-P. (2003) The IMGT strategy for the automatic annotation of IG and TR cDNA sequences: IMGT/Automat. Proceedings of the European Conference on Computational Biology (ECCB'2003) Paris, France, 27–30 September, INRIA (DISC/Spid), DKB-31, pp. 103–104.

Giudicelli,V., Chaume, and Lefranc,M.-P. (2004) IMGT/V-QUEST, an integrated software for immunoglobulin and T cell receptor V–J and V–D–J rearrangement analysis. *Nucleic Acids Res.* (in press).

Lafaille,J.J., DeCloux,A., Bonneville,M., Takagaki,Y. and Tonegawa,S. (1989) Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V–(D)–J joining. *Cell*, **59**, 859–870.

Landau,N.R., St John,T.P., Weissman,I.L., Wolf,S.C., Silverstone,A.E. and Baltimore,D. (1984) Cloning of terminal transferase cDNA by antibody screening. *Proc. Natl Acad. Sci., USA*, **81**, 5836–5840.

Lefranc,M.-P. (2001) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, **29**, 207–209.

Lefranc,M.-P. (2003a) IMGT® databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, http://imgt.cines.fr. *Leukemia*, **17**, 260–266.

Lefranc,M.-P. (2003b) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, **31**, 307–310.

Lefranc,M.-P. (2003c) IMGT, the international ImMunoGeneTics information system®, http://imgt.cines.fr. In: Bock,G. and Goode,J. (ed.), Immunoinformatics: Bioinformatics Strategies for Better Understanding of Immune Function. *Novartis Foundation Symposium 254*. John Wiley and sons, Chichester, UK, pp. 126–142.

Lefranc,M.-P. (2003d) IMGT, the international ImMunoGeneTics information system®, http://imgt.cines.fr. In: Lo,B.K.C. (ed.), Antibody Engineering Methods and Protocols, 2nd edition, *Methods in Molecular Biology*. Humana Press, Totowa, NJ, Vol. 248, Chap. 3, pp. 27–49.

Lefranc,M.-P. (2004) IMGT-ONTOLOGY and IMGT databases, tools and web resources for immunogenetics and immunoinformatics. *Mol. Immunol.*, **40**, 647–659.

Lefranc,M.-P. and Lefranc,G. (2001a) *The Immunoglobulin FactsBook*. Academic Press, London, UK, 458 pp., ISBN:012441351X.

Lefranc,M.-P. and Lefranc,G. (2001b) *The T cell Receptor FactsBook*. Academic Press, London, UK, 398 pp., ISBN:0124413528.

Lefranc,M.-P., Giudicelli,V., Busin,C., Bodmer,J., Muller,W., Bontrop,R., Lemaitre,M., Malik,A. and Chaume,D. (1998) IMGT, the International ImMunoGeneTics database. *Nucleic Acids Res.*, **26**, 297–303.

Lefranc,M.-P., Giudicelli,V., Ginestoux,C., Bodmer,J., Muller,W., Bontrop,R., Lemaitre,M., Malik,A., Barbié, V. and Chaume,D. (1999) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, **27**, 209–212.

Lefranc,M.-P., Giudicelli,V., Ginestoux,C., Bosc,N., Folch,G., Guiraudou,D., Jabado-Michaloud,J., Magris,S., Scaviner,D., Thouvenin,V., Combres,K., Girod,D. *et al*. (2003a) IMGT-ONTOLOGY for immunogenetics and immunoinformatics (http://imgt.cines.fr). *In Silico Biol.*, **4**, 0004, http://www.bioinfo.de/isb/2003/04/2004.

Lefranc,M.-P., Pommié,C., Ruiz,M., Giudicelli,V., Foulquier,E., Truong,L., Thouvenin-Contet,V. and Lefranc,G. (2003b) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, **27**, 55–77.

Miyazaki,S., Sugawara,H., Gojobori,T. and Tateno,Y. (2003) DNA Data Bank of Japan (DDBJ) in XML. *Nucleic Acids Res.*, **31**, 13–16.

Pommié,C., Sabatier,R., Lefranc,G. and Lefranc,M.-P. (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J. Mol. Recognit.*, **17**, 17–32.

Ruiz,M., Giudicelli,V., Ginestoux,C., Stoehr,P., Robinson,J., Bodmer,J., Marsh,S.G., Bontrop,R., Lemaitre,M., Lefranc,G., Chaume,D. and Lefranc,M.-P. (2000) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, **28**, 219–221.

Sakano,H., Huppi,K., Heinrich,G. and Tonegawa,S. (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature*, **280**, 288–294.

Stoesser,G., Baker,W., van den Broek,A., Garcia-Pastor,M., Kanz,C., Kulikova,T., Leinonen,R., Lin,Q., Lombard,V. and Lopez,R. (2003) The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res.*, **31**, 17–22.