http://www.imgt.org

Systèmes d'information
-
Ontologies
Bases de connaissances

UE FMBS312 Bioinformatique (TC3 du MR2 BioMed)

Véronique Giudicelli
LIGM, IGH, UPR CNRS 1142
141 rue de la Cardonille
34296 Montpellier Cedex 5
tel: 04.99.61.99.65 fax: 04.99.61.99.01
e-mail: Veronique.Giudicelli@igh.cnrs.fr

27 septembre 2010

Système d'information

Ensemble des moyens

organisation (CNRS,…),
acteurs,
systèmes informatiques

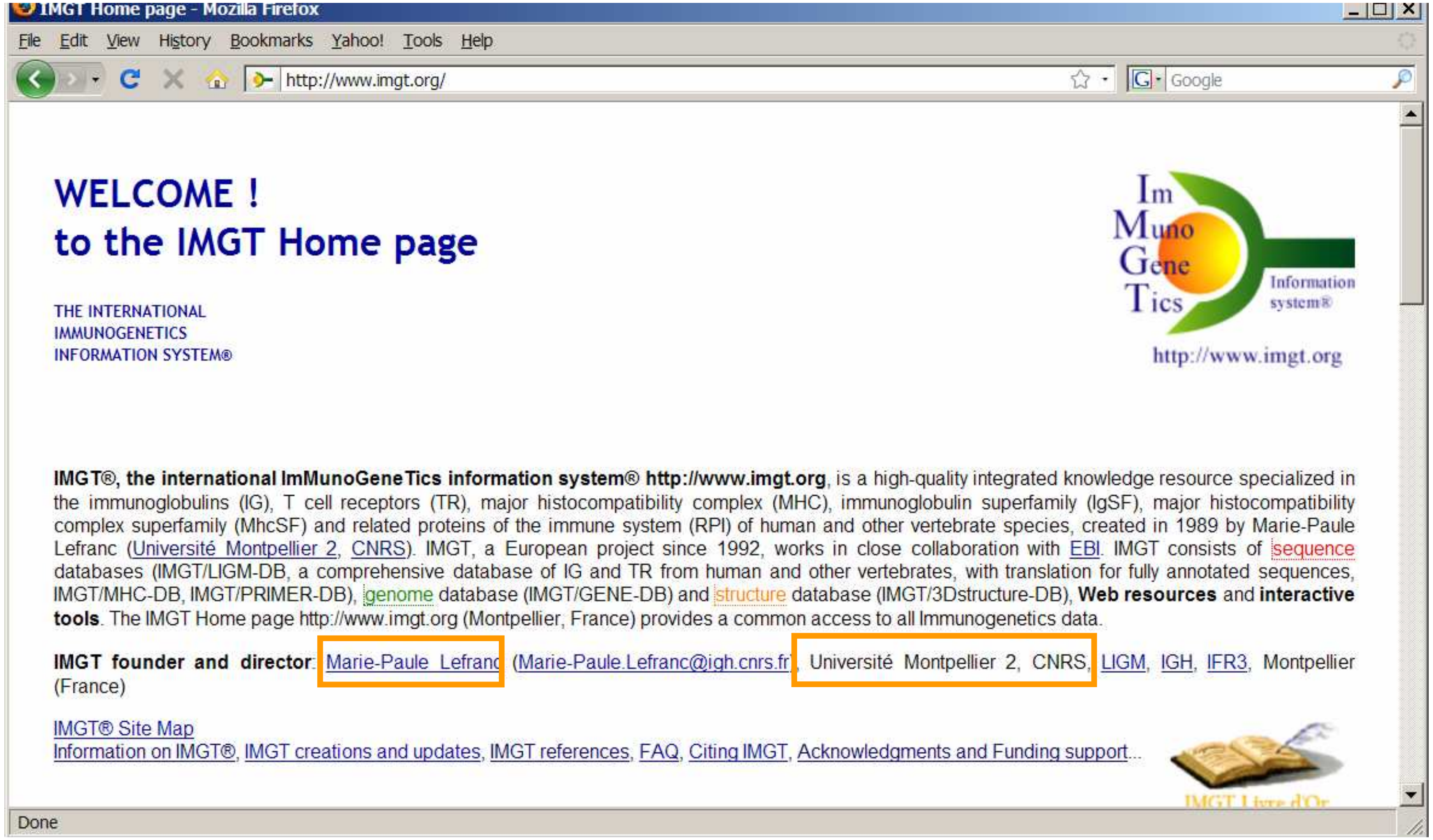
.

nécessaires au traitement et à l'exploitation des informations dans le cadre d'objectifs définis

IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org)

Source unique de connaissances en immunogénétique et immunoinformatique, au niveau international.

IMGT® est spécialisé dans les séquences, structures et données génétiques des immunoglobulines (IG), des récepteurs T (TR), du complexe majeur d'histocompatibilité (MHC) des vertébrés, des protéines des superfamilles IgSF et MhcSF, et des protéines apparentées du système immunitaire.



WELCOME !
to the IMGT Home page

THE INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

IMGT®, the international ImMunoGeneTics information system® http://www.imgt.org, is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC), immunoglobulin superfamily (IgSF), major histocompatibility complex superfamily (MhcSF) and related proteins of the immune system (RPI) of human and other vertebrate species, created in 1989 by Marie-Paule Lefranc (Université Montpellier 2, CNRS). IMGT, a European project since 1992, works in close collaboration with EBI. IMGT consists of sequence databases (IMGT/LIGM-DB, a comprehensive database of IG and TR from human and other vertebrates, with translation for fully annotated sequences, IMGT/MHC-DB, IMGT/PRIMER-DB), genome database (IMGT/GENE-DB) and structure database (IMGT/3Dstructure-DB), **Web resources** and **interactive tools**. The IMGT Home page http://www.imgt.org (Montpellier, France) provides a common access to all Immunogenetics data.

IMGT founder and director: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr), Université Montpellier 2, CNRS, LIGM, IGH, IFR3, Montpellier (France)

IMGT® Site Map
Information on IMGT®, IMGT creations and updates, IMGT references, FAQ, Citing IMGT, Acknowledgments and Funding support...

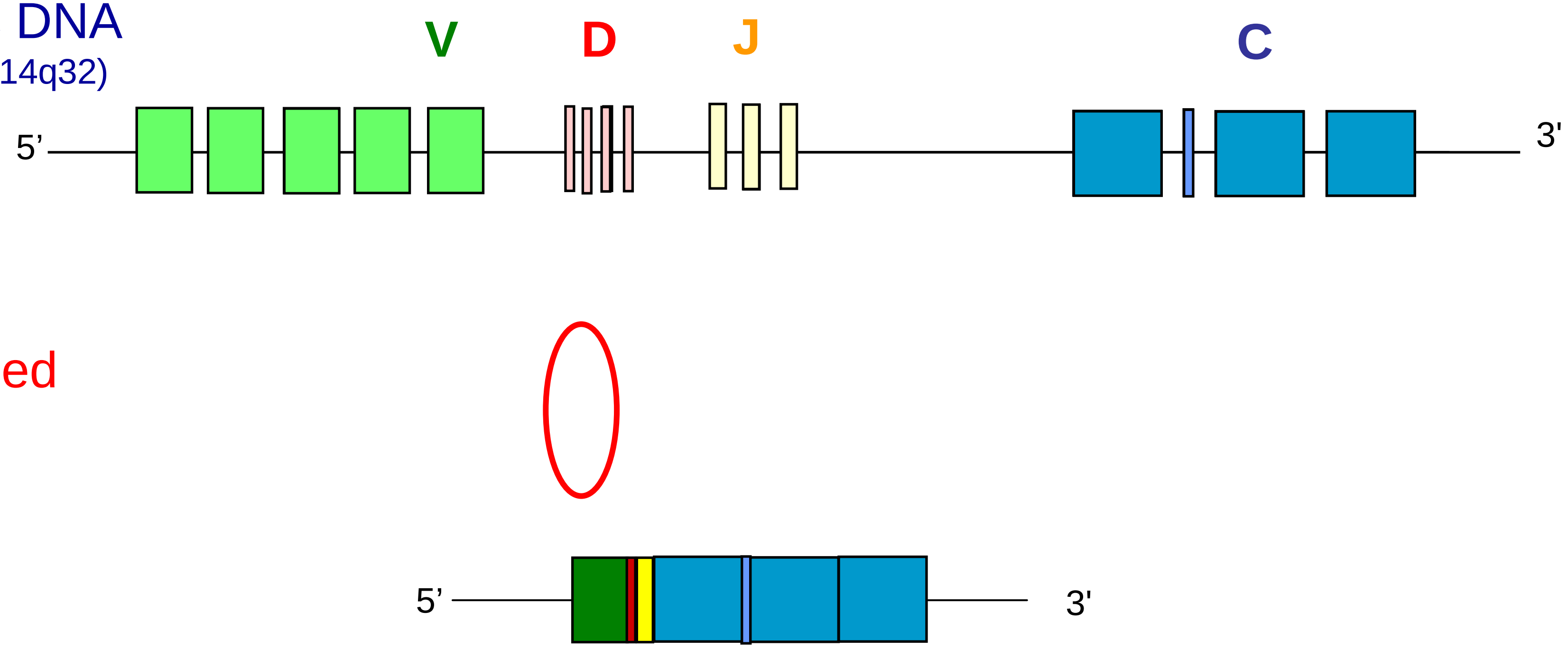
Le contexte biologique:
la synthèse des IG et des TR

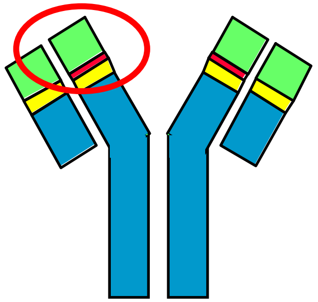
http://www.imgt.org

genomic DNA
(IGH Locus 14q32)

rearranged
DNA

mRNA





Potentiellement 2 x 10 ^{12}
differentes IG
Par individu

Diversité des junctions :
exemple d'une chaîne lourde d'IG

http://www.imgt.org

3'V-REGION N1-REGION **D-REGION** N2-REGION **5'J-REGION**

tgtgcgaaaga tacc agcatattgtggtggtgactgctattcc gat acaactggttcgacccc**tgg**

2nd-CYS

JUNCTION

J-TRP

C A P Y R G D T Y D Y S W

tgt gcg cca tac cgg ggt gac act tat gat tac tcc tgg

Système informatique (SI)

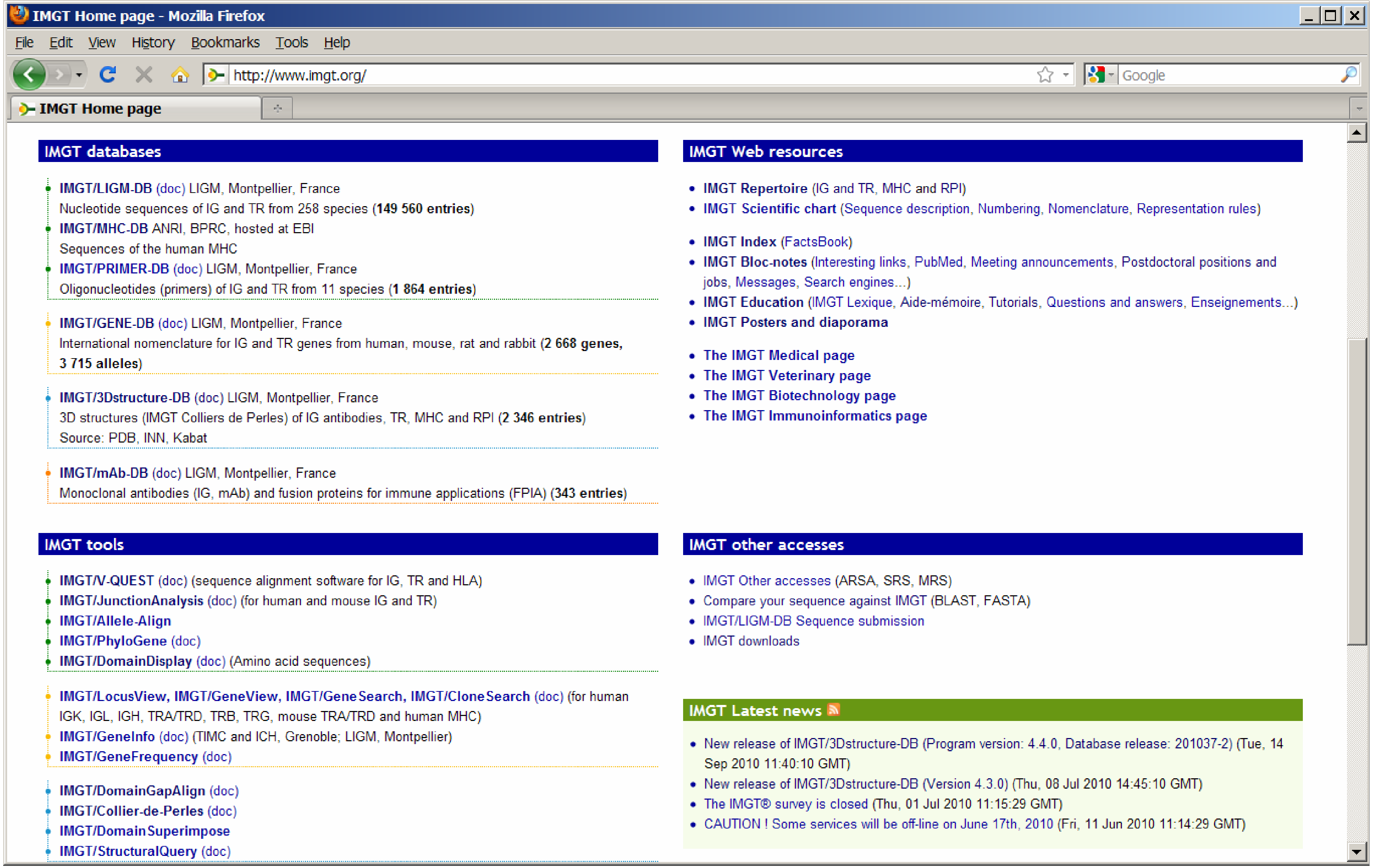
Partie automatisée d'un système d'information.

Le **système informatique** regroupe :

- Le matériel physique et câblage (PC, réseau,….)
- Le réseau (adresse IP, noms, …)
- Le(s) système(s) d'exploitation (Windows, Linux, …)
- Sauvegardes

- **Les applications :**
  bases de données,
  logiciels de développement,
  interfaces d'exploitation, …

http://www.imgt.org



IMGT databases

- **IMGT/LIGM-DB** (doc) LIGM, Montpellier, France
  Nucleotide sequences of IG and TR from 258 species (**149 560 entries**)
- **IMGT/MHC-DB** ANRI, BPRC, hosted at EBI
  Sequences of the human MHC
- **IMGT/PRIMER-DB** (doc) LIGM, Montpellier, France
  Oligonucleotides (primers) of IG and TR from 11 species (**1 864 entries**)

- **IMGT/GENE-DB** (doc) LIGM, Montpellier, France
  International nomenclature for IG and TR genes from human, mouse, rat and rabbit (**2 668 genes, 3 715 alleles**)

- **IMGT/3Dstructure-DB** (doc) LIGM, Montpellier, France
  3D structures (IMGT Colliers de Perles) of IG antibodies, TR, MHC and RPI (**2 346 entries**)
  Source: PDB, INN, Kabat

- **IMGT/mAb-DB** (doc) LIGM, Montpellier, France
  Monoclonal antibodies (IG, mAb) and fusion proteins for immune applications (FPIA) (**343 entries**)

IMGT Web resources

- **IMGT Repertoire** (IG and TR, MHC and RPI)
- **IMGT Scientific chart** (Sequence description, Numbering, Nomenclature, Representation rules)

- **IMGT Index** (FactsBook)
- **IMGT Bloc-notes** (Interesting links, PubMed, Meeting announcements, Postdoctoral positions and jobs, Messages, Search engines...)
- **IMGT Education** (IMGT Lexique, Aide-mémoire, Tutorials, Questions and answers, Enseignements...)
- **IMGT Posters and diaporama**

- **The IMGT Medical page**
- **The IMGT Veterinary page**
- **The IMGT Biotechnology page**
- **The IMGT Immunoinformatics page**

IMGT tools

- **IMGT/V-QUEST** (doc) (sequence alignment software for IG, TR and HLA)
- **IMGT/JunctionAnalysis** (doc) (for human and mouse IG and TR)
- **IMGT/Allele-Align**
- **IMGT/PhyloGene** (doc)
- **IMGT/DomainDisplay** (doc) (Amino acid sequences)

- **IMGT/LocusView, IMGT/GeneView, IMGT/GeneSearch, IMGT/CloneSearch** (doc) (for human IGK, IGL, IGH, TRA/TRD, TRB, TRG, mouse TRA/TRD and human MHC)
- **IMGT/GeneInfo** (doc) (TIMC and ICH, Grenoble; LIGM, Montpellier)
- **IMGT/GeneFrequency** (doc)

- **IMGT/DomainGapAlign** (doc)
- **IMGT/Collier-de-Perles** (doc)
- **IMGT/DomainSuperimpose**
- **IMGT/StructuralQuery** (doc)

IMGT other accesses

- IMGT Other accesses (ARSA, SRS, MRS)
- Compare your sequence against IMGT (BLAST, FASTA)
- IMGT/LIGM-DB Sequence submission
- IMGT downloads

IMGT Latest news

- New release of IMGT/3Dstructure-DB (Program version: 4.4.0, Database release: 201037-2) (Tue, 14 Sep 2010 11:40:10 GMT)
- New release of IMGT/3Dstructure-DB (Version 4.3.0) (Thu, 08 Jul 2010 14:45:10 GMT)
- The IMGT® survey is closed (Thu, 01 Jul 2010 11:15:29 GMT)
- CAUTION ! Some services will be off-line on June 17th, 2010 (Fri, 11 Jun 2010 11:14:29 GMT)

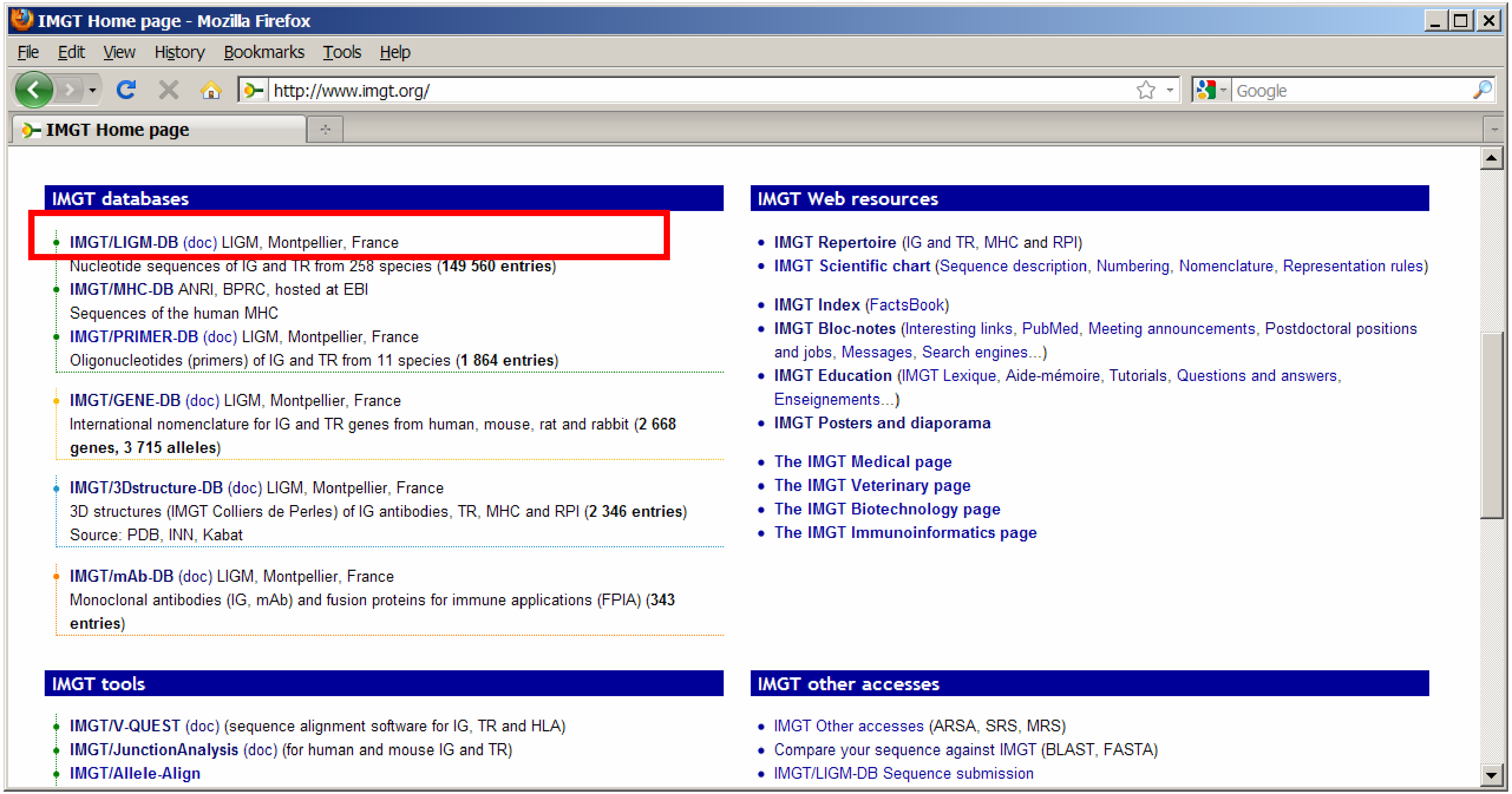
Une application informatique distribuée

utilise plusieurs machines

L'architecture d'une application distribuée est donnée par:

- la description des machines utilisées,
- les fonctions de ces machines pour l'application:
  - fonction stockage des données
  - fonction traitements des données
  - fonction interface utilisateur

http://www.imgt.org



IMGT databases

- **IMGT/LIGM-DB** (doc) LIGM, Montpellier, France
  Nucleotide sequences of IG and TR from 258 species (**149 560 entries**)
- **IMGT/MHC-DB** ANRI, BPRC, hosted at EBI
  Sequences of the human MHC
- **IMGT/PRIMER-DB** (doc) LIGM, Montpellier, France
  Oligonucleotides (primers) of IG and TR from 11 species (**1 864 entries**)
- **IMGT/GENE-DB** (doc) LIGM, Montpellier, France
  International nomenclature for IG and TR genes from human, mouse, rat and rabbit (**2 668 genes, 3 715 alleles**)
- **IMGT/3Dstructure-DB** (doc) LIGM, Montpellier, France
  3D structures (IMGT Colliers de Perles) of IG antibodies, TR, MHC and RPI (**2 346 entries**)
  Source: PDB, INN, Kabat
- **IMGT/mAb-DB** (doc) LIGM, Montpellier, France
  Monoclonal antibodies (IG, mAb) and fusion proteins for immune applications (FPIA) (**343 entries**)

IMGT tools

- **IMGT/V-QUEST** (doc) (sequence alignment software for IG, TR and HLA)
- **IMGT/JunctionAnalysis** (doc) (for human and mouse IG and TR)
- **IMGT/Allele-Align**

IMGT Web resources

- **IMGT Repertoire** (IG and TR, MHC and RPI)
- **IMGT Scientific chart** (Sequence description, Numbering, Nomenclature, Representation rules)
- **IMGT Index** (FactsBook)
- **IMGT Bloc-notes** (Interesting links, PubMed, Meeting announcements, Postdoctoral positions and jobs, Messages, Search engines...)
- **IMGT Education** (IMGT Lexique, Aide-mémoire, Tutorials, Questions and answers, Enseignements...)
- **IMGT Posters and diaporama**
- **The IMGT Medical page**
- **The IMGT Veterinary page**
- **The IMGT Biotechnology page**
- **The IMGT Immunoinformatics page**

IMGT other accesses

- IMGT Other accesses (ARSA, SRS, MRS)
- Compare your sequence against IMGT (BLAST, FASTA)
- IMGT/LIGM-DB Sequence submission

Les objectifs de IMGT/LIGM-DB

http://www.imgt.org

- Contenir toutes les séquences IG et TR des bases EMBL/GenBank/DDBJ

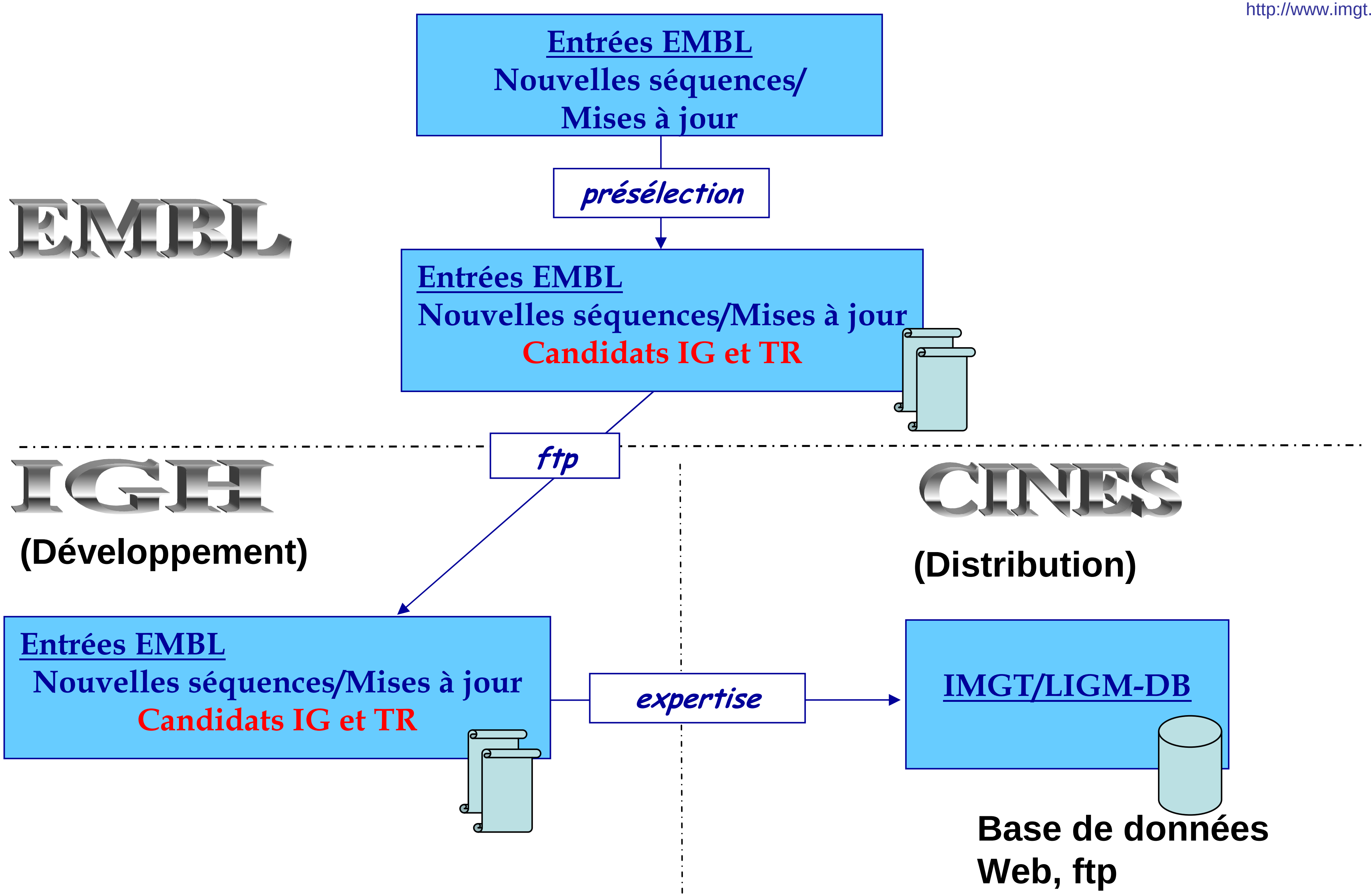
  \+ de 149.500 séquences de 235 espèces

- Expertiser les séquences en fonction des connaissances en immunogénétique
  - identification des gènes,
  - description des motifs (régions codantes, ...)

- Suivi et intégration des nouvelles connaissances dans le système d'information IMGT®

IMGT/LIGM-DB: application distribuée

http://www.imgt.org



Entrées EMBL
Nouvelles séquences/
Mises à jour

présélection

EMBL

Entrées EMBL
Nouvelles séquences/Mises à jour
Candidats IG et TR

ftp

IGH

(Développement)

CINES

(Distribution)

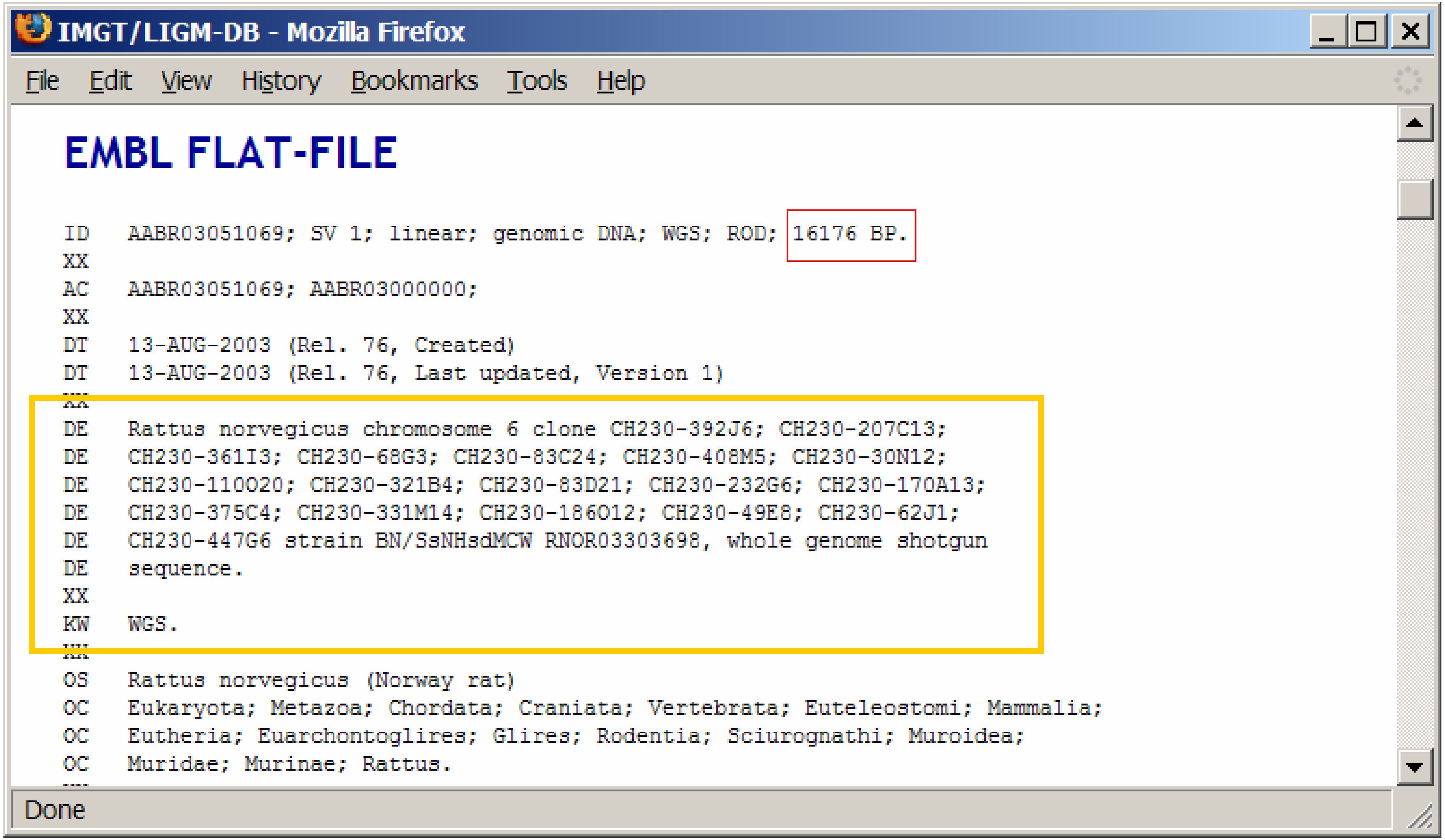
Entrées EMBL
Nouvelles séquences/Mises à jour
Candidats IG et TR

expertise

IMGT/LIGM-DB

Base de données
Web, ftp

A l’arrivée dans IMGT®

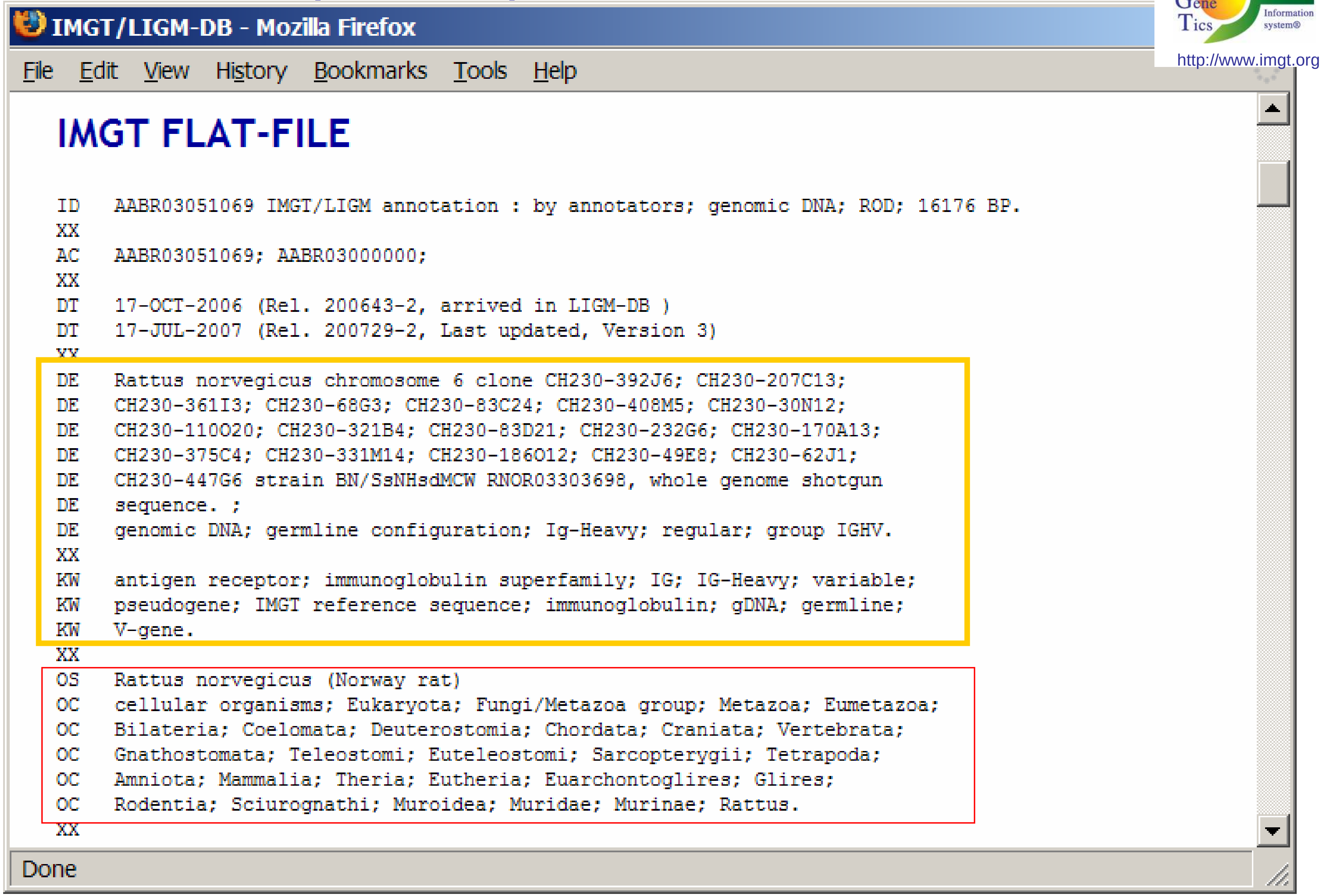


EMBL FLAT-FILE

```
ID   AABR03051069; SV 1; linear; genomic DNA; WGS; ROD; 16176 BP.
XX
AC   AABR03051069; AABR03000000;
XX
DT   13-AUG-2003 (Rel. 76, Created)
DT   13-AUG-2003 (Rel. 76, Last updated, Version 1)
XX
DE   Rattus norvegicus chromosome 6 clone CH230-392J6; CH230-207C13;
DE   CH230-361I3; CH230-68G3; CH230-83C24; CH230-408M5; CH230-30N12;
DE   CH230-110O20; CH230-321B4; CH230-83D21; CH230-232G6; CH230-170A13;
DE   CH230-375C4; CH230-331M14; CH230-186O12; CH230-49E8; CH230-62J1;
DE   CH230-447G6 strain BN/SsNHsdMCW RNOR03303698, whole genome shotgun
DE   sequence.
XX
KW   WGS.
XX
OS   Rattus norvegicus (Norway rat)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea;
OC   Muridae; Murinae; Rattus.
```

Après expertise dans IMGT®

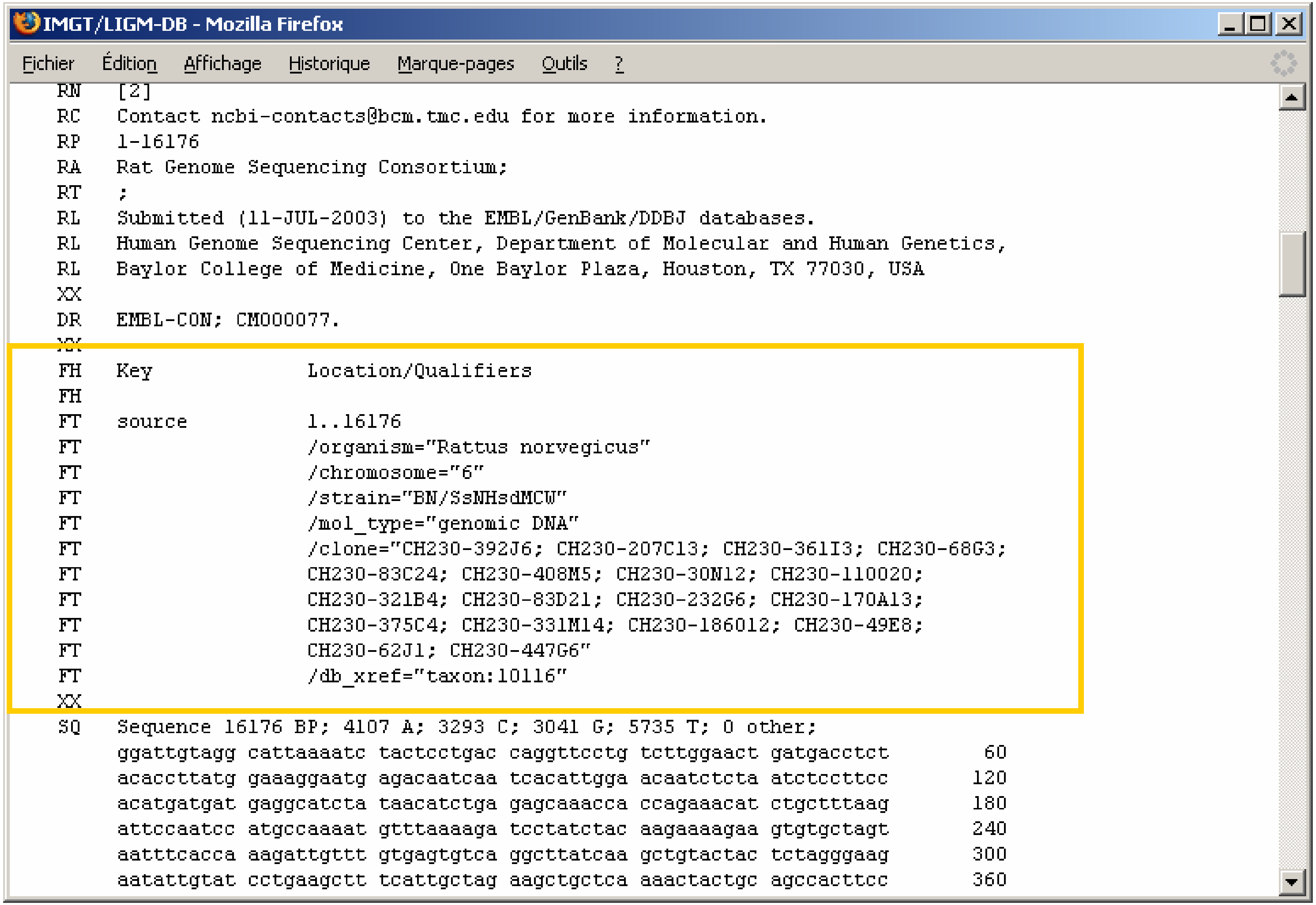
http://www.imgt.org



IMGT FLAT-FILE

```
ID   AABR03051069 IMGT/LIGM annotation : by annotators; genomic DNA; ROD; 16176 BP.
XX
AC   AABR03051069; AABR03000000;
XX
DT   17-OCT-2006 (Rel. 200643-2, arrived in LIGM-DB )
DT   17-JUL-2007 (Rel. 200729-2, Last updated, Version 3)
XX
DE   Rattus norvegicus chromosome 6 clone CH230-392J6; CH230-207C13;
DE   CH230-361I3; CH230-68G3; CH230-83C24; CH230-408M5; CH230-30N12;
DE   CH230-110O20; CH230-321B4; CH230-83D21; CH230-232G6; CH230-170A13;
DE   CH230-375C4; CH230-331M14; CH230-186O12; CH230-49E8; CH230-62J1;
DE   CH230-447G6 strain BN/SsNHsdMCW RNOR03303698, whole genome shotgun
DE   sequence. ;
DE   genomic DNA; germline configuration; Ig-Heavy; regular; group IGHV.
XX
KW   antigen receptor; immunoglobulin superfamily; IG; IG-Heavy; variable;
KW   pseudogene; IMGT reference sequence; immunoglobulin; gDNA; germline;
KW   V-gene.
XX
OS   Rattus norvegicus (Norway rat)
OC   cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;
OC   Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata;
OC   Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda;
OC   Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Glires;
OC   Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
XX
```

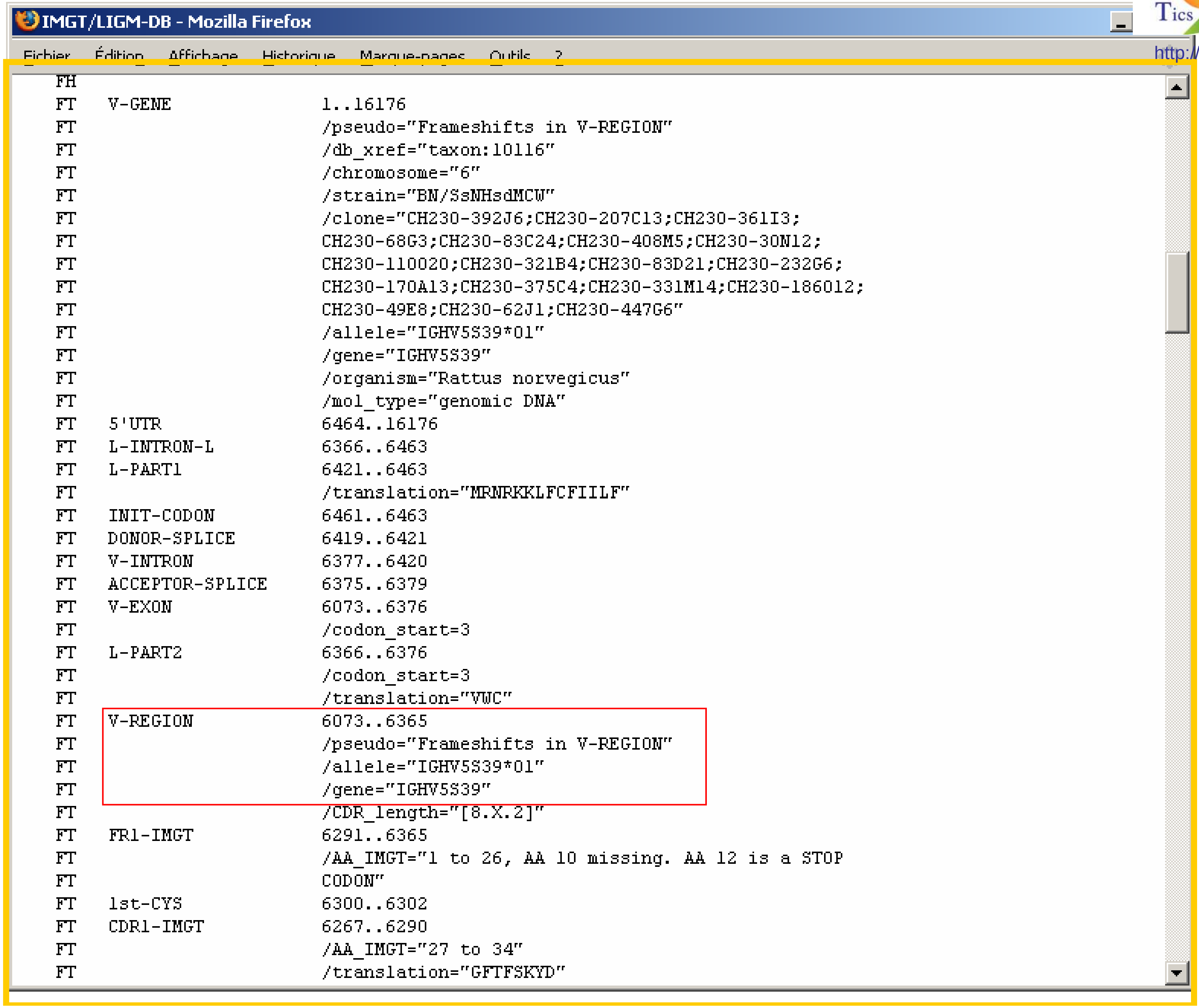
A l’arrivée dans IMGT®



```
RN   [2]
RC   Contact ncbi-contacts@bcm.tmc.edu for more information.
RP   1-16176
RA   Rat Genome Sequencing Consortium;
RT   ;
RL   Submitted (11-JUL-2003) to the EMBL/GenBank/DDBJ databases.
RL   Human Genome Sequencing Center, Department of Molecular and Human Genetics,
RL   Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
XX
DR   EMBL-CON; CM000077.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..16176
FT                   /organism="Rattus norvegicus"
FT                   /chromosome="6"
FT                   /strain="BN/SsNHsdMCW"
FT                   /mol_type="genomic DNA"
FT                   /clone="CH230-392J6; CH230-207C13; CH230-361I3; CH230-68G3;
FT                   CH230-83C24; CH230-408M5; CH230-30N12; CH230-110020;
FT                   CH230-321B4; CH230-83D21; CH230-232G6; CH230-170A13;
FT                   CH230-375C4; CH230-331M14; CH230-186012; CH230-49E8;
FT                   CH230-62J1; CH230-447G6"
FT                   /db_xref="taxon:10116"
XX
SQ   Sequence 16176 BP; 4107 A; 3293 C; 3041 G; 5735 T; 0 other;
     ggattgtagg cattaaaatc tactcctgac caggttcctg tcttggaact gatgacctct        60
     acaccttatg gaaaggaatg agacaatcaa tcacattgga acaatctcta atctccttcc       120
     acatgatgat gaggcatcta taacatctga gagcaaacca ccagaaacat ctgctttaag       180
     attccaatcc atgccaaaat gtttaaaaga tcctatctac aagaaaagaa gtgtgctagt       240
     aatttcacca aagattgttt gtgagtgtca ggcttatcaa gctgtactac tctagggaag       300
     aatattgtat cctgaagctt tcattgctag aagctgctca aaactactgc agccacttcc       360
```

Après expertise dans IMGT®

http://www.imgt.org



```
FH
FT   V-GENE               1..16176
FT                        /pseudo="Frameshifts in V-REGION"
FT                        /db_xref="taxon:10116"
FT                        /chromosome="6"
FT                        /strain="BN/SsNHsdMCW"
FT                        /clone="CH230-392J6;CH230-207C13;CH230-361I3;
FT                        CH230-68G3;CH230-83C24;CH230-408M5;CH230-30N12;
FT                        CH230-110020;CH230-321B4;CH230-83D21;CH230-232G6;
FT                        CH230-170A13;CH230-375C4;CH230-331M14;CH230-186012;
FT                        CH230-49E8;CH230-62J1;CH230-447G6"
FT                        /allele="IGHV5S39*01"
FT                        /gene="IGHV5S39"
FT                        /organism="Rattus norvegicus"
FT                        /mol_type="genomic DNA"
FT   5'UTR                6464..16176
FT   L-INTRON-L           6366..6463
FT   L-PART1              6421..6463
FT                        /translation="MRNRKKLFCFIILF"
FT   INIT-CODON           6461..6463
FT   DONOR-SPLICE         6419..6421
FT   V-INTRON             6377..6420
FT   ACCEPTOR-SPLICE      6375..6379
FT   V-EXON               6073..6376
FT                        /codon_start=3
FT   L-PART2              6366..6376
FT                        /codon_start=3
FT                        /translation="VWC"
FT   V-REGION             6073..6365
FT                        /pseudo="Frameshifts in V-REGION"
FT                        /allele="IGHV5S39*01"
FT                        /gene="IGHV5S39"
FT                        /CDR_length="[8.X.2]"
FT   FR1-IMGT             6291..6365
FT                        /AA_IMGT="1 to 26, AA 10 missing. AA 12 is a STOP
FT                        CODON"
FT   1st-CYS              6300..6302
FT   CDR1-IMGT            6267..6290
FT                        /AA_IMGT="27 to 34"
FT                        /translation="GFTFSKYD"
```

Information et connaissance

- **Une Information** comprend des données primaires issues l'expérimentation (mesures, images, séquences) ainsi que les données secondaires qui comprennent aussi ce qu'il est nécessaire de connaître pour leur analyse.

  (résultats + matériels & méthodes).

- **La connaissance** inclut tout ce qu'il est nécessaire pour réaliser l'annotation des données, telle qu'elle peut être réalisée par des experts dans un domaine particulier.

=> vocabulaire standardisé, contrôlé,
règles d'annotation, dépendances entre les termes

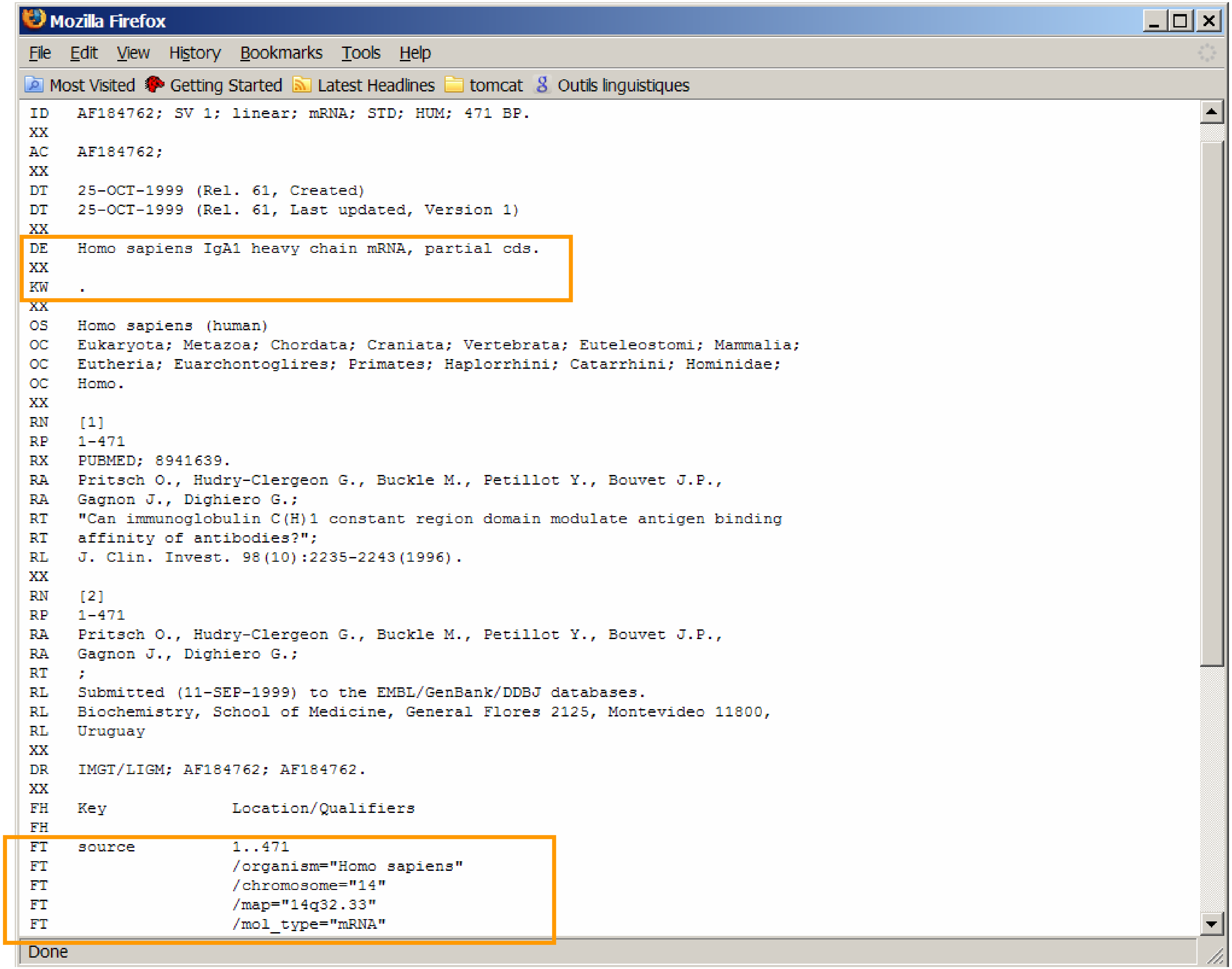
Un exemple de question

Qu'est-ce qu'un gène?
=> plusieurs définitions

Dans une base de données:

- établissement d'un dictionnaire des données
- mise en place de contraintes à travers un vocabulaire contrôlé

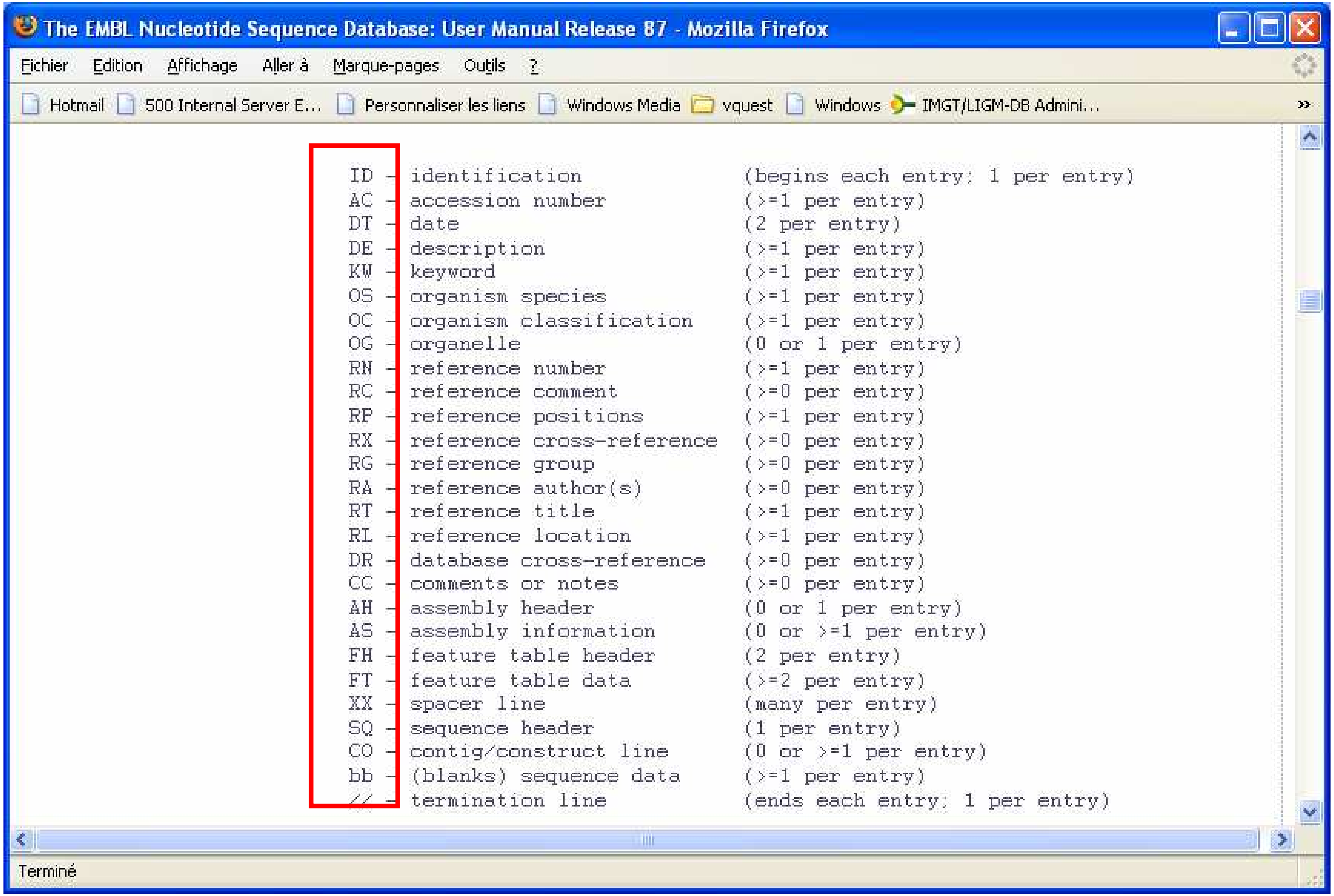
Recherches bibliographiques à travers des thesaurus de mots clés



```
ID   AF184762; SV 1; linear; mRNA; STD; HUM; 471 BP.
XX
AC   AF184762;
XX
DT   25-OCT-1999 (Rel. 61, Created)
DT   25-OCT-1999 (Rel. 61, Last updated, Version 1)
XX
DE   Homo sapiens IgA1 heavy chain mRNA, partial cds.
XX
KW   .
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae;
OC   Homo.
XX
RN   [1]
RP   1-471
RX   PUBMED; 8941639.
RA   Pritsch O., Hudry-Clergeon G., Buckle M., Petillot Y., Bouvet J.P.,
RA   Gagnon J., Dighiero G.;
RT   "Can immunoglobulin C(H)1 constant region domain modulate antigen binding
RT   affinity of antibodies?";
RL   J. Clin. Invest. 98(10):2235-2243(1996).
XX
RN   [2]
RP   1-471
RA   Pritsch O., Hudry-Clergeon G., Buckle M., Petillot Y., Bouvet J.P.,
RA   Gagnon J., Dighiero G.;
RT   ;
RL   Submitted (11-SEP-1999) to the EMBL/GenBank/DDBJ databases.
RL   Biochemistry, School of Medicine, General Flores 2125, Montevideo 11800,
RL   Uruguay
XX
DR   IMGT/LIGM; AF184762; AF184762.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..471
FT                   /organism="Homo sapiens"
FT                   /chromosome="14"
FT                   /map="14q32.33"
FT                   /mol_type="mRNA"
```

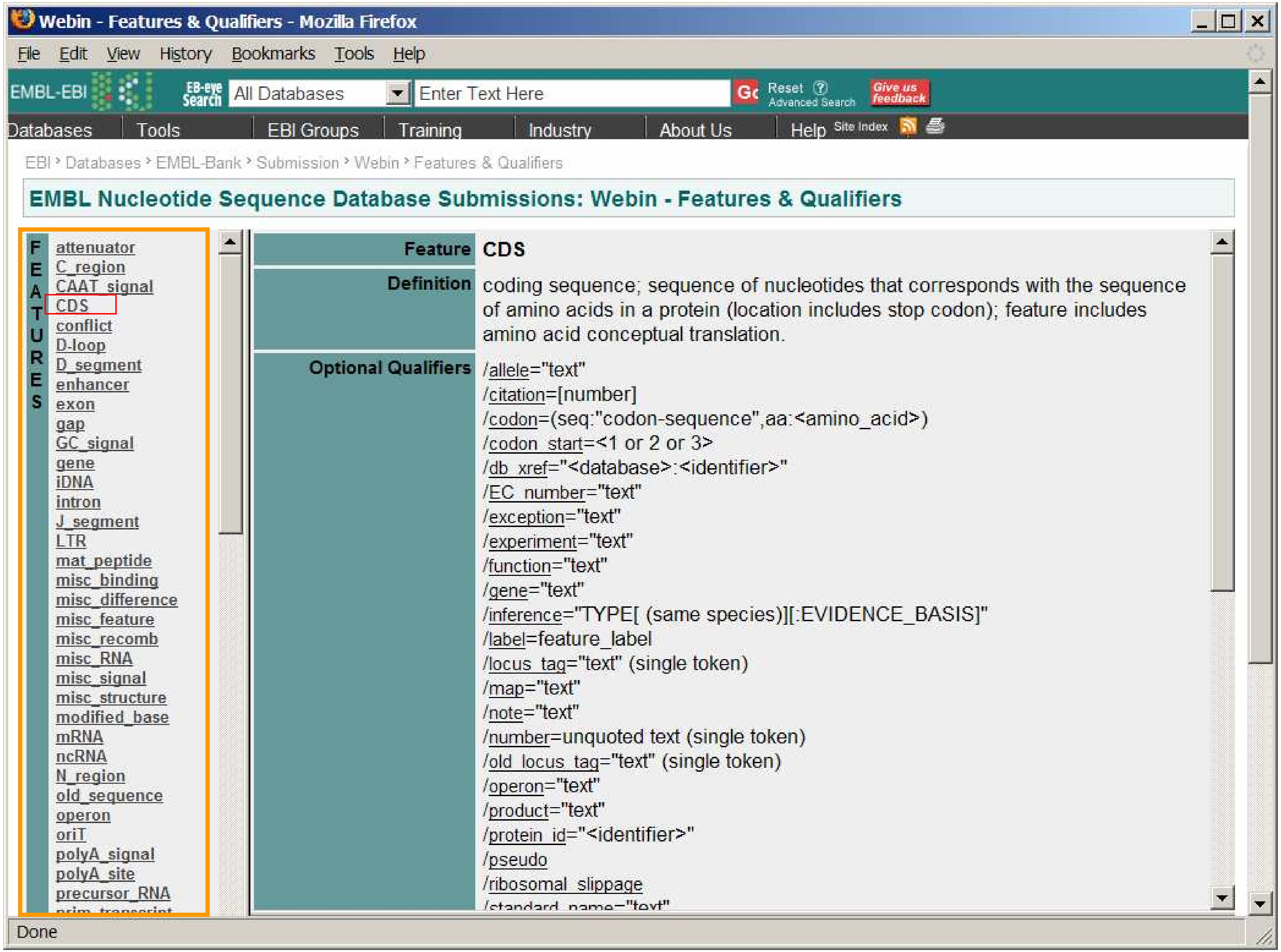
Le vocabulaire contrôlé des bases de données généralistes DDBJ/EMBL/GenBank

Des codes pour indiquer les types d'information



```
ID - identification              (begins each entry; 1 per entry)
AC - accession number            (>=1 per entry)
DT - date                        (2 per entry)
DE - description                 (>=1 per entry)
KW - keyword                     (>=1 per entry)
OS - organism species            (>=1 per entry)
OC - organism classification     (>=1 per entry)
OG - organelle                   (0 or 1 per entry)
RN - reference number            (>=1 per entry)
RC - reference comment           (>=0 per entry)
RP - reference positions         (>=1 per entry)
RX - reference cross-reference   (>=0 per entry)
RG - reference group             (>=0 per entry)
RA - reference author(s)         (>=0 per entry)
RT - reference title             (>=1 per entry)
RL - reference location          (>=1 per entry)
DR - database cross-reference    (>=0 per entry)
CC - comments or notes           (>=0 per entry)
AH - assembly header             (0 or 1 per entry)
AS - assembly information        (0 or >=1 per entry)
FH - feature table header        (2 per entry)
FT - feature table data          (>=2 per entry)
XX - spacer line                 (many per entry)
SQ - sequence header             (1 per entry)
CO - contig/construct line       (0 or >=1 per entry)
bb - (blanks) sequence data      (>=1 per entry)
// - termination line            (ends each entry; 1 per entry)
```

EMBL Feature labels (http://www.ebi.ac.uk/embl/WebFeat/index.html)



EBI › Databases › EMBL-Bank › Submission › Webin › Features & Qualifiers

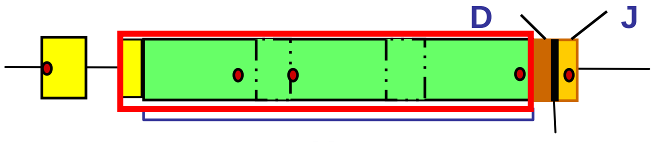
EMBL Nucleotide Sequence Database Submissions: Webin - Features & Qualifiers

FEATURES

- attenuator
- C_region
- CAAT_signal
- CDS
- conflict
- D-loop
- D_segment
- enhancer
- exon
- gap
- GC_signal
- gene
- iDNA
- intron
- J_segment
- LTR
- mat_peptide
- misc_binding
- misc_difference
- misc_feature
- misc_recomb
- misc_RNA
- misc_signal
- misc_structure
- modified_base
- mRNA
- ncRNA
- N_region
- old_sequence
- operon
- oriT
- polyA_signal
- polyA_site
- precursor_RNA
- prim_transcript

Feature	**CDS**
Definition	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation.
Optional Qualifiers	/allele="text" /citation=[number] /codon=(seq:"codon-sequence",aa:<amino_acid>) /codon_start=<1 or 2 or 3> /db_xref="<database>:<identifier>" /EC_number="text" /exception="text" /experiment="text" /function="text" /gene="text" /inference="TYPE[(same species)][:EVIDENCE_BASIS]" /label=feature_label /locus_tag="text" (single token) /map="text" /note="text" /number=unquoted text (single token) /old_locus_tag="text" (single token) /operon="text" /product="text" /protein_id="<identifier>" /pseudo /ribosomal_slippage /standard_name="text"

EMBL Feature labels



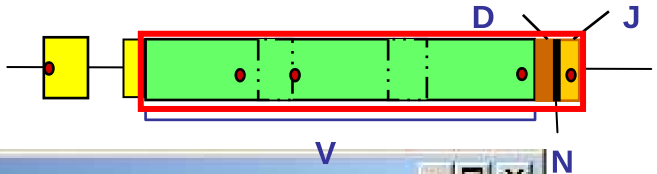
EMBL Nucleotide Sequence Database Submissions: Webin - Features & Qualifiers

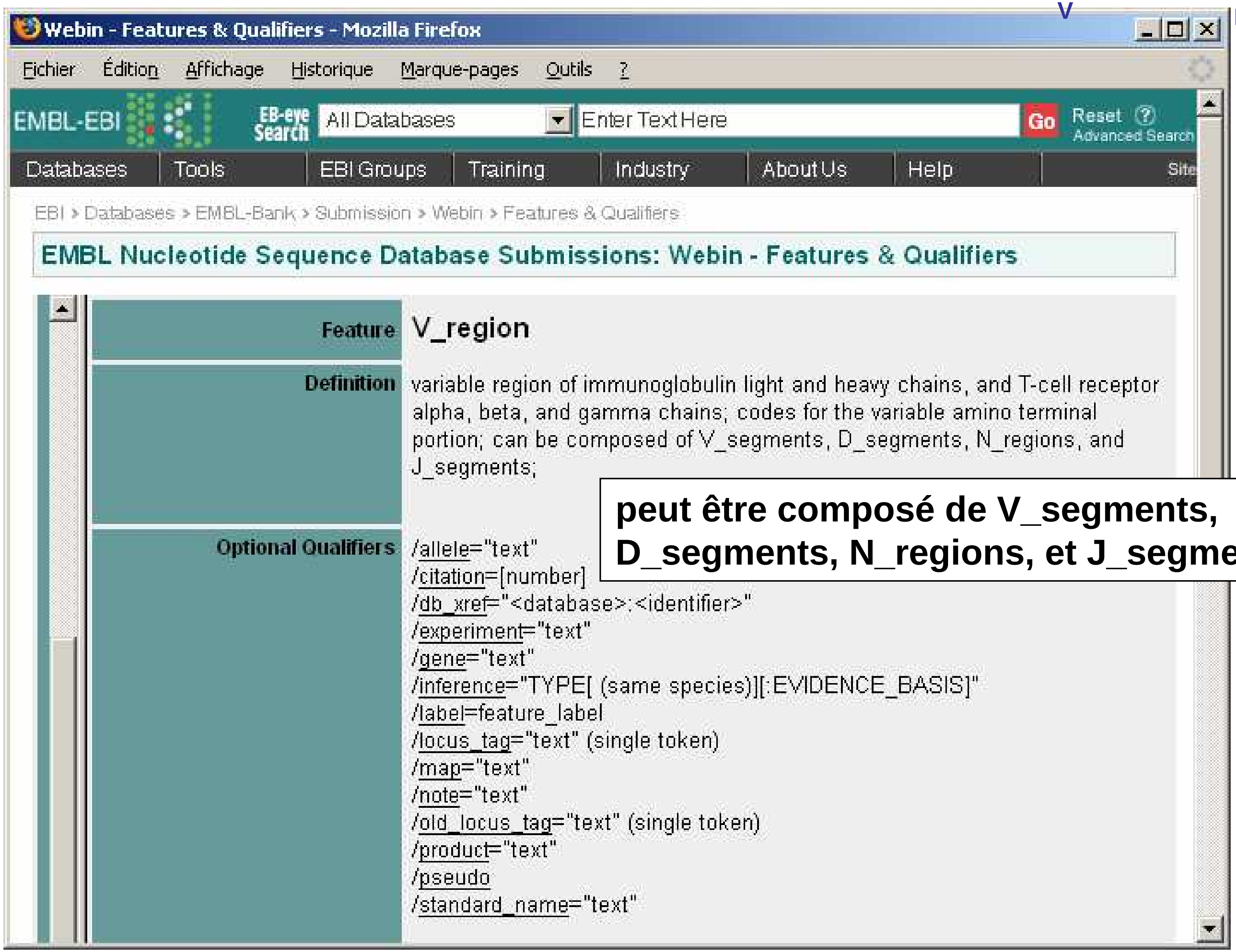
Feature	**V_segment**
Definition	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide;
Optional Qualifiers	/allele="text" /citation=[number] /db_xref="<databas /experiment="text" /gene="text" /inference="TYPE[(same species)][:EVIDENCE_BASIS]" /label=feature_label /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /standard_name="text"

code la majeur partie de la région variable (V_region) et les quelques acides aminés du peptide signal

?

EMBL Feature labels





EBI > Databases > EMBL-Bank > Submission > Webin > Features & Qualifiers

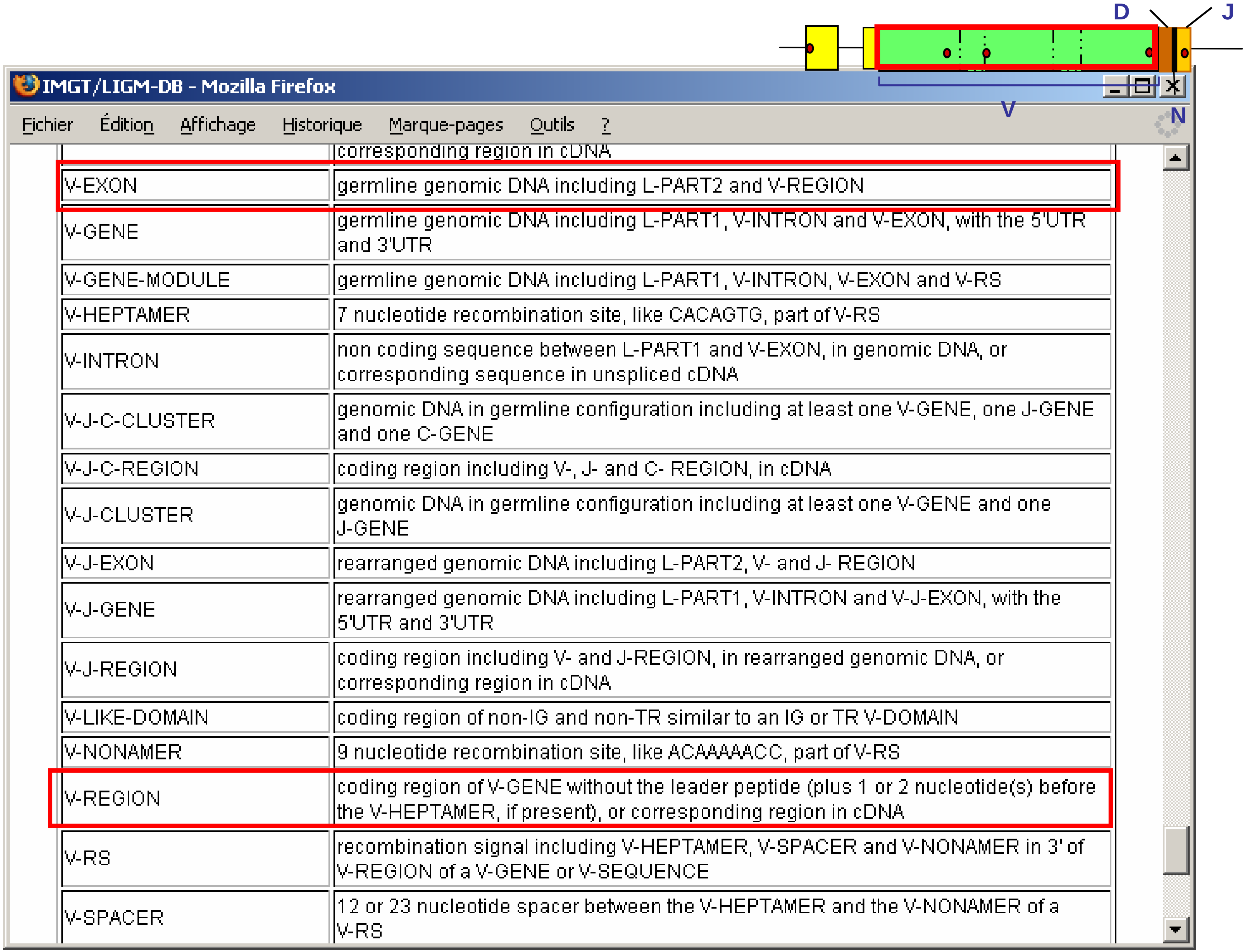
EMBL Nucleotide Sequence Database Submissions: Webin - Features & Qualifiers

Feature	V_region
Definition	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be composed of V_segments, D_segments, N_regions, and J_segments;
Optional Qualifiers	/allele="text" /citation=[number] /db_xref="<database>:<identifier>" /experiment="text" /gene="text" /inference="TYPE[(same species)][:EVIDENCE_BASIS]" /label=feature_label /locus_tag="text" (single token) /map="text" /note="text" /old_locus_tag="text" (single token) /product="text" /pseudo /standard_name="text"

peut être composé de V_segments, D_segments, N_regions, et J_segments

?

List and definition of IMGT standardized labels



	corresponding region in cDNA
V-EXON	germline genomic DNA including L-PART2 and V-REGION
V-GENE	germline genomic DNA including L-PART1, V-INTRON and V-EXON, with the 5'UTR and 3'UTR
V-GENE-MODULE	germline genomic DNA including L-PART1, V-INTRON, V-EXON and V-RS
V-HEPTAMER	7 nucleotide recombination site, like CACAGTG, part of V-RS
V-INTRON	non coding sequence between L-PART1 and V-EXON, in genomic DNA, or corresponding sequence in unspliced cDNA
V-J-C-CLUSTER	genomic DNA in germline configuration including at least one V-GENE, one J-GENE and one C-GENE
V-J-C-REGION	coding region including V-, J- and C- REGION, in cDNA
V-J-CLUSTER	genomic DNA in germline configuration including at least one V-GENE and one J-GENE
V-J-EXON	rearranged genomic DNA including L-PART2, V- and J- REGION
V-J-GENE	rearranged genomic DNA including L-PART1, V-INTRON and V-J-EXON, with the 5'UTR and 3'UTR
V-J-REGION	coding region including V- and J-REGION, in rearranged genomic DNA, or corresponding region in cDNA
V-LIKE-DOMAIN	coding region of non-IG and non-TR similar to an IG or TR V-DOMAIN
V-NONAMER	9 nucleotide recombination site, like ACAAAAACC, part of V-RS
V-REGION	coding region of V-GENE without the leader peptide (plus 1 or 2 nucleotide(s) before the V-HEPTAMER, if present), or corresponding region in cDNA
V-RS	recombination signal including V-HEPTAMER, V-SPACER and V-NONAMER in 3' of V-REGION of a V-GENE or V-SEQUENCE
V-SPACER	12 or 23 nucleotide spacer between the V-HEPTAMER and the V-NONAMER of a V-RS

OK

OK

Quelle est la signification des ces termes dans le contexte de la biologie?

Comment ces termes sont-ils organisés?

Comment sont-il reliés?

Ontologies

Une **ontologie** définit formellement les **termes** employés pour **décrire et représenter** un **domaine de connaissance**.

Les ontologies sont destinées à être utilisées par:

- des personnes
- des bases de données
- des applications

ayant besoin de partager des informations

Ontologies

Au sein d’une ontologie, les termes sont regroupés sous forme de **concepts** (ou classes) sémantiques: ils définissent un groupe d’individus possédant des propriétés similaires.

Les ontologies incluent les **définitions**, informatiquement exploitables des **concepts** élémentaires et de leurs **relations**.

Les ontologies doivent permettre le **partage** et la **réutilisation des connaissances.**

Bases de connaissances:

Une ontologie ainsi que l'ensemble des **instances** individuelles des **concepts** constituent une base de connaissances. Une frontière subtile marque la fin d'une ontologie et le début d'une base de connaissances.

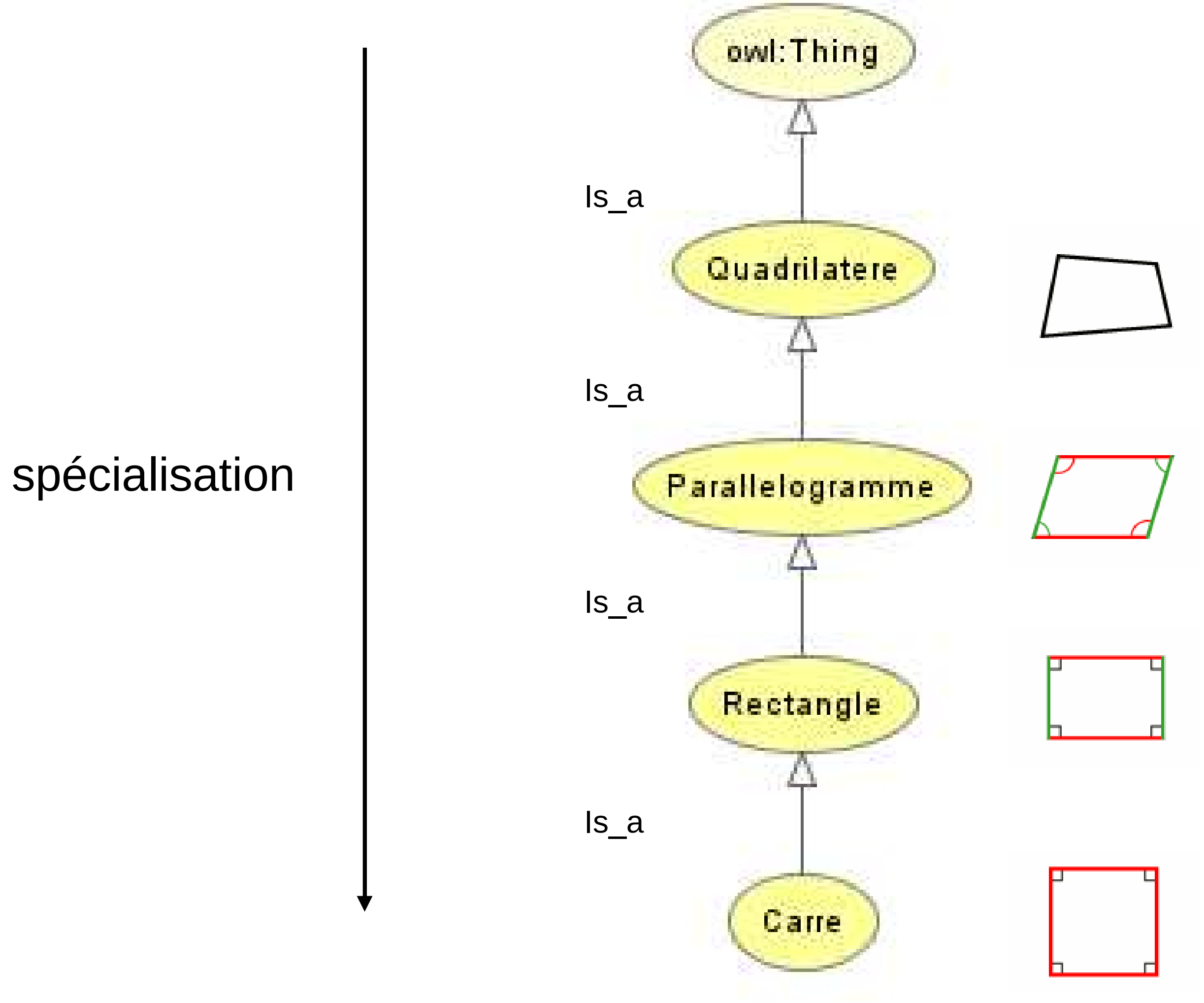
Exemple simple: les formes géométriques
représentation hiérarchique de concepts avec la relation « is_a »

(OWL, Web Ontology Language)

Notion de Classe :

Définit un groupe d'individus possédant des propriétés similaires.

Thing est la classe mère.



spécialisation

Is_a

Is_a

Is_a

Is_a

Def : polygone à 4 côtés

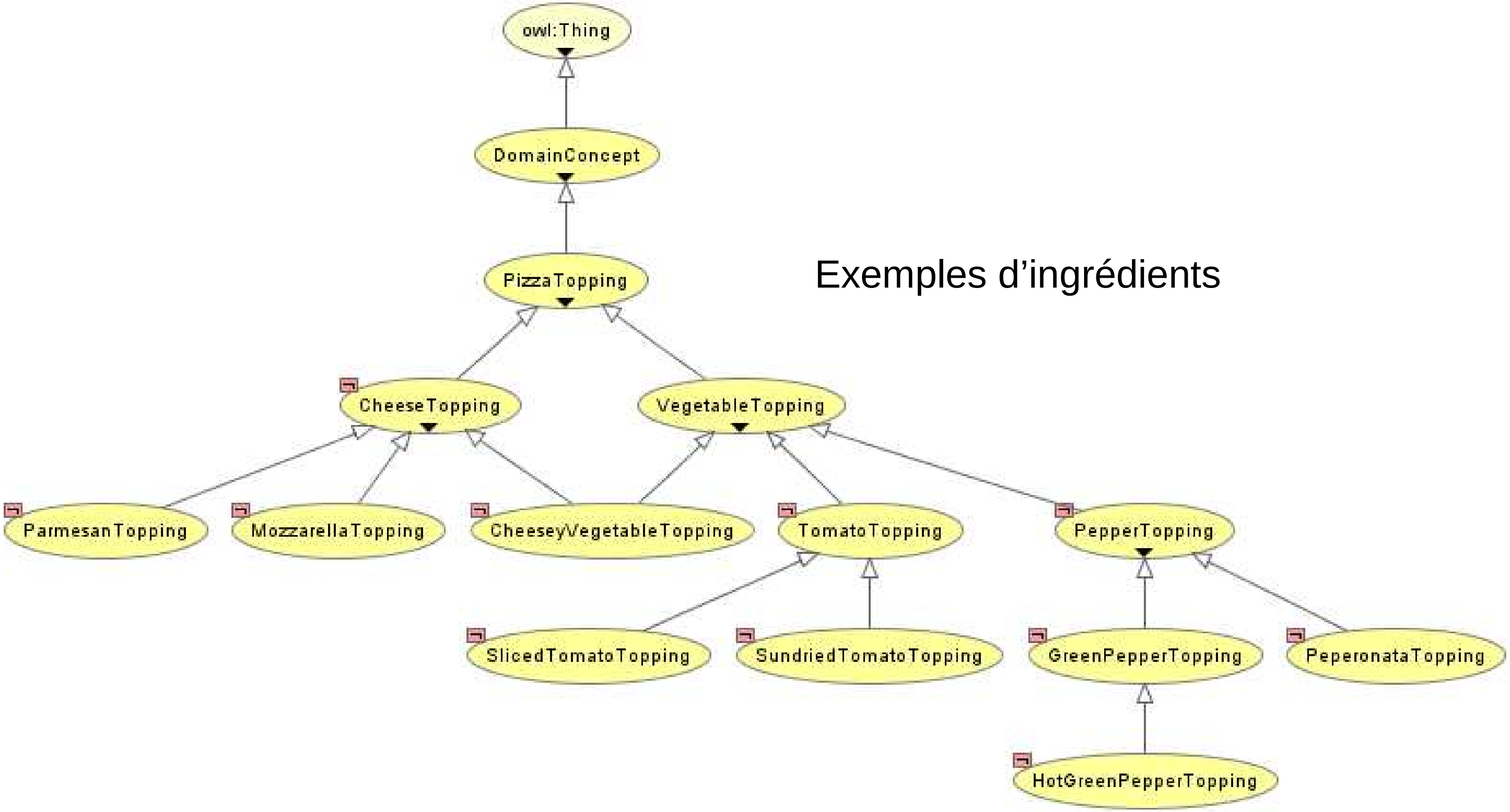
Def: quadrilatère (convexe) dont les côtés sont parallèles deux à deux

Def: Parallélogramme à angles droits

Def: rectangle dont les cotés sont égaux

Exemple de hiérarchie : ontologie des pizzas

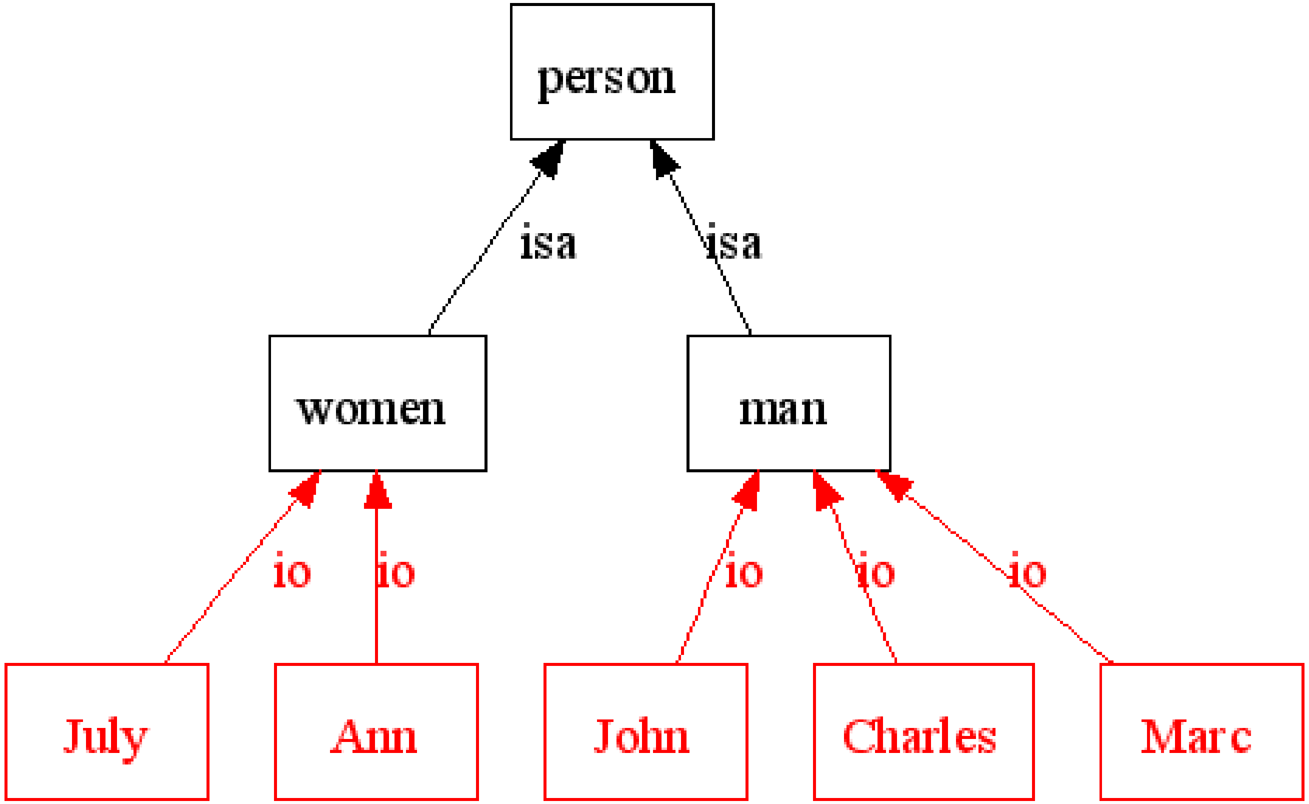
(exemple fournit avec l'éditeur d'ontologies Protégé http://protege.stanford.edu/)



Exemples d'ingrédients

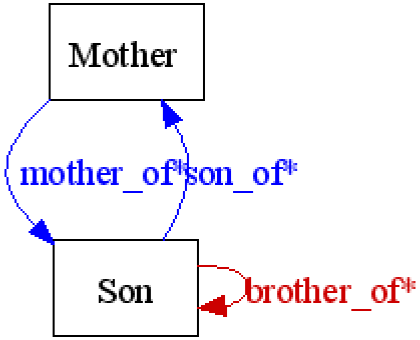
Instances

Ce sont les « individus » qui peuplent les classes

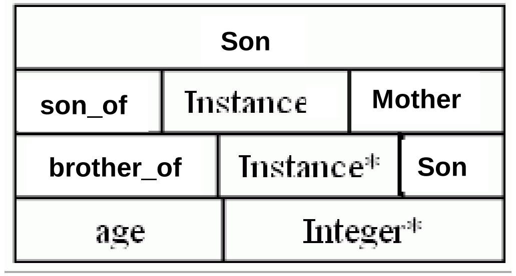


Propriétés/Relations
(OWL, Web Ontology Language)

- Propriété d’objet : relie une instance à une autre instance



- Propriété de type de données: relie une instance à des valeurs.



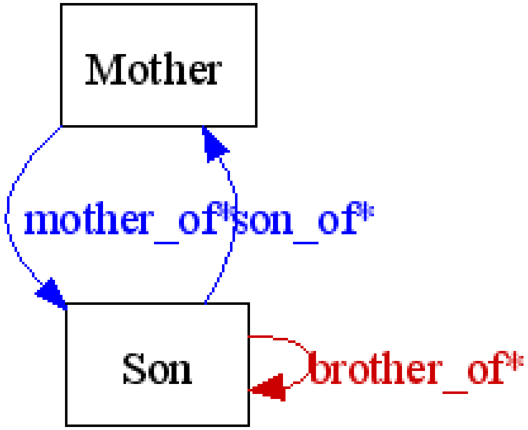
Son		
son_of	Instance	Mother
brother_of	Instance*	Son
age	Integer*	

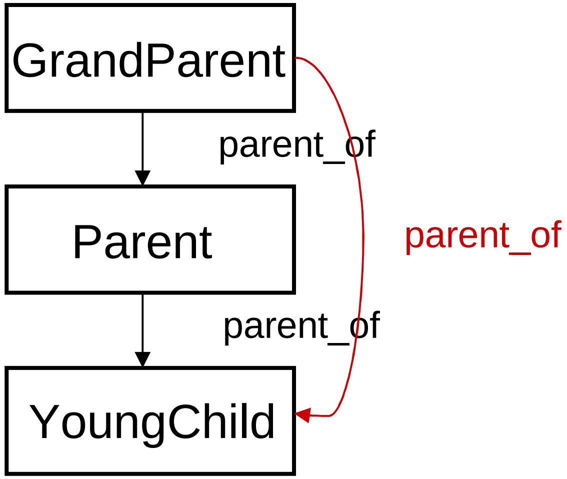
Propriétés/Relations
(OWL, Web Ontology Language)

domain : classes pour lesquelles est définie la propriété,
range : classes reliées par la propriété au domain.

domain	relation	range
Son	brother_of	Son
Mother	mother_of	Son
Son	son_of	Mother

Caractéristiques des Propriétés/relations: symétriques, transitives, inverses et fonctionnelles

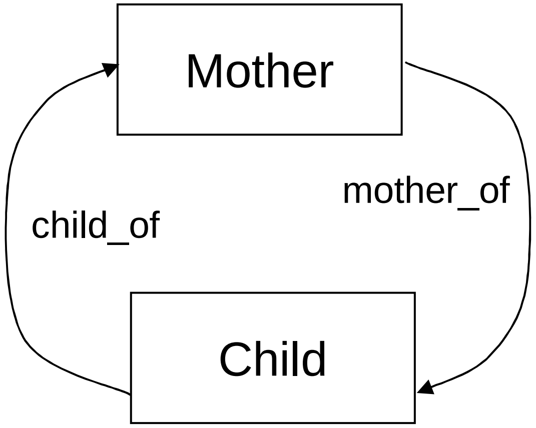




(familles de 3 générations avec de jeunes enfants)

Symétrique: brother_of
Transitive: parent_of
Inverse: son_of / mother_of
Fonctionnelle : son_of

Restriction sur les Propriétés/Relations : cardinalités



Ex de familles nombreuses de moins de 10 enfants

- Une mère a au moins 3 enfants
- Une mère a au plus 10 enfants
- Un enfant a une et une seule mère

minCardinality: toute instance de la classe sera relié par la propriété à au moins x individus Exemple: propriété mother_of de la classe Mother : minCardinality=3

maxCardinality: toute instance de la classe sera relié par la propriété à au plus x individus Exemple: propriété mother_of de la classe Mother : maxCardinality=10

cardinality: toute instance de la classe sera relié par la propriété à exactement x individus Exemple: propriété son_of de la classe Mother : cardinality=1

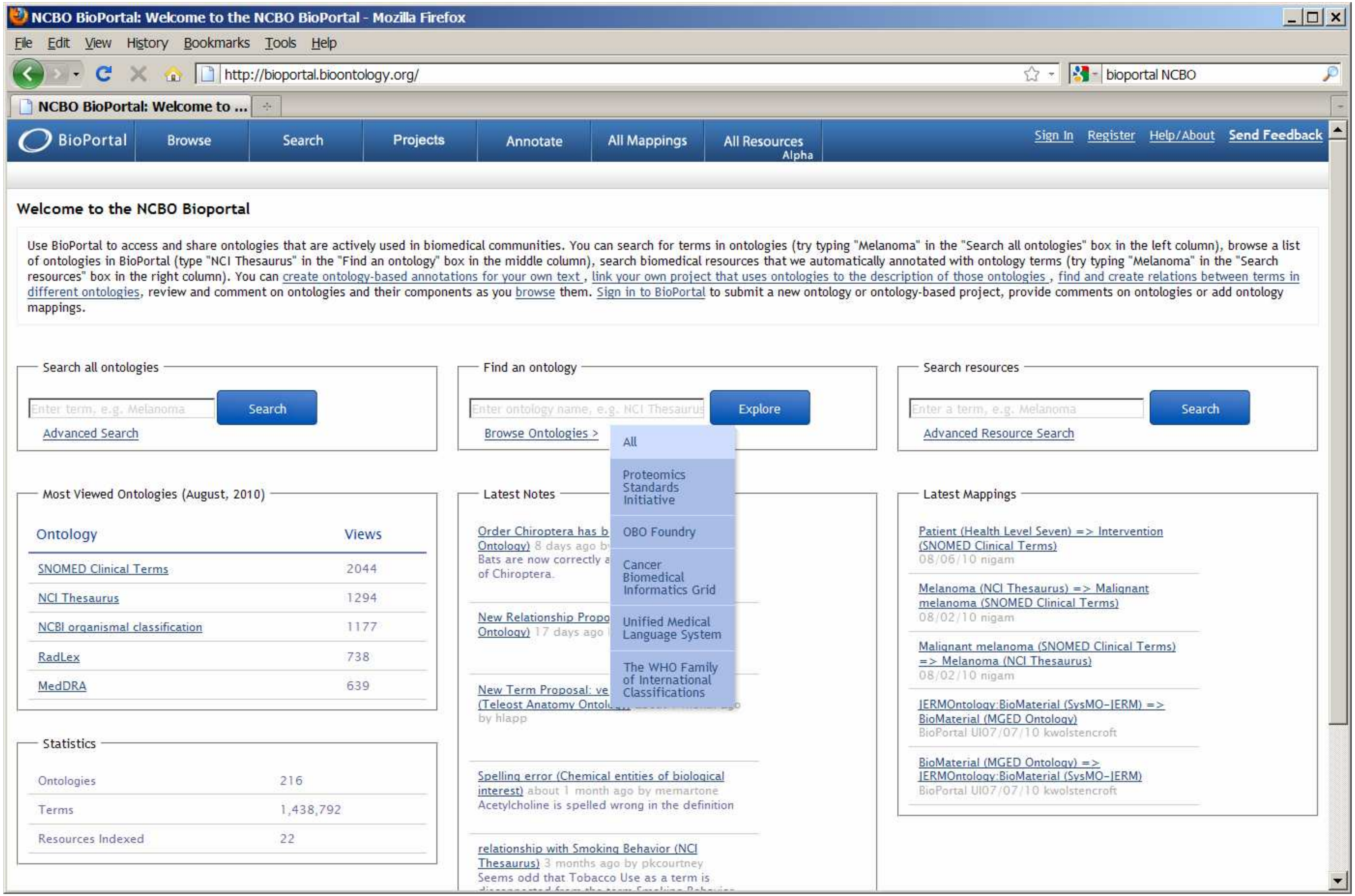
Les ontologies en Biologie

- On s'intéresse aux ontologies qui sont du domaine public.
- Leur nombre augmente régulièrement
  (besoin de définir, de partager).
- Elles couvrent des sujets et domaines différents.

NCBO Bioportal (http://bioportal.bioontology.org/)
OBO (Open Biological Ontologies) (http://www.obofoundry.org/)

recensent les ontologies en biologie.

NCBO Bioportal (http://bioportal.bioontology.org/)



Welcome to the NCBO Bioportal

Use BioPortal to access and share ontologies that are actively used in biomedical communities. You can search for terms in ontologies (try typing "Melanoma" in the "Search all ontologies" box in the left column), browse a list of ontologies in BioPortal (type "NCI Thesaurus" in the "Find an ontology" box in the middle column), search biomedical resources that we automatically annotated with ontology terms (try typing "Melanoma" in the "Search resources" box in the right column). You can create ontology-based annotations for your own text , link your own project that uses ontologies to the description of those ontologies , find and create relations between terms in different ontologies, review and comment on ontologies and their components as you browse them. Sign in to BioPortal to submit a new ontology or ontology-based project, provide comments on ontologies or add ontology mappings.

Most Viewed Ontologies (August, 2010)

Ontology	Views
SNOMED Clinical Terms	2044
NCI Thesaurus	1294
NCBI organismal classification	1177
RadLex	738
MedDRA	639

Statistics

Ontologies	216
Terms	1,438,792
Resources Indexed	22

Latest Notes

Order Chiroptera has b
Ontology) 8 days ago b
Bats are now correctly a
of Chiroptera.

New Relationship Propo
Ontology) 17 days ago

New Term Proposal: ve
(Teleost Anatomy Ontol
by hlapp

Spelling error (Chemical entities of biological interest) about 1 month ago by memartone
Acetylcholine is spelled wrong in the definition

relationship with Smoking Behavior (NCI Thesaurus) 3 months ago by pkcourtney
Seems odd that Tobacco Use as a term is

Latest Mappings

Patient (Health Level Seven) => Intervention (SNOMED Clinical Terms)
08/06/10 nigam

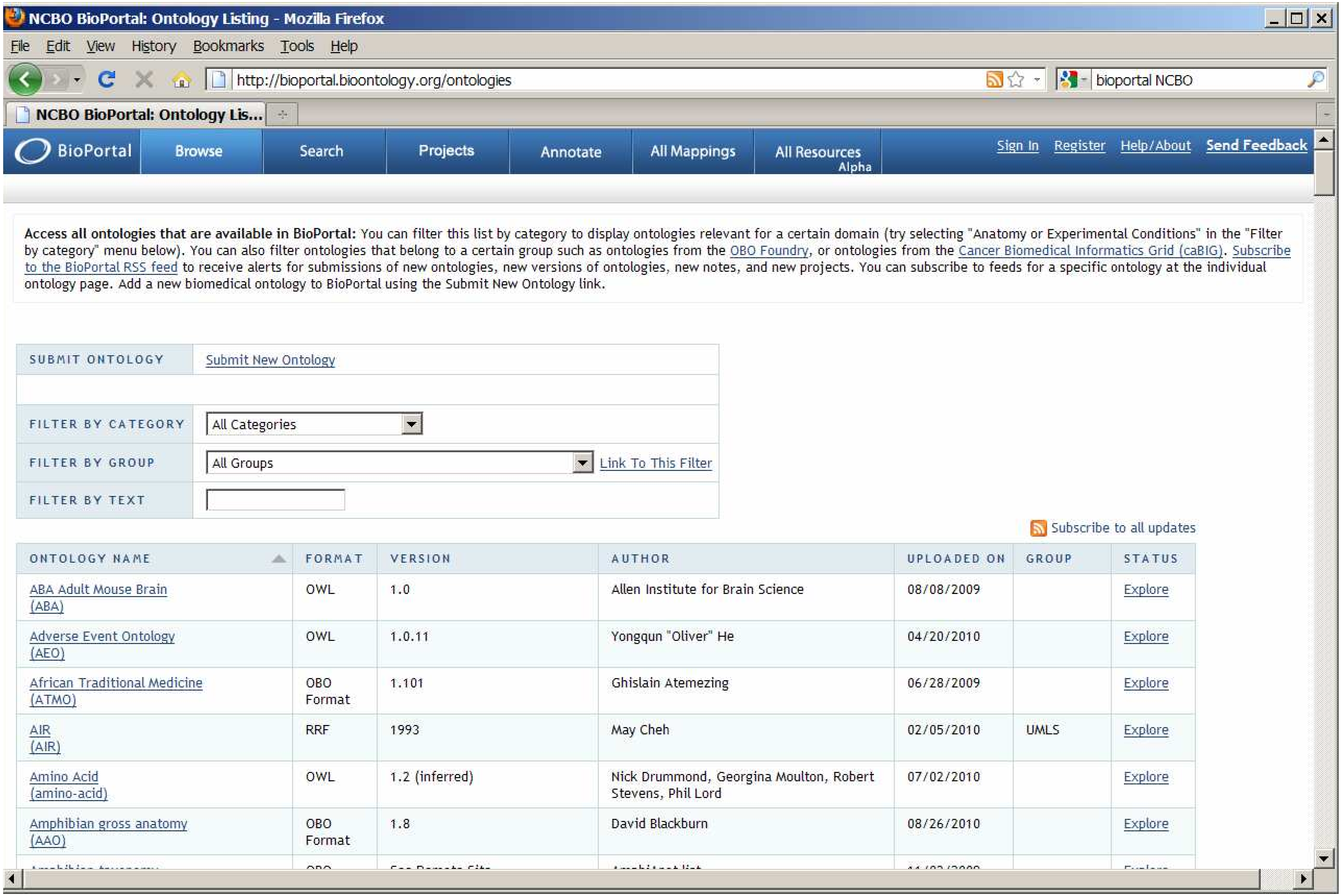
Melanoma (NCI Thesaurus) => Malignant melanoma (SNOMED Clinical Terms)
08/02/10 nigam

Malignant melanoma (SNOMED Clinical Terms) => Melanoma (NCI Thesaurus)
08/02/10 nigam

JERMOntology:BioMaterial (SysMO-JERM) => BioMaterial (MGED Ontology)
BioPortal UI07/07/10 kwolstencroft

BioMaterial (MGED Ontology) => JERMOntology:BioMaterial (SysMO-JERM)
BioPortal UI07/07/10 kwolstencroft

NCBO Bioportal



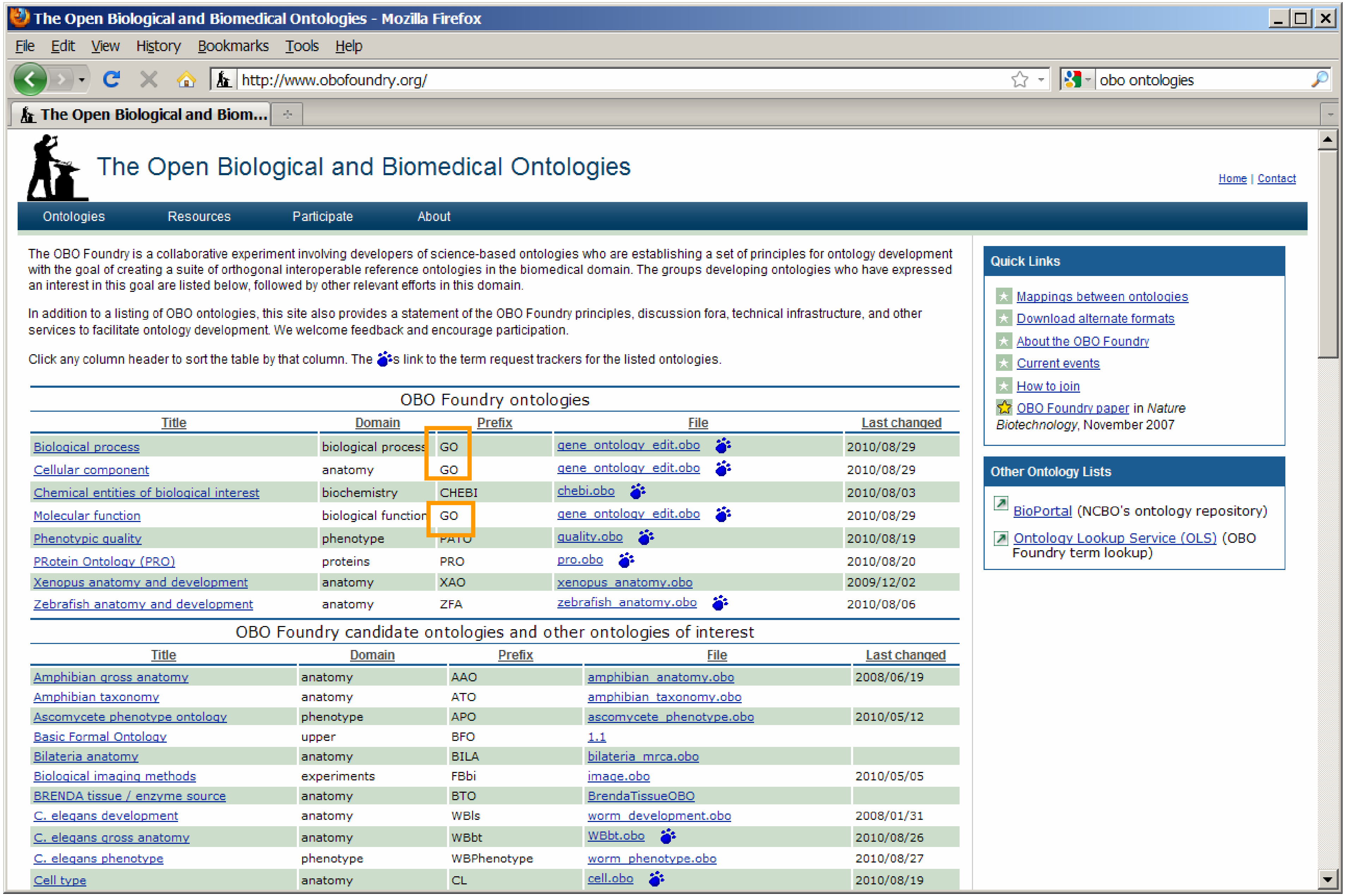
Access all ontologies that are available in BioPortal: You can filter this list by category to display ontologies relevant for a certain domain (try selecting "Anatomy or Experimental Conditions" in the "Filter by category" menu below). You can also filter ontologies that belong to a certain group such as ontologies from the OBO Foundry, or ontologies from the Cancer Biomedical Informatics Grid (caBIG). Subscribe to the BioPortal RSS feed to receive alerts for submissions of new ontologies, new versions of ontologies, new notes, and new projects. You can subscribe to feeds for a specific ontology at the individual ontology page. Add a new biomedical ontology to BioPortal using the Submit New Ontology link.

SUBMIT ONTOLOGY	Submit New Ontology
FILTER BY CATEGORY	All Categories
FILTER BY GROUP	All Groups Link To This Filter
FILTER BY TEXT	

Subscribe to all updates

ONTOLOGY NAME	FORMAT	VERSION	AUTHOR	UPLOADED ON	GROUP	STATUS
ABA Adult Mouse Brain (ABA)	OWL	1.0	Allen Institute for Brain Science	08/08/2009		Explore
Adverse Event Ontology (AEO)	OWL	1.0.11	Yongqun "Oliver" He	04/20/2010		Explore
African Traditional Medicine (ATMO)	OBO Format	1.101	Ghislain Atemezing	06/28/2009		Explore
AIR (AIR)	RRF	1993	May Cheh	02/05/2010	UMLS	Explore
Amino Acid (amino-acid)	OWL	1.2 (inferred)	Nick Drummond, Georgina Moulton, Robert Stevens, Phil Lord	07/02/2010		Explore
Amphibian gross anatomy (AAO)	OBO Format	1.8	David Blackburn	08/26/2010		Explore

OBO (http://www.obofoundry.org/)



The Open Biological and Biomedical Ontologies

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion fora, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The 🐾s link to the term request trackers for the listed ontologies.

OBO Foundry ontologies

Title	Domain	Prefix	File	Last changed
Biological process	biological process	GO	gene_ontology_edit.obo 🐾	2010/08/29
Cellular component	anatomy	GO	gene_ontology_edit.obo 🐾	2010/08/29
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo 🐾	2010/08/03
Molecular function	biological function	GO	gene_ontology_edit.obo 🐾	2010/08/29
Phenotypic quality	phenotype	PATO	quality.obo 🐾	2010/08/19
PRotein Ontology (PRO)	proteins	PRO	pro.obo 🐾	2010/08/20
Xenopus anatomy and development	anatomy	XAO	xenopus_anatomy.obo	2009/12/02
Zebrafish anatomy and development	anatomy	ZFA	zebrafish_anatomy.obo 🐾	2010/08/06

OBO Foundry candidate ontologies and other ontologies of interest

Title	Domain	Prefix	File	Last changed
Amphibian gross anatomy	anatomy	AAO	amphibian_anatomy.obo	2008/06/19
Amphibian taxonomy	anatomy	ATO	amphibian_taxonomy.obo	
Ascomycete phenotype ontology	phenotype	APO	ascomycete_phenotype.obo	2010/05/12
Basic Formal Ontology	upper	BFO	1.1	
Bilateria anatomy	anatomy	BILA	bilateria_mrca.obo	
Biological imaging methods	experiments	FBbi	image.obo	2010/05/05
BRENDA tissue / enzyme source	anatomy	BTO	BrendaTissueOBO	
C. elegans development	anatomy	WBls	worm_development.obo	2008/01/31
C. elegans gross anatomy	anatomy	WBbt	WBbt.obo 🐾	2010/08/26
C. elegans phenotype	phenotype	WBPhenotype	worm_phenotype.obo	2010/08/27
Cell type	anatomy	CL	cell.obo 🐾	2010/08/19

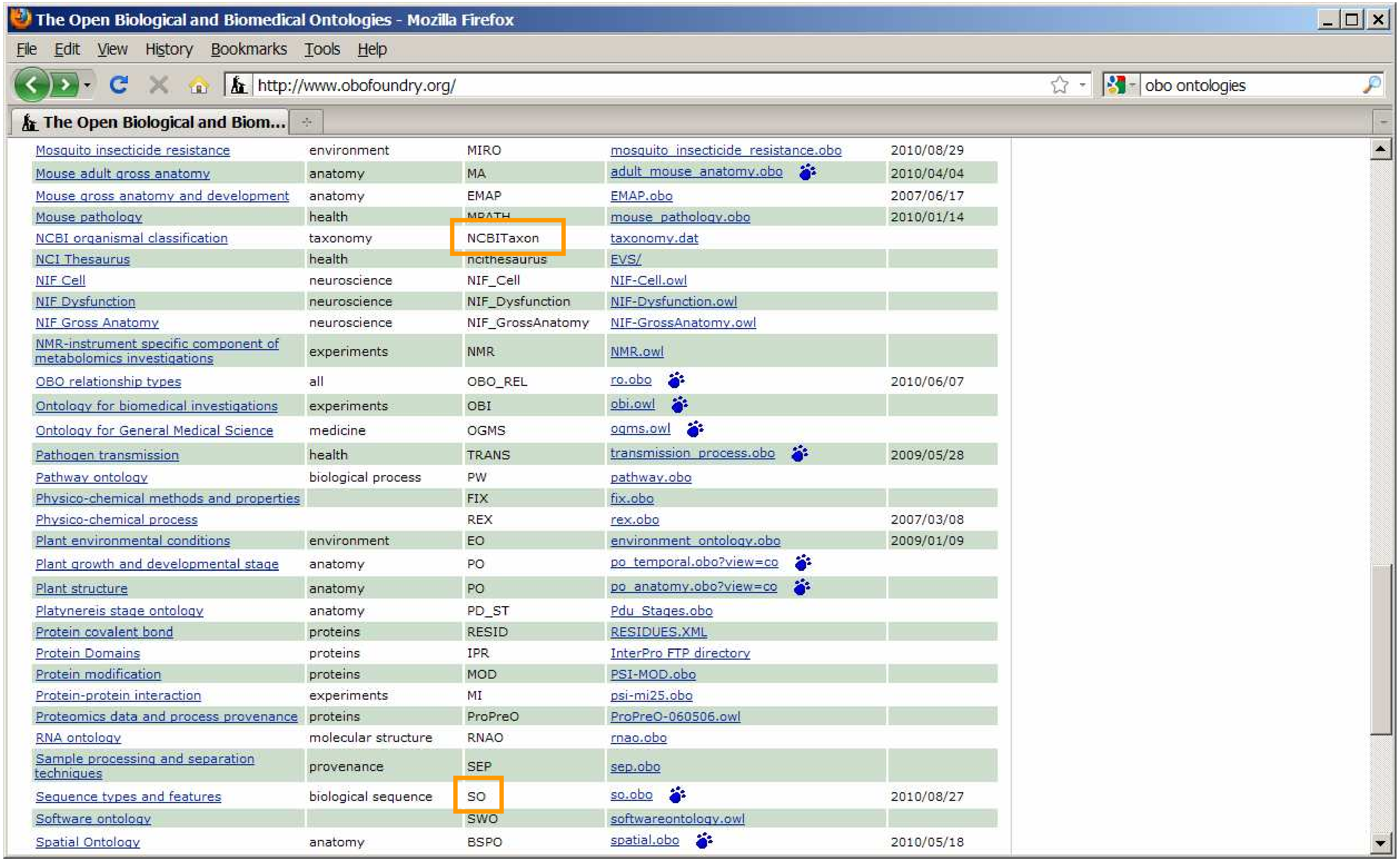
Quick Links

- Mappings between ontologies
- Download alternate formats
- About the OBO Foundry
- Current events
- How to join
- OBO Foundry paper in *Nature Biotechnology*, November 2007

Other Ontology Lists

- BioPortal (NCBO's ontology repository)
- Ontology Lookup Service (OLS) (OBO Foundry term lookup)

OBO (http://www.obofoundry.org/)



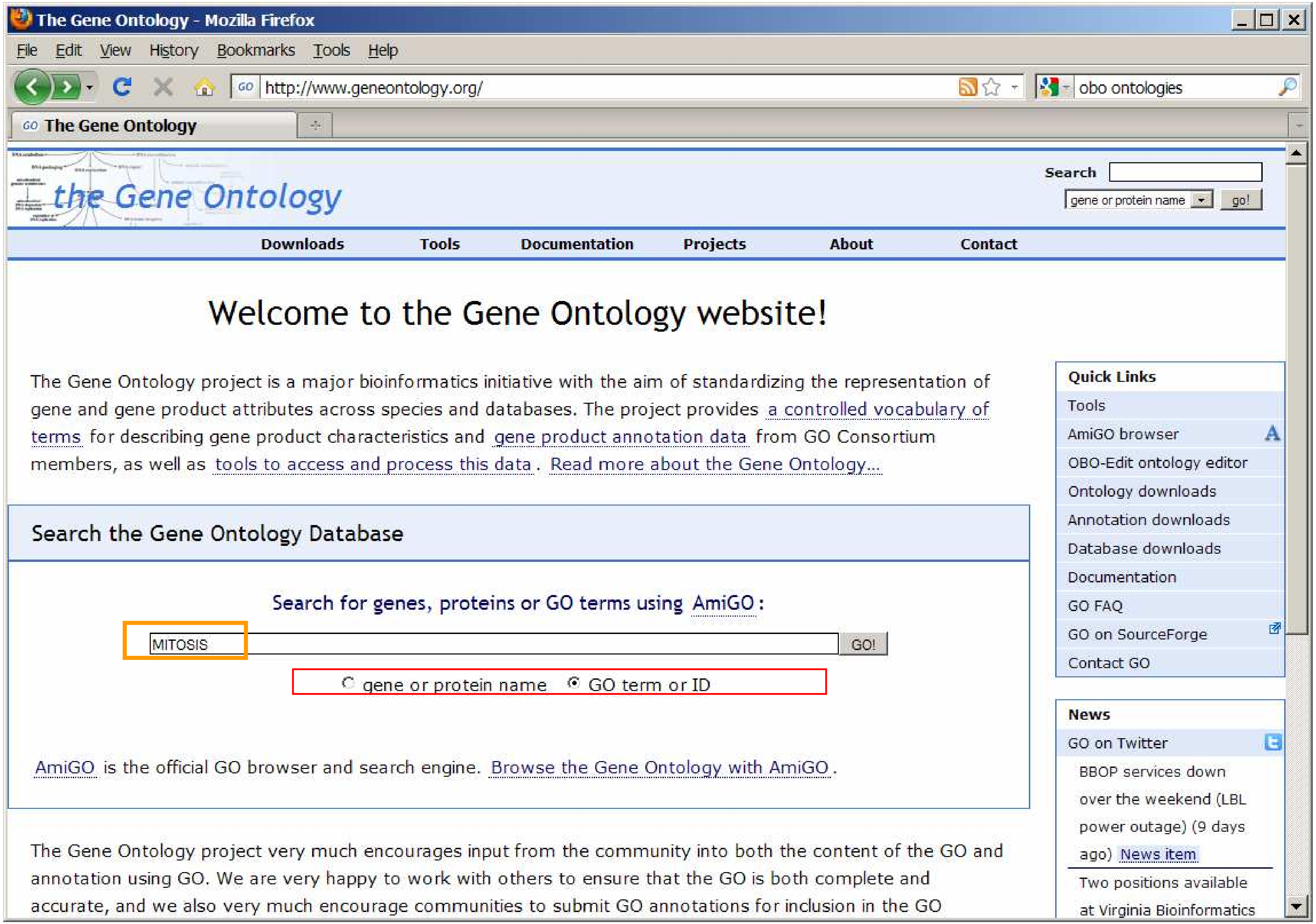
Mosquito insecticide resistance	environment	MIRO	mosquito_insecticide_resistance.obo	2010/08/29
Mouse adult gross anatomy	anatomy	MA	adult_mouse_anatomy.obo	2010/04/04
Mouse gross anatomy and development	anatomy	EMAP	EMAP.obo	2007/06/17
Mouse pathology	health	MPATH	mouse_pathology.obo	2010/01/14
NCBI organismal classification	taxonomy	NCBITaxon	taxonomy.dat	
NCI Thesaurus	health	ncithesaurus	EVS/	
NIF Cell	neuroscience	NIF_Cell	NIF-Cell.owl	
NIF Dysfunction	neuroscience	NIF_Dysfunction	NIF-Dysfunction.owl	
NIF Gross Anatomy	neuroscience	NIF_GrossAnatomy	NIF-GrossAnatomy.owl	
NMR-instrument specific component of metabolomics investigations	experiments	NMR	NMR.owl	
OBO relationship types	all	OBO_REL	ro.obo	2010/06/07
Ontology for biomedical investigations	experiments	OBI	obi.owl	
Ontology for General Medical Science	medicine	OGMS	ogms.owl	
Pathogen transmission	health	TRANS	transmission_process.obo	2009/05/28
Pathway ontology	biological process	PW	pathway.obo	
Physico-chemical methods and properties		FIX	fix.obo	
Physico-chemical process		REX	rex.obo	2007/03/08
Plant environmental conditions	environment	EO	environment_ontology.obo	2009/01/09
Plant growth and developmental stage	anatomy	PO	po_temporal.obo?view=co	
Plant structure	anatomy	PO	po_anatomy.obo?view=co	
Platynereis stage ontology	anatomy	PD_ST	Pdu_Stages.obo	
Protein covalent bond	proteins	RESID	RESIDUES.XML	
Protein Domains	proteins	IPR	InterPro FTP directory	
Protein modification	proteins	MOD	PSI-MOD.obo	
Protein-protein interaction	experiments	MI	psi-mi25.obo	
Proteomics data and process provenance	proteins	ProPreO	ProPreO-060506.owl	
RNA ontology	molecular structure	RNAO	rnao.obo	
Sample processing and separation techniques	provenance	SEP	sep.obo	
Sequence types and features	biological sequence	SO	so.obo	2010/08/27
Software ontology		SWO	softwareontology.owl	
Spatial Ontology	anatomy	BSPO	spatial.obo	2010/05/18

Gene Ontology.

- GO a été créée en 1998. GO résulte d'une collaboration entre plusieurs bases de données: FlyBase (drosophile), the Saccharomyces Genome Database, et des bases de données de génomes (homme et souris), etc.

- GO comprend 3 parties axées sur :
  - **la fonction moléculaire** (fonction des gènes exprimés ex: ATPase activity.
  - **le processus biologique** (rôles biologique généraux de fonctions moléculaire complexes ex: la mitose).
  - **les composants cellulaires** (structures subcellulaires, localisation des complexes macromoleculaires ex: le noyau, le télomère).

Gene Ontology (http://www.geneontology.org/)



Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data. Read more about the Gene Ontology...

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

MITOSIS

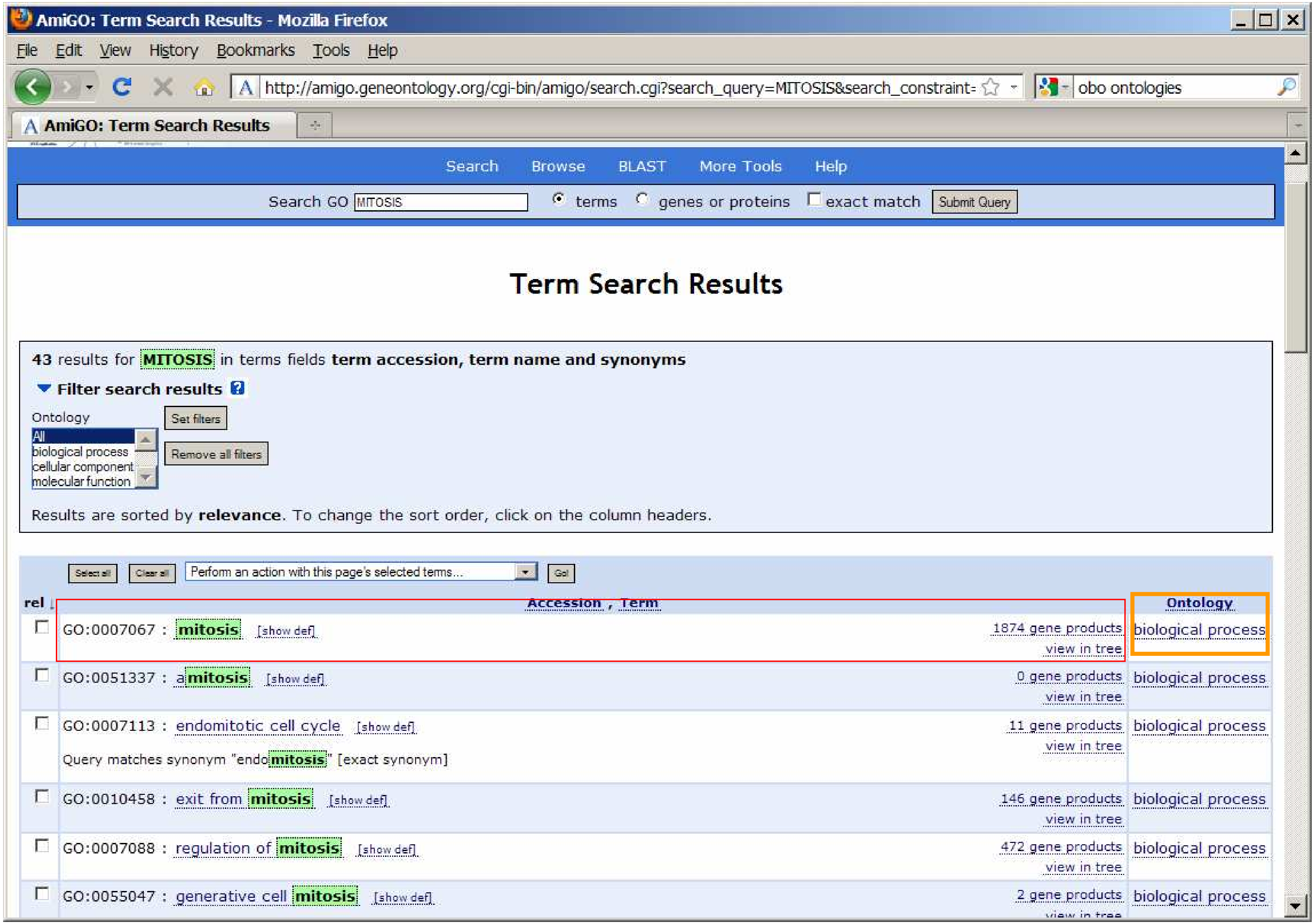
○ gene or protein name ● GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO.

The Gene Ontology project very much encourages input from the community into both the content of the GO and annotation using GO. We are very happy to work with others to ensure that the GO is both complete and accurate, and we also very much encourage communities to submit GO annotations for inclusion in the GO

BBOP services down over the weekend (LBL power outage) (9 days ago) News item

Two positions available at Virginia Bioinformatics



Search GO MITOSIS ● terms ○ genes or proteins ☐ exact match

Term Search Results

43 results for **MITOSIS** in terms fields **term accession, term name and synonyms**

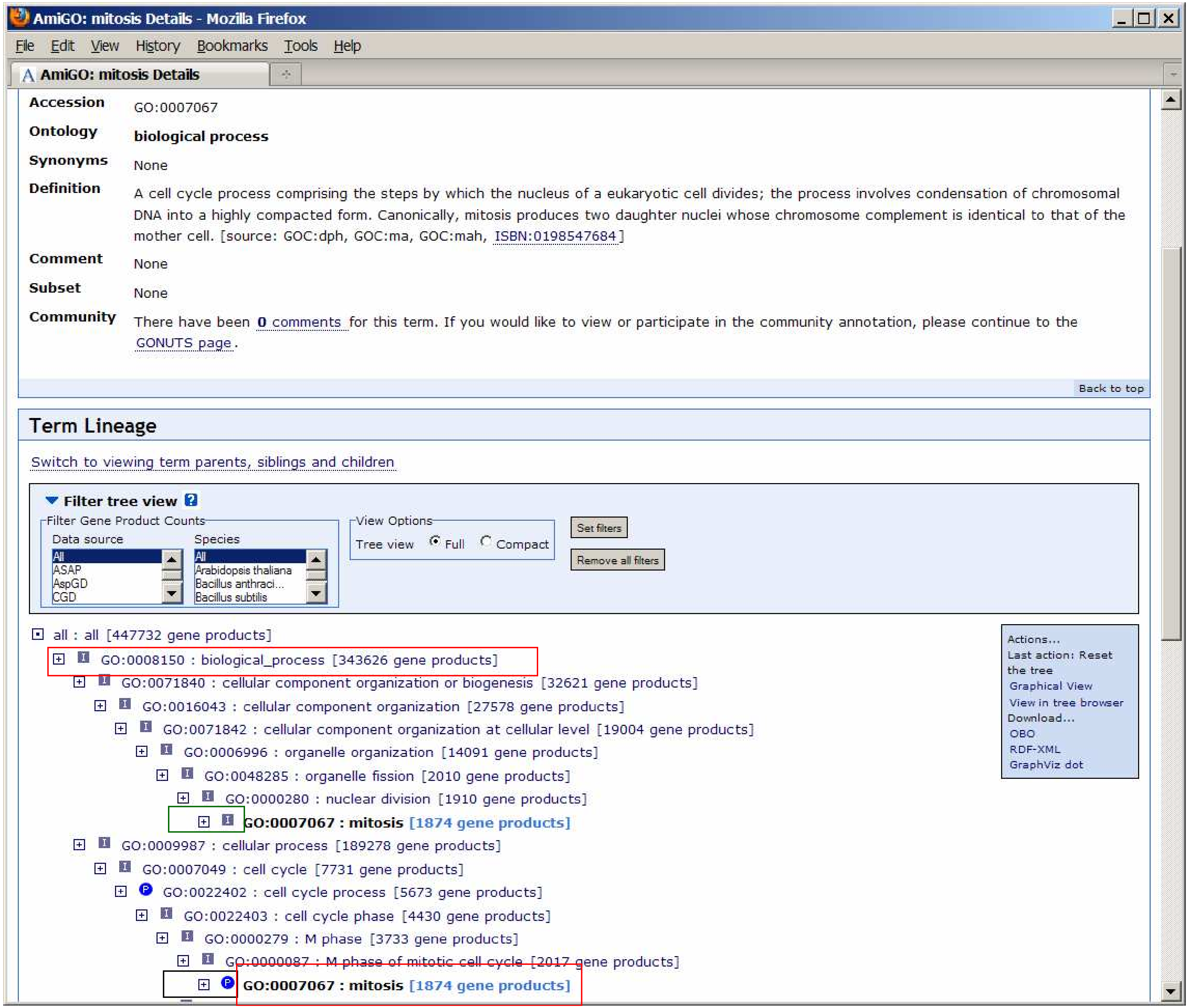
Filter search results

Ontology

All
biological process
cellular component
molecular function

Results are sorted by **relevance**. To change the sort order, click on the column headers.

rel	Accession , Term		Ontology
☐	GO:0007067 : **mitosis** [show def]	1874 gene products view in tree	biological process
☐	GO:0051337 : a**mitosis** [show def]	0 gene products view in tree	biological process
☐	GO:0007113 : endomitotic cell cycle [show def] Query matches synonym "endo**mitosis**" [exact synonym]	11 gene products view in tree	biological process
☐	GO:0010458 : exit from **mitosis** [show def]	146 gene products view in tree	biological process
☐	GO:0007088 : regulation of **mitosis** [show def]	472 gene products view in tree	biological process
☐	GO:0055047 : generative cell **mitosis** [show def]	2 gene products	biological process



Accession	GO:0007067
Ontology	**biological process**
Synonyms	None
Definition	A cell cycle process comprising the steps by which the nucleus of a eukaryotic cell divides; the process involves condensation of chromosomal DNA into a highly compacted form. Canonically, mitosis produces two daughter nuclei whose chromosome complement is identical to that of the mother cell. [source: GOC:dph, GOC:ma, GOC:mah, ISBN:0198547684]
Comment	None
Subset	None
Community	There have been 0 comments for this term. If you would like to view or participate in the community annotation, please continue to the GONUTS page.

Term Lineage

Switch to viewing term parents, siblings and children

Filter tree view

Filter Gene Product Counts

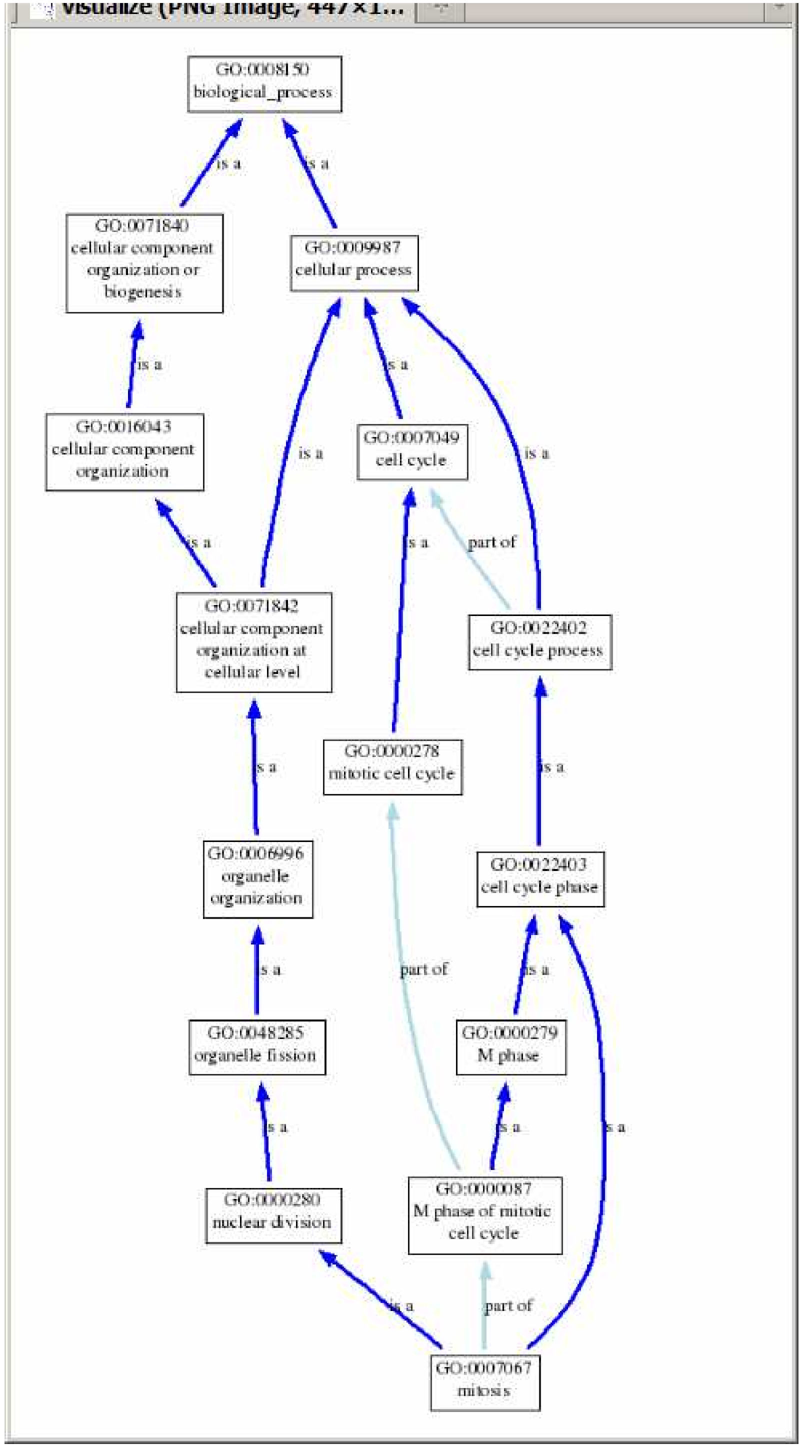
Data source: All, ASAP, AspGD, CGD

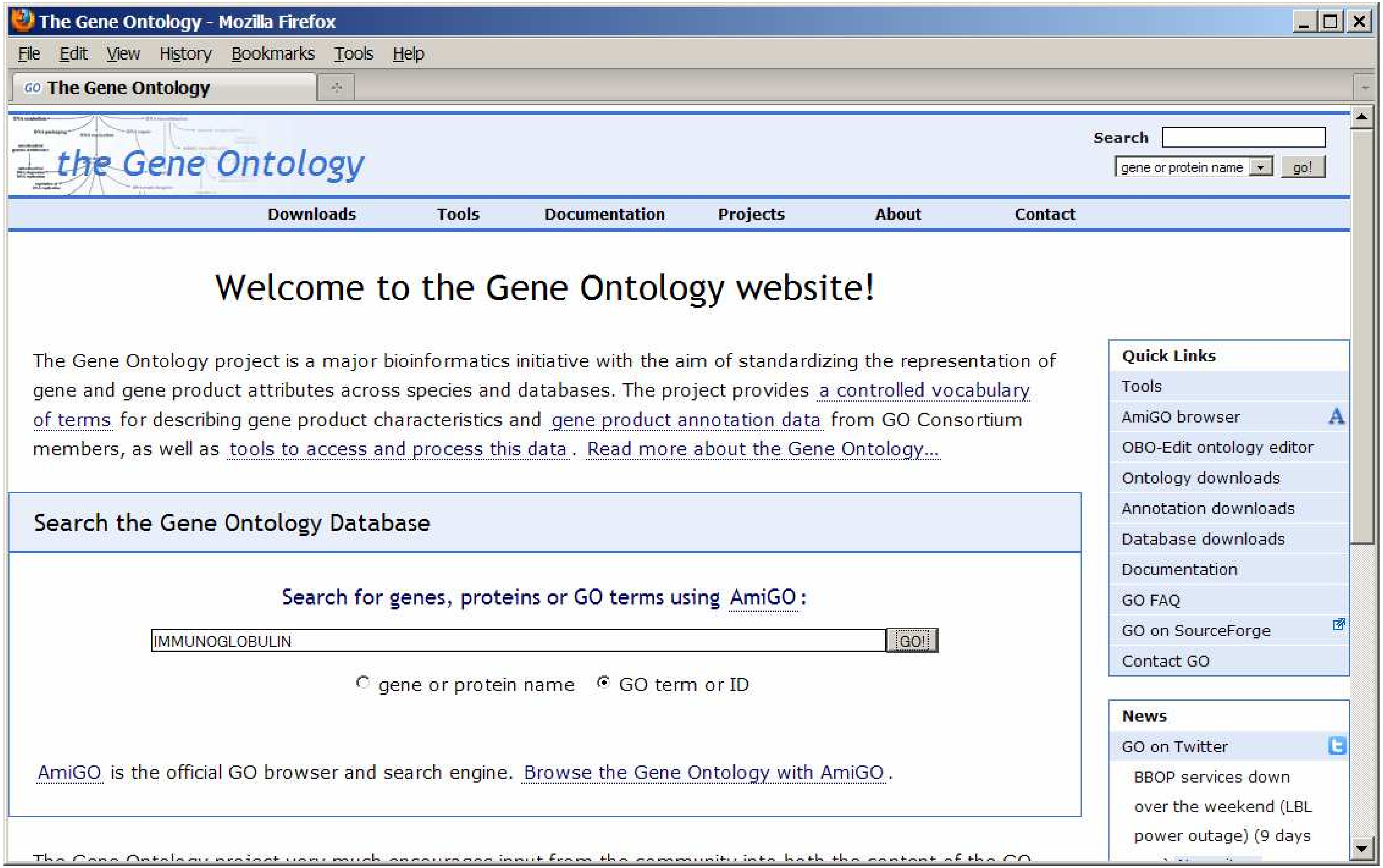
Species: All, Arabidopsis thaliana, Bacillus anthraci..., Bacillus subtilis

View Options

Tree view ● Full ○ Compact

- all : all [447732 gene products]
  - GO:0008150 : biological_process [343626 gene products]
    - GO:0071840 : cellular component organization or biogenesis [32621 gene products]
      - GO:0016043 : cellular component organization [27578 gene products]
        - GO:0071842 : cellular component organization at cellular level [19004 gene products]
          - GO:0006996 : organelle organization [14091 gene products]
            - GO:0048285 : organelle fission [2010 gene products]
              - GO:0000280 : nuclear division [1910 gene products]
                - **GO:0007067 : mitosis** [1874 gene products]
    - GO:0009987 : cellular process [189278 gene products]
      - GO:0007049 : cell cycle [7731 gene products]
        - GO:0022402 : cell cycle process [5673 gene products]
          - GO:0022403 : cell cycle phase [4430 gene products]
            - GO:0000279 : M phase [3733 gene products]
              - GO:0000087 : M phase of mitotic cell cycle [2017 gene products]
                - **GO:0007067 : mitosis** [1874 gene products]


Visualize (PNG Image, 447×1...
GO:0008150 biological_process
is a
is a
GO:0071840 cellular component organization or biogenesis
GO:0009987 cellular process
is a
is a
is a
is a
GO:0016043 cellular component organization
GO:0007049 cell cycle
is a
is a
part of
GO:0071842 cellular component organization at cellular level
GO:0022402 cell cycle process
is a
GO:0000278 mitotic cell cycle
is a
GO:0006996 organelle organization
GO:0022403 cell cycle phase
is a
part of
is a
GO:0048285 organelle fission
GO:0000279 M phase
is a
is a
is a
GO:0000280 nuclear division
GO:0000087 M phase of mitotic cell cycle
is a
part of
GO:0007067 mitosis



Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data. Read more about the Gene Ontology...

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

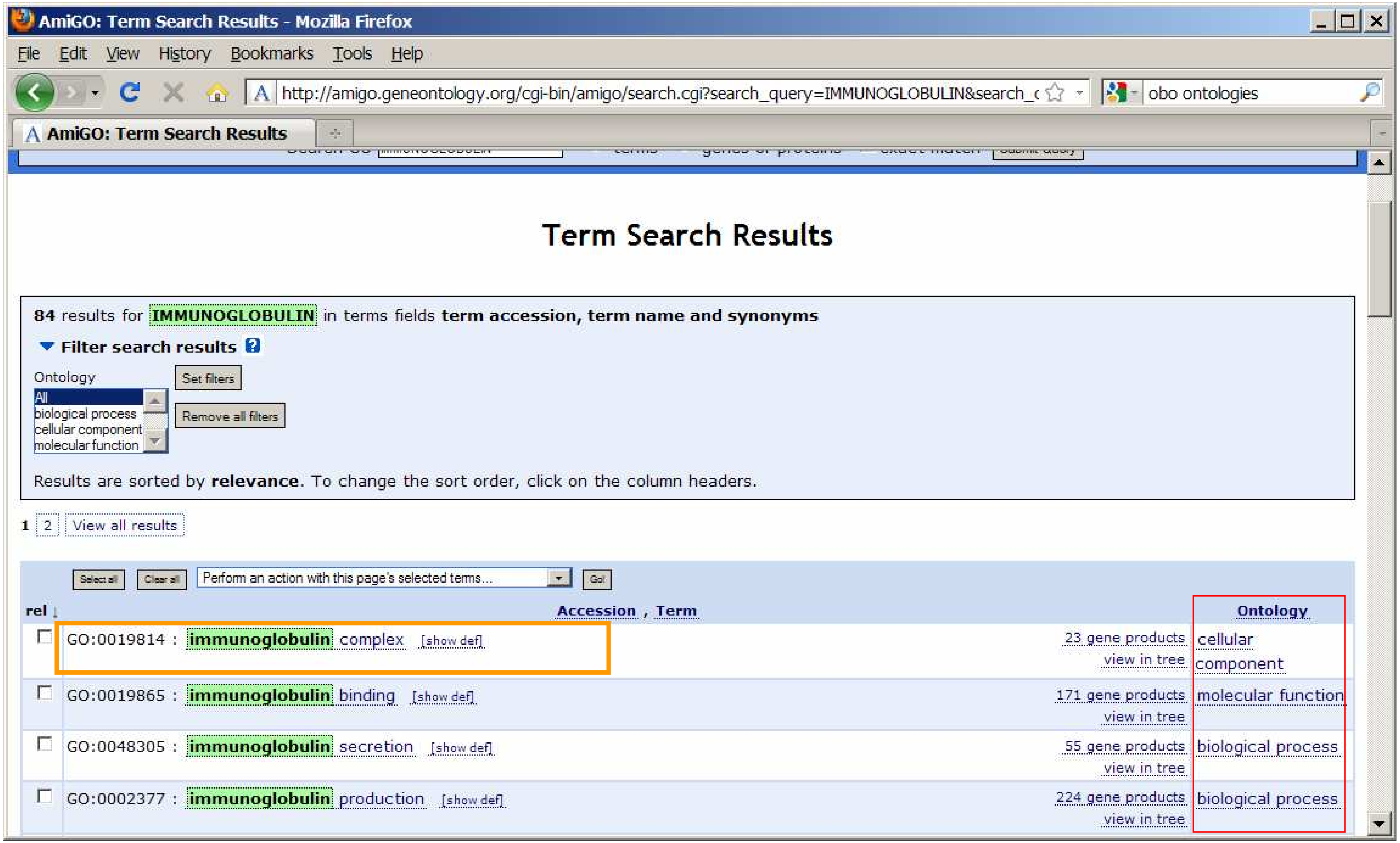
IMMUNOGLOBULIN

○ gene or protein name ● GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO.

The Gene Ontology project very much encourages input from the community into both the content of the GO

BBOP services down over the weekend (LBL power outage) (9 days



Term Search Results

84 results for **IMMUNOGLOBULIN** in terms fields **term accession, term name and synonyms**

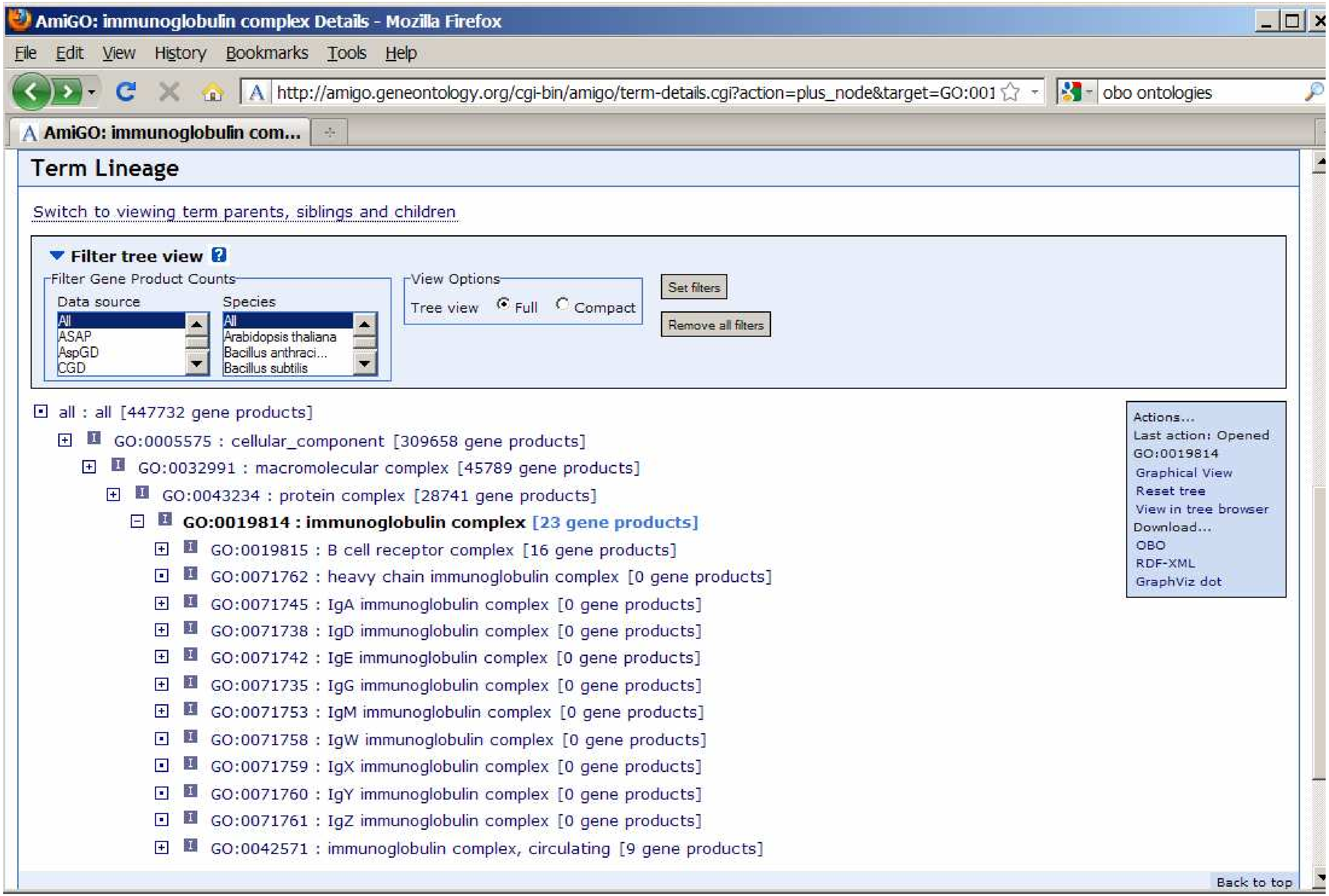
Filter search results

Ontology
All
biological process
cellular component
molecular function

Results are sorted by **relevance**. To change the sort order, click on the column headers.

1 2 View all results

rel	Accession , Term		Ontology
☐	GO:0019814 : immunoglobulin complex [show def]	23 gene products view in tree	cellular component
☐	GO:0019865 : immunoglobulin binding [show def]	171 gene products view in tree	molecular function
☐	GO:0048305 : immunoglobulin secretion [show def]	55 gene products view in tree	biological process
☐	GO:0002377 : immunoglobulin production [show def]	224 gene products view in tree	biological process



Term Lineage

Switch to viewing term parents, siblings and children

Filter tree view

Filter Gene Product Counts

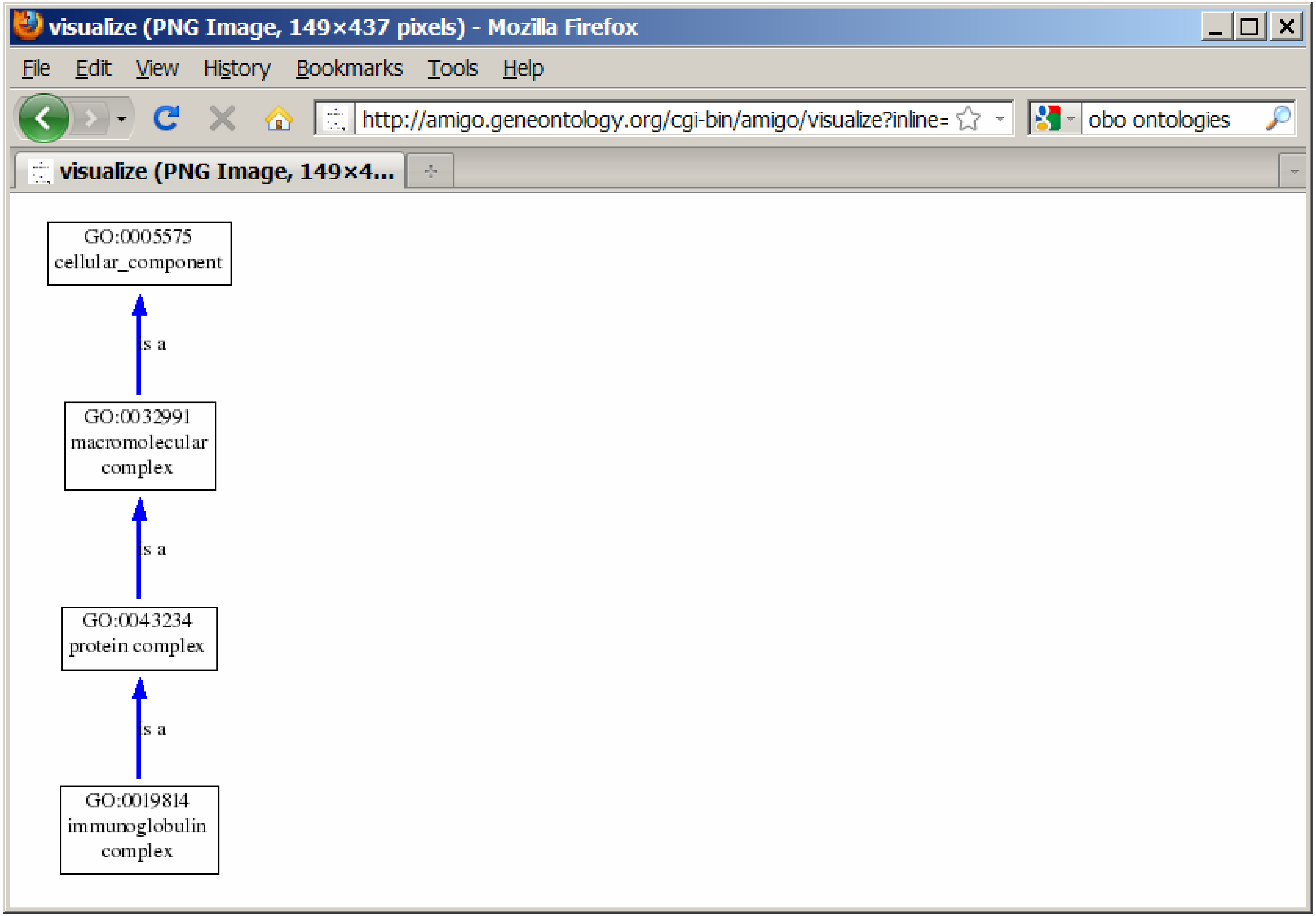
Data source	Species
All	All
ASAP	Arabidopsis thaliana
AspGD	Bacillus anthraci...
CGD	Bacillus subtilis

View Options

Tree view ● Full ○ Compact

```
⊡ all : all [447732 gene products]
  ⊞ GO:0005575 : cellular_component [309658 gene products]
    ⊞ GO:0032991 : macromolecular complex [45789 gene products]
      ⊞ GO:0043234 : protein complex [28741 gene products]
        ⊟ GO:0019814 : immunoglobulin complex [23 gene products]
          ⊞ GO:0019815 : B cell receptor complex [16 gene products]
          ⊡ GO:0071762 : heavy chain immunoglobulin complex [0 gene products]
          ⊞ GO:0071745 : IgA immunoglobulin complex [0 gene products]
          ⊞ GO:0071738 : IgD immunoglobulin complex [0 gene products]
          ⊞ GO:0071742 : IgE immunoglobulin complex [0 gene products]
          ⊞ GO:0071735 : IgG immunoglobulin complex [0 gene products]
          ⊞ GO:0071753 : IgM immunoglobulin complex [0 gene products]
          ⊡ GO:0071758 : IgW immunoglobulin complex [0 gene products]
          ⊡ GO:0071759 : IgX immunoglobulin complex [0 gene products]
          ⊡ GO:0071760 : IgY immunoglobulin complex [0 gene products]
          ⊡ GO:0071761 : IgZ immunoglobulin complex [0 gene products]
          ⊞ GO:0042571 : immunoglobulin complex, circulating [9 gene products]
```

Actions...
Last action: Opened GO:0019814
Graphical View
Reset tree
View in tree browser
Download...
OBO
RDF-XML
GraphViz dot


visualize (PNG Image, 149×437 pixels) - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
http://amigo.geneontology.org/cgi-bin/amigo/visualize?inline=
obo ontologies
visualize (PNG Image, 149×4...
GO:0005575
cellular_component
s a
GO:0032991
macromolecular
complex
s a
GO:0043234
protein complex
s a
GO:0019814
immunoglobulin
complex

Dans le domaine de l’immunogénétique

Problèmes et limites de GO:

- Comment se compose le site de reconnaissance d’un anticorps?

- Quels sont les motifs constitutifs importants?

- Comment caractériser précisément des séquences d’IG (.. et des TR): identification, classification des gènes, description, numérotation des acides aminés, obtention ?

Sequence Ontology

(http://www.sequenceontology.org/)

- Vocabulaire contrôlé pour l’annotation des séquences et l'annotation des génomes

- proposer une représentation structurée des annotations

- vocabulaire pour la description des mutations

Sequence Ontology:
une ontologie pour décrire les séquences biologiques

Welcome to the Sequence Ontology

This is the home page of the Sequence Ontology Project (SO), a joint effort by genome annotation centres, and other groups using sequence annotation data, including: WormBase, FlyBase, the Mouse Genome Informatics group, and the Sanger Institute. We are a part of the Gene Ontology Project and the Open Biomedical Ontologies (OBO) . Our aim is to develop an ontology suitable for describing biological sequences. For questions, please send mail to the SO developers mailing list

Introduction

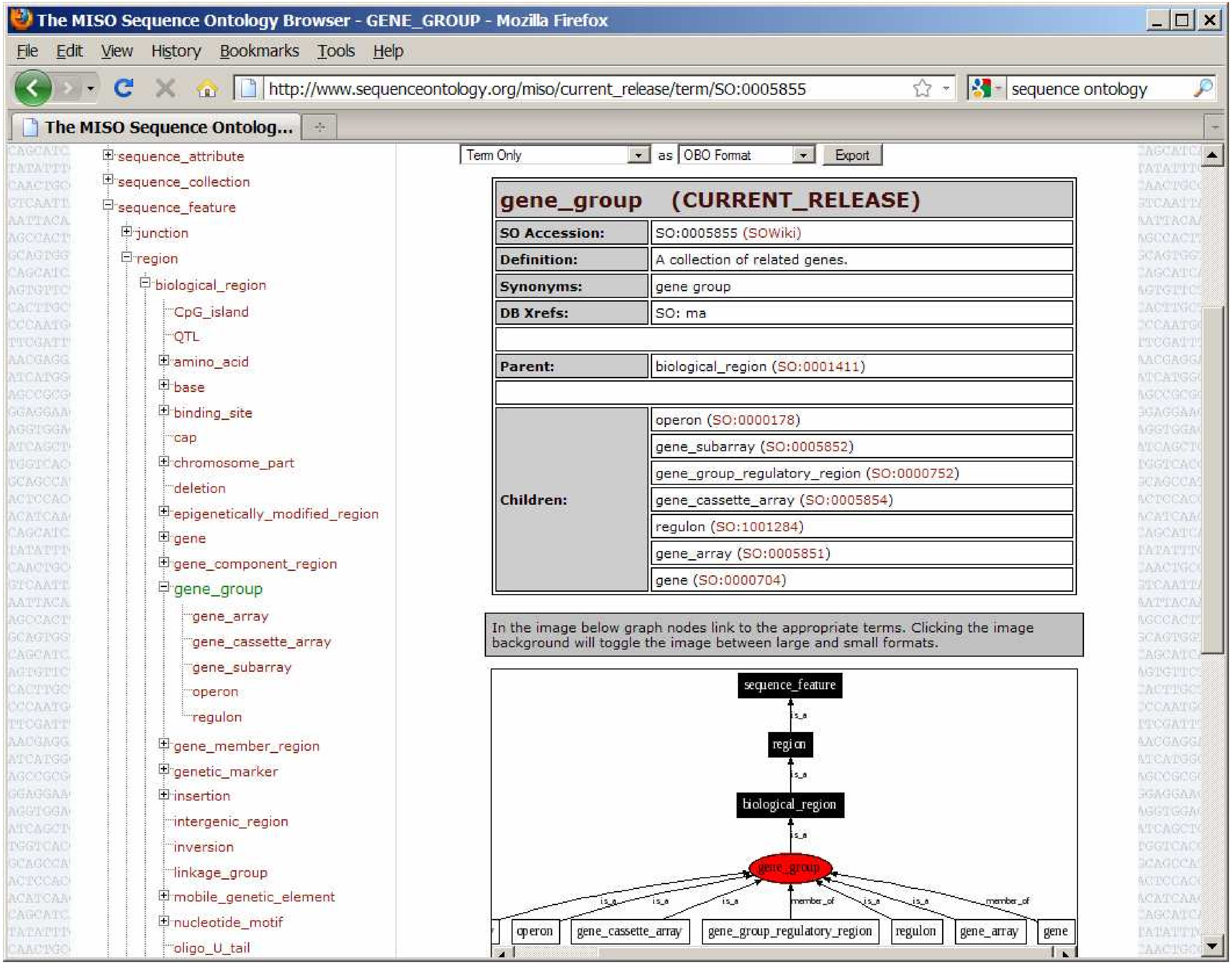
The Sequence Ontology is a set of terms and relationships used to describe the features and attributes of biological sequence. SO includes different kinds of features which can be located on the sequence. Biological features are those which are defined by their disposition to be involved in a biological process. Examples are **binding_site** and **exon**. Biomaterial features are those which are intended for use in an experiment such as **aptamer** and **PCR_product**. There are also experimental features which are the result of an experiment. SO also provides a rich set of attributes to describe these features such as

News

- **August 2010** 'Toward a Richer Representation of Sequence Variation in the Sequence Ontology' with Mike Bada accepted by AIMM Workshop.
- **June 2010** Release 2.4.3 available.
- **May 2010** SOBA paper available from Nucleic Acids Research.
- **April 2010** Release 2.4.2 available.
- **March 2010** New SO paper out: 'Evolution of the

Welcome to MISO, the Sequence Ontology Browser

- Search for a SO term by entering a SO term name or synonym in the query box above;
- Explore the structure of SO and browse for SO terms using the expandable, cascading tree on the left;
- Go to the detail page for a term where you can:
  - Get details about a term, it's definition and relationships;
  - See graphical views of a term's place in the ontology and link to it's neighbors;
  - Export details about a term in a variety of formats;
  - And access and contribute detailed documentation about a term and it's history by linking through to the SO wiki

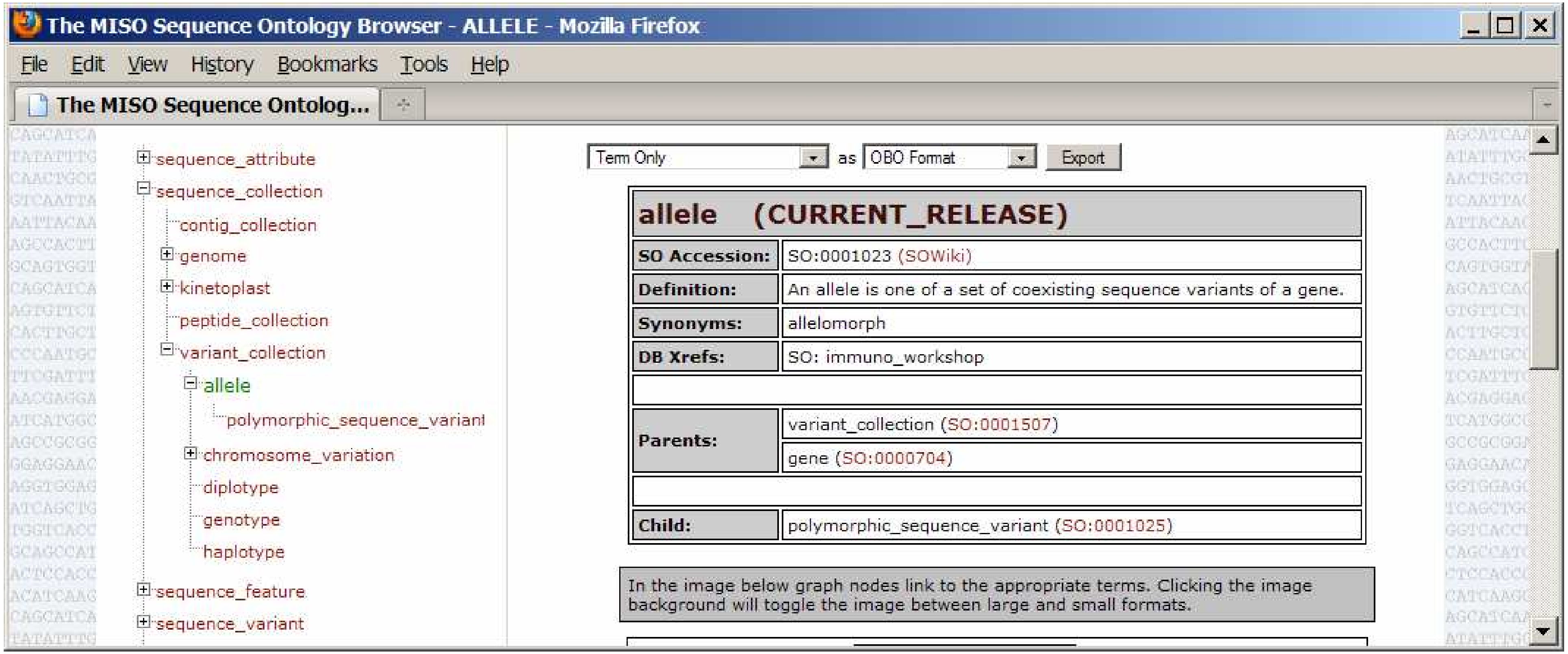


- sequence_attribute
- sequence_collection
- sequence_feature
  - junction
  - region
    - biological_region
      - CpG_island
      - QTL
      - amino_acid
      - base
      - binding_site
      - cap
      - chromosome_part
      - deletion
      - epigenetically_modified_region
      - gene
      - gene_component_region
      - gene_group
        - gene_array
        - gene_cassette_array
        - gene_subarray
        - operon
        - regulon
      - gene_member_region
      - genetic_marker
      - insertion
      - intergenic_region
      - inversion
      - linkage_group
      - mobile_genetic_element
      - nucleotide_motif
      - oligo_U_tail

gene_group (CURRENT_RELEASE)

SO Accession:	SO:0005855 (SOWiki)
Definition:	A collection of related genes.
Synonyms:	gene group
DB Xrefs:	SO: ma
Parent:	biological_region (SO:0001411)
Children:	operon (SO:0000178) gene_subarray (SO:0005852) gene_group_regulatory_region (SO:0000752) gene_cassette_array (SO:0005854) regulon (SO:1001284) gene_array (SO:0005851) gene (SO:0000704)

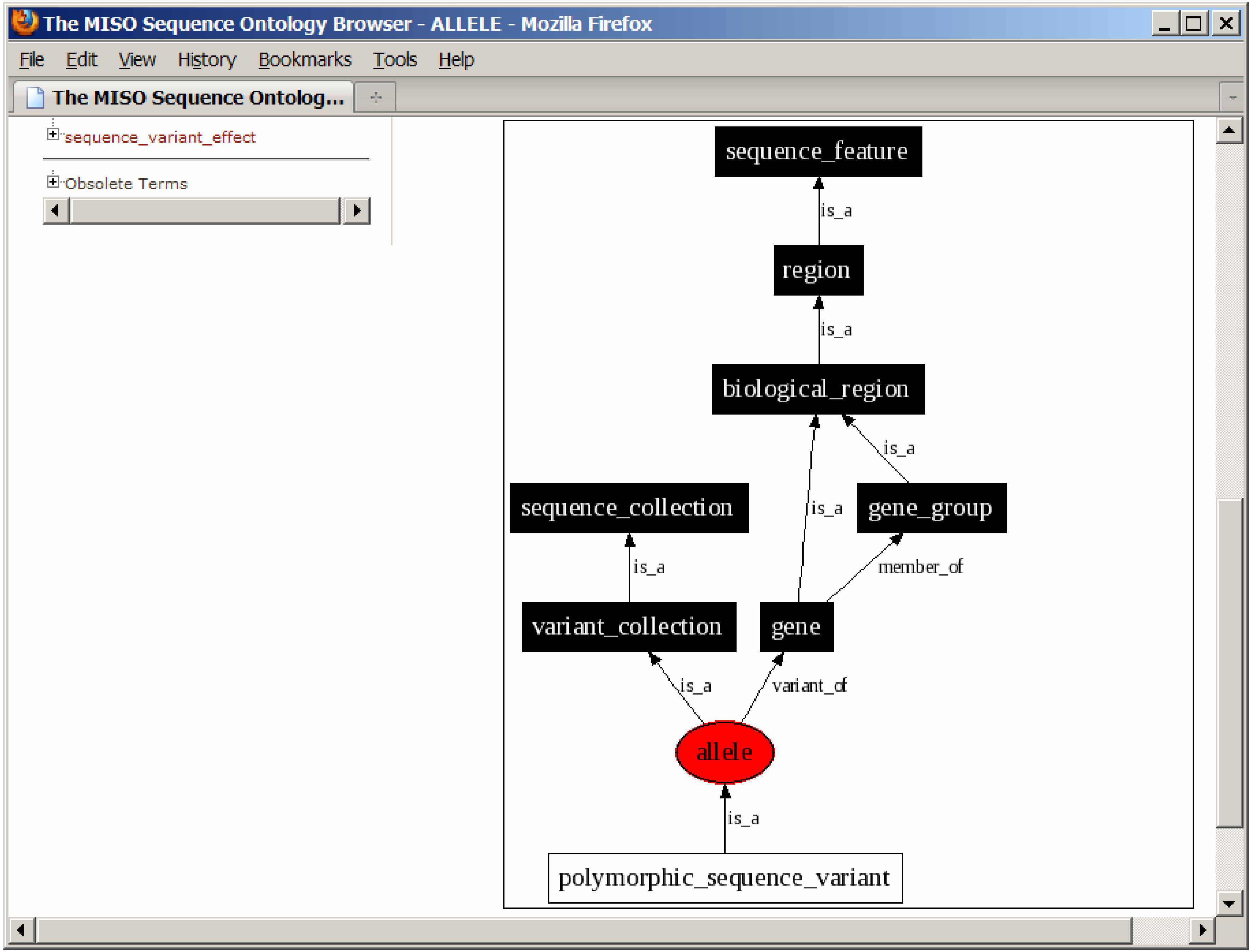
In the image below graph nodes link to the appropriate terms. Clicking the image background will toggle the image between large and small formats.

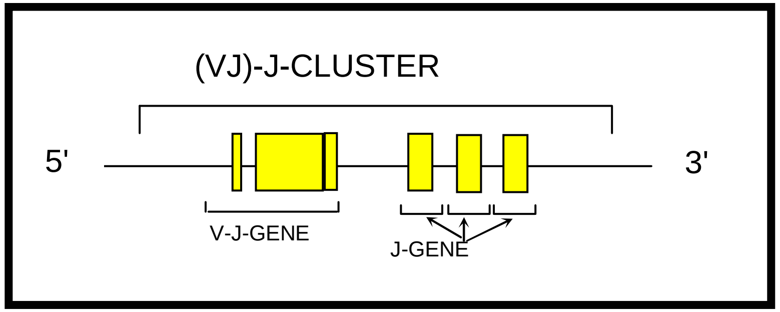


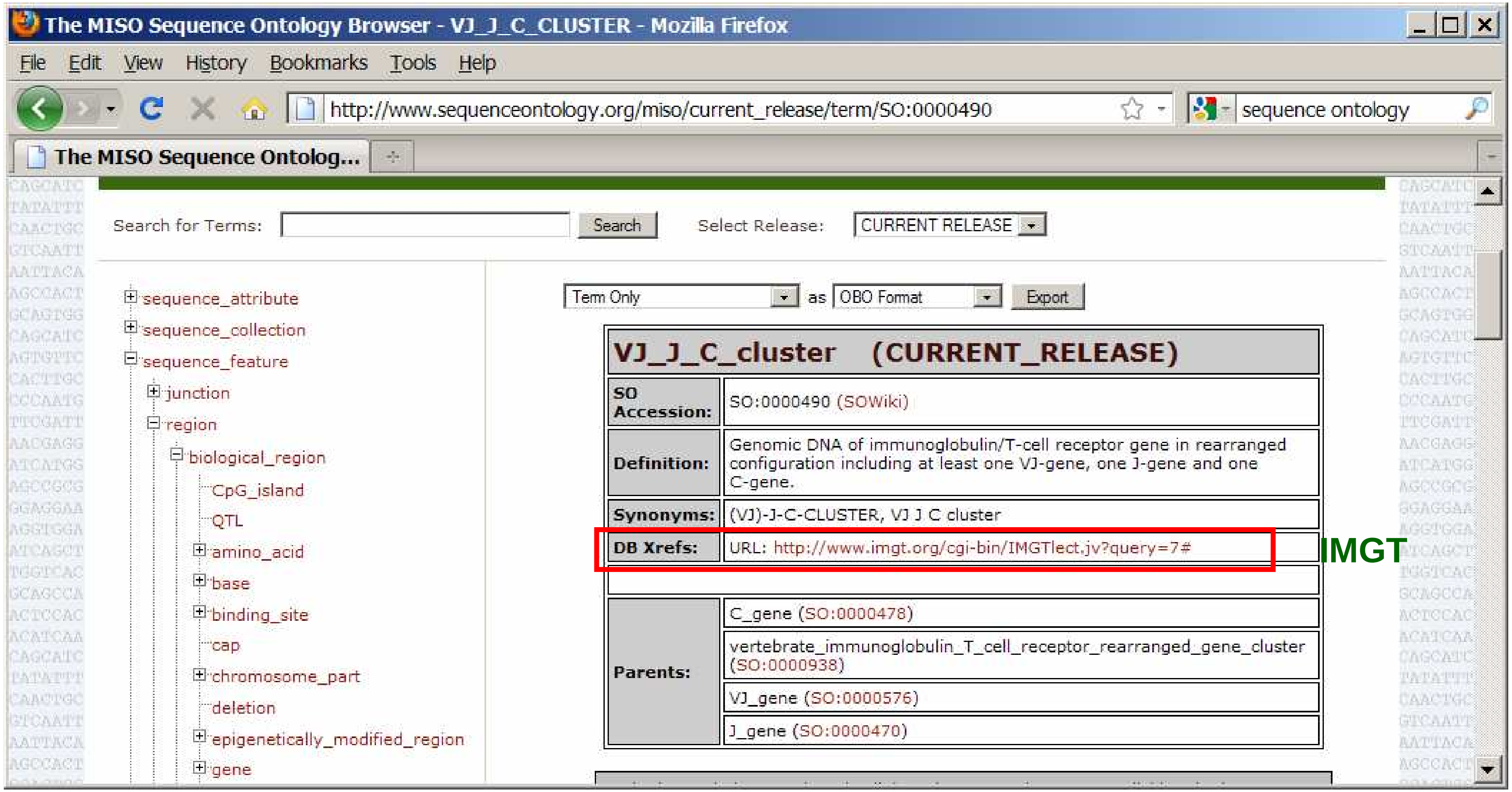
Term Only as OBO Format

allele (CURRENT_RELEASE)	
SO Accession:	SO:0001023 (SOWiki)
Definition:	An allele is one of a set of coexisting sequence variants of a gene.
Synonyms:	allelomorph
DB Xrefs:	SO: immuno_workshop
Parents:	variant_collection (SO:0001507)
	gene (SO:0000704)
Child:	polymorphic_sequence_variant (SO:0001025)

In the image below graph nodes link to the appropriate terms. Clicking the image background will toggle the image between large and small formats.

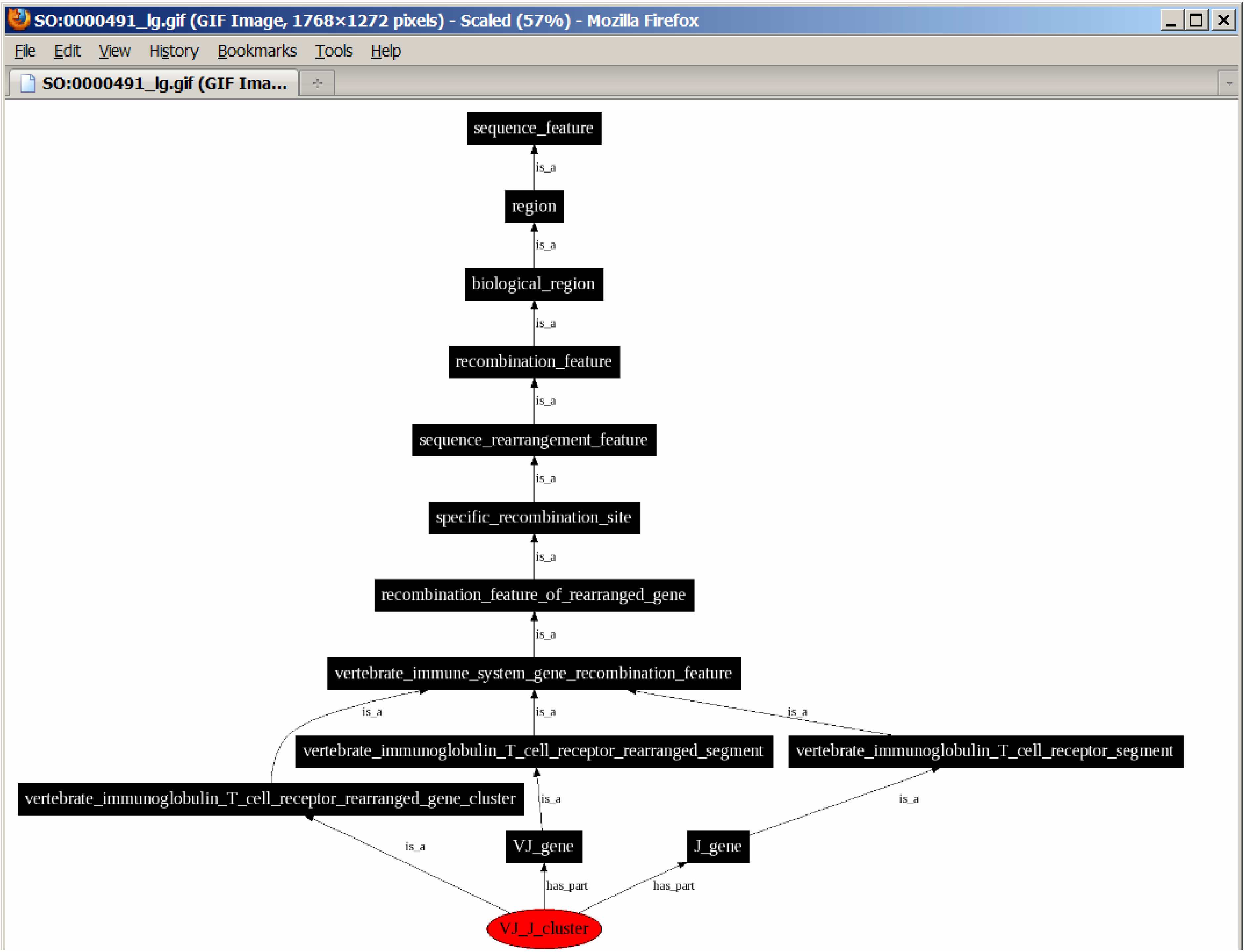

The MISO Sequence Ontology Browser - ALLELE - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
The MISO Sequence Ontolog...
sequence_variant_effect
Obsolete Terms
sequence_feature
is_a
region
is_a
biological_region
is_a
sequence_collection
is_a
gene_group
is_a
member_of
variant_collection
gene
is_a
variant_of
allele
is_a
polymorphic_sequence_variant

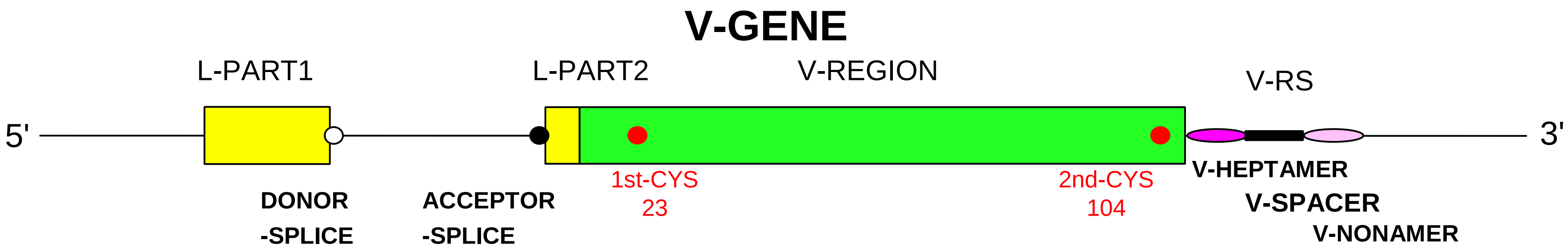




VJ_J_C_cluster (CURRENT_RELEASE)

SO Accession:	SO:0000490 (SOWiki)
Definition:	Genomic DNA of immunoglobulin/T-cell receptor gene in rearranged configuration including at least one VJ-gene, one J-gene and one C-gene.
Synonyms:	(VJ)-J-C-CLUSTER, VJ J C cluster
DB Xrefs:	URL: http://www.imgt.org/cgi-bin/IMGTlect.jv?query=7#
Parents:	C_gene (SO:0000478) vertebrate_immunoglobulin_T_cell_receptor_rearranged_gene_cluster (SO:0000938) VJ_gene (SO:0000576) J_gene (SO:0000470)


SO:0000491_lg.gif (GIF Image, 1768×1272 pixels) - Scaled (57%) - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
SO:0000491_lg.gif (GIF Ima...
sequence_feature
is_a
region
is_a
biological_region
is_a
recombination_feature
is_a
sequence_rearrangement_feature
is_a
specific_recombination_site
is_a
recombination_feature_of_rearranged_gene
is_a
vertebrate_immune_system_gene_recombination_feature
is_a
is_a
is_a
vertebrate_immunoglobulin_T_cell_receptor_rearranged_segment
vertebrate_immunoglobulin_T_cell_receptor_segment
vertebrate_immunoglobulin_T_cell_receptor_rearranged_gene_cluster
is_a
is_a
is_a
VJ_gene
J_gene
has_part
has_part
VJ_J_cluster



V_heptamer (CURRENT_RELEASE)

SO Accession:	SO:0000533 (SOWiki)
Definition:	7 nucleotide recombination site (e.g. CACAGTG), part of V-gene recombination feature of an immunoglobulin/T-cell receptor gene.
Synonyms:	V_heptamer, V-HEPTAMER
DB Xrefs:	URL: http://www.imgt.org/cgi-bin/IMGTlect.jv?query=7#
Parents:	heptamer_of_recombination_feature_of_vertebrate_immune_system_gene (SO:0000561) V_gene_recombination_feature (SO:0000538)

Dans le domaine de l'immunogénétique

Limites de SO:

- Il est nécessaire de prendre en compte d'autres informations:
  type de gènes, type de chaînes, …

- Il faut des relations plus précises que « part_of » entre les motifs

- Comment caractériser précisément des séquences d'IG (.. et des TR): identification, classification des gènes, description, numérotation des acides aminés, obtention ?

=> nécessité d'une ontologie spécifique

IMGT-ONTOLOGY

http://www.imgt.org

Pour standardiser, partager, réutiliser et représenter les connaissances en immunogénétique

IMGT-ONTOLOGY repose sur 7 axiomes:

IDENTIFICATION

CLASSIFICATION

DESCRIPTION

NUMEROTATION

LOCALIZATION

ORIENTATION

OBTENTION

IMGT-ONTOLOGY

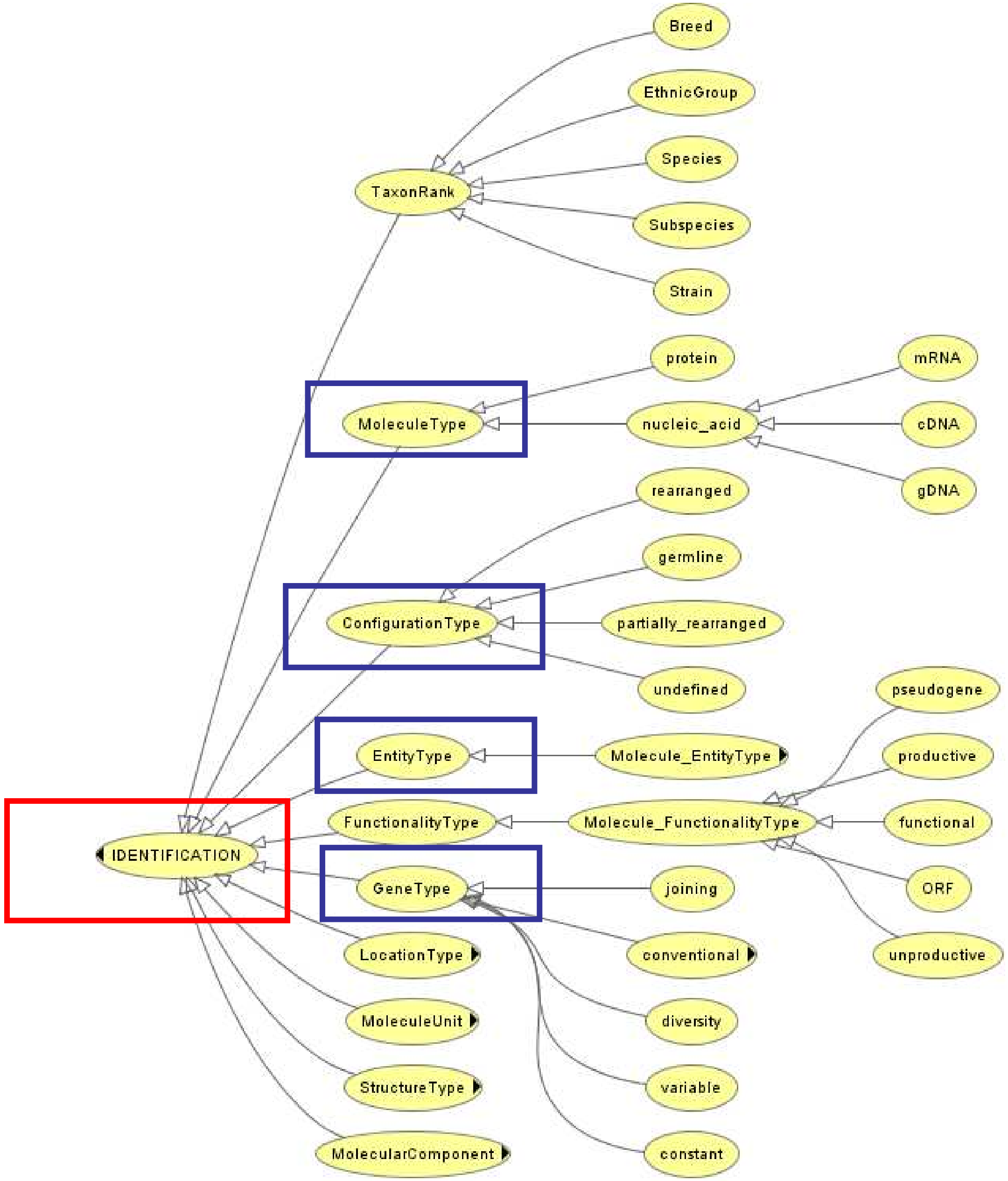
http://www.imgt.org

IMGT Scientific chart :
IMGT-ONTOLOGY en *langage naturel* pour les biologistes
http://imgt.cines.fr/textes/IMGTScientificChart/

En cours : édition en langage OWL
Version v1.0.0 sur le site BioPortal
(http://bioportal.bioontology.org/visualize/42685)

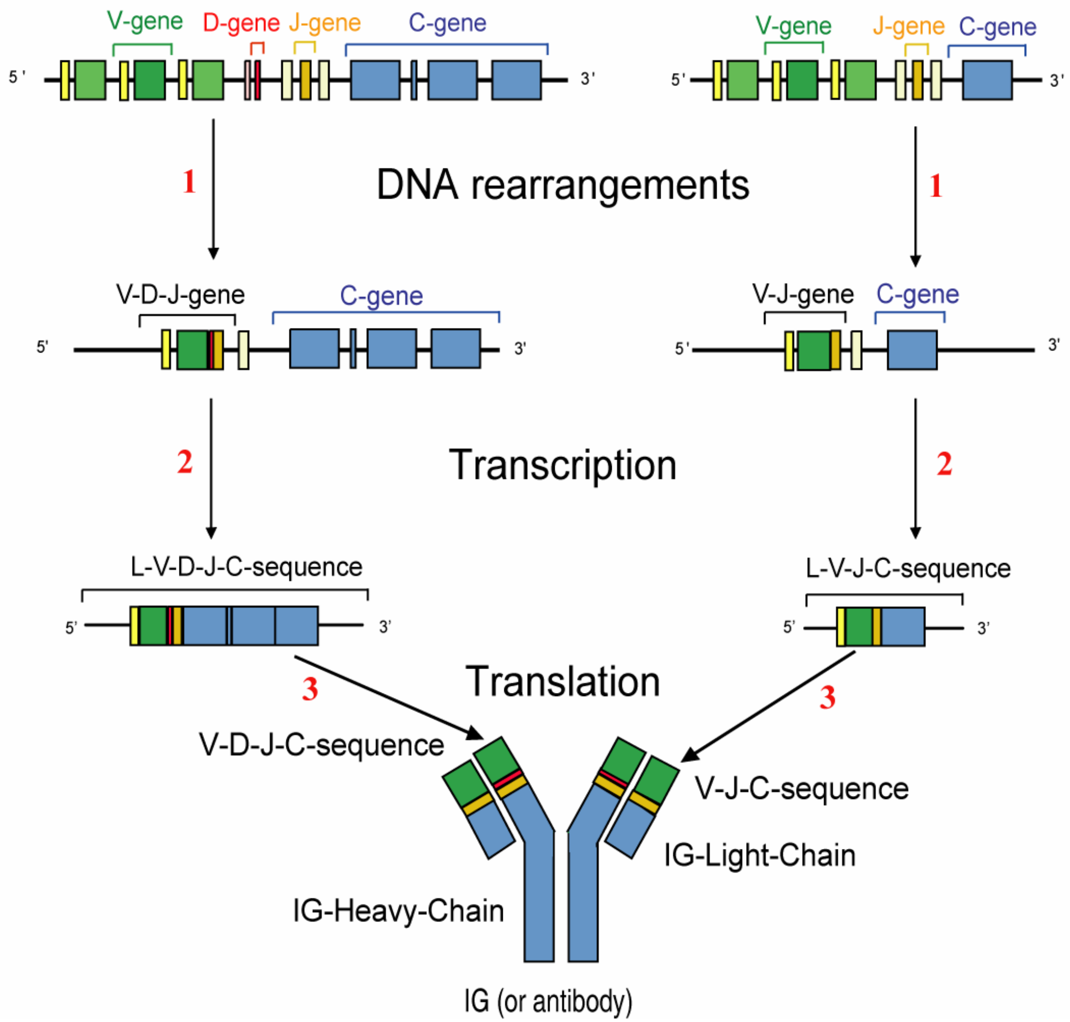
Concepts d’Indentification

http://www.imgt.org



Synthèse des immunoglobulines

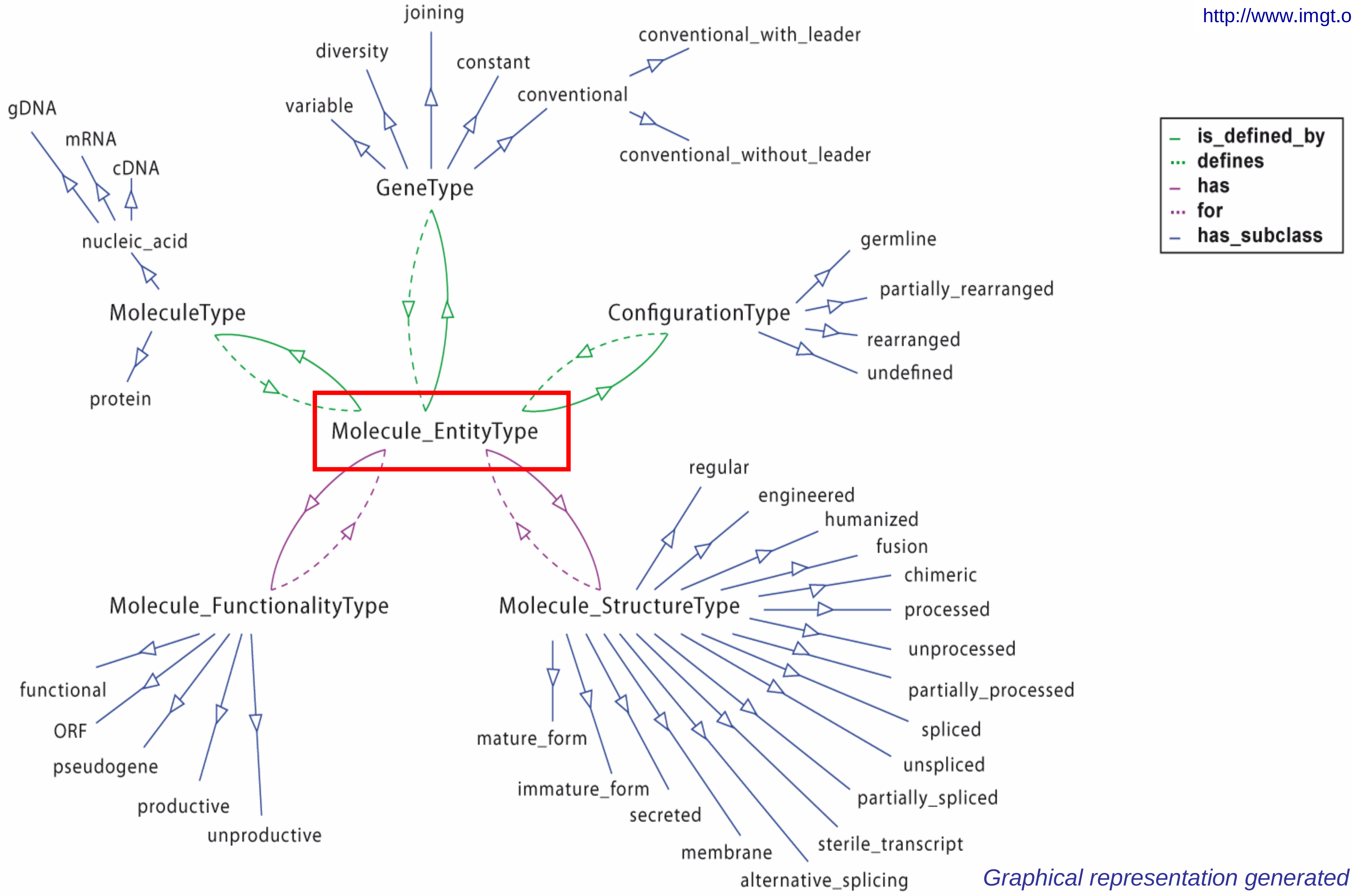
http://imgt.cines.fr
http://www.imgt.org



	"MoleculeType" concept	"ConfigurationType" concept
Genome	gDNA	germline
	gDNA	rearranged
Transcriptome	mRNA (or *in vitro* cDNA)	rearranged
Proteome	protein	rearranged

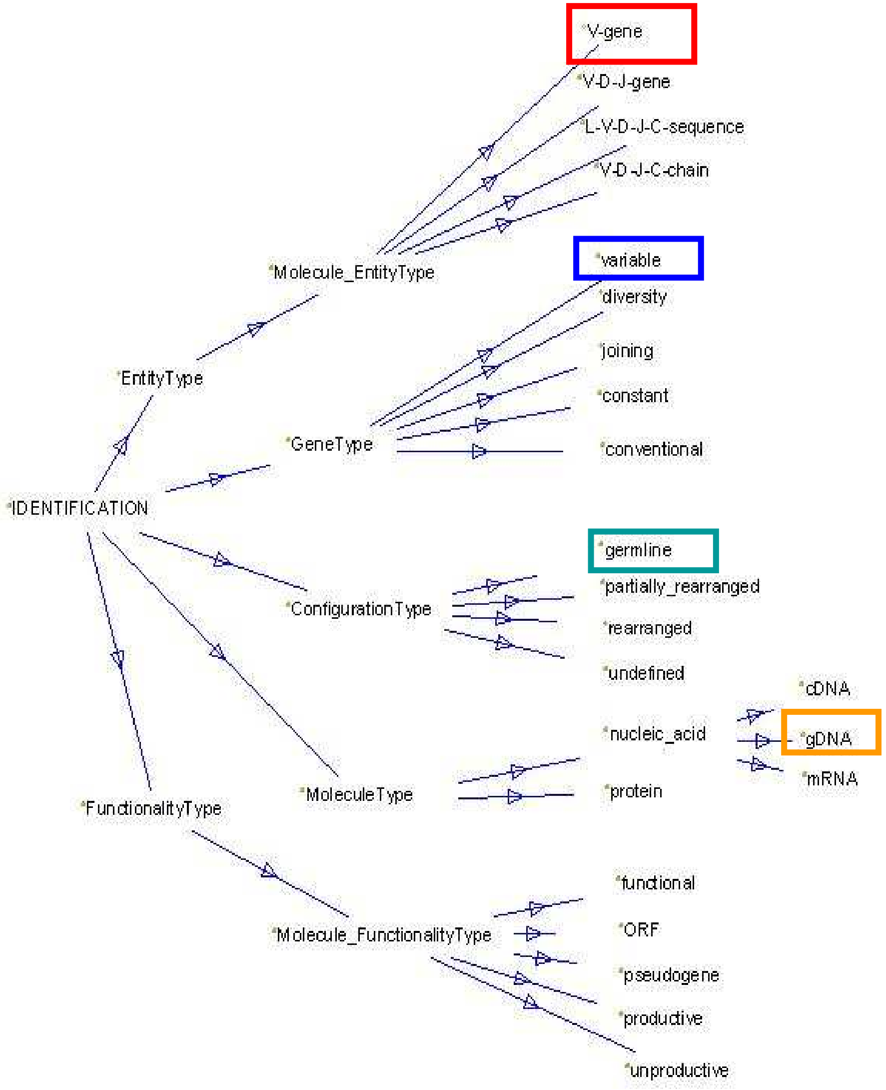
Relationships between Molecule_EntityType and the other IDENTIFICATION concepts

http://www.imgt.org



Graphical representation generated with
the Jambalaya plugin of Protégé

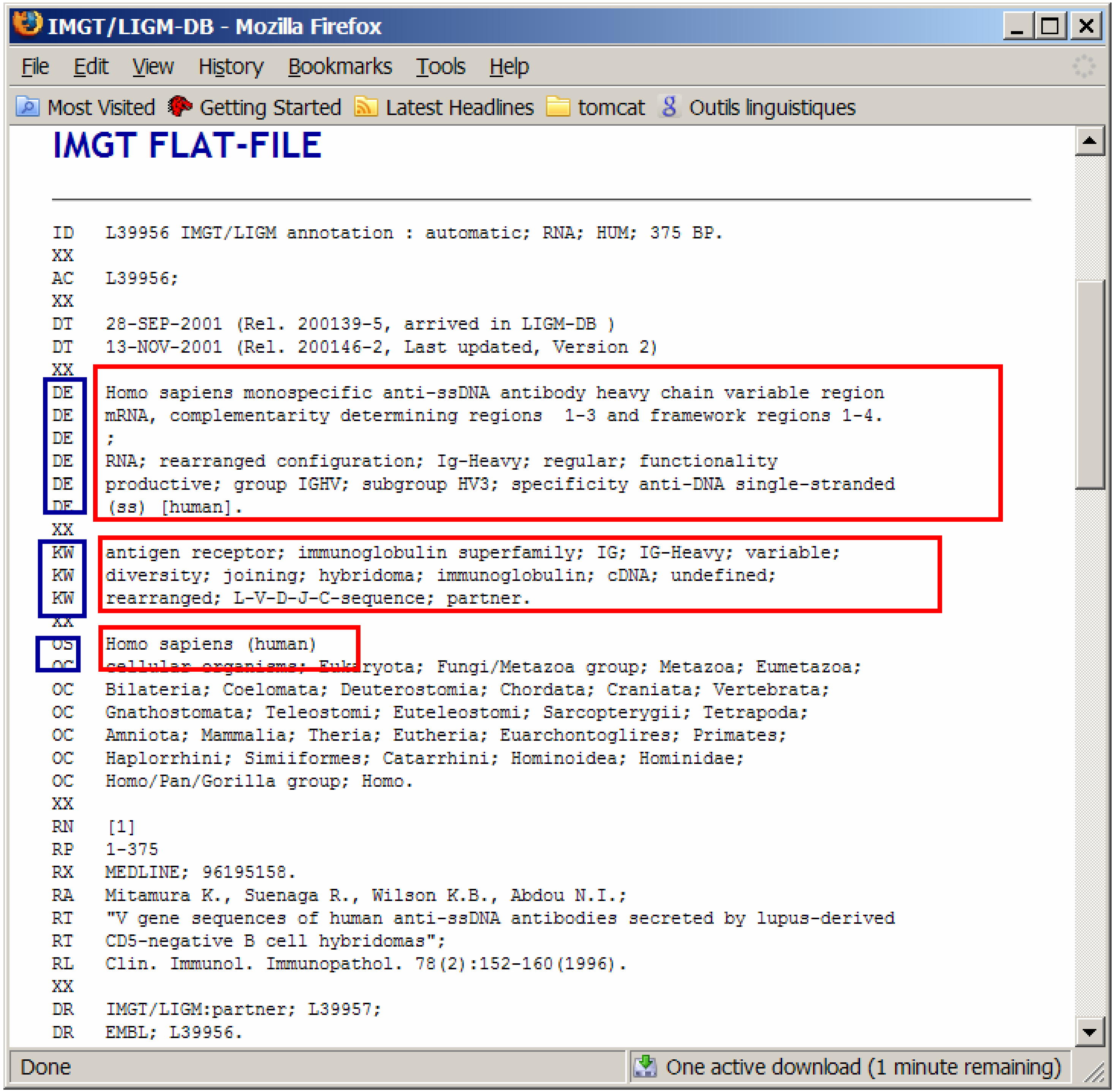
Result: the formalization of the relationships between concepts of IDENTIFICATION allows to represent the rules of the dependency between keywords used in IMGT® databases



IMGT/LIGM-DB flatfile

```
ID   X07448 IMGT/LIGM annotation : by annotators; genomic DNA; HUM; 618 BP.
XX
AC   X07448;
XX
DT   15-MAY-1995 (Rel. 2, arrived in LIGM-DB )
DT   20-OCT-2008 (Rel. 200843-1, Last updated, Version 10)
XX
DE   Human V35 gene for Ig heavy chain  ;
DE   genomic DNA; germline configuration; Ig-Heavy; regular; functionality
DE   functional; group IGHV; subgroup HV1.
XX
KW   antigen receptor; Immunoglobulin superfamily (IgSF);
KW   immunoglobulin (IG); IG-Heavy; variable; IMGT reference sequence; gDNA;
KW   germline; functional; V-gene.
XX
OS   Homo sapiens (human)
OC   cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;
OC   Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata;
OC   Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda;
OC   Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates;
OC   Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae;
OC   Homo/Pan/Gorilla group; Homo.
XX
RN   [1]
RP   1-618
RX   PUBMED; 2841108.
RA   Matsuda F., Lee K.H., Nakai S., Sato T., Kodaira M., Zong S.Q., Ohno
H.,
RA   Fukuhara S., Honjo T.;
RT   "Dispersed localization of D segments in the human immunoglobulin
RT   heavy-chain locus";
RL   EMBO J. 7(4):1047-1051(1988).
XX
```

http://www.imgt.org



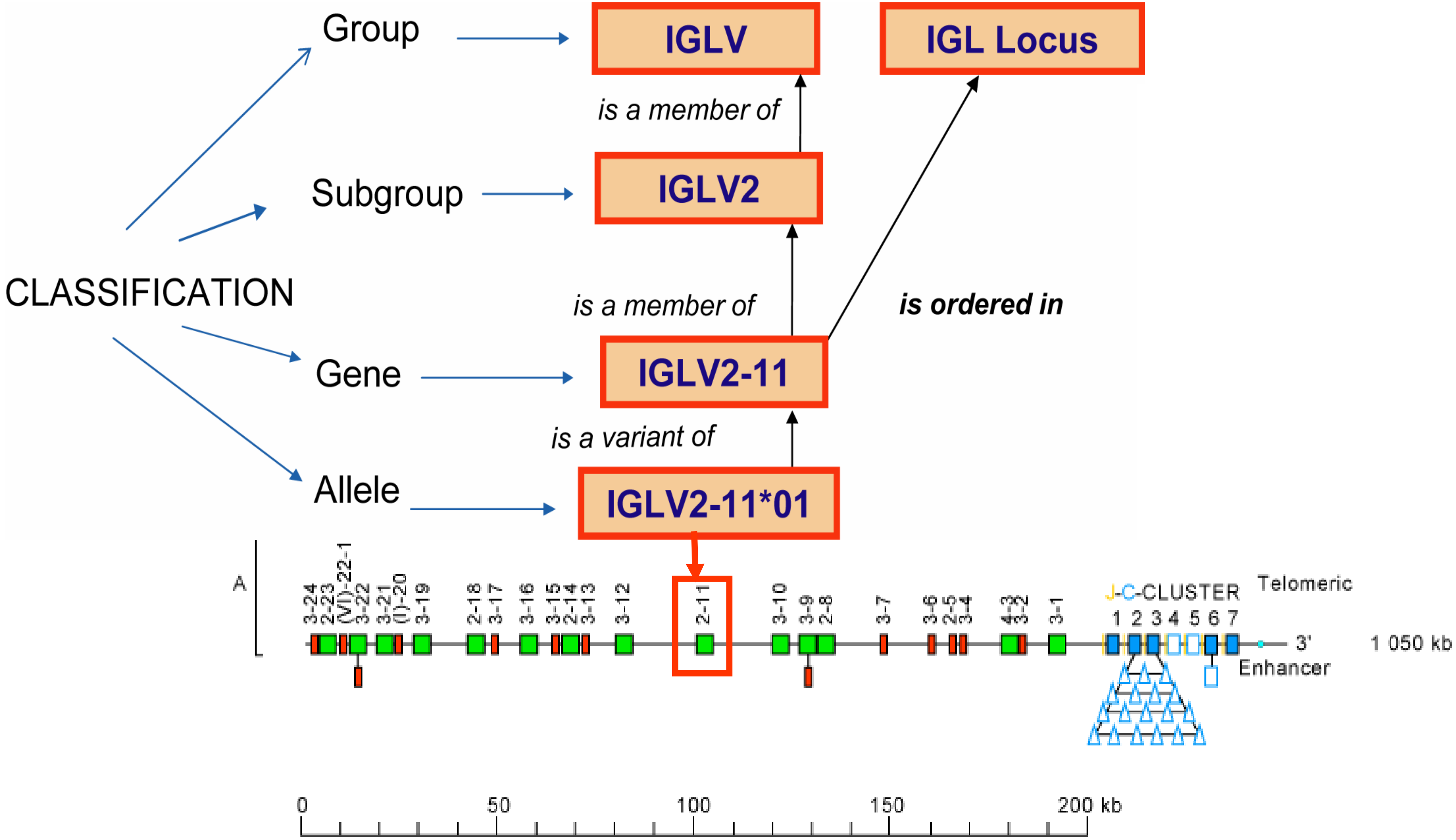
IMGT FLAT-FILE

```
ID   L39956 IMGT/LIGM annotation : automatic; RNA; HUM; 375 BP.
XX
AC   L39956;
XX
DT   28-SEP-2001 (Rel. 200139-5, arrived in LIGM-DB )
DT   13-NOV-2001 (Rel. 200146-2, Last updated, Version 2)
XX
DE   Homo sapiens monospecific anti-ssDNA antibody heavy chain variable region
DE   mRNA, complementarity determining regions  1-3 and framework regions 1-4.
DE   ;
DE   RNA; rearranged configuration; Ig-Heavy; regular; functionality
DE   productive; group IGHV; subgroup HV3; specificity anti-DNA single-stranded
DE   (ss) [human].
XX
KW   antigen receptor; immunoglobulin superfamily; IG; IG-Heavy; variable;
KW   diversity; joining; hybridoma; immunoglobulin; cDNA; undefined;
KW   rearranged; L-V-D-J-C-sequence; partner.
XX
OS   Homo sapiens (human)
OC   cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;
OC   Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata;
OC   Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda;
OC   Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates;
OC   Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae;
OC   Homo/Pan/Gorilla group; Homo.
XX
RN   [1]
RP   1-375
RX   MEDLINE; 96195158.
RA   Mitamura K., Suenaga R., Wilson K.B., Abdou N.I.;
RT   "V gene sequences of human anti-ssDNA antibodies secreted by lupus-derived
RT   CD5-negative B cell hybridomas";
RL   Clin. Immunol. Immunopathol. 78(2):152-160(1996).
XX
DR   IMGT/LIGM:partner; L39957;
DR   EMBL; L39956.
```

Concepts of CLASSIFICATION: IMGT® gene nomenclature

http://www.imgt.org

Human IGL locus (22q11.2)



CLASSIFICATION

Group → **IGLV**

Subgroup → **IGLV2**

Gene → **IGLV2-11**

Allele → **IGLV2-11*01**

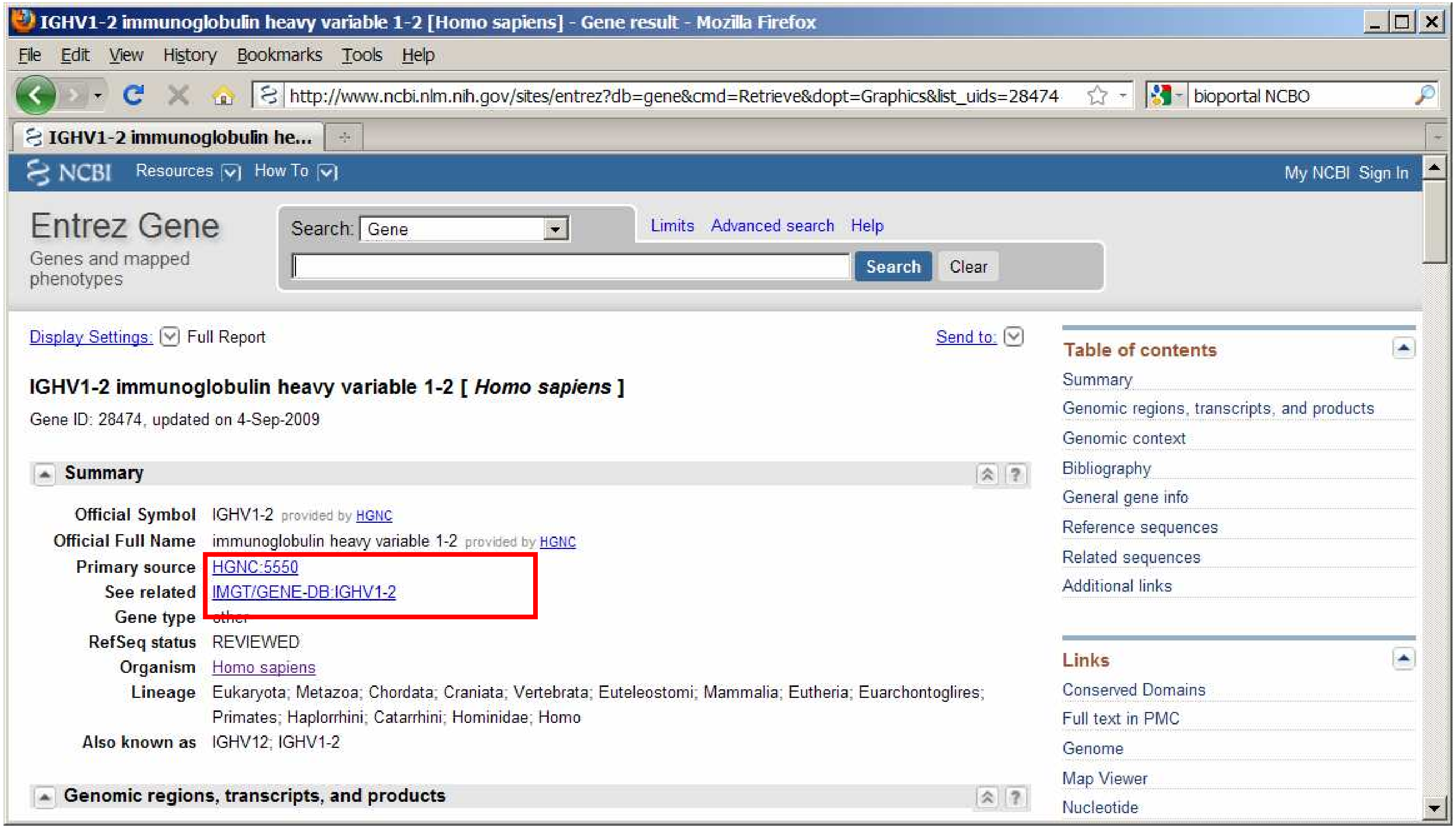
is a member of

is a member of

is a variant of

IGL Locus

is ordered in



IGHV1-2 immunoglobulin heavy variable 1-2 [*Homo sapiens*]

Gene ID: 28474, updated on 4-Sep-2009

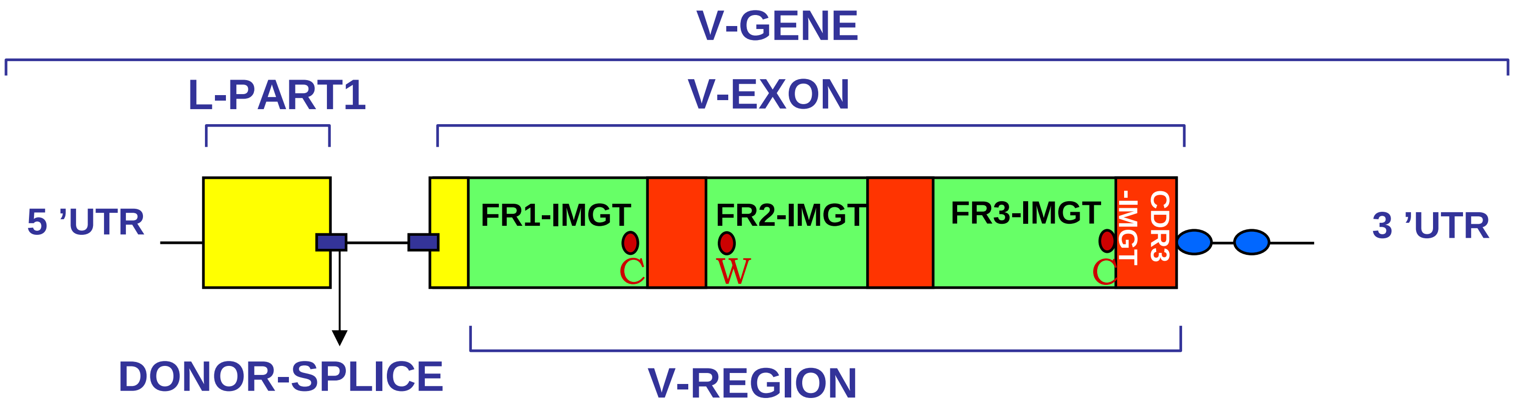
Summary

Official Symbol	IGHV1-2 provided by HGNC
Official Full Name	immunoglobulin heavy variable 1-2 provided by HGNC
Primary source	HGNC:5550
See related	IMGT/GENE-DB:IGHV1-2
Gene type	other
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	IGHV12; IGHV1-2

Genomic regions, transcripts, and products

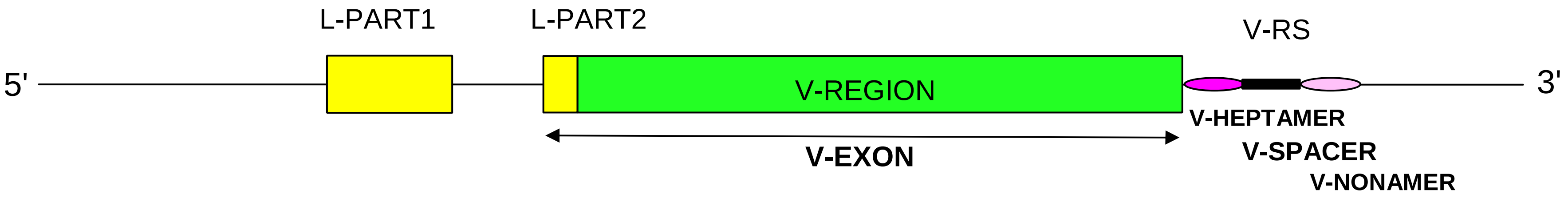
"DESCRIPTION"

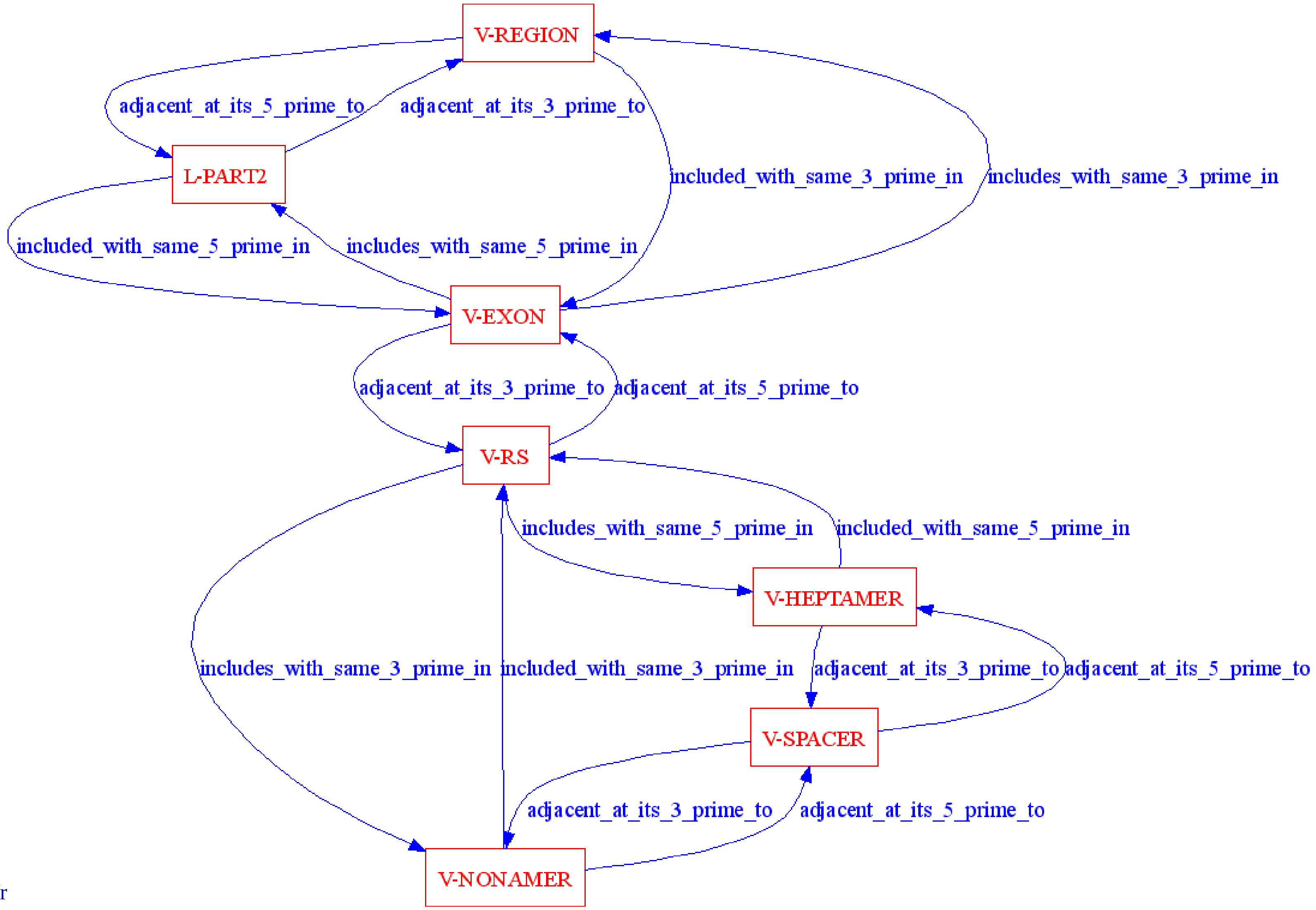
http://www.imgt.org



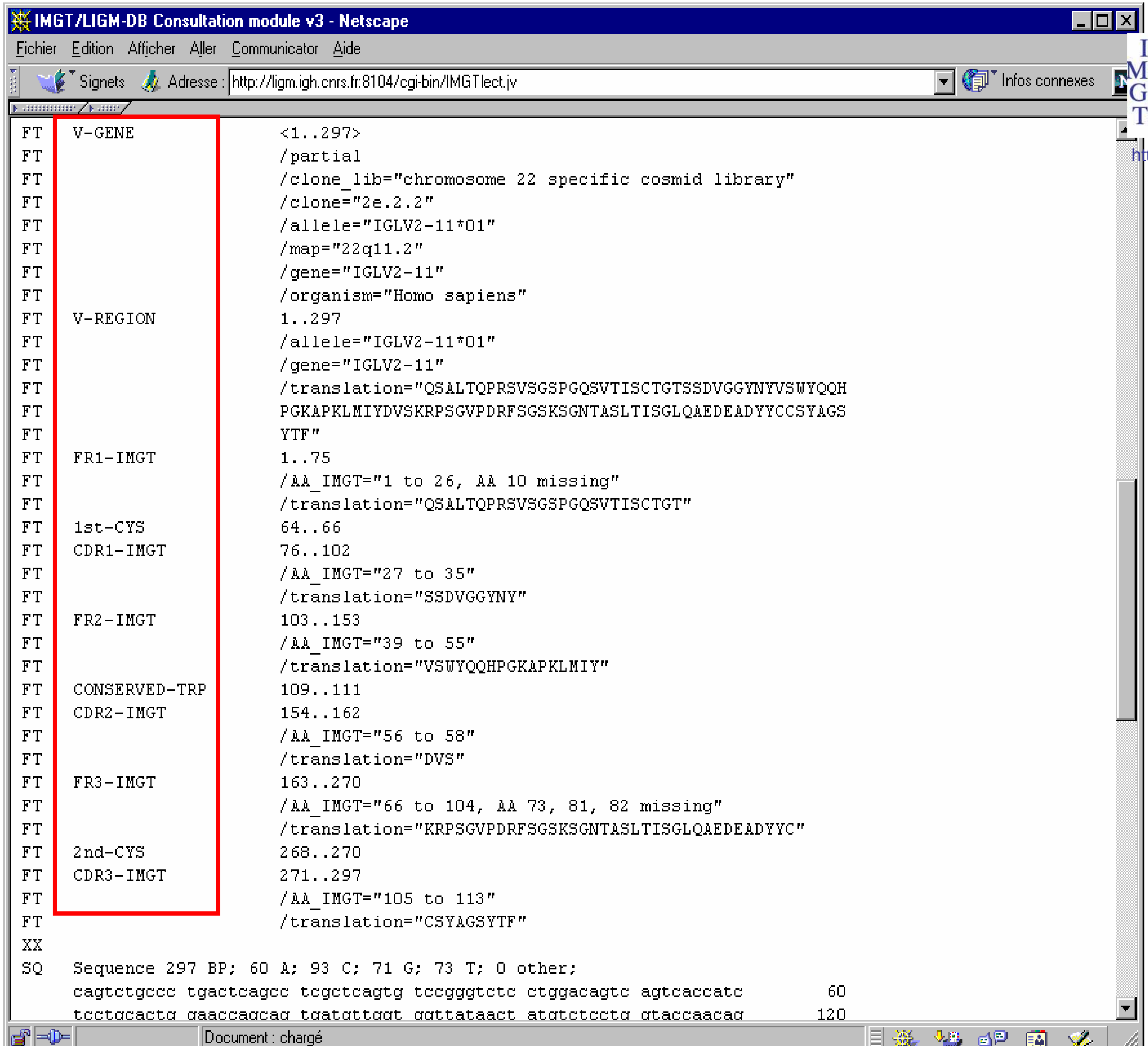
Label 1	**Label**	**Label relations**
V-GENE	V-EXON	
FR3-IMGT	CDR3-IMGT	
L-PART1	DONOR-SPLICE	
V-REGION	FR1-IMGT	
V-REGION	CDR3-IMGT	

V-GENE





http://imgt.cines.fr



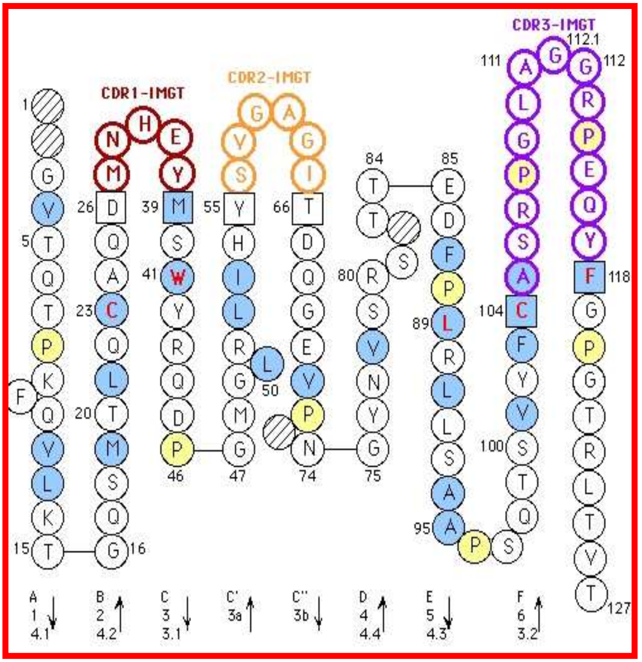
```
FT   V-GENE          <1..297>
FT                   /partial
FT                   /clone_lib="chromosome 22 specific cosmid library"
FT                   /clone="2e.2.2"
FT                   /allele="IGLV2-11*01"
FT                   /map="22q11.2"
FT                   /gene="IGLV2-11"
FT                   /organism="Homo sapiens"
FT   V-REGION        1..297
FT                   /allele="IGLV2-11*01"
FT                   /gene="IGLV2-11"
FT                   /translation="QSALTQPRSVSGSPGQSVTISCTGTSSDVGGYNYVSWYQQH
FT                   PGKAPKLMIYDVSKRPSGVPDRFSGSKSGNTASLTISGLQAEDEADYYCCSYAGS
FT                   YTF"
FT   FR1-IMGT        1..75
FT                   /AA_IMGT="1 to 26, AA 10 missing"
FT                   /translation="QSALTQPRSVSGSPGQSVTISCTGT"
FT   1st-CYS         64..66
FT   CDR1-IMGT       76..102
FT                   /AA_IMGT="27 to 35"
FT                   /translation="SSDVGGYNY"
FT   FR2-IMGT        103..153
FT                   /AA_IMGT="39 to 55"
FT                   /translation="VSWYQQHPGKAPKLMIY"
FT   CONSERVED-TRP   109..111
FT   CDR2-IMGT       154..162
FT                   /AA_IMGT="56 to 58"
FT                   /translation="DVS"
FT   FR3-IMGT        163..270
FT                   /AA_IMGT="66 to 104, AA 73, 81, 82 missing"
FT                   /translation="KRPSGVPDRFSGSKSGNTASLTISGLQAEDEADYYC"
FT   2nd-CYS         268..270
FT   CDR3-IMGT       271..297
FT                   /AA_IMGT="105 to 113"
FT                   /translation="CSYAGSYTF"
XX
SQ   Sequence 297 BP; 60 A; 93 C; 71 G; 73 T; 0 other;
     cagtctgccc tgactcagcc tcgctcagtg tccgggtctc ctggacagtc agtcaccatc        60
     tcctgcactg gaaccagcag tgatgttggt ggttataact atgtctcctg gtaccaacag       120
```

http://www.imgt.org

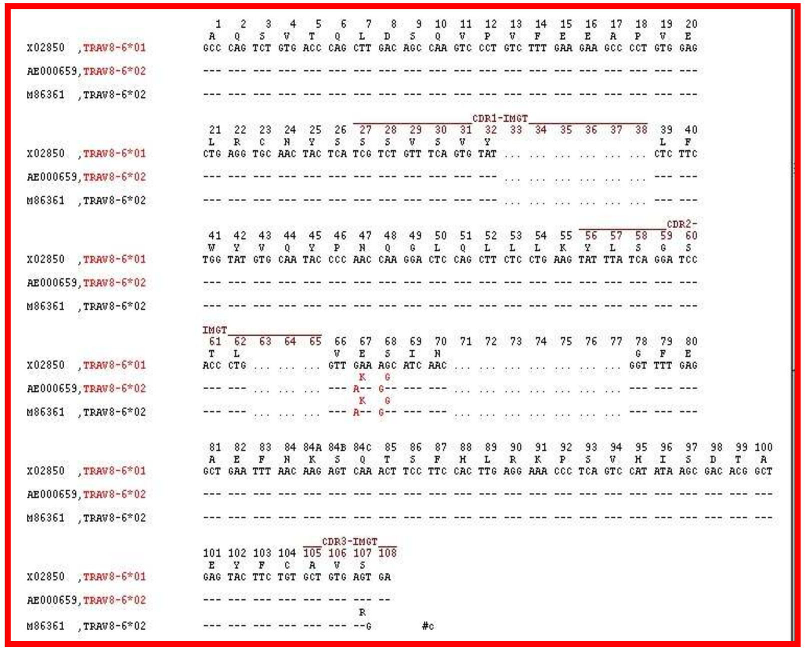
"NUMEROTATION"

http://imgt.cines.fr
http://www.imgt.org

Collier De Perles

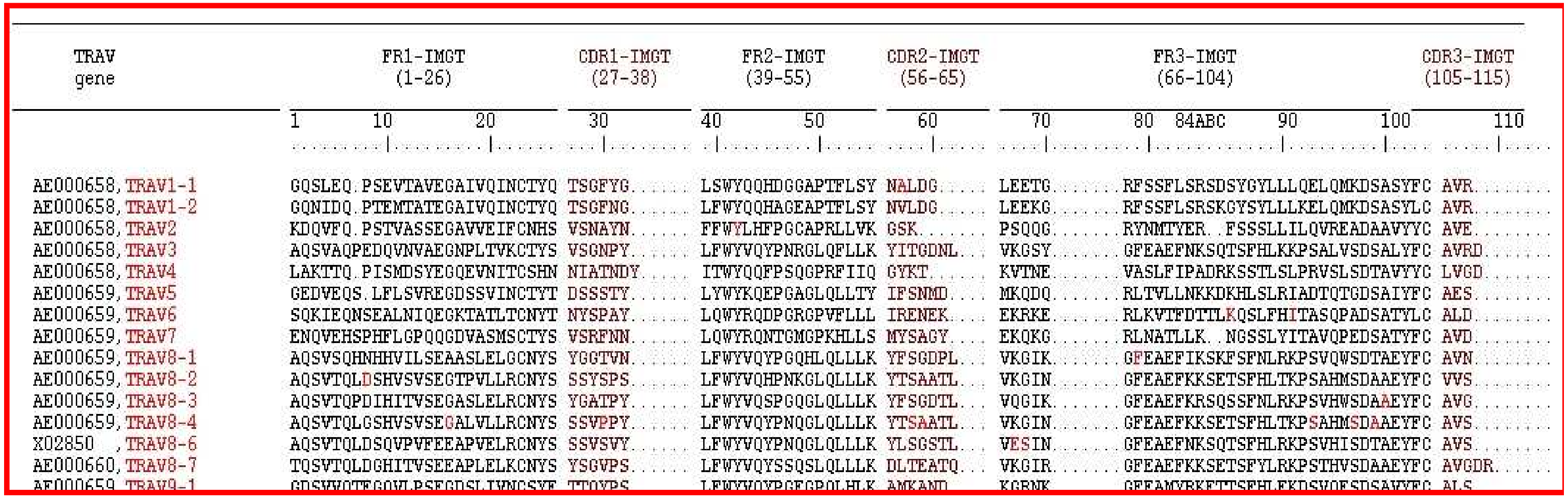


Alignment of alleles



```
                       1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
                       A   Q   S   V   T   Q   L   D   S   Q   V   P   V   F   E   E   A   P   V   E
X02850   ,TRAV8-6*01  GCC CAG TCT GTG ACC CAG CTT GAC AGC CAA GTC CCT GTC TTT GAA GAA GCC CCT GTG GAG

AE000659,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M86361   ,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                              CDR1-IMGT
                      21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
                       L   R   C   N   Y   S   S   S   V   S   V   Y                           L   F
X02850   ,TRAV8-6*01  CTG AGG TGC AAC TAC TCA TCG TCT GTT TCA GTG TAT ... ... ... ... ... ... CTC TTC

AE000659,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- ... ... ... ... ... ... --- ---

M86361   ,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- ... ... ... ... ... ... --- ---

                                                                                           CDR2-
                      41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
                       W   Y   V   Q   Y   P   N   Q   G   L   Q   L   L   L   K   Y   L   S   G   S
X02850   ,TRAV8-6*01  TGG TAT GTG CAA TAC CCC AAC CAA GGA CTC CAG CTT CTC CTG AAG TAT TTA TCA GGA TCC

AE000659,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M86361   ,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                     IMGT
                      61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80
                       T   L               V   E   S   I   N                               G   F   E
X02850   ,TRAV8-6*01  ACC CTG ... ... ... GTT GAA AGC ATC AAC ... ... ... ... ... ... ... GGT TTT GAG
                                               K   G
AE000659,TRAV8-6*02  --- --- ... ... ... --- A-- G-- --- --- ... ... ... ... ... ... ... --- --- ---
                                               K   G
M86361   ,TRAV8-6*02  --- --- ... ... ... --- A-- G-- --- --- ... ... ... ... ... ... ... --- --- ---

                      81  82  83  84 84A 84B 84C  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100
                       A   E   F   N   K   S   Q   T   S   F   H   L   R   K   P   S   V   H   I   S   D   T   A
X02850   ,TRAV8-6*01  GCT GAA TTT AAC AAG AGT CAA ACT TCC TTC CAC TTG AGG AAA CCC TCA GTC CAT ATA AGC GAC ACG GCT

AE000659,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M86361   ,TRAV8-6*02  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                          CDR3-IMGT
                     101 102 103 104 105 106 107 108
                       E   Y   F   C   A   V   S
X02850   ,TRAV8-6*01  GAG TAC TTC TGT GCT GTG AGT GA

AE000659,TRAV8-6*02  --- --- --- --- --- --- --- --
                                               R
M86361   ,TRAV8-6*02  --- --- --- --- --- --- --G        #c
```

Protein Display



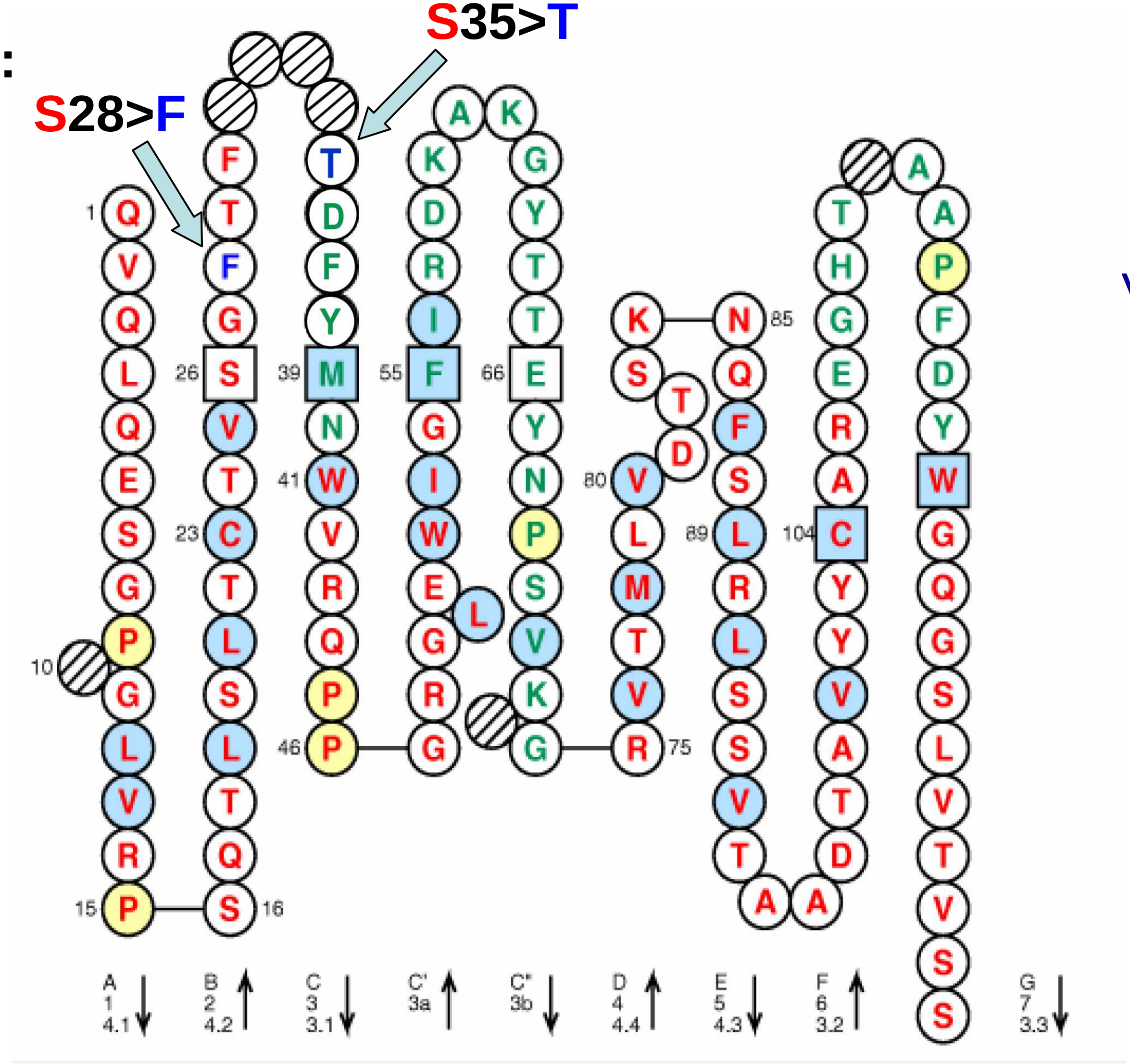
```
    TRAV                        FR1-IMGT          CDR1-IMGT     FR2-IMGT      CDR2-IMGT               FR3-IMGT                          CDR3-IMGT
    gene                         (1-26)            (27-38)      (39-55)        (56-65)                (66-104)                          (105-115)

                          1        10        20      30        40        50      60        70       80  84ABC    90        100        110
                          .........|.........|...... ...|......... .|.........|..... ....|..... ....|.........|.............|.........|.... .....|.....

AE000658,TRAV1-1    GQSLEQ.PSEVTAVEGAIVQINCTYQ TSGFYG...... LSWYQQHDGGAPTFLSY NALDG..... LEETG.......RFSSFLSRSDSYGYLLLQELQMKDSASYFC AVR........
AE000658,TRAV1-2    GQNIDQ.PTEMTATEGAIVQINCTYQ TSGFNG...... LFWYQQHAGEAPTFLSY NVLDG..... LEEKG.......RFSSFLSRSKGYSYLLLKELQMKDSASYLC AVR........
AE000658,TRAV2      KDQVFQ.PSTVASSEGAVVEIFCNHS VSNAYN...... FFWYLHFPGCAPRLLVK GSK....... PSQQG.......RYNMTYER..FSSSLLILQVREADAAVYYC AVE........
AE000658,TRAV3      AQSVAQPEDQVNVAEGNPLTVKCTYS VSGNPY...... LFWYVQYPNRGLQFLLK YITGDNL... VKGSY.......GFEAEFNKSQTSFHLKKPSALVSDSALYFC AVRD.......
AE000658,TRAV4      LAKTTQ.PISMDSYEGQEVNITCSHN NIATNDY..... ITWYQQFPSQGPRFIIQ GYKT...... KVTNE.......VASLFIPADRKSSTLSLPRVSLSDTAVYYC LVGD.......
AE000659,TRAV5      GEDVEQS.LFLSVREGDSSVINCTYT DSSSTY...... LYWYKQEPGAGLQLLTY IFSNMD.... MKQDQ.......RLTVLLNKKDKHLSLRIADTQTGDSAIYFC AES........
AE000659,TRAV6      SQKIEQNSEALNIQEGKTATLTCNYT NYSPAY...... LQWYRQDPGRGPVFLLL IRENEK.... EKRKE.......RLKVTFDTTLKQSLFHITASQPADSATYLC ALD........
AE000659,TRAV7      ENQVEHSPHFLGPQQGDVASMSCTYS VSRFNN...... LQWYRQNTGMGPKHLLS MYSAGY.... EKQKG.......RLNATLLK..NGSSLYITAVQPEDSATYFC AVD........
AE000659,TRAV8-1    AQSVSQHNHHVILSEAASLELGCNYS YGGTVN...... LFWYVQYPGQHLQLLLK YFSGDPL... VKGIK.......GFEAEFIKSKFSFNLRKPSVQWSDTAEYFC AVN........
AE000659,TRAV8-2    AQSVTQLDSHVSVSEGTPVLLRCNYS SSYSPS...... LFWYVQHPNKGLQLLLK YTSAATL... VKGIN.......GFEAEFKKSETSFHLTKPSAHMSDAAEYFC VVS........
AE000659,TRAV8-3    AQSVTQPDIHITVSEGASLELRCNYS YGATPY...... LFWYVQSPGQGLQLLLK YFSGDTL... VQGIK.......GFEAEFKRSQSSFNLRKPSVHWSDAAEYFC AVG........
AE000659,TRAV8-4    AQSVTQLGSHVSVSEGALVLLRCNYS SSVPPY...... LFWYVQYPNQGLQLLLK YTSAATL... VKGIN.......GFEAEFKKSETSFHLTKPSAHMSDAAEYFC AVS........
X02850  ,TRAV8-6    AQSVTQLDSQVPVFEEAPVELRCNYS SSVSVY...... LFWYVQYPNQGLQLLLK YLSGSTL... VESIN.......GFEAEFNKSQTSFHLRKPSVHISDTAEYFC AVS........
AE000660,TRAV8-7    TQSVTQLDGHITVSEEAPLELKCNYS YSGVPS...... LFWYVQYSSQSLQLLLK DLTEATQ... VKGIR.......GFEAEFKKSETSFYLRKPSTHVSDAAEYFC AVGDR......
AE000659,TRAV9-1    GDSVVQTEGQVLPSEGDSLIVNCSYE TTQYPS       LFWYVQYPGEGPQLHLK AMKAND     KGRNK       GFEAMYRKETTSFHLEKDSVQESDSAVYFC ALS
```

Antibody humanization and engineering

http://www.imgt.org

Alemtuzumab (CAMPATH®)

2 mutations:



S28>F

S35>T

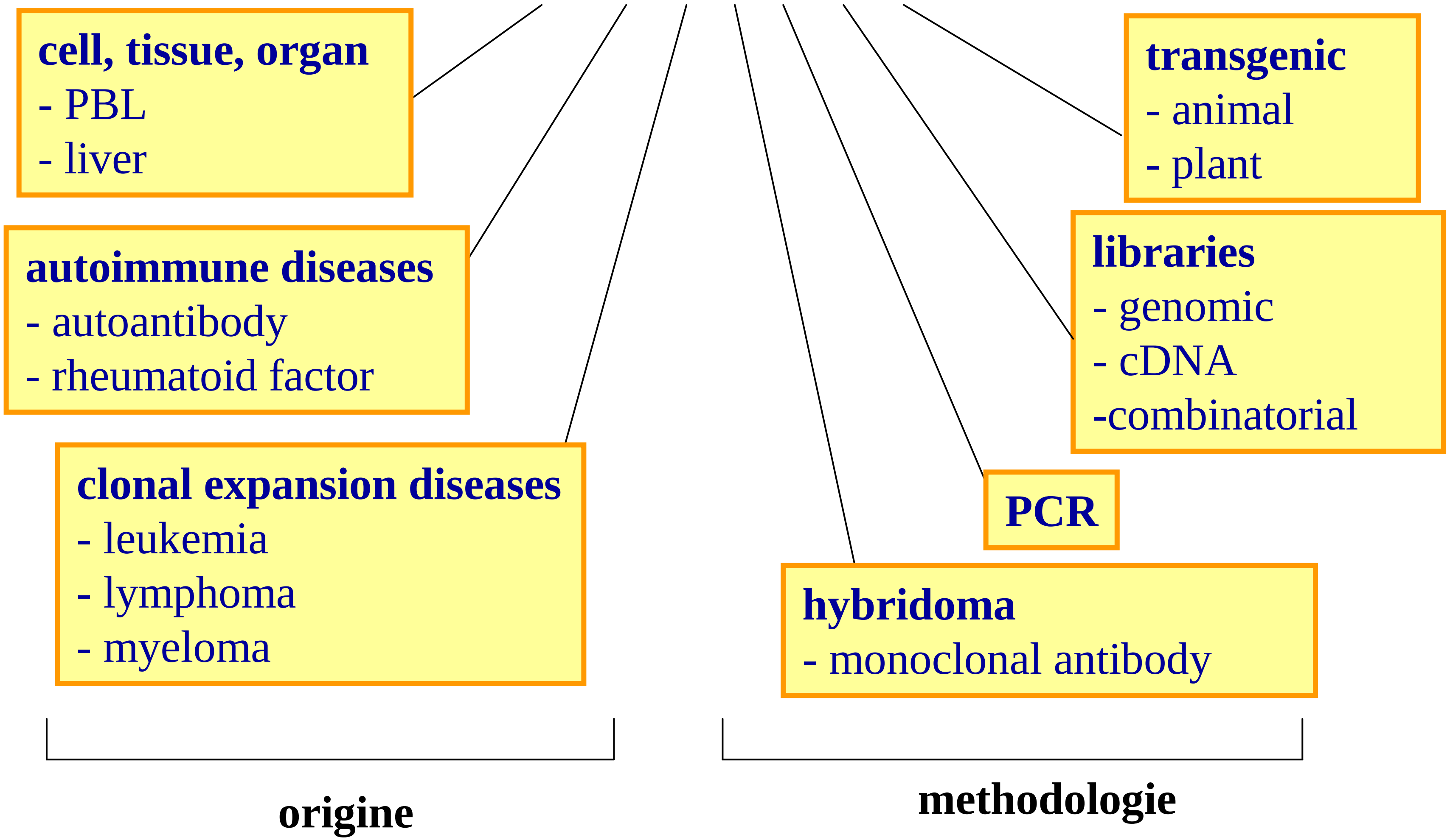
VH domain
[8.10.12]

🟥 human

🟩 rat

http://www.imgt.org

"OBTENTION"



cell, tissue, organ
- PBL
- liver

autoimmune diseases
- autoantibody
- rheumatoid factor

clonal expansion diseases
- leukemia
- lymphoma
- myeloma

origine

transgenic
- animal
- plant

libraries
- genomic
- cDNA
- combinatorial

PCR

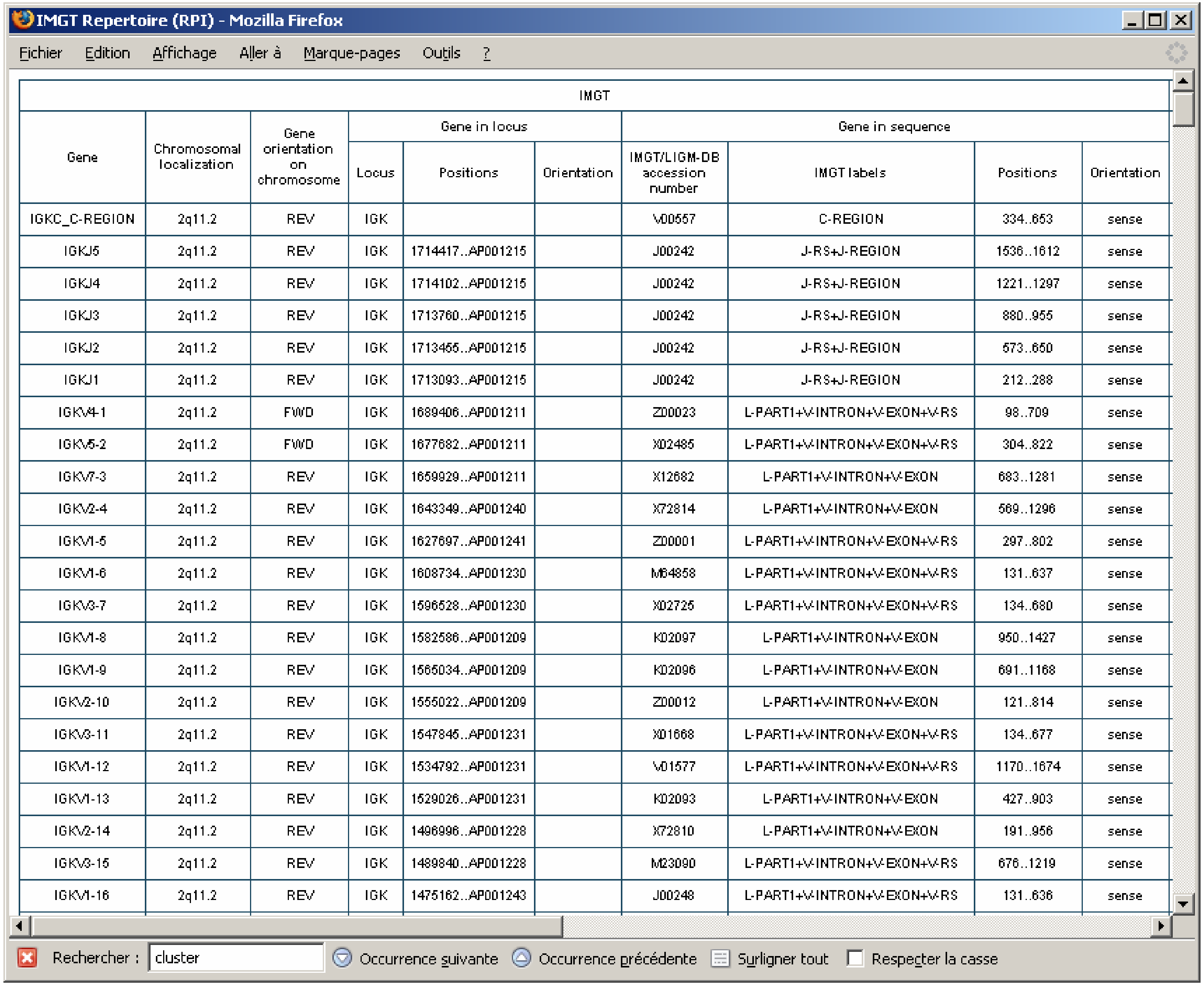
hybridoma
- monoclonal antibody

methodologie

The "OBTENTION" concepts specify the origin and methodology

"ORIENTATION"

http://www.imgt.org

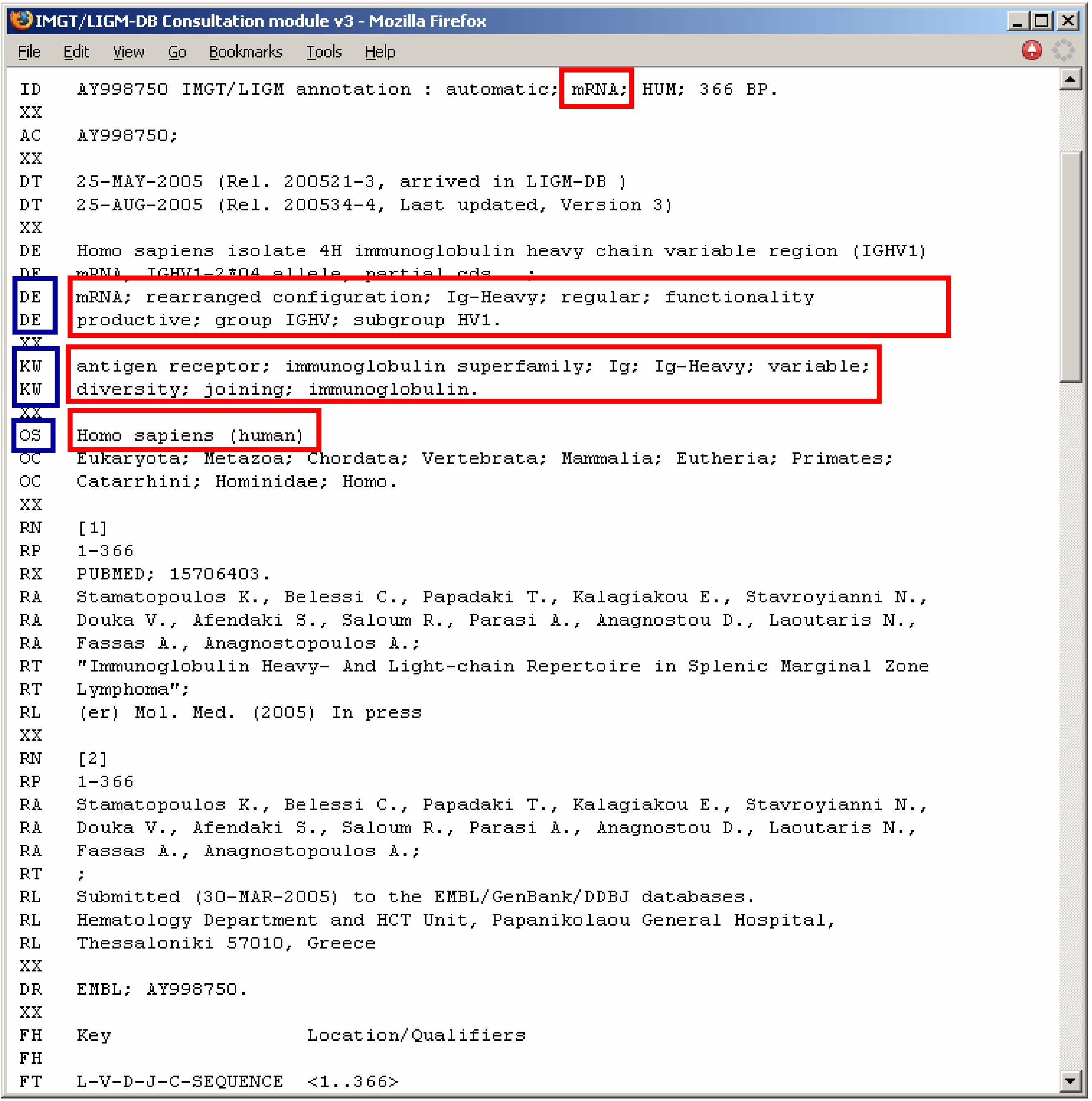


IMGT

Gene	Chromosomal localization	Gene orientation on chromosome	Gene in locus			Gene in sequence			
			Locus	Positions	Orientation	IMGT/LIGM-DB accession number	IMGT labels	Positions	Orientation
IGKC_C-REGION	2q11.2	REV	IGK			V00557	C-REGION	334..653	sense
IGKJ5	2q11.2	REV	IGK	1714417..AP001215		J00242	J-RS+J-REGION	1536..1612	sense
IGKJ4	2q11.2	REV	IGK	1714102..AP001215		J00242	J-RS+J-REGION	1221..1297	sense
IGKJ3	2q11.2	REV	IGK	1713760..AP001215		J00242	J-RS+J-REGION	880..955	sense
IGKJ2	2q11.2	REV	IGK	1713455..AP001215		J00242	J-RS+J-REGION	573..650	sense
IGKJ1	2q11.2	REV	IGK	1713093..AP001215		J00242	J-RS+J-REGION	212..288	sense
IGKV4-1	2q11.2	FWD	IGK	1689406..AP001211		Z00023	L-PART1+V-INTRON+V-EXON+V-RS	98..709	sense
IGKV5-2	2q11.2	FWD	IGK	1677682..AP001211		X02485	L-PART1+V-INTRON+V-EXON+V-RS	304..822	sense
IGKV7-3	2q11.2	REV	IGK	1659929..AP001211		X12682	L-PART1+V-INTRON+V-EXON	683..1281	sense
IGKV2-4	2q11.2	REV	IGK	1643349..AP001240		X72814	L-PART1+V-INTRON+V-EXON	569..1296	sense
IGKV1-5	2q11.2	REV	IGK	1627697..AP001241		Z00001	L-PART1+V-INTRON+V-EXON+V-RS	297..802	sense
IGKV1-6	2q11.2	REV	IGK	1608734..AP001230		M64858	L-PART1+V-INTRON+V-EXON+V-RS	131..637	sense
IGKV3-7	2q11.2	REV	IGK	1596528..AP001230		X02725	L-PART1+V-INTRON+V-EXON+V-RS	134..680	sense
IGKV1-8	2q11.2	REV	IGK	1582586..AP001209		K02097	L-PART1+V-INTRON+V-EXON	950..1427	sense
IGKV1-9	2q11.2	REV	IGK	1565034..AP001209		K02096	L-PART1+V-INTRON+V-EXON	691..1168	sense
IGKV2-10	2q11.2	REV	IGK	1555022..AP001209		Z00012	L-PART1+V-INTRON+V-EXON	121..814	sense
IGKV3-11	2q11.2	REV	IGK	1547845..AP001231		X01668	L-PART1+V-INTRON+V-EXON+V-RS	134..677	sense
IGKV1-12	2q11.2	REV	IGK	1534792..AP001231		V01577	L-PART1+V-INTRON+V-EXON+V-RS	1170..1674	sense
IGKV1-13	2q11.2	REV	IGK	1529026..AP001231		K02093	L-PART1+V-INTRON+V-EXON	427..903	sense
IGKV2-14	2q11.2	REV	IGK	1496996..AP001228		X72810	L-PART1+V-INTRON+V-EXON	191..956	sense
IGKV3-15	2q11.2	REV	IGK	1489840..AP001228		M23090	L-PART1+V-INTRON+V-EXON+V-RS	676..1219	sense
IGKV1-16	2q11.2	REV	IGK	1475162..AP001243		J00248	L-PART1+V-INTRON+V-EXON+V-RS	131..636	sense

Annotation and IMGT-ONTOLOGY concepts

IDENTIFICATION



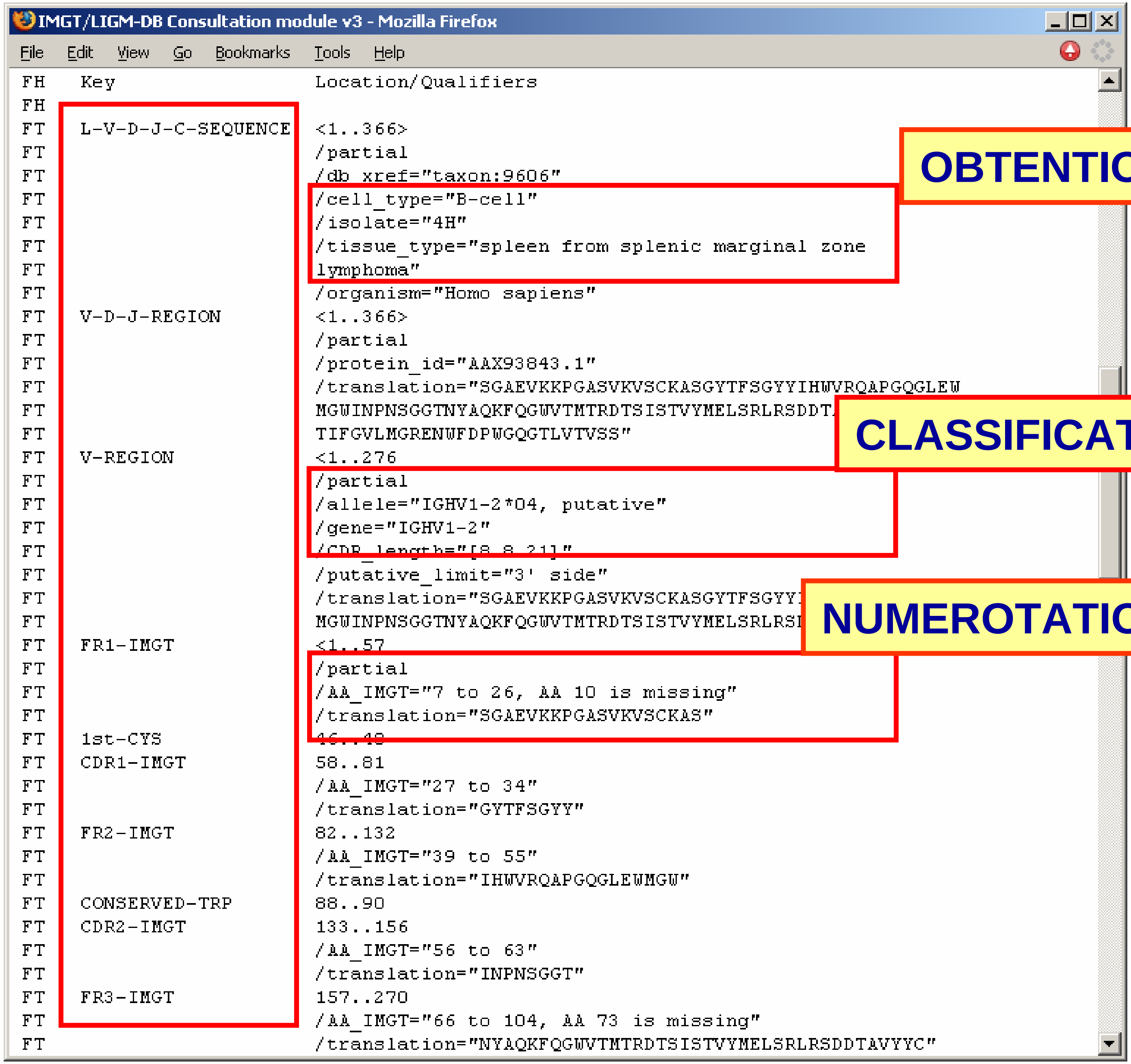
```
ID   AY998750 IMGT/LIGM annotation : automatic; mRNA; HUM; 366 BP.
XX
AC   AY998750;
XX
DT   25-MAY-2005 (Rel. 200521-3, arrived in LIGM-DB )
DT   25-AUG-2005 (Rel. 200534-4, Last updated, Version 3)
XX
DE   Homo sapiens isolate 4H immunoglobulin heavy chain variable region (IGHV1)
DE   mRNA, IGHV1-2*04 allele, partial cds.   ;
DE   mRNA; rearranged configuration; Ig-Heavy; regular; functionality
DE   productive; group IGHV; subgroup HV1.
XX
KW   antigen receptor; immunoglobulin superfamily; Ig; Ig-Heavy; variable;
KW   diversity; joining; immunoglobulin.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-366
RX   PUBMED; 15706403.
RA   Stamatopoulos K., Belessi C., Papadaki T., Kalagiakou E., Stavroyianni N.,
RA   Douka V., Afendaki S., Saloum R., Parasi A., Anagnostou D., Laoutaris N.,
RA   Fassas A., Anagnostopoulos A.;
RT   "Immunoglobulin Heavy- And Light-chain Repertoire in Splenic Marginal Zone
RT   Lymphoma";
RL   (er) Mol. Med. (2005) In press
XX
RN   [2]
RP   1-366
RA   Stamatopoulos K., Belessi C., Papadaki T., Kalagiakou E., Stavroyianni N.,
RA   Douka V., Afendaki S., Saloum R., Parasi A., Anagnostou D., Laoutaris N.,
RA   Fassas A., Anagnostopoulos A.;
RT   ;
RL   Submitted (30-MAR-2005) to the EMBL/GenBank/DDBJ databases.
RL   Hematology Department and HCT Unit, Papanikolaou General Hospital,
RL   Thessaloniki 57010, Greece
XX
DR   EMBL; AY998750.
XX
FH   Key                 Location/Qualifiers
FH
FT   L-V-D-J-C-SEQUENCE  <1..366>
```

http://www.imgt.org

Annotation and IMGT-ONTOLOGY concepts

IMGT® the international ImMunoGeneTics Information system®

http://www.imgt.org



```
FH   Key                 Location/Qualifiers
FH
FT   L-V-D-J-C-SEQUENCE  <1..366>
FT                       /partial
FT                       /db_xref="taxon:9606"
FT                       /cell_type="B-cell"
FT                       /isolate="4H"
FT                       /tissue_type="spleen from splenic marginal zone
FT                       lymphoma"
FT                       /organism="Homo sapiens"
FT   V-D-J-REGION        <1..366>
FT                       /partial
FT                       /protein_id="AAX93843.1"
FT                       /translation="SGAEVKKPGASVKVSCKASGYTFSGYYIHWVRQAPGQGLEW
FT                       MGWINPNSGGTNYAQKFQGWVTMTRDTSISTVYMELSRLRSDDT
FT                       TIFGVLMGRENWFDPWGQGTLVTVSS"
FT   V-REGION            <1..276
FT                       /partial
FT                       /allele="IGHV1-2*04, putative"
FT                       /gene="IGHV1-2"
FT                       /CDR_length="[8.8.21]"
FT                       /putative_limit="3' side"
FT                       /translation="SGAEVKKPGASVKVSCKASGYTFSGYYI
FT                       MGWINPNSGGTNYAQKFQGWVTMTRDTSISTVYMELSRLRSI
FT   FR1-IMGT            <1..57
FT                       /partial
FT                       /AA_IMGT="7 to 26, AA 10 is missing"
FT                       /translation="SGAEVKKPGASVKVSCKAS"
FT   1st-CYS             46..48
FT   CDR1-IMGT           58..81
FT                       /AA_IMGT="27 to 34"
FT                       /translation="GYTFSGYY"
FT   FR2-IMGT            82..132
FT                       /AA_IMGT="39 to 55"
FT                       /translation="IHWVRQAPGQGLEWMGW"
FT   CONSERVED-TRP       88..90
FT   CDR2-IMGT           133..156
FT                       /AA_IMGT="56 to 63"
FT                       /translation="INPNSGGT"
FT   FR3-IMGT            157..270
FT                       /AA_IMGT="66 to 104, AA 73 is missing"
FT                       /translation="NYAQKFQGWVTMTRDTSISTVYMELSRLRSDDTAVYYC"
```

DESCRIPTION

OBTENTION

CLASSIFICATION

NUMEROTATION

Les atouts de IMGT-ONTOLOGY

1 - Pour les utilisateurs:

- Définir un vocabulaire précis et spécifique du domaine qui permette de décrire en détail toutes les caractéristiques des IG et des TR
- Proposer ce vocabulaire standardisé comme critères de sélection dans la base

Les atouts de IMGT-ONTOLOGY

2 - Pour les annotateurs du laboratoire:

- Faciliter la communication au sein de l'équipe
- Faciliter la formation du nouveau personnel
- Etablir les règles d'annotation qui expriment les dépendances entre les termes du vocabulaire

Les atouts de IMGT-ONTOLOGY

3 - Automatisation des procédures:

- Alléger le travail des annotateurs
- Développer des outil d'annotation

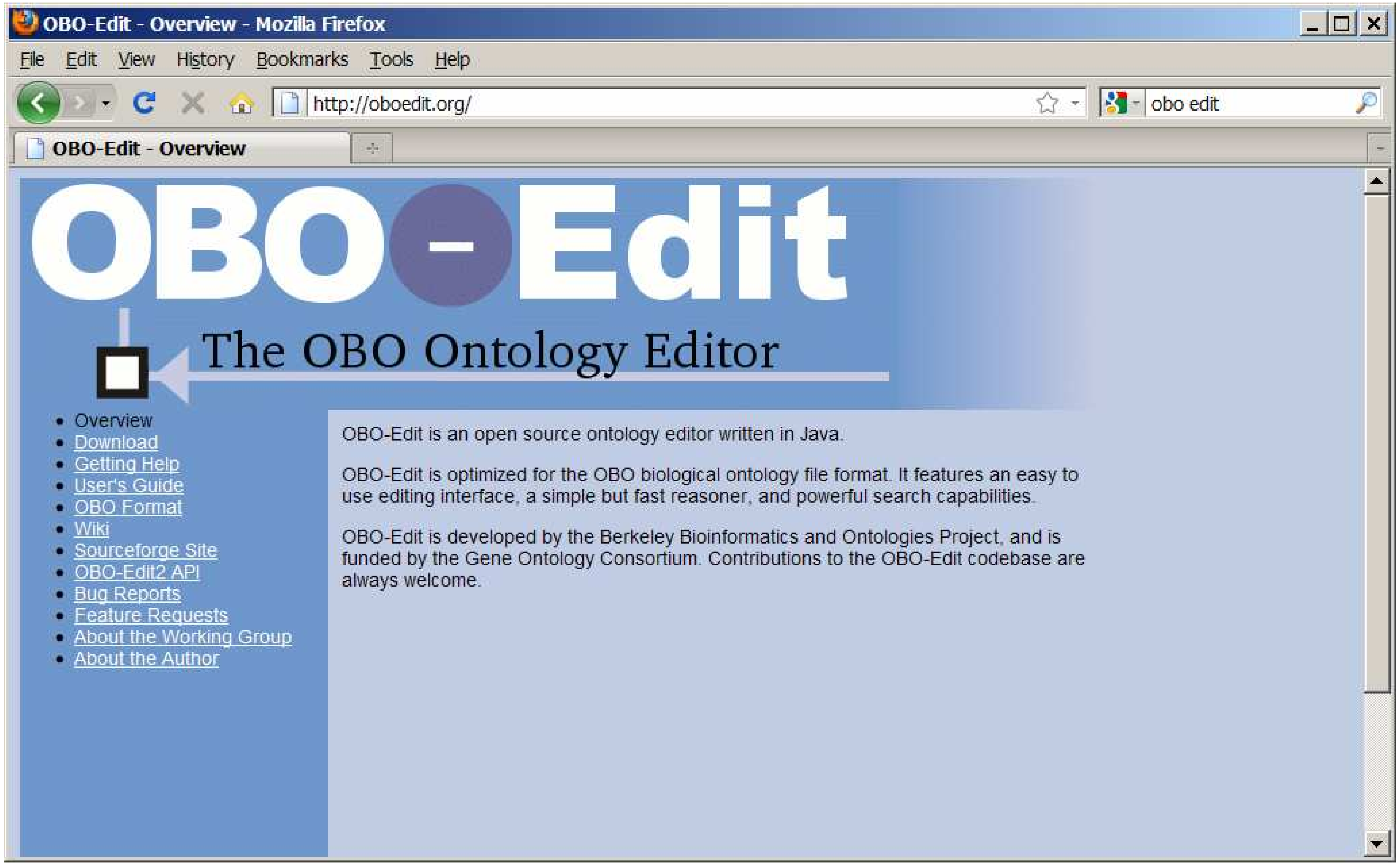
Les atouts de IMGT-ONTOLOGY

4 – Intégration des nouvelles connaissances en immunogénétique

- Modifier les règles existantes
- Appliquer les modifications aux données préexistantes

Exemple d’éditeur d’ontologie: OBO-Edit (GO, SO, …)

http://oboedit.org/



- Overview
- Download
- Getting Help
- User's Guide
- OBO Format
- Wiki
- Sourceforge Site
- OBO-Edit2 API
- Bug Reports
- Feature Requests
- About the Working Group
- About the Author

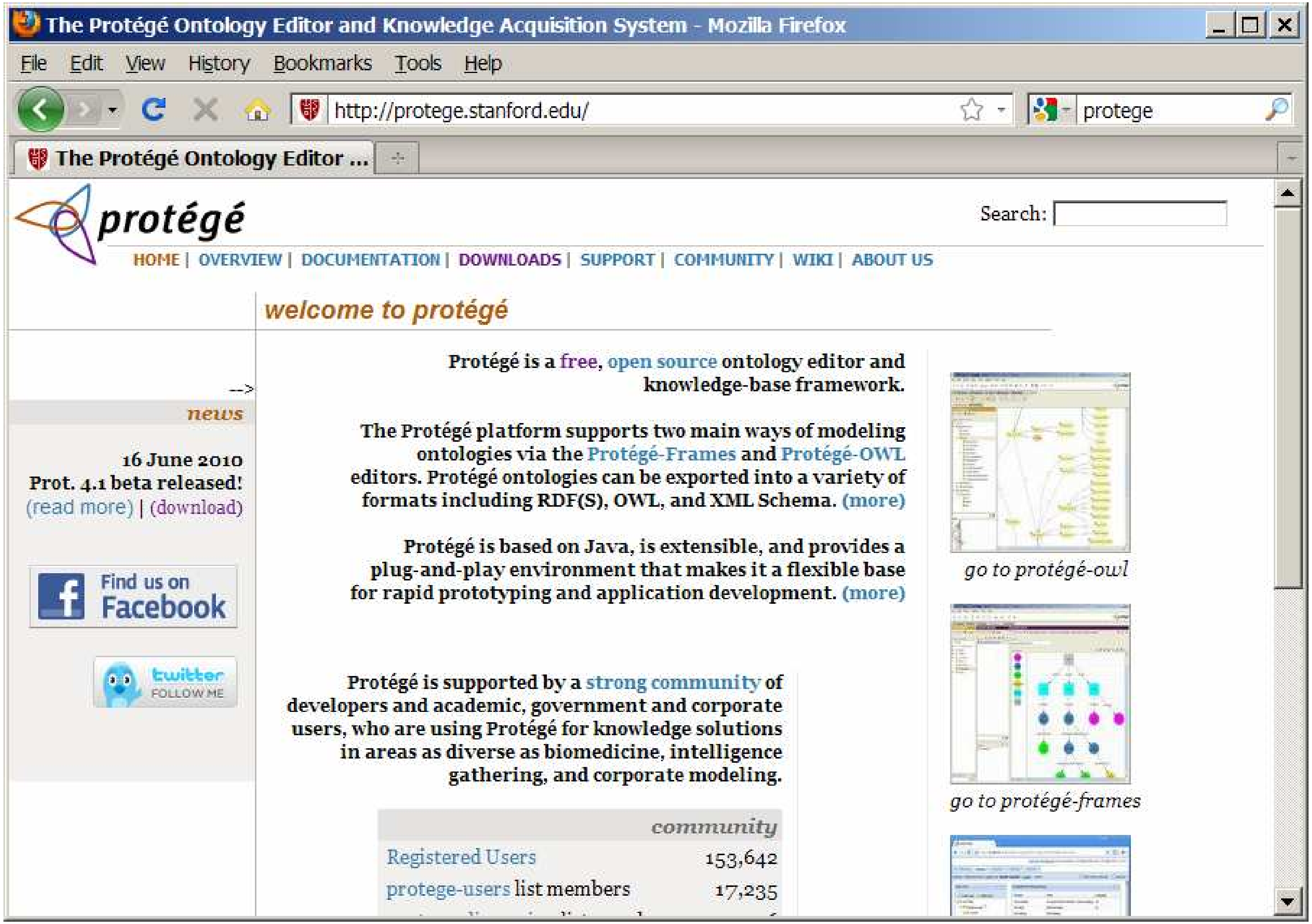
OBO-Edit is an open source ontology editor written in Java.

OBO-Edit is optimized for the OBO biological ontology file format. It features an easy to use editing interface, a simple but fast reasoner, and powerful search capabilities.

OBO-Edit is developed by the Berkeley Bioinformatics and Ontologies Project, and is funded by the Gene Ontology Consortium. Contributions to the OBO-Edit codebase are always welcome.

Exemple d'éditeur d'ontologie: Protégé

http://protege.stanford.edu/



welcome to protégé

-->

news

16 June 2010
Prot. 4.1 beta released!
(read more) | (download)

Protégé is a free, open source ontology editor and knowledge-base framework.

The Protégé platform supports two main ways of modeling ontologies via the Protégé-Frames and Protégé-OWL editors. Protégé ontologies can be exported into a variety of formats including RDF(S), OWL, and XML Schema. (more)

Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development. (more)

Protégé is supported by a strong community of developers and academic, government and corporate users, who are using Protégé for knowledge solutions in areas as diverse as biomedicine, intelligence gathering, and corporate modeling.

community	
Registered Users	153,642
protege-users list members	17,235

Administration et contrôle de données dans un système d'information: exemple de IMGT/LIGM-DB

Im
Muno
Gene
Tics
Information system®

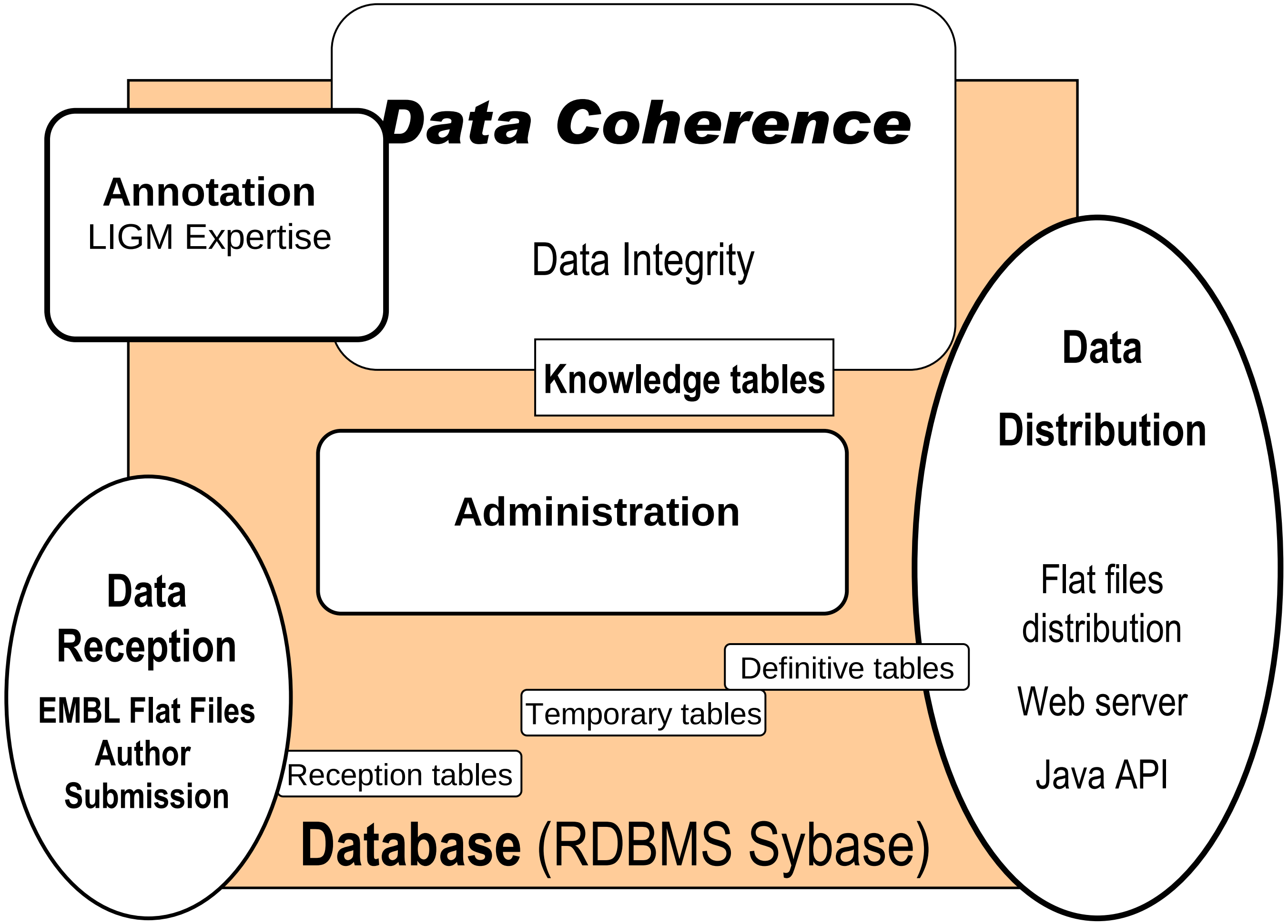
http://www.imgt.org

1- Coordination des flux des données
de différentes sources dans la base: comment gérer
l'entrée des séquences (700 arrivées par semaine) ,
des annotations et leur distribution?

2- Gestion des accès aux données: qui annote les séquences,
quelles informations peuvent être consultées, extraites
et modifiées, et par qui?

1- Coordination des flux de données

IMGT® Information system® http://www.imgt.org



Data Coherence

Data Integrity

Annotation
LIGM Expertise

Knowledge tables

Administration

Data Reception
EMBL Flat Files
Author Submission

Data Distribution

Flat files distribution

Web server

Java API

Definitive tables

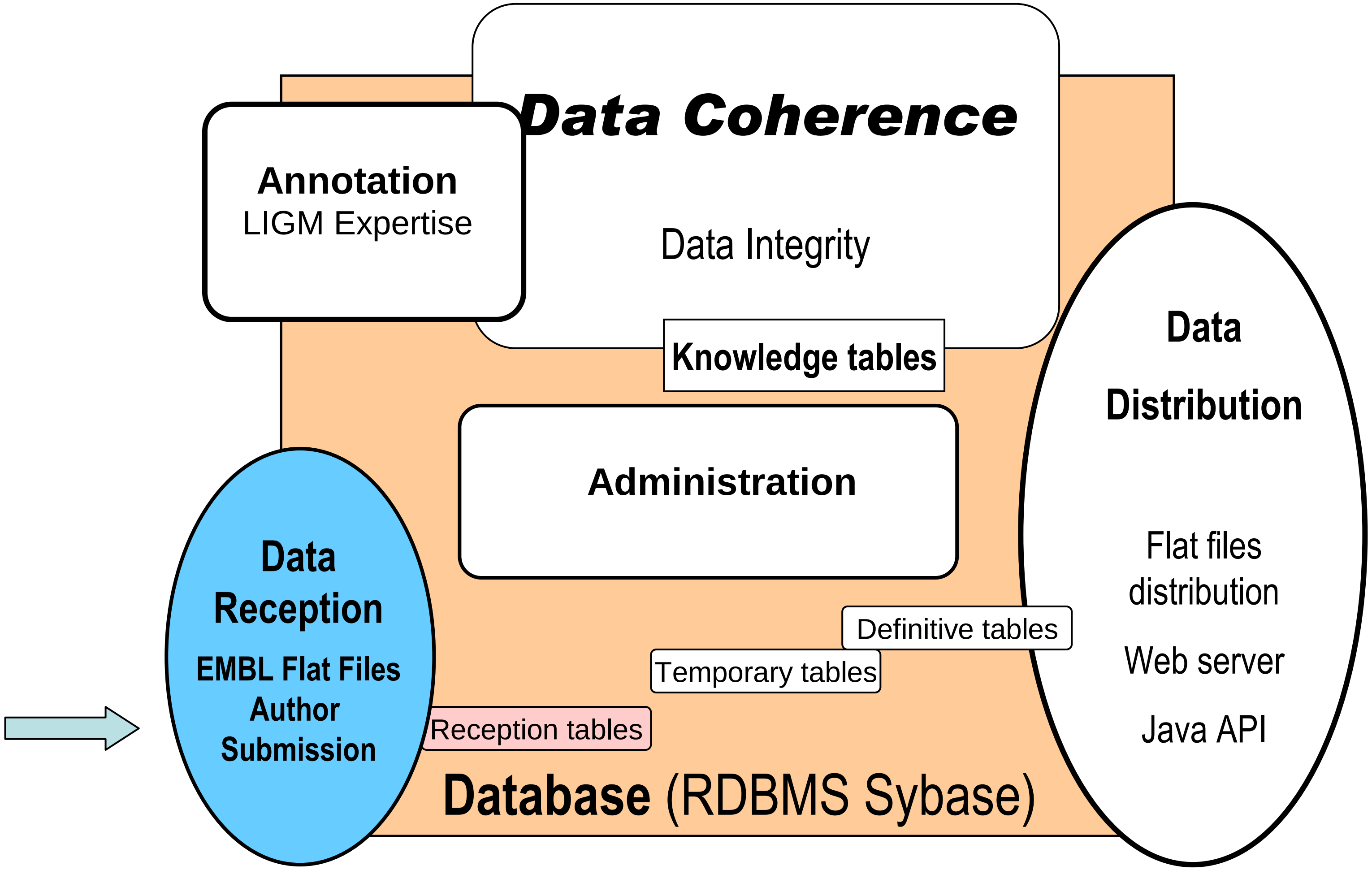
Temporary tables

Reception tables

Database (RDBMS Sybase)

1- Coordination des flux de données

http://www.imgt.org



Data Coherence

Data Integrity

Knowledge tables

Annotation
LIGM Expertise

Administration

Data Reception
EMBL Flat Files
Author Submission

Reception tables

Temporary tables

Definitive tables

Data Distribution

Flat files distribution

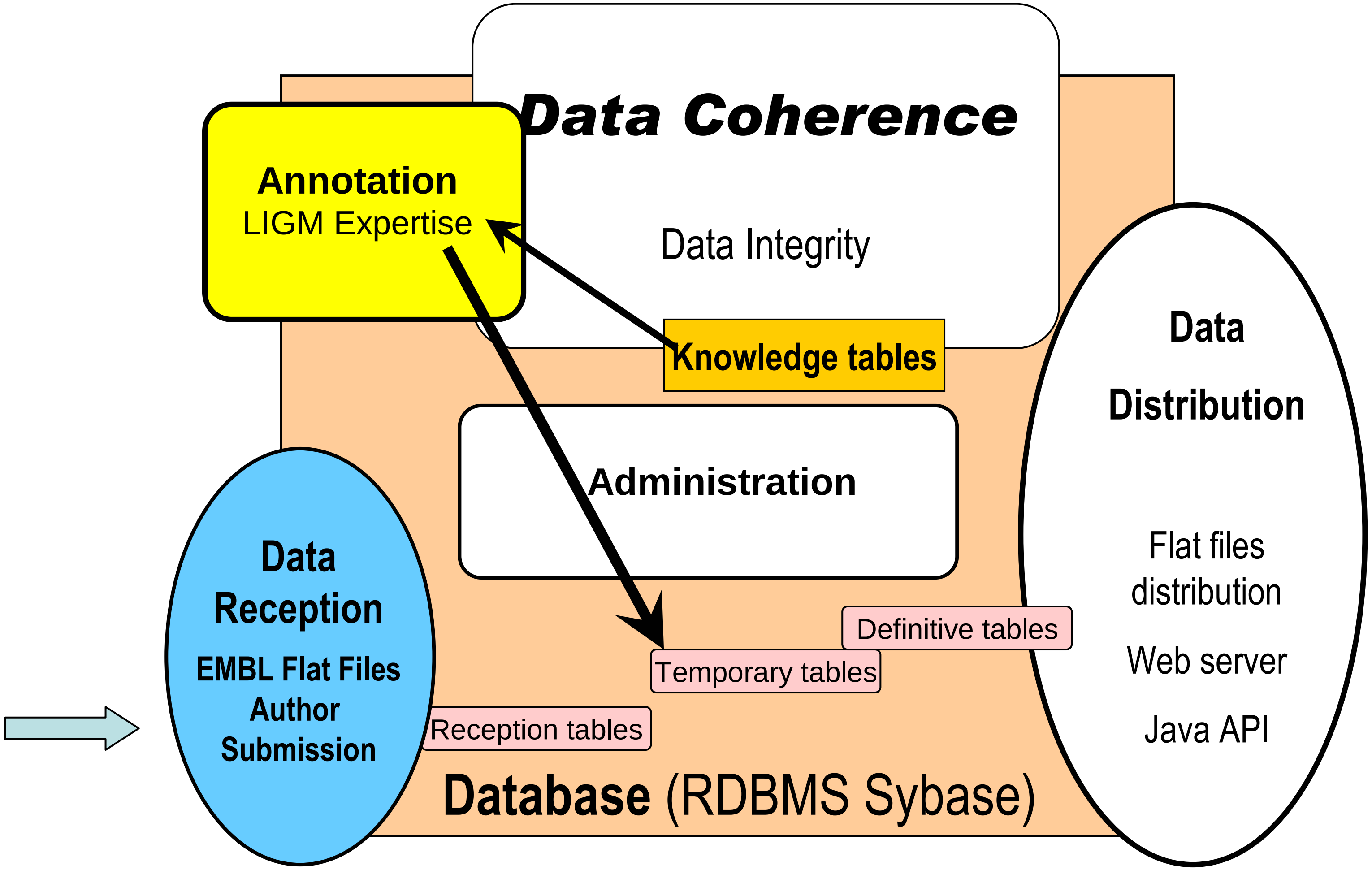
Web server

Java API

Database (RDBMS Sybase)

1- Coordination des flux de données

http://www.imgt.org



Data Coherence

Data Integrity

Annotation
LIGM Expertise

Knowledge tables

Administration

Data Reception
EMBL Flat Files
Author Submission

Reception tables

Temporary tables

Definitive tables

Data Distribution

Flat files distribution

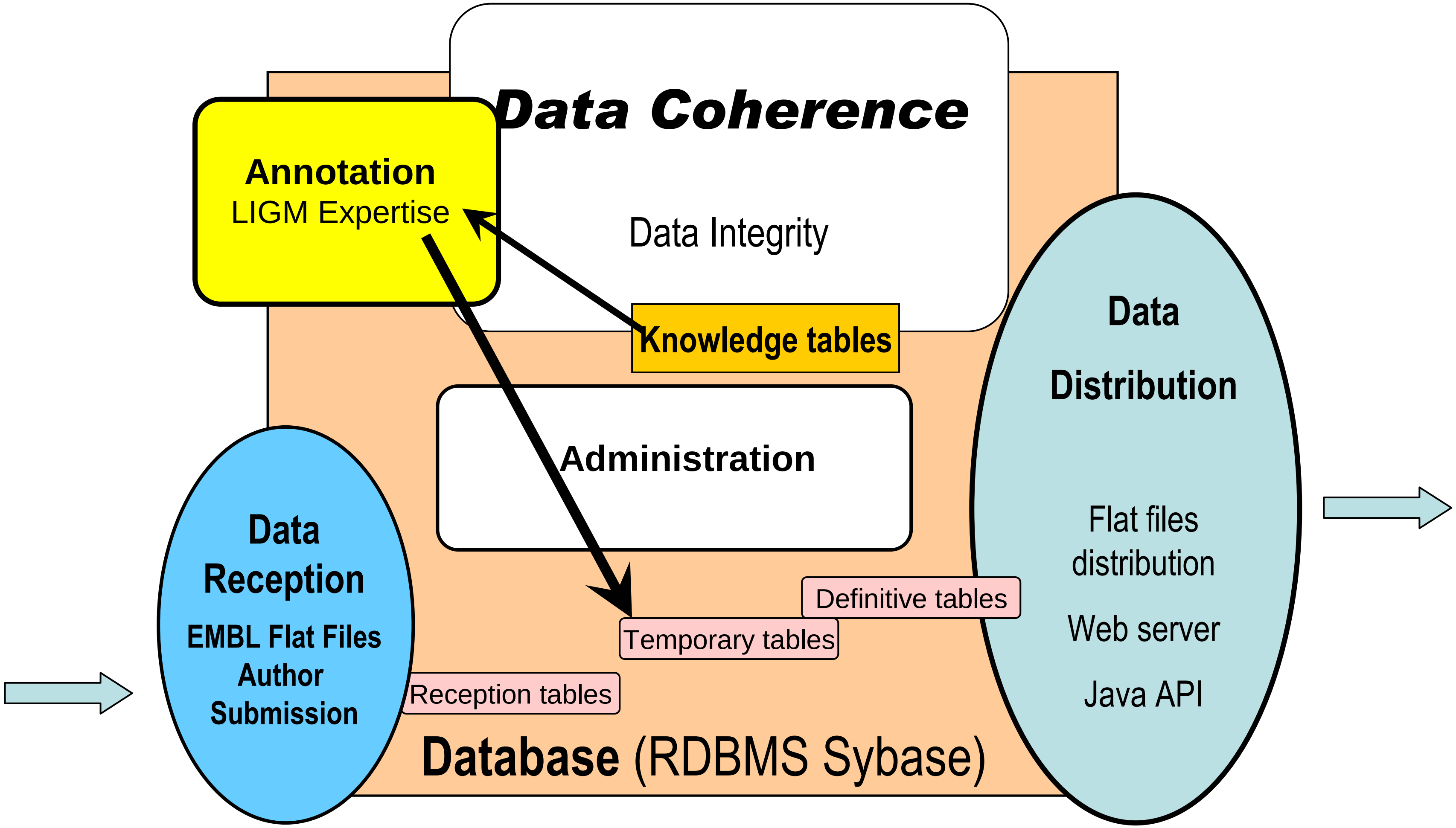
Web server

Java API

Database (RDBMS Sybase)

1- Coordination des flux de données

http://www.imgt.org



Data Coherence

Data Integrity

Knowledge tables

Annotation
LIGM Expertise

Administration

Data
Reception
EMBL Flat Files
Author
Submission

Reception tables

Temporary tables

Definitive tables

Database (RDBMS Sybase)

Data
Distribution

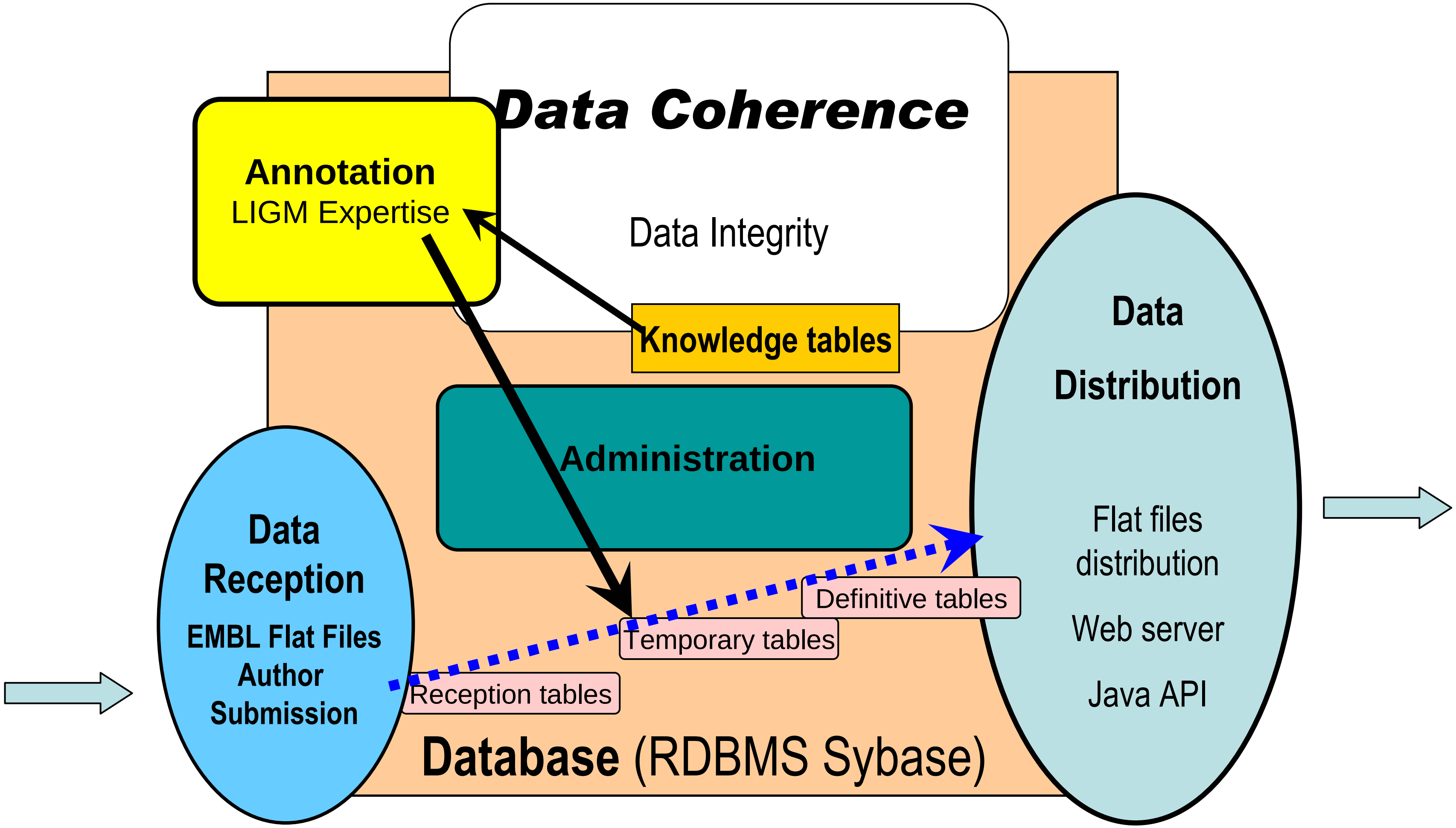
Flat files
distribution

Web server

Java API

1- Coordination des flux de données

http://www.imgt.org



Data Coherence

Data Integrity

Annotation
LIGM Expertise

Knowledge tables

Administration

Data
Reception
EMBL Flat Files
Author
Submission

Reception tables

Temporary tables

Definitive tables

Data
Distribution

Flat files
distribution

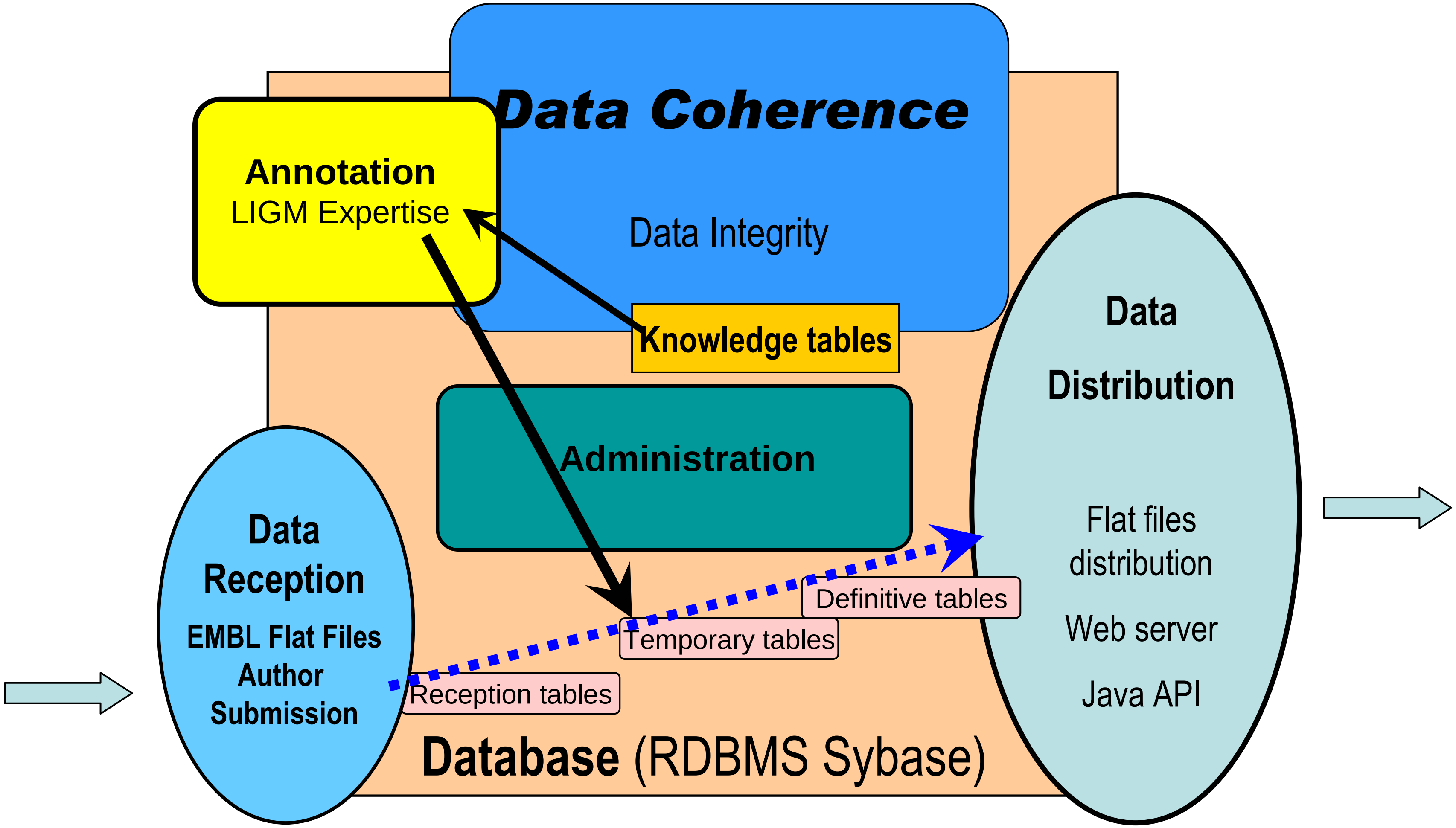
Web server

Java API

Database (RDBMS Sybase)

1- Coordination des flux de données

http://www.imgt.org



Data Coherence

Data Integrity

Annotation
LIGM Expertise

Knowledge tables

Administration

Data Reception

EMBL Flat Files
Author Submission

Reception tables

Temporary tables

Definitive tables

Data Distribution

Flat files distribution

Web server

Java API

Database (RDBMS Sybase)

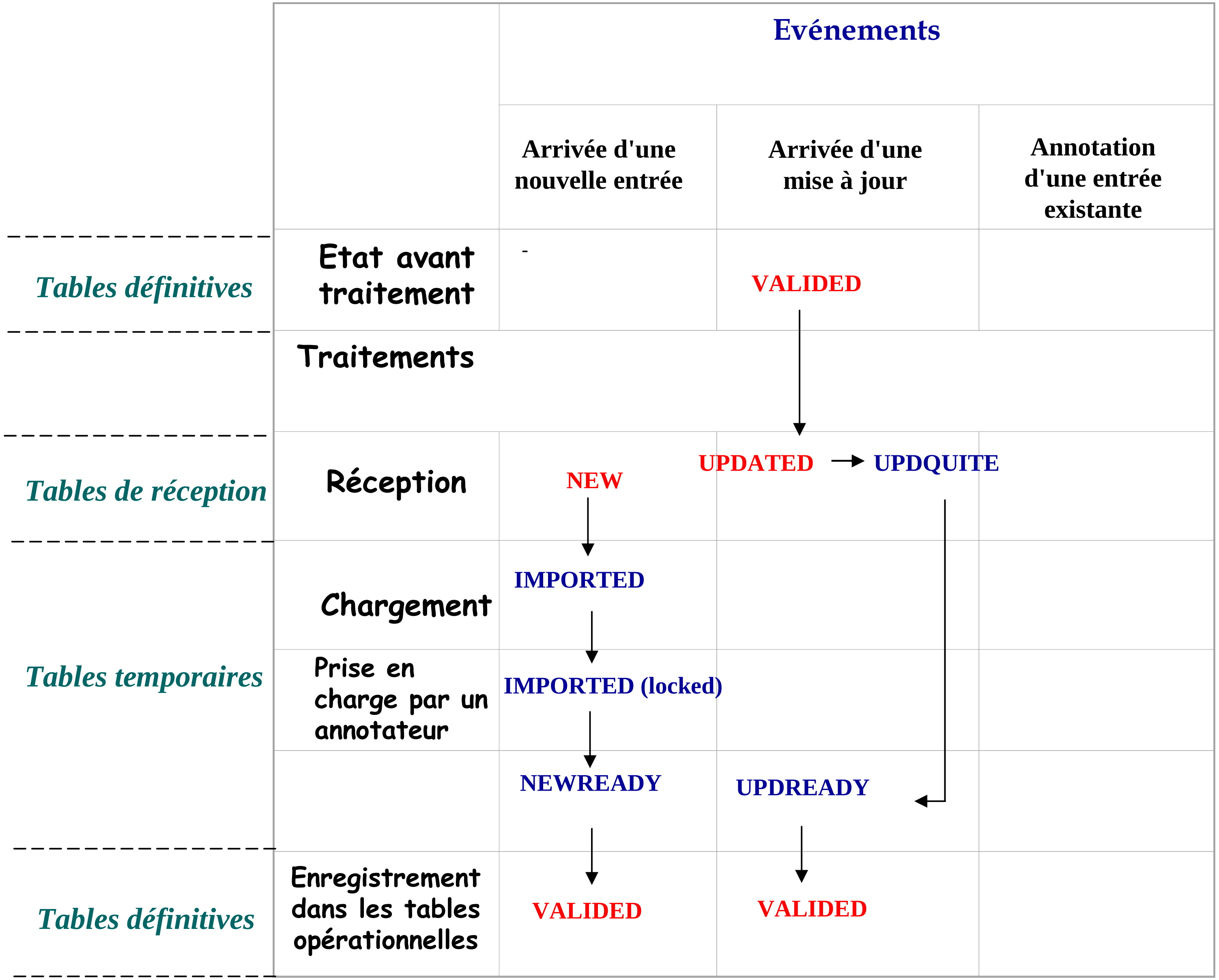
2 - Gestion accès aux données

http://www.imgt.org

		Evénements		
		Arrivée d'une nouvelle entrée	Arrivée d'une mise à jour	Annotation d'une entrée existante
Tables définitives	**Etat avant traitement**	-	VALIDED	
	Traitements			
Tables de réception	**Réception**	NEW	UPDATED	
Tables temporaires	**Chargement**	IMPORTED		
	Prise en charge par un annotateur	IMPORTED (locked)		
		NEWREADY		
Tables définitives	**Enregistrement dans les tables opérationnelles**	VALIDED		

(locked) : indique que l'entrée est prise en charge par un annotateur

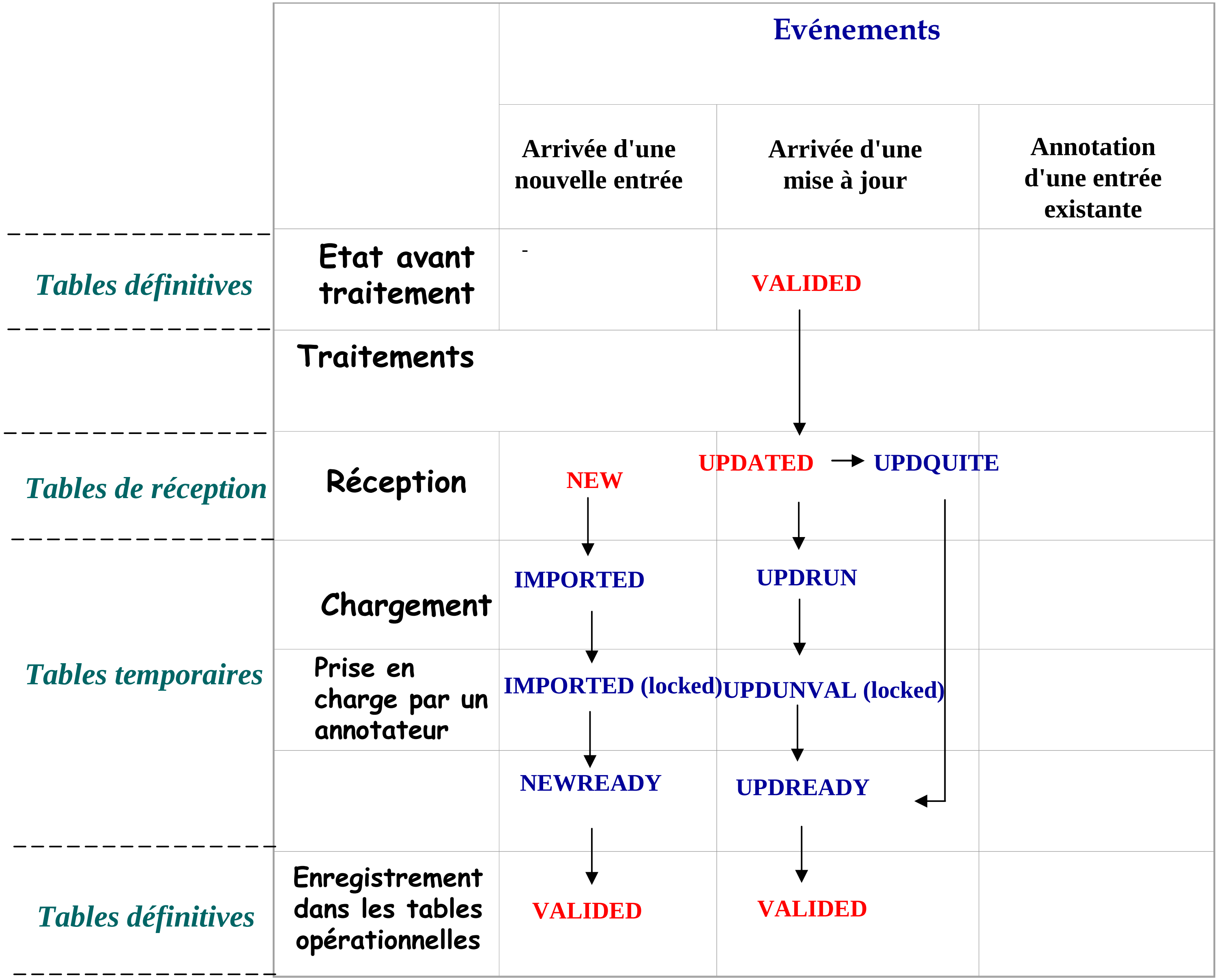
http://www.imgt.org



		Evénements		
		Arrivée d'une nouvelle entrée	Arrivée d'une mise à jour	Annotation d'une entrée existante
Tables définitives	Etat avant traitement	-	VALIDED	
	Traitements		↓	
Tables de réception	Réception	NEW	UPDATED → UPDQUITE	
Tables temporaires	Chargement	↓ IMPORTED		
	Prise en charge par un annotateur	↓ IMPORTED (locked)		
		↓ NEWREADY	UPDREADY ←	
Tables définitives	Enregistrement dans les tables opérationnelles	↓ VALIDED	↓ VALIDED	

(locked) : indique que l'entrée est prise en charge par un annotateur

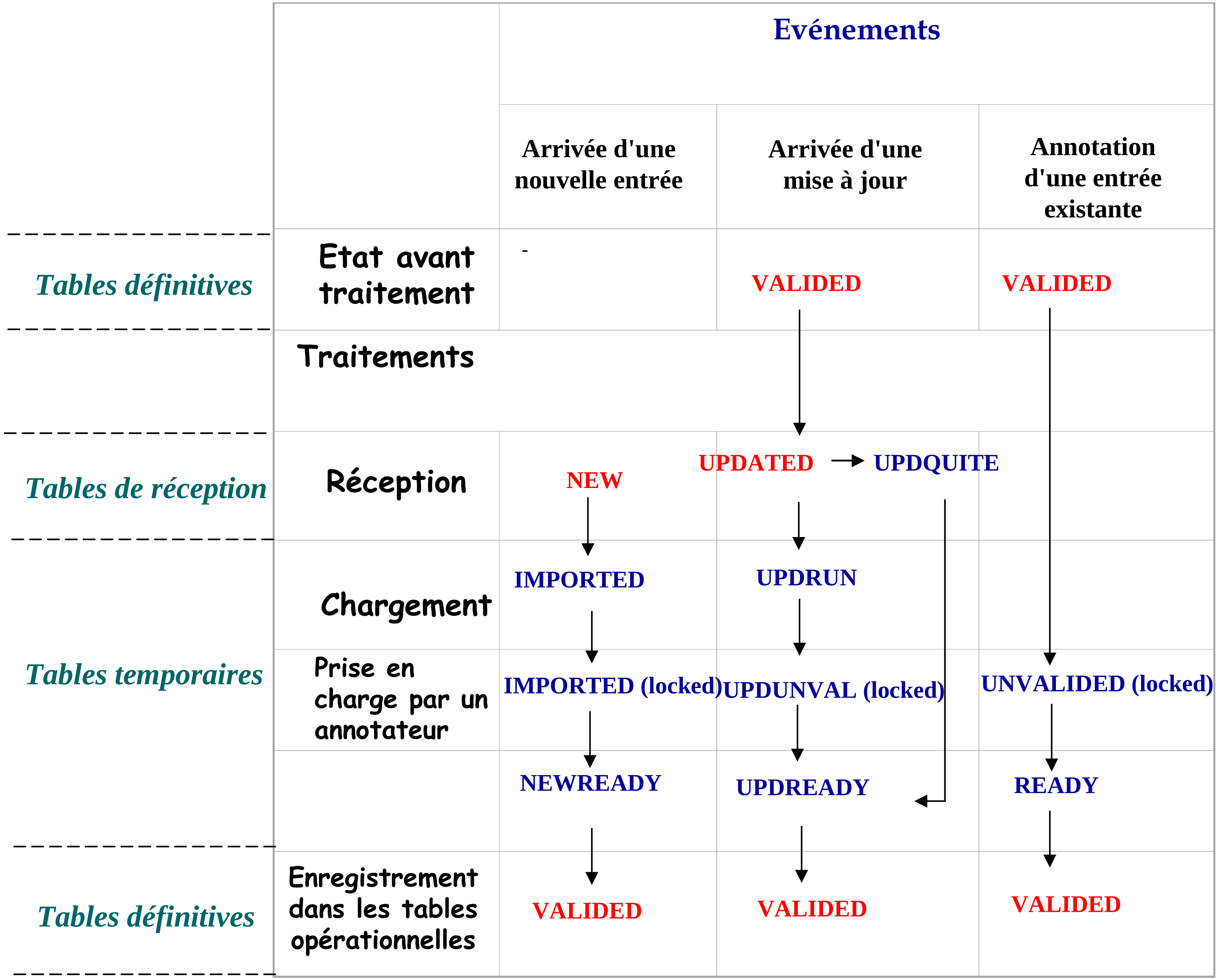
http://www.imgt.org



		Evénements		
		Arrivée d'une nouvelle entrée	Arrivée d'une mise à jour	Annotation d'une entrée existante
Tables définitives	**Etat avant traitement**	-	VALIDED	
	Traitements			
Tables de réception	**Réception**	NEW	UPDATED → UPDQUITE	
Tables temporaires	**Chargement**	IMPORTED	UPDRUN	
	Prise en charge par un annotateur	IMPORTED (locked)	UPDUNVAL (locked)	
		NEWREADY	UPDREADY	
Tables définitives	**Enregistrement dans les tables opérationnelles**	VALIDED	VALIDED	

(locked) : indique que l'entrée est prise en charge par un annotateur

http://www.imgt.org

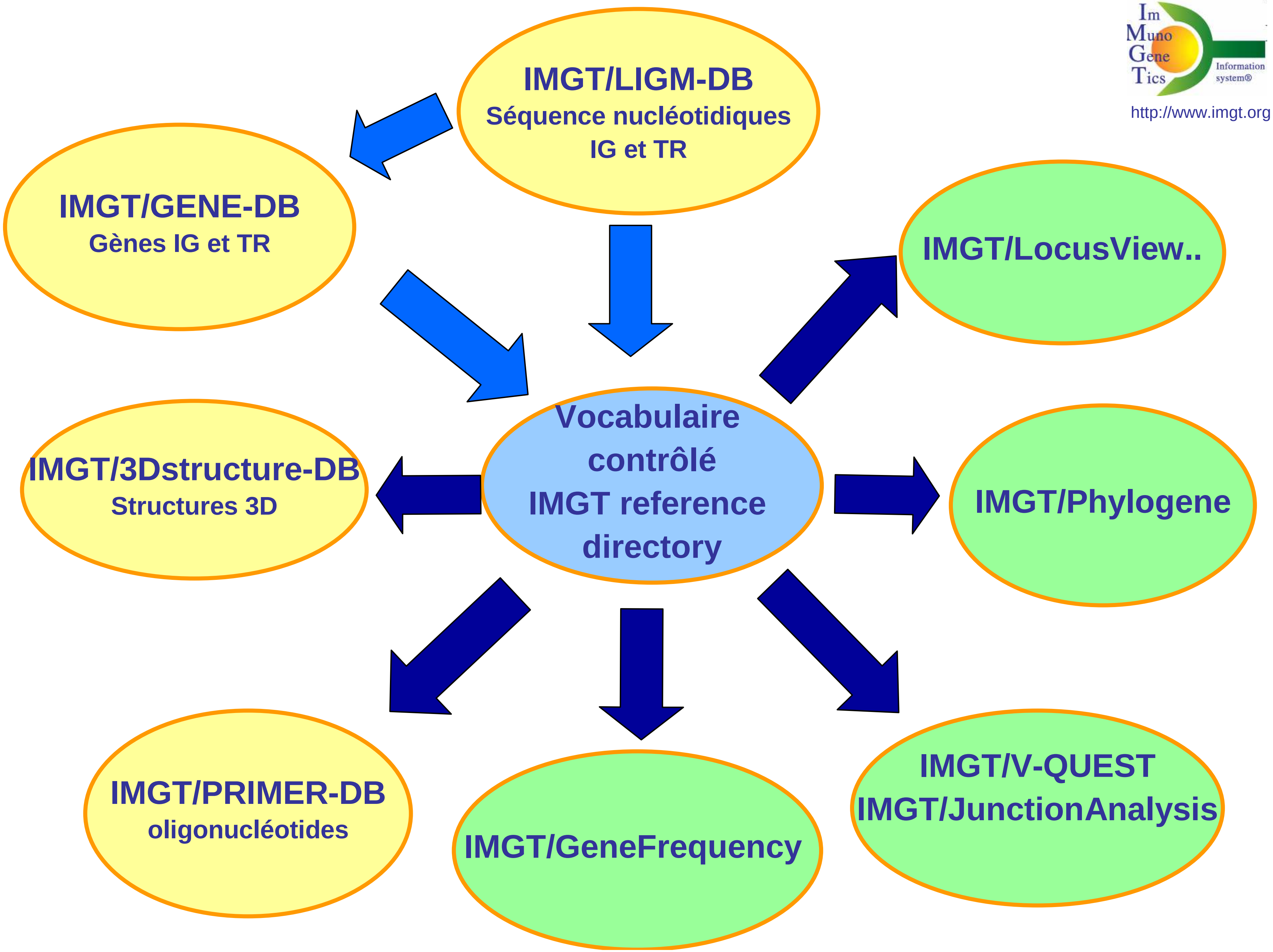


		Evénements		
		Arrivée d'une nouvelle entrée	Arrivée d'une mise à jour	Annotation d'une entrée existante
Tables définitives	**Etat avant traitement**	-	VALIDED	VALIDED
	Traitements			
Tables de réception	**Réception**	NEW	UPDATED → UPDQUITE	
Tables temporaires	**Chargement**	IMPORTED	UPDRUN	
	Prise en charge par un annotateur	IMPORTED (locked)	UPDUNVAL (locked)	UNVALIDED (locked)
		NEWREADY	UPDREADY	READY
Tables définitives	**Enregistrement dans les tables opérationnelles**	VALIDED	VALIDED	VALIDED

(locked) : indique que l'entrée est prise en charge par un annotateur

Im
Muno
Gene
Tics
Information system®

http://www.imgt.org



IMGT/LIGM-DB
Séquence nucléotidiques
IG et TR

IMGT/GENE-DB
Gènes IG et TR

IMGT/3Dstructure-DB
Structures 3D

IMGT/PRIMER-DB
oligonucléotides

Vocabulaire contrôlé
IMGT reference directory

IMGT/LocusView..

IMGT/Phylogene

IMGT/GeneFrequency

IMGT/V-QUEST
IMGT/JunctionAnalysis

Rôles des responsables de SI

Assurer la cohérence d'un ensemble d'informations qui comprend des données (dans une base de données...), des applications, interfaces pour un ensemble d'utilisateurs

On demande en plus:

- la disponibilité de l'information le plus rapidement possible
- simultanément sous différents formats
- à travers diverses interfaces
- "tout le monde" veut offrir des services

Les utilisateurs ont des demandes difficiles à anticiper, qui évoluent rapidement

Complexité des SI

- Les systèmes d'information n'ont jamais été conçus pour durer si longtemps ou pour être résistants au temps qui passe (bug de l'an 2000)
- Il faudrait adapter certaines méthodes, standards, protocoles dépassés/oubliés.
- Les versions des outils, standards, middleware changent plusieurs fois au cours d'un même projet.
- “La toute dernière release, (ou la version beta)" des outils disponibles comprend souvent de nombreux bugs.

Quand c'est enfin mature, c'est déjà obsolète !

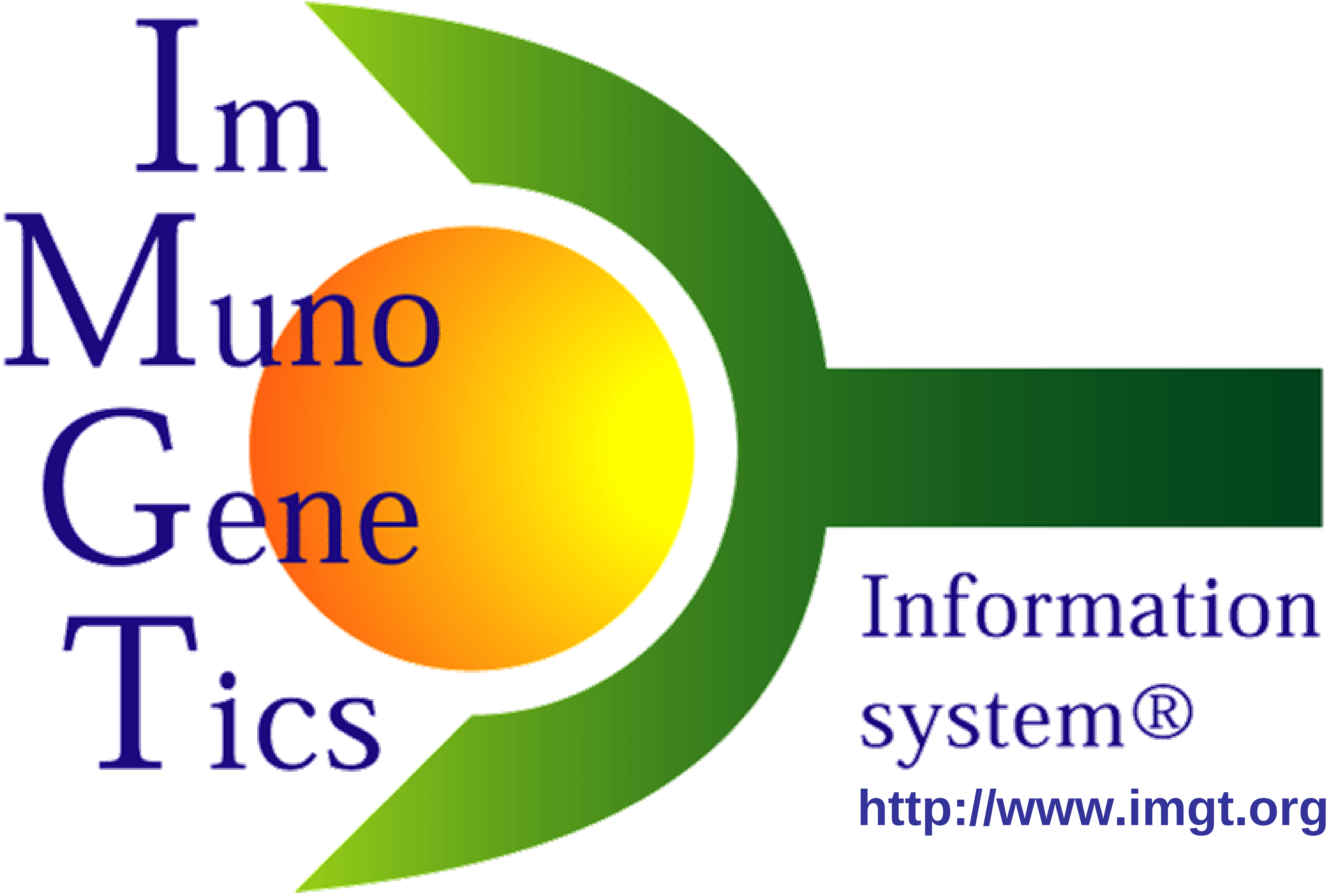
La durée de vie d’un SI dépend de

- Portabilité: il existe plusieurs plate-formes alternatives qui peuvent faire tourner le système
- Evolutivité: un système est évolutif, on peut:
  - ajouter des fonctionnalités
  - augmenter ses capacités maximum
  - l'adapter à des besoins différents/nouveaux
  - corriger des dysfonctionnements
- Réutilisation: Un système est réutilisable s'il peut être intégré (en totalité ou en partie) dans un nouveau projet.
- Standardisation: **ensemble de propriétés décrites dans une norme, supportée ou certifiée par un organisme officiel.**

L’équilibre des SI

introduire une innovation ↔ rester standard/compatible

augmenter l'intégration avec d'autres systèmes ↔ préserver la capacité à évoluer du système


Im
Muno
Gene
Tics
Information
system®
http://www.imgt.org